**Population Genomics**
*Editor-in-Chief:* Om P. Rajora

Paul A. Hohenlohe
Om P. Rajora *Editors*

# Population Genomics: Wildlife

Springer

# Population Genomics

**Editor-in-Chief**
Om P. Rajora
Faculty of Forestry and Environmental Management
University of New Brunswick
Fredericton, NB, Canada

This pioneering *Population Genomics Series* deals with the concepts and approaches of population genomics and their applications in addressing fundamental and applied topics in a wide variety of organisms. Population genomics is a fast emerging discipline, which has created a paradigm shift in many fields of life and medical sciences, including population biology, ecology, evolution, conservation, agriculture, horticulture, forestry, fisheries, human health and medicine.

Population genomics has revolutionized various disciplines of biology including population, evolutionary, ecological and conservation genetics, plant and animal breeding, human health, genetic medicine, and pharmacology by allowing to address novel and long-standing intractable questions with unprecedented power and accuracy. It employs large-scale or genome-wide genetic information across individuals and populations and bioinformatics, and provides a comprehensive genome-wide perspective and new insights that were not possible before.

Population genomics has provided novel conceptual approaches, and is tremendously advancing our understanding the roles of evolutionary processes, such as mutation, genetic drift, gene flow, and natural selection, in shaping up genetic variation at individual loci and across the genome and populations, disentangling the locus-specific effects from the genome-wide effects, detecting and localizing the functional genomic elements, improving the assessment of population genetic parameters or processes such as adaptive evolution, effective population size, gene flow, admixture, inbreeding and outbreeding depression, demography, and biogeography, and resolving evolutionary histories and phylogenetic relationships of extant and extinct species. Population genomics research is also providing key insights into the genomic basis of fitness, local adaptation, ecological and climate acclimation and adaptation, speciation, complex ecologically and economically important traits, and disease and insect resistance in plants, animals and/or humans. In fact, population genomics research has enabled the identification of genes and genetic variants associated with many disease conditions in humans, and it is facilitating genetic medicine and pharmacology. Furthermore, application of population genomics concepts and approaches facilitates plant and animal breeding, forensics, delineation of conservation genetic units, understanding evolutionary and genetic impacts of resource management practices and climate and environmental change, and conservation and sustainable management of plant and animal genetic resources.

The volume editors in this Series have been carefully selected and topics written by leading scholars from around the world.

More information about this series at http://www.springer.com/series/13836

Paul A. Hohenlohe • Om P. Rajora
Editors

# Population Genomics: Wildlife

Springer

*Editors*
Paul A. Hohenlohe
Department of Biological Sciences,
Institute for Bioinformatics and
Evolutionary Studies
University of Idaho
Moscow, ID, USA

Om P. Rajora
Faculty of Forestry and Environmental
Management
University of New Brunswick
Fredericton, NB, Canada

ISSN 2364-6764 ISSN 2364-6772 (electronic)
Population Genomics
ISBN 978-3-030-63488-9 ISBN 978-3-030-63489-6 (eBook)
https://doi.org/10.1007/978-3-030-63489-6



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

***Population Genomics Book Series***

*This Population Genomics book series is dedicated to my (late) parents, and my wife Malti and children Apoorva, Anu, and Maneesha.*

*Om P. Rajora*

***Population Genomics: Wildlife***

*This book is dedicated to my family, and wildlife lovers, researchers, educators, and managers.*

*Om P. Rajora*

*This book is dedicated to my family, and to all others who love the diversity of wildlife and seek to conserve it for future generations.*

*Paul A. Hohenlohe*

# Preface

Spurred by recent advances in DNA sequencing technology, genomics has reshaped multiple fields within biology since the turn of this century. Population genomics approaches can be applied far beyond traditional model organisms to reveal new insights into natural populations around the globe, including wildlife. While traditional genetics tools have been increasingly applied to wildlife populations, genomics provides not just increased power but also entirely new avenues of research. Population genomics can reveal not only diverse aspects of population biology, including various population genetic characteristics, overall patterns of demography, or relationships among populations, but also the genetic basis of adaptation, particularly in the face of rapid environmental changes.

Driven by advances in population genomics, and as part of the pioneering Population Genomics book series covering all aspects of population genomics, the current volume focuses on wildlife population genomics. Here, we consider wildlife to include primarily terrestrial, vertebrate animal species that are of concern for conservation or management. We intend this volume to serve as a resource for researchers in wildlife biology, who are using or may consider using population genomics approaches, and also for conservation practitioners, wildlife managers, and others outside the field to gain insights into the many ways that population genomics can contribute to our understanding and management of wildlife species.

Despite our focus on wildlife population genomics here, many of the approaches and topics covered in this book apply broadly, and wildlife biology can also learn from applications of population genomics in all other organisms. For instance, the genetics of adaptation to heterogeneous environments and adaptive differentiation among populations is addressed across animals, plants, and other organisms. Many population genomics approaches that are applicable in wildlife have been well-developed in fish, particularly in species that are the target of intensive harvest or management. Examples and illustrative case studies from these other taxa can be found throughout this volume.

This volume also focuses specifically on challenges and questions that are central to population genomics in wildlife. For instance, it is important in many wildlife taxa

to make use of non-invasive samples, collected from hair, feathers, scat, or even environmental DNA, and techniques for population genomics research using such samples are quickly advancing. This includes techniques, such as sequence capture or amplicon sequencing, that are well-suited for using low-quality genetic/genomic samples in wildlife studies. Many wildlife species are also limited to very small populations, which limits potential sample sizes and sequencing approaches and also drives the research questions that are most important for population genomics to address. In small populations, gathering the maximum amount of genetic information, such as whole-genome sequencing, from a small number of individuals may be the best way at present to draw meaningful conclusions. Important questions in small wildlife populations include loss of genetic diversity, the role of deleterious variation, and inbreeding depression, and all of these can be examined at a fine scale from population genomic data.

The chapters of this volume are arranged in five parts covering major components of wildlife population genomics. Following an introductory overview of wildlife population genomics applications and approaches, the sections cover sampling and sequencing approaches for wildlife, wildlife population biology, specific challenges and threats to wildlife populations, and applications of population genomics to management and conservation. In all cases, we attempt to balance discussion of general issues and concepts with chapters focused on particular taxa, from ungulates and amphibians to individual bird species. Case studies of population genomics applied to wildlife populations illustrate the potential of the field across all these chapters, and each chapter also provides a prospective view of future progress. The chapters are written by leading and emerging research scholars in wildlife population genomics.

Threatened wildlife species are at the heart of an ongoing biodiversity crisis, thus expanding the scientific toolkit for management and conservation is critical. A wide range of approaches is needed, from social science to ecology, and interdisciplinary connections among these fields are critical for science to guide solutions. Population genomics will play an important role in this network by providing general understanding of wildlife biology, specific information to guide management decisions in wildlife species, and efficient approaches for monitoring wildlife populations and the effects of our actions. We envision this book to be suitable for a wide readership, including undergraduate and graduate students, research scholars and professionals, and we hope that it will serve as a valuable resource for the field of wildlife population genomics.

Moscow, ID, USA — Paul A. Hohenlohe
Fredericton, NB, Canada — Om P. Rajora

# Contents

# List of Contributors

**Cameron L. Aldridge** Natural Resource Ecology Laboratory and Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA

**Kimberly R. Andrews** Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID, USA
Genetics and Genomics Group, University of Washington JISAO and NOAA Pacific Marine Environmental Lab, Seattle, WA, USA

**Niko Balkenhol** Faculty of Forest Sciences, Wildlife Sciences, University of Goettingen, Göttingen, Germany

**Soraia Barbosa** Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA

**Marc A. Beer** School of Biological Sciences, Washington State University, Pullman, WA, USA

**Leonardo Campagna** Cornell Lab of Ornithology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

**Andrew J. Crawford** Department of Biological Sciences, Universidad de Los Andes, Bogotá, Colombia

**Marta De Barba** Laboratoire d'Ecologie Alpine (LECA), CNRS, University of Grenoble Alpes, University of Savoie Mont Blanc, Grenoble, France

**Luke T. Dunning** Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

**Sarah W. Fitzpatrick** W.K. Kellogg Biological Station, Department of Integrative Biology, Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, Hickory Corners, MI, USA

**Brenna R. Forester** Department of Biology, Colorado State University, Fort Collins, CO, USA

**W. Chris Funk** Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA

**Caren S. Goldberg** School of the Environment, Washington State University, Pullman, WA, USA

**Brian K. Hand** Flathead Lake Biological Station, University of Montana, Polson, MT, USA

**Sarah A. Hendricks** Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA

**Paul M. Hime** Department of Biology, University of Kentucky, Lexington, KY, USA
Biodiversity Institute, University of Kansas, Lawrence, KS, USA

**Jacob Höglund** Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

**Paul A. Hohenlohe** Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA

**Scott Hotaling** Department of Biology, University of Kentucky, Lexington, KY, USA
School of Biological Sciences, Washington State University, Pullman, WA, USA

**Evelyn L. Jensen** Department of Biology, The University of British Columbia, Kelowna, BC, Canada

**Kara S. Jones** Department of Biology, University of Kentucky, Lexington, KY, USA

**Christopher P. Kozakiewicz** School of Biological Sciences, Washington State University, Pullman, WA, USA

**Justin D. Kratovil** Department of Biology, University of Kentucky, Lexington, KY, USA
Department of Entomology, University of Kentucky, Lexington, KY, USA

**Christopher J. Kyle** Environmental and Life Sciences, Trent University, Peterborough, ON, Canada
Department of Forensics, Trent University, Peterborough, ON, Canada

**Erin L. Landguth** School of Public and Community Health Sciences, University of Montana, Missoula, MT, USA

**Emily K. Latch** Behavioral and Molecular Ecology Research Group, Department of Biological Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

**Anssi Laurila** Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

**Philip Lavretsky** Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX, USA

**Daria Martchenko** Environmental and Life Sciences, Trent University, Peterborough, ON, Canada

**Mason O. Murphy** Department of Biology, University of Kentucky, Lexington, KY, USA
Department of Biology, Miami University, Oxford, OH, USA

**Schyler O. Nunziata** Department of Biology, University of Kentucky, Lexington, KY, USA
Department of Entomology, University of Maryland, College Park, MD, USA

**Kevin P. Oh** U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO, USA

**Sara J. Oyler-McCance** U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO, USA

**Meghan B. Parsley** School of the Environment, Washington State University, Pullman, WA, USA

**Erin Prewer** Environmental and Life Sciences, Trent University, Peterborough, ON, Canada

**Om P. Rajora** Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, NB, Canada

**Kristina M. Ramstad** Department of Biology and Geology, University of South Carolina Aiken, Aiken, SC, USA

**Patrik Rödin-Mörch** Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

**Michael A. Russello** Department of Biology, The University of British Columbia, Kelowna, BC, Canada

**Anna E. Savage** Department of Biology, University of Central Florida, Orlando, FL, USA

**Michael K. Schwartz** US Department of Agriculture, Forest Service, Rocky Mountain Research Station, National Genomics Center for Wildlife and Fish Conservation, Missoula, MT, USA

**Aaron B. A. Shafer** Environmental and Life Sciences, Trent University, Peterborough, ON, Canada
Department of Forensics, Trent University, Peterborough, ON, Canada

**Andrew Storfer** School of Biological Sciences, Washington State University, Pullman, WA, USA

**David P. L. Toews** Cornell Lab of Ornithology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

**Lisette P. Waits** Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID, USA

**Jennifer Walsh** Cornell Lab of Ornithology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

**Ashley D. Walters** US Department of Agriculture, Forest Service, Rocky Mountain Research Station, National Genomics Center for Wildlife and Fish Conservation, Missoula, MT, USA

**David W. Weisrock** Department of Biology, University of Kentucky, Lexington, KY, USA

**Kelly R. Zamudio** Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

**Shawna J. Zimmerman** Natural Resource Ecology Laboratory and Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA

# Part I
# Introduction

# Wildlife Population Genomics: Applications and Approaches


**Soraia Barbosa, Sarah A. Hendricks, W. Chris Funk, Om P. Rajora, and Paul A. Hohenlohe**

**Abstract** Population genomics provides a powerful and growing set of approaches for wildlife biology, revealing new insights into demographic history, population structure, adaptation, and the consequences of genetic diversity. Given the multiple threats faced by global biodiversity, it is critical for researchers to advance efforts to manage and conserve wildlife populations. In this chapter we provide an overview of the research questions that can be addressed in wildlife population genomics, applications to specific conservation and management issues, and the variety of technical methods at all stages from sampling to sequencing and data analysis. Wildlife species, here defined as vertebrate species of specific conservation or management concern, present unique challenges and opportunities. These include not only the necessity of using poor-quality samples from non-invasive or archival collections, but also the availability of genomic reference data from closely related domestic species. We highlight a number of case studies in particular taxa that illustrate recent progress in wildlife population genomics, including how population genomics approaches have been applied to date, and also how the field can continue to connect research to urgent conservation actions in wildlife populations. We also discuss prospects for applications of population epigenomics, transcriptomics, metagenomics, and eDNA approaches in wildlife.

Soraia Barbosa and Sarah A. Hendricks contributed equally to this work.

S. Barbosa · S. A. Hendricks · P. A. Hohenlohe (✉)
Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA
e-mail: hohenlohe@uidaho.edu

W. C. Funk
Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA

O. P. Rajora
Faculty of Forestry and Environmental Management, University of New Brunswick, Fredericton, NB, Canada

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)], https://doi.org/10.1007/13836_2020_83, 

**Keywords** Adaptive potential · Conservation units · Effective population size · Genetic rescue · Inbreeding depression · Population epigenomics · Population genomics · Population transcriptomics · Whole-genome sequencing · Wildlife

## 1 Introduction

Wildlife species face a number of threats, such as habitat loss and fragmentation, direct mortality from exploitation, invasive species, pollution, and climate change. The genomic revolution has democratized the field of population genomics, allowing high-throughput sequencing to be applied in nearly any organism, including natural populations of rare or difficult-to-study wildlife species (Luikart et al. 2019; Rajora 2019). Wildlife biology can benefit from population genomics in several ways: first, by improving our basic understanding of wildlife species and populations, including their evolutionary history and relationships, adaptation to local environments and ecological interactions, and population dynamics and viability. Second, this information can inform management and conservation efforts, such as delineating population units for conservation, maintaining genetic diversity in captive or wild populations, or predicting adaptive potential. The importance of genetic variation in conservation, including its role in setting conservation targets and monitoring the status of biodiversity, is increasingly recognized and can benefit from genomics tools (Hoban et al. 2020). Population genomics studies can provide efficient genetic approaches for monitoring and managing populations, and a number of technical improvements specifically make genomics more applicable to wildlife. These include methods for using non-invasive DNA samples or environmental DNA, and sequencing tools that can be used in the field.

Traditional wildlife population genetics has focused on mitochondrial, and less frequently, nuclear DNA sequences for population-level analyses, and microsatellites for individual-level analyses (Schwartz et al. 2007; Allendorf 2017). Mitochondrial DNA has been the most widely used molecular marker for population genetic diversity, phylogeography, conservation units, and species identification (DeYoung and Honeycutt 2005; Hajibabaei et al. 2007). Microsatellite analyses have focused mostly on identifying fine-scale genetic population structure and connectivity, population origin, estimating kinship, abundance and dispersal, as well as studying behavior, such as determining mating systems (Carroll et al. 2018). A great deal of wildlife conservation research has been dedicated to evaluate population connectivity and individual dispersal, improved by the ability of performing individual identification through non-invasive samples, for example in Cabrera voles (*Microtus cabrerae*; Ferreira et al. 2018) to tigers (*Panthera tigris*; Thatte et al. 2018). Here, integration with landscape ecology has enabled great insights into the identification of dispersal corridors and barriers to gene flow, which has shown the vulnerability of isolated small populations in many species of conservation concern (Proctor et al. 2005; Shah and McRae 2008; Waits et al.

2016). Very important were also the studies aimed at identifying links between variation at microsatellites associated with immune genes and selection and fitness, particularly using Major Histocompatibility Complex genes (Oliver et al. 2009; Palomares et al. 2012). Traditional markers like microsatellites continue to be extremely valuable for wildlife genetics studies, for instance for estimating relatedness in captive populations of Iberian lynx (*Lynx pardinus*; Kleinman-Ruiz et al. 2019) or fragmented wild populations of African leopards (*Panthera pardus pardus*; Naude et al. 2020).

Wildlife biology has much to gain from the transition from population genetics to population genomics methods, as over the past decade, the techniques of population genomics have been applied widely across the fields of biology (Rajora 2019). Applications to conservation have been slow to develop because of hurdles in translating research into concrete actions, due to limitations of costs, sample quality, or applicability of the methods. The power of emerging genomics methods to answer different questions is central to their application in wildlife. For instance, the use of RAD sequencing to detect adaptive variation has been under debate because the technique samples a subset of loci across the genome, and some important regions of the genome may be missed. This potentially limits inferences about the genetics of adaptation, but progress is also being made in how to quantify and assess trade-offs among methods (Catchen et al. 2017; Lowry et al. 2017a, b; McKinney et al. 2017; Hohenlohe et al. 2020). However, recent years have seen accelerating progress in translating genomics research into management (Allendorf et al. 2010; Steiner et al. 2013; Shafer et al. 2015; Breed et al. 2019; Walters and Schwartz 2020, this volume). For example in salmonid fish, multiple approaches, including SNP arrays, RAD sequencing, and whole-genome analysis, have been used to identify conservation units, quantify genetic diversity, detect local adaptation, and determine genotype–phenotype associations, all with consequences for the intensive management and conservation efforts in these fish (Waples and Lindley 2018; Waples et al. 2020).

A critical need in many wildlife studies is to gather genetic data from non-invasive samples, such as feces and hair. Mitochondrial DNA and microsatellites have been extensively applied in these situations, which promoted a rapid expansion of their use in wildlife conservation (Waits and Paetkau 2005; Beja-Pereira et al. 2009; Andrews et al. 2018, this volume). This was especially useful for threatened and elusive species, for which non-invasive genetic sampling provided more accurate estimates of species presence, density, kinship, and dispersal, often at a lower cost (Barbosa et al. 2013; Hedges et al. 2013; Ferreira et al. 2018). Due to issues of sample quality, the use of non-invasive samples in the genomic era has lagged and so have genetic monitoring studies that deal with threatened and elusive species (Carroll et al. 2018). Nonetheless, an expanding set of genomics tools can now be applied to non-invasive and low-quality DNA samples (Carroll et al. 2018; Andrews et al. 2018, this volume). For example, active research areas in genomic research for wildlife monitoring include the use of in situ sequence amplification, which has been used for bird sexing from blood and feather samples, DNA barcoding, and for single species detection from environmental DNA samples

(Centeno-Cuadros et al. 2017; Williams et al. 2019; Watsa et al. 2020). Environmental DNA samples may also have potential beyond species detection, for assessing population-level characteristics of genetic diversity or population structure (Goldberg and Parsley 2020, this volume).

In this chapter we provide an overview of the field of wildlife population genomics: the range of techniques and resources available for genomic studies, the biological questions that can be addressed, and applications of population genomics to wildlife management and conservation. We highlight a few key areas, such as whole-genome sequencing, that are emerging as central to the field. We also discuss approaches with future potential for applications to wildlife, such as population epigenomics, population transcriptomics, metagenomics, and eDNA for population genomics.

## 2 Addressing Research Questions in Wildlife Biology

### *2.1 Population Genetics Versus Genomics in Wildlife*

Traditional population genetics has applied techniques like allozyme and microsatellite genotyping or sequencing of mitochondrial and chloroplast genes to provide a wealth of knowledge about natural populations (Allendorf 2017). However, these techniques provide data on a limited number of genetic markers across individuals, and a common assumption is that this sample of markers represents the action of neutral processes that affect the whole genome. Statistical power of these traditional genetics approaches is also limited by the sample size of loci or markers. Advances in next-generation sequencing technology have led to a proliferation of techniques for population genomics studies, all of which have the potential to provide fine-scale genetic data across the genome of multiple individuals. The central advance of next-generation sequencing is that heterogeneous pools of DNA fragments can be sequenced together, rather than requiring individual fragments to be isolated and amplified (Mardis 2008). This means that data can be gathered across thousands of loci, or even across the whole genome, in a single sequencing library. Critical for applications to wildlife, many approaches in population genomics are suitable even in taxa with little or no existing genomic resources.

Many basic questions in wildlife populations were addressed with traditional genetic tools, and these can be addressed with genomics techniques as well. An advantage of the number of loci that genomics approaches provide is much higher precision in estimating population genetic statistics or detecting patterns, such as genetic differentiation among populations or phylogenetic relationships among taxa (Hohenlohe et al. 2019). This use of high-throughput sequencing to address longstanding questions, but with greater precision or statistical power, has been called “broad-sense genomics” (Garner et al. 2016). For example, Zimmerman et al. (2020) compared microsatellite genotyping with reduced representation sequencing in Gunnison sage-grouse (*Centrocercus minimus*) and found finer-scale detection of

population structure with the genomics approach. Additionally, population genomics opens the door to address new questions in wildlife biology that were previously intractable with traditional genetic tools, what has been called "narrow-sense genomics" (Garner et al. 2016), particularly when genetic information can be arrayed along a map of the genome (Allendorf 2017). For instance, the scale of genomics tools can reveal features of neutral processes such as fine-scale historical reconstructions of inbreeding in small wildlife populations (e.g., Grossen et al. 2020). Narrow-sense population genomics enables the fundamental advance of being able to detect specific genomic regions or loci that are under natural selection or associated with ecologically important traits (Garner et al. 2016; Allendorf 2017).

Surveying examples of recent work in applying population genomics to wildlife reveal a few basic conclusions (Table 1). First, a wide range of scientific questions have been addressed, and these can be very roughly divided into those that affect the genome as a whole (e.g., demographic patterns, population relationships, and other "neutral" processes) and those that relate to a subset of the genome containing genetic variation related to adaptation, fitness, or important phenotypes. Second, these distinctions among types of research questions or genetics versus genomics techniques are often not clear. For example, many studies listed in Table 1 and described below address multiple questions at once, such as population structure and detection of adaptive variation, that span the "broad-sense" and "narrow-sense" aspects of genomics (e.g., Saremi et al. 2019; Oyler-McCance et al. 2020, this volume). Several studies also combine techniques, such as using next-generation sequencing tools to efficiently identify a set of marker loci that can be consistently genotyped over time, for instance in long-term monitoring of wildlife populations. The resulting marker panels may have relatively few loci and not constitute a "genomic" dataset in terms of representation across the genome. Nonetheless, such marker panels may target adaptive variation and represent a substantial advance in wildlife population genomics (Meek et al. 2016; Förster et al. 2018; Eriksson et al. 2020).

Studies in wildlife population genomics can occur across a wide range of taxonomic and spatial scales, and these factors drive the sampling design as well as choice of sequencing techniques and analysis tools (Fig. 1). At one extreme, questions about phylogenetic relationships or species presence in a community cut across related species, while requiring relatively few individual samples. At the opposite extreme, studies focused on individual relatedness or inbreeding can occur within a single population, sampling a large number of potentially related individuals. In the middle, studies of population structure or local adaptation gain statistical power by sampling individuals within a species across a broad range of populations, locations, or environmental factors. In all cases, the number of loci required varies widely among research questions, depending on whether a study needs a smaller representative sample of loci, or more comprehensive sampling to reveal factors like selection affecting individual genes.

**Table 1** Examples of research goals that can be addressed in wildlife using population genomics, applications to conservation or management efforts, and recent illustrative studies

| Research goal | Conservation or management application | Published examples |
|---|---|---|
| *Demographic processes and population relationships* | | |
| Estimate phylogenetic relationships among taxa | Understand evolutionary relationships among threatened species | Wolves (*Canis* spp.); Sinding et al. (2018) |
| Estimate effective population size ($N_e$) | Assess ongoing loss of genetic diversity; identify conservation priorities | Gorillas (*Gorilla beringei* subspp.); van der Valk et al. (2019)<br>Ibex (*Capra ibex*); Grossen et al. (2018) |
| Reconstruct historical trends in $N_e$ | Understand historic influences on current genetic diversity | Tasmanian devil (*Sarcophilus harrisii*); Patton et al. (2019)<br>Eurasian lynx (*Lynx lynx*); Lucena-Perez et al. (2020) |
| Identify geographic population structure | Identify population units for conservation | Pandas (*Ailuropoda melanoleuca*); Zhao et al. (2013)<br>Yellow-legged frogs (*Rana boylii*); McCartney-Melstad et al. (2018)<br>Pangolins (*Manis* spp.); Hu et al. (2020) |
| Quantify population distinctiveness | Establish whether populations meet criteria for conservation status listing | Rockfish (*Sebastes* spp.); Andrews et al. (2018) |
| Estimate population connectivity and levels of gene flow | Manage migration among populations to maintain genetic diversity | Polar bears (*Ursus maritimus*); Jensen et al. (2020) |
| Estimate levels of hybridization | Maintain locally adapted genotypes; characterize the spread of hybridizing invasive species | Westslope cutthroat trout (*Oncorhynchus clarki lewisi*); Muhlfeld et al. (2017) |
| Estimate current levels and historic trends of inbreeding | Identify priority populations for conservation action | Pumas (*Felis concolor*); Saremi et al. (2019) |
| *Adaptive and functional variation* | | |
| Estimate heritability of phenotypic traits | Quantify the adaptive potential of populations to respond to selection | Hihi (*Notiomystis cincta*); de Villemereuil et al. (2019) |
| Test for inbreeding depression | Quantify population-level impacts of inbreeding; identify targets for genetic rescue | Red deer (*Cervus elaphus*); Huisman et al. (2016) |
| Assess the fitness impacts of deleterious mutations in small populations | Quantify the effects of genetic drift and purging on population fitness; identify targets for assisted gene flow | Island foxes (*Urocyon littoralis*); Robinson et al. (2018)<br>Alpine ibex (*Capra ibex*); Grossen et al. (2020) |

(continued)

**Table 1** (continued)

| Research goal | Conservation or management application | Published examples |
|---|---|---|
| Identify loci associated with adaptive differentiation, with either outlier or GEA approaches | Evaluate adaptive differences among populations; inform potential translocations or assisted gene flow | Thick-billed murres (*Uria lomvia*); Tigano et al. (2017)<br>Pikas (*Ochotona princeps*); Waterhouse et al. (2018) |
| Test for contemporary genomic responses to selection | Identify populations currently adapting to environmental change | Chipmunks (*Tamias* spp.); Bi et al. (2019) |
| Identify loci associated with phenotypic traits (GWAS) | Manage populations for ecologically important phenotypes | Tasmanian devils (*Sarcophilus harrisii*); Margres et al. (2018) |
| Estimate adaptive potential | Assess the capacity of populations to adapt to environmental change without intervention | Willow flycatchers (*Empidonax traillii*); Ruegg et al. (2018) |
| Estimate genomic vulnerability | Identify populations that may be genetically maladapted to future environmental conditions and warrant management actions | Yellow warblers (*Setophaga petechia*); Bay et al. (2018) |
| Develop genetic marker panels for high-throughput genotyping | Genetic monitoring of natural populations, including tracking adaptive responses | Eurasian lynx (*Lynx lynx*); Förster et al. (2018) |

## 2.2 Populations, Demography, and Neutral Processes

A central feature of wildlife biology is the size, distribution, and relationships of populations across a species' range. Population genomics tools provide an abundance of genetic data that can be used to understand wildlife populations (Hohenlohe et al. 2020). The size of a population strongly influences its viability, including its genetic capacity to adapt to environmental change, with implications for wildlife conservation and management actions. The genetic consequences of small population size are captured by $N_e$, the effective population size. $N_e$ captures the rate of genetic drift in a population; formally, it is the size of an idealized population with the same rate of genetic drift as the population under study (Charlesworth 2009). $N_e$ can be estimated with genetic data and multiple genomic data sources (Nunziata and Weisrock 2018; Grossen et al. 2018). Genetic and genomic data are also applied to delineate populations according to different criteria for conservation or management (Funk et al. 2012).

For example, Grossen et al. (2018) estimated $N_e$ in several populations of Alpine ibex (*Capra ibex*) compared to the closely related Iberian ibex (*Capra pyrenaica*) and domestic goat (*Capra hircus*), using over 100,000 SNP loci identified through RAD sequencing. These data provide precise estimates of individual-level

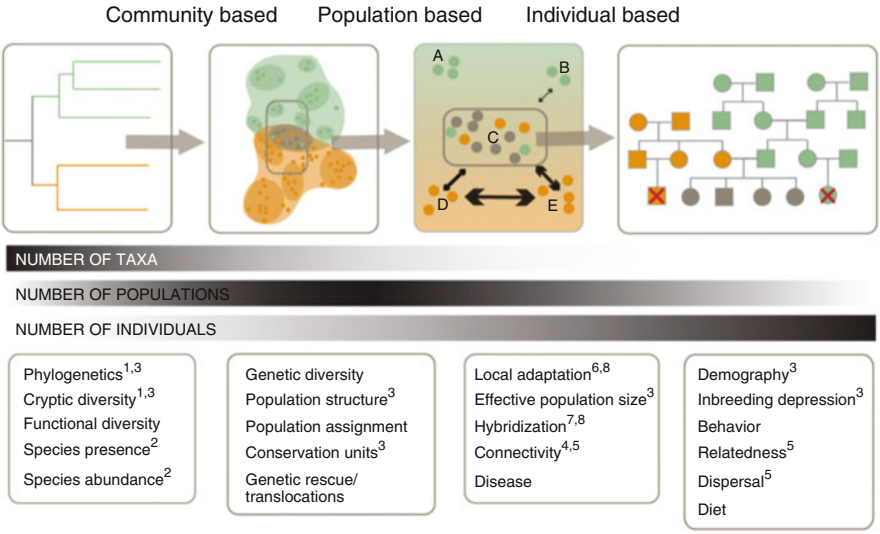


**Fig. 1** Research questions in wildlife biology can be addressed with population genomics across a range of scales, from groups of related species (left) to populations within a species (middle), to individuals within a single population (right). At these different scales, the relative numbers of taxa, populations, or individuals that should be sampled for any particular study vary (darker gray represents relatively more sampling at this level). At each scale, different research questions may require different numbers of loci to adequately sample the genome, in order to provide statistical power for particular analyses or fine-grained genomic information. For all of these aspects, this figure is meant as a rough, non-quantitative guide. Case studies for research questions: (1) Barbosa et al. (2018); (2) Marshall and Stepien (2019); (3) Hu et al. (2020); (4) Eriksson et al. (2020); (5) Escoda et al. (2019); (6) Rellstab et al. (2019); (7) Peek et al. (2019); (8) Mills et al. (2018)

heterozygosity and standing genetic variation across the genome for each population (Fig. 2a). These were translated into estimates of $N_e$ using linkage disequilibrium-based methods (Fig. 2b). Despite the large amount of genomic data, note that the estimates of $N_e$ still have large confidence intervals, particularly on the upper limit. Nonetheless, these estimates of $N_e$ in Alpine ibex populations, many of which were established by a series of reintroductions during the last century, provide useful information for understanding these populations. Further, genomic data can reveal the consequences of low $N_e$, such as inbreeding and accumulation of deleterious alleles, that provide a more detailed picture of the genetic status and viability of wildlife populations (Kardos et al. 2018; Robinson et al. 2019; Grossen et al. 2020).

Understanding the history of populations can be important for wildlife, including historical effects on genetic diversity. Even in the absence of historical samples, population genomics tools can provide a detailed picture of population demography, including current and historical trends in $N_e$ (Fig. 3a; Salmona et al. 2019; Lucena-Perez et al. 2020). Genome-wide data can be used to estimate time scales of population bottlenecks and expansions as well as infer the severity of demographic changes, which can help explain current levels of genetic diversity. The combination

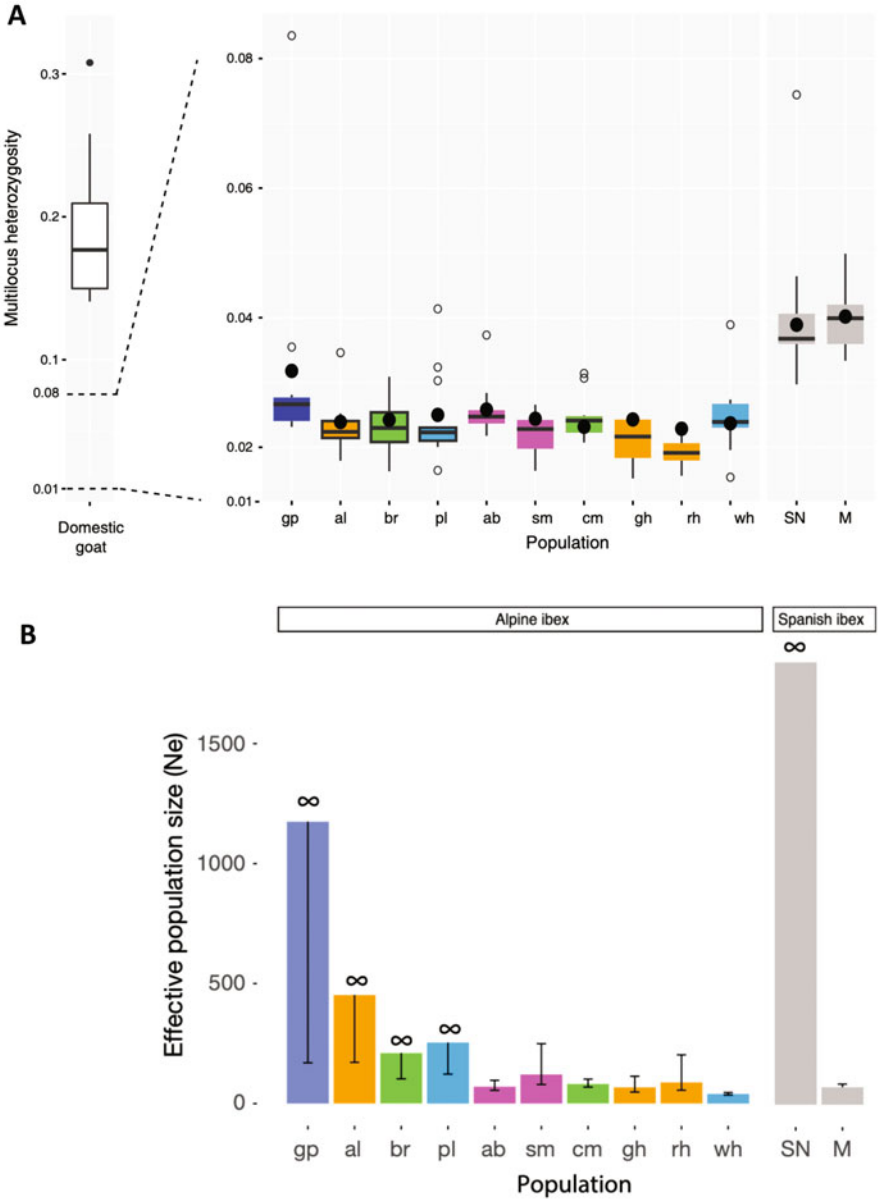


**Fig. 2** Estimates of (**a**) genetic diversity and (**b**) effective population sizes across reintroduced populations of Alpine ibex (*Capra ibex*) and Iberian ibex (*C. pyrenaica*) based on RAD sequencing data. Reproduced from Grossen et al. (2018)

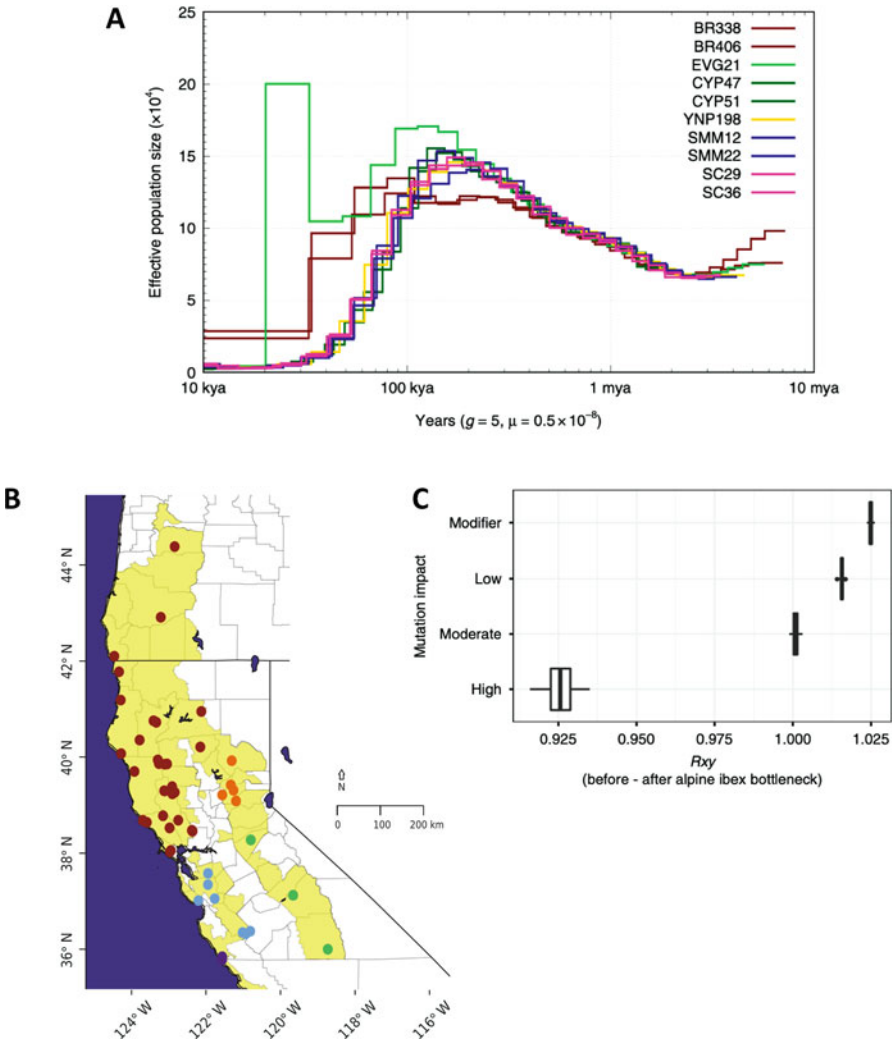


**Fig. 3** Examples of population genomics to understand demographics and population structure in wildlife: (**a**) Reconstruction of historical effective population sizes using sequentially Markovian coalescent (SMC) analysis of whole-genome data from 10 individual pumas (*Puma concolor*) from populations across North and South America. Reproduced from Saremi et al. (2019). (**b**) Population structure of the foothill yellow-legged frog (*Rana boylii*) in the western USA, estimated by phylogenetic analysis of genomic data from a RAD sequencing approach, with colors indicating different population groups that could be managed as genetically distinct units. Reproduced from McCartney-Melstad et al. (2018). (**c**) The effects of population bottlenecks and inbreeding on deleterious mutations in Alpine ibex (*Capra ibex*), assessed by change in allele frequency of the derived allele ($R_{xy}$). $R_{xy} < 1$ indicates a decrease in frequency of these mutations after bottlenecks in Alpine ibex compared to Iberian ibex; $R_{xy} > 1$ indicates an increase in frequency after bottlenecks. The excess of modifier and low-impact mutations shows the influence of genetic drift, while the reduced frequency of high-impact mutations indicates purging of these mutations during population bottlenecks. Reproduced from Grossen et al. (2020)

of demographic methods to estimate historical population size as well as diversity metric estimation can further reveal the relationship between historical and current levels of genetic diversity in light of demographic history. In the case of African wild dogs (*Lycaon pictus*), whole-genome analysis indicated that despite historically low effective population sizes, heterozygosity remains high in the current population (Armstrong et al. 2019). A study using WGS of both museum and contemporary samples by van der Valk et al. (2019) showed that over the past century the mountain gorilla (*Gorilla beringei beringei*) population has remained small, but genetically stable, while the Graur's gorilla (*Gorilla beringei graueri*) underwent population declines which led to increased inbreeding and loss of genetic diversity.

## 2.3 Population Structure and Connectivity

The distribution of genetic variation, population structure, and connectivity can be estimated using population genomics tools. These analyses can be critical for addressing key questions in wildlife conservation and management, as described below. For instance, population structure analysis of the foothill yellow-legged frog (*Rana boylii*) in the western USA, estimated by phylogenetic analysis of genomic data from a RAD sequencing approach, indicated that different population groups could be managed as genetically distinct units (Fig. 3b; McCartney-Melstad et al. 2018). Because of greater statistical power and sensitivity, genomic data can often reveal population structure that is not apparent with mitochondrial sequencing or fewer microsatellite loci, as observed, for example, in Gunnison sage-grouse (*Centrocercus minimus*; Zimmerman et al. 2020). In species with specific environmental threats, like polar bears (*Ursus maritimus*) facing climate change, assessment of population structure is a necessary first step to understand how population might respond (Jensen et al. 2020).

Combined with environmental data, demographic studies can assess how geography and climatic history influence geographic ranges, population sizes, gene flow, divergence, and speciation (Salmona et al. 2019). Ancient demography is important for understanding the driving factors, such as environmental changes, behind past population fluctuations and factors involved in historical connections or barriers to connectivity among populations. For example, analyses of demographic history of Malayan pangolins (*Manis javanica*) showed the effects of long-term environmental changes, including climate (as measured in surface temperature) and sea-level oscillations, revealing multiple population size changes in their evolutionary history (Hu et al. 2020). In another example, Zhao et al. (2013) analyzed WGS data of wild giant pandas (*Ailuropoda melanoleuca*) finding the occurrence of multiple demographic events such as population expansion, bottlenecks, and population divergence. They inferred that the decline of pandas in the last 3,000 years is likely due to anthropogenic disturbances. Timing of historical splits between populations can also be identified by divergence in population size estimates, as observed between European and Asian populations of Eurasian lynx (*Lynx lynx*; Lucena-Perez et al. 2020). Historical demographic reconstruction has the power to assess population

changes in light of the past environmental and anthropogenic shifts and may be able to inform the effect of ongoing changes on spatial distribution of genetic diversity (Prates et al. 2016).

Many wildlife species have reduced and fragmented populations, leading to loss of genetic diversity and inbreeding, and the potential for reduced fitness from inbreeding depression. Overall levels of genetic diversity can be estimated with genetic tools, and genomic data provide precise estimates from densely sampling the genome. Population genomics tools can precisely estimate individual inbreeding coefficients and pairwise genetic relatedness – and the relationship between them – to test for inbreeding depression (Huisman et al. 2016). They can also be used to map loci associated with individual inbreeding or reduced fitness, similar to approaches for mapping adaptive variation as discussed below.

## *2.4 Hybridization*

Biological aspects of some wildlife systems present particular challenges for management that can be addressed with population genomics; for instance, Toews et al. (2018, this volume) document how hybridization among bird species has been an important source of variation and possibly led to the formation of new species. In mallard ducks and their relatives (*Anas* spp.), hybridization has occurred between feral mallards introduced widely by humans and native *Anas* species across the globe (Lavretsky 2020, this volume). This creates challenges for management of native biodiversity in this group. Genomics methods for identifying hybrid individuals and monitoring the extent of hybridization across a landscape can answer basic questions that may inform management decisions. Additionally, identifying genomic regions or genes associated with admixture – the flow of allelic variation into a hybridized population – can reveal how selection operates in these populations. Alleles from an invasive species that spread rapidly into a native population may be considered "invasive alleles." For instance, some alleles from introduced rainbow trout (*Oncorhynchus mykiss*) move preferentially into native westslope cutthroat trout (*O. clarkii lewisi*), although selection predominantly acts against introduced alleles across most of the genome (Kovach et al. 2016). The spread of hybridization across the native species range in this system depends on a large number of factors, including water temperature and precipitation as well as proximity to sources (stocking locations) for the invasive rainbow trout (Muhlfeld et al. 2017). Climate change will continue to impact the spread of hybridization in this and other systems (Muhlfeld et al. 2017).

In addition to more accurately quantifying hybridization across individuals and populations, and identifying loci that are responding to selection in hybridized populations, population genomics can also reveal the history of hybridization. For instance, North American canids have a complex pattern of hybridization among taxa, both over evolutionary history and more recently in response to anthropogenic factors (vonHoldt et al. 2016; Sinding et al. 2018). Red wolves (*Canis rufus*), native

to the southeastern United States, have been subject to extensive hybridization with coyotes, which expanded their range eastward from the western US following European settlement. However, they also show evidence of historical hybridization with wolf and/or coyote lineages earlier in their evolutionary history (vonHoldt et al. 2016; Hohenlohe et al. 2017). Understanding not only the extent of hybrid ancestry but also the time scale over which hybridization occurred in red wolves is important for management decisions (Waples et al. 2018). Conversely, conservation policy can be informed by our growing understanding of the role of hybridization in wildlife taxa (vonHoldt et al. 2018; Heppenheimer et al. 2018; Funk et al. 2019).

## *2.5 Adaptive Variation*

A central feature of wildlife populations is their adaptation to local environmental and ecological conditions, the genetic variation that facilitates adaptive responses, and how these affect population size, growth rates, dispersal, and the long-term viability of wildlife populations. Population genomics tools provide multiple approaches for assessing adaptive genetic variation in wildlife populations. At the phenotypic level, genomics tools can be used to estimate individual relatedness and heritability of phenotypic traits, including traits linked to ecological functions or even fitness as a phenotype (Gienapp et al. 2017; de Villemereuil et al. 2019). This provides a quantitative genetic assessment of the ability of populations to adapt (Reed et al. 2011).

Increasingly, population genomics tools are being used to detect specific loci associated with fitness, adaptation, or ecological functions (Luikart et al. 2019). One approach is outlier tests that identify loci that are strongly differentiated among populations, indicating a signature of local adaptation (Beaumont and Nichols 1996). Outlier tests have the advantage of relying only on sampling individuals from different populations, without requiring other data on phenotypes or environmental variables. However, several other factors such as recombination rate heterogeneity and demographic fluctuations can have strong effects on errors in outlier analyses, including high rates of false positives (Lotterhos and Whitlock 2015; Hoban et al. 2016). If both genomic data and phenotypic measurements are available on a set of individuals, loci can be associated with phenotypic variation using genome-wide association studies (GWAS; Wellenreuther and Hansson 2016). However, the power of GWAS in wildlife systems is often limited. Even for model species and those with high levels of genome-wide heterozygosity, sample sizes need to be thousands of individuals for sufficient power to detect loci of small or moderate effect (Joshi et al. 2015). However, GWAS with limited sample sizes in wildlife can still reveal important features of the genetic basis of key traits, even if the specific effects of individual loci cannot be quantified with precision (Margres et al. 2018). In other cases, for abundant or commercially harvested species, sample sizes are sufficient to unravel the genetic basis of complex traits in wildlife using GWAS (Sinclair-Waters et al. 2020).

For many wildlife species, populations are distributed across a variable landscape, and combining genomic data with measurements of environmental variables reveals insights into adaptation and ecological factors affecting wildlife populations (Manel et al. 2003; Forester et al. 2018, this volume). This approach is called landscape genomics. The field can be divided into neutral landscape genomics, which focuses on understanding gene flow and connectivity, and adaptive landscape genomics, which focuses on characterizing the genetic basis of adaptation and how natural selection structures the distribution of adaptive genetic variation across the range of a species (Balkenhol et al. 2019). However, both neutral and adaptive information are available from most genomic datasets, so many studies can address both concurrently. The central approach of adaptive landscape genomics is genotype–environment association (GEA) analysis, which links allelic variation to environmental variables. Forester et al. (2018, this volume) provide guidance for applying GEA analysis, including design of sampling, genomic and environmental data production, and specific issues that can be addressed in wildlife populations.

Some case studies illustrate the complementary use of multiple techniques in applying population genomics to wildlife, such as combining whole-genome sequencing with genotyping at a small panel of markers. This approach can address multiple questions, such as identifying population structure and conservation units along with adaptive differentiation. For example, researchers working on sage-grouse (Oh et al. 2019; Zimmerman et al. 2020; Oyler-McCance et al. 2020, this volume) used WGS to determine genome-wide differentiation between two species (greater and Gunnison) of sage-grouse (*Centrocercus* spp.) and found intraspecific population structure consistent with genetic drift due to limited gene flow among populations. Further, they used a high-density marker panel to probe SNPs exhibiting extreme population differentiation. They found candidate genes associated with local dietary adaptations, which calls for conservation strategies that account for the specific chemistry of local sagebrush on sage-grouse.

## *2.6 Deleterious Variation and Inbreeding Depression*

Many wildlife populations have reduced and fragmented populations, leading to loss of genetic diversity and inbreeding, and the potential for reduced fitness from inbreeding depression. Population genomics tools can precisely estimate individual inbreeding coefficients and pairwise genetic relatedness to test for inbreeding depression. In red deer (*Cervus elaphus*), inbreeding coefficients estimated using SNPs were compared to several different fitness metrics (Huisman et al. 2016). Strong evidence for inbreeding depression was found including associations between annual breeding success, offspring survival, and juvenile birthweight and survival. Robinson et al. (2019) found evidence of severe inbreeding depression in the gray wolves (*Canis lupus*) of Isle Royale. They used population genetic simulations, comparison of inbreeding coefficients estimated from runs of homozygosity (ROH) from wolves from a variety of demographic histories, and morphological

analysis to determine that this population of wolves has undergone an increase in homozygosity of strongly deleterious recessive mutations. The use of WGS data to estimate ROH is particularly useful to both quantify inbreeding coefficients and identify causal loci for inbreeding depression (Kardos et al. 2018; Hohenlohe et al. 2020).

Population genomics tools can also be used to map loci associated with individual inbreeding or reduced fitness, although this inevitably suffers from small sample size in small wildlife populations. However, as with GWAS studies for identifying adaptive loci, mapping of loci associated with inbreeding depression or loss of fitness in small, bottlenecked populations can be facilitated by other sources of information. For example, a more powerful approach includes functional data on mutations identified across the genome from WGS data compared with well-annotated reference data (Fig. 3c; Robinson et al. 2018; Grossen et al. 2020). The functional consequences of mutations can be predicted based on where they occur in well-annotated genomic sequences, which are often available either for focal wildlife species or for close relatives.

## *2.7 Specific Threats and Adaptive Potential*

A major threat to viability and persistence of wildlife populations is their response to environmental change. Two related questions determine whether wildlife populations may be able to adapt and persist under environmental change: what is the mismatch between the genetic state of a population and future environmental conditions (genomic vulnerability; Bay et al. 2018), and how much genetic variation exists in a population to allow it to adapt to changing conditions (adaptive potential; Dawson et al. 2011; Nicotra et al. 2015; Funk et al. 2019). The ability to assess the adaptive genetic variation present in a wildlife population is a key goal of wildlife population genomics. An emerging, powerful approach combines data on adaptive genetic variation from approaches like GEA with environmental data and climate modeling (Fig. 4; Ruegg et al. 2018; Funk et al. 2019; Forester et al. 2018, this volume). This includes genetic responses to climate change, where future changes can be compared to current adaptive genetic variation across climate variables (Bay et al. 2018; Razgour et al. 2019). It could be applied in other cases where genetic responses have been observed to other human-caused environmental changes such as habitat modification, changes in the ecological community, population fragmentation, or effects on behavior (Benazzo et al. 2017). These applications can combine multiple techniques that focus on the genetic basis of particular phenotypes, in addition to fitness under environmental conditions, along with multiple sources of non-genomic information (Funk et al. 2019). This can help identify populations, regions, or protected areas that contain hotspots of adaptive genetic diversity for evolutionary response to environmental change (Mills et al. 2018).

Wildlife population can persist in the face of climate change through a combination of genetic adaptation, phenotypic plasticity, range shifts, and management

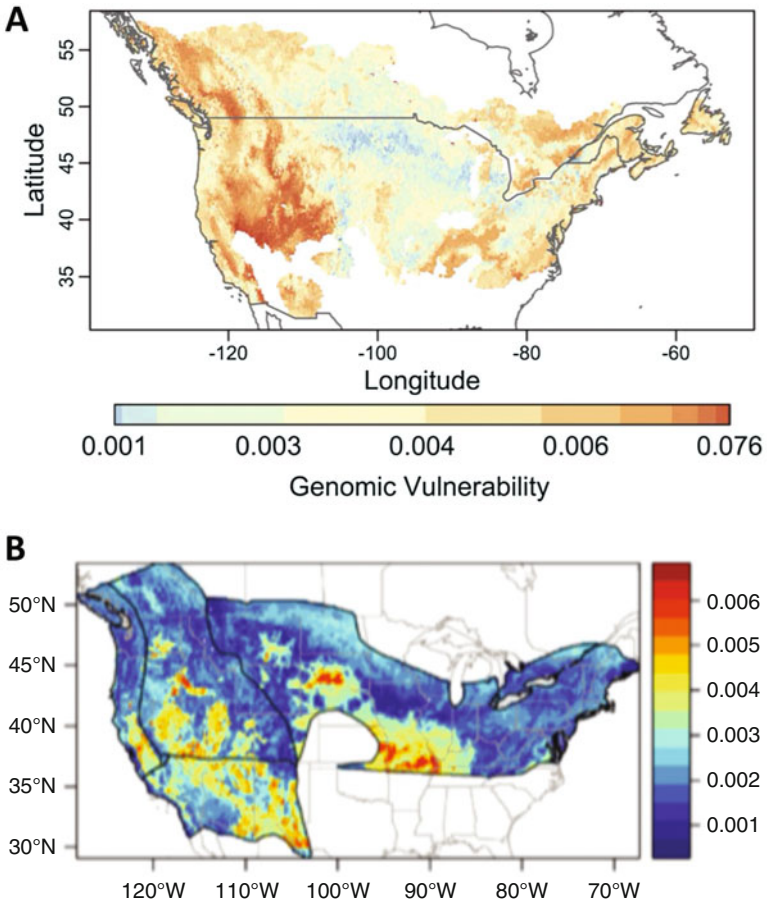


**Fig. 4** Genomic vulnerability to climate change in two North American bird species, estimated by comparing current patterns of local adaptation to climate conditions with future predictions under climate change scenarios. (**a**) In yellow warblers (*Setophaga petechia*), areas of recent population declines corresponded with areas of highest genomic vulnerability to future climate change, shown here. Reproduced from Bay et al. (2018). (**b**) In willow flycatchers (*Empidonax traillii*), climate vulnerability is high in the endangered southwestern subspecies (*E. t. extimus*), shown on the map as the southwestern portion of the range bordered by dark lines. Reproduced from Ruegg et al. (2018)

intervention. Population genomic data can be combined with phenotypic and environmental information to understand a wide range of potential ecological and evolutionary responses (Waldvogel et al. 2020). By combining analysis of local adaptation with projections of future conditions, population genomics tools can also assess the vulnerability of wildlife populations to future change (Fig. 4; Bay et al. 2018; Ruegg et al. 2018). Climate change directly interacts with a number of key phenotypes, and population genomics tools can identify the genetic basis of this adaptive variation (Razgour et al. 2019; Höglund et al. 2019, this volume). For

example, climate change affects ecological interactions with invasive species, with consequences for competition or hybridization with native wildlife populations (Chown et al. 2015). Increasing water temperatures can increase the spread of invasive rainbow trout (*Oncorhynchus mykiss*) and hybridization with native westslope cutthroat trout (*O. clarki lewisii*), with genomic consequences for the native populations (Muhlfeld et al. 2017).

Another threat to many wildlife species is disease, which can be facilitated or exacerbated by other anthropogenic influences and potentially affects large swaths of biodiversity (for example, chytridiomycosis in amphibians; Scheele et al. 2019; Funk et al. 2018, this volume). Population genomics can assess the variation that may permit wildlife population to adapt to emerging diseases (Epstein et al. 2016; Gupta et al. 2020; Auteri and Knowles 2020; Storfer et al. 2020, this volume). Population genomics can be applied to wildlife disease in multiple ways, including pathogen detection, inferring disease transmission and predicting spread, as well as assessing genetic variation for resistance (Blanchong et al. 2016). Storfer et al. (2020, this volume) assess the applications of population genomics to disease in wildlife, focusing on four case studies: colony collapse in honeybees (*Apis mellifera*), chytridiomycosis in amphibians, whitenose syndrome in bats, and transmissible cancer in Tasmanian devils (*Sarcophilus harrisii*). All four of these diseases have arisen relatively recently, have spread widely across host species, represent major threats to population persistence, and include complex interactions among hosts, pathogens, and ecological communities that can be addressed with population genomics tools. A specific group of diseases, cancer, is poorly understood in wildlife species but may have widespread impacts as a result of genetic and environmental changes (Pesavento et al. 2018; Hendricks et al. 2020, this volume). The ability of wildlife populations to withstand cancer and other diseases is closely tied to their genetic diversity, demographic history, and inbreeding, factors that are tractable with population genomic data.

## 3 Applications in Genetic Management and Conservation of Wildlife

Population genomics approaches have multiple applications to wildlife conservation and management actions (Walters and Schwartz 2020, this volume). There has been criticism of the broader field of conservation genomics and its slow pace in achieving its potential for connecting research to direct conservation action (Shafer et al. 2015; Garner et al. 2016). But within the last several years, applications of population genomics tools to wildlife provide a rapidly growing set of examples illustrating the breadth of issues that can be addressed (Table 1). While many of the published examples in Table 1 still have not been implemented in management actions, they give wildlife managers and policymakers an overview of the types of information that can inform decisions. The issues addressed include basic features of wildlife

populations that have long been confronted with genetic data, such as microsatellite loci, as well as new issues focused on the genetic basis of adaptive or deleterious traits. Different strategies of sampling, data collection, and analysis are appropriate at different scales (Fig. 1).

Population genomics can be applied simply by providing basic information about wildlife populations without intervention: for example, estimating phylogenetic relationships, delineating population units, estimating population size and genetic diversity, determining whether populations meet criteria for conservation listing, or assessing population vulnerability to threats. Ongoing monitoring can estimate trends in these features over time. Additionally, population genomics tools can also inform decision-making for more intensive management actions, such as translocations of individuals or captive breeding, and monitor the consequences of these actions after they are carried out. In all of these cases, the power of population genomics to identify both genome-wide patterns and also identify and assess loci with adaptive significance can improve the utility of genetic data for conservation and management applications in wildlife.

## *3.1 Delineating Population Units for Management*

Population genomics provides basic information on population sizes, distribution, and connectivity. This allows the delineation of conservation units for management and assessment of their size and distribution (Fig. 3b; Funk et al. 2012). It also facilitates prioritization of populations for conservation on the basis of genetic factors by quantifying the effects of current and historical population dynamics on genetic diversity, inbreeding, population fitness, and adaptive potential. For instance, some laws designed to protect endangered wildlife, such as the U.S. Endangered Species Act, take adaptive potential into consideration in endangered species listing and delisting decisions (Funk et al. 2019). As a result, different management goals rely on different criteria for delineating populations. For example, Evolutionarily Significant Units (ESUs) are defined by reproductive isolation and adaptive difference from other populations, so that an ESU represents a significant evolutionary or genetic component of the species (Funk et al. 2012). Management Units (MUs) are local populations that are demographically independent, so that management goals based on population size, such as regulation of harvest levels, may be designed at this level. Multiple MUs, defined by demographic connectivity, may be present within an ESU, defined by genetic connectivity. This can be understood in the context of basic population genetic models: demographic connectivity generally depends on the migration rate ($m$), the proportion of individuals that migrate among populations per generation, while genetic connectivity generally relies on the absolute number of individuals migrating ($Nm$, the product of population size and migration rate) (Lowe and Allendorf 2010). Specific adaptive differences may further lead to specific delineation of populations as adaptive units (Prince et al. 2017).

As an example, this framework was applied to the Iberian endemic and near-threatened Cabrera vole (*Microtus cabrerae*). This case study illustrates an important role of genomics to resolve gaps or inconsistencies from previous, smaller-scale genetic datasets. Early analysis of microsatellite and nuclear DNA sequencing data of Cabrera vole populations revealed little variation across the species distribution, contrary to mitochondrial DNA that showed a clear division in at least two genetic groups (Alasaad et al. 2013; Barbosa et al. 2017). Subsequent analysis of a subset of these samples with reduced representation genomic sequencing allowed for the identification of four ESUs, while the identification of neutral and outlier variation further led to the identification of six MUs and three adaptive units, respectively (Barbosa et al. 2018). Similar studies have also provided a better understanding of genetic population structure leading to changes in conservation listing and management, such as among rockfish species (*Sebastes* spp.; Andrews et al. 2018; Walters and Schwartz 2020, this volume).

## *3.2 Monitoring*

Genetic monitoring of wildlife populations can address a number of basic issues, including abundance, effective population size, genetic diversity, vital rates, hybridization, as well as temporal trends in all of these factors (Carroll et al. 2018; Flanagan et al. 2018; Hoban et al. 2020). Genetic monitoring in wildlife populations has often used microsatellite markers, in part because a relatively small number of loci are typically sufficient to estimate individual identity, relatedness, dispersal, and metrics of genetic diversity and population differentiation (de Barba et al. 2010). Microsatellites can be genotyped with non-invasive and low-quality DNA samples, facilitating long-term monitoring of wildlife populations (Waits and Paetkau 2005; Selkoe and Toonen 2006; de Barba et al. 2016). Genomics techniques have overtaken microsatellites in many respects, with advantages in numbers of loci, cost per sample, consistency in genotyping, as well as advances in using genomics techniques for non-invasive samples in wildlife (Hunter et al. 2018). Nonetheless, microsatellites remain a key tool for genetic monitoring in wildlife populations, particularly where a panel of microsatellite loci has been used for long-term monitoring and maintaining a consistent dataset is important for understanding long-term trends.

Monitoring requires a set of genetic markers that can be consistently genotyped across many individual samples over time, using a standardized protocol that is rapid and cost-effective. Population genomics approaches can be most effective in providing a large set of loci from which to choose an optimal set of markers that can then be rapidly genotyped using another technique (Förster et al. 2018; von Thaden et al. 2020). For instance, Förster et al. (2018) started by designing targeted capture probes from the domestic cat (*Felis catus*) reference genome and using them to gather sequence data for 809 nuclear coding regions in Eurasian lynx (*Lynx lynx*). From these sequences, they optimized a marker panel of 96 SNP loci that could be

genotyped on a high-throughput Fluidigm platform. The 96-marker panel was able to identify individuals, assign individuals to source populations, and detect population structure. In contrast to panels of microsatellite loci that are often assumed to be neutral, marker panels developed from genomic datasets can specifically include loci that have adaptive or functional significance. This allows monitoring efforts to track genetic variation at specific adaptive loci, for instance to understand population responses to environmental stress or management actions, and to identify populations that lack adaptive variation (Flanagan et al. 2018; Leroy et al. 2018). Some powerful complementary approaches in wildlife population genomics would combine monitoring of genetic diversity at particular loci with an understanding of the consequences for population viability, both in terms of functional consequences of specific alleles and population-level consequences like inbreeding depression (Robinson et al. 2018; Leroy et al. 2018; Grossen et al. 2020).

### *3.3 Genetic Management of Wild Populations*

Many wildlife populations are primarily managed by regulating harvest levels. This has genetic implications based on the resulting effective population size and potential loss of genetic variation through genetic drift in small populations (although the relationship between selective harvest and $N_e$ is complex; Kuparinen et al. 2016). Genetic monitoring of $N_e$ and levels of genetic variation can be informative, by tracking both average levels of variation across the genome and also maintenance of variation at adaptive loci. If adaptive loci are known for a harvested wildlife species, for instance through GEA tests, these should be included on genetic marker panels designed for monitoring. Still, panels should always also include a genome-wide set of loci to track average levels of genetic variation. This is because any genomic information on the genetic basis of adaptation will necessarily be incomplete, especially relative to future environmental change. Harvest levels may be set with a goal of maintaining target levels of variation, both in genome-wide average and at specific adaptive loci, to support future population persistence.

In many cases, selective harvest of wildlife populations leads to changes in particular phenotypes (Kvalnes et al. 2016). However, it is difficult to separate the effects of non-genetic factors, such as phenotypic plasticity, from genetic evolution in response to harvest (Kuparinen and Festa-Bianchet 2017). Genomic identification of loci associated with phenotypic variation, for instance with GWAS, and inclusion of these loci in monitoring panels could resolve this issue, by directly observing a response to selection at the genetic level. Some phenotypes, such as horn size (Miller et al. 2018; Sim and Coltman 2019), may be more tractable than others such as behavior (Leclerc et al. 2019). As above, harvest levels or regulations on harvest with respect to age, sex, or phenotype could be designed with the goal of maintaining genetic variation for particular phenotypes or to minimize genetic evolution in response to harvest. Alternatively, genomic monitoring of harvested populations

provides another means to identify the genetic basis of phenotypes subject to harvest-induced selection (e.g., Bowles et al. 2020).

In other cases, wildlife conservation efforts in natural populations are more intensive, involving movement of individuals among populations or reintroduction to unoccupied habitat. Individuals may be translocated into a population with the goal of genetic rescue, which is an increase in population fitness and decrease in extinction probability caused by the genetic variation added to the population. Fitzpatrick and Funk (2019, this volume) outline a variety of ways in which population genomics can help managers with decisions about genetic rescue. Genetic rescue may occur by reducing inbreeding depression via masking deleterious alleles expressed in the homozygous state, or by infusing additive genetic variation on which selection can act so that populations can adapt to changing environments (Bell et al. 2019). Genomics tools can help identify populations suffering from low genetic variation and inbreeding depression (Table 1). They can also help identify the best potential source populations that are not too adaptively divergent from the target recipient population, in order to avoid outbreeding depression, a loss of fitness caused by genetic mixing. Finally, if and when genetic rescue is implemented, genomic data can be used to monitor changes in genetic variation and the relative fitness of immigrants, residents, and hybrids to test whether gene flow is increasing fitness as desired (Miller et al. 2012; Flanagan et al. 2018; Fitzpatrick et al. 2020).

Ferchaud et al. (2018) provide a case study for using population genomics tools to quantify the genetic effects of population supplementation in lake trout (*Salvelinus namaycush*). The researchers used a reduced representation sequencing approach to genotype nearly 5,000 SNP markers in several stocked and unstocked populations. They found higher levels of neutral genetic diversity in stocked populations. They also used functional information from the related rainbow trout (*Oncorhynchus mykiss*) to identify deleterious alleles among the SNP loci that were genotyped, and found that deleterious alleles were more abundant in unstocked populations. These results suggest that supplementation not only adds genetic variation but may also improve the ability of selection to purge deleterious alleles in supplemented populations. However, the researchers also identified fixed deleterious alleles in a source population, emphasizing the role of genomic data in identifying suitable source populations for translocations.

## *3.4 Captive Breeding*

Population genomics is also being incorporated into intensive management of captive wildlife populations (Russello and Jensen 2018, this volume). Captive breeding has typically relied on pedigree-based management, but population genomics tools can provide more accurate estimates of genetic relatedness (Kardos et al. 2015) to guide breeding decisions, as well as critical information on functional genetic diversity in captive populations (Brandies et al. 2019; Russello and Jensen

2018, this volume). One example is using genomics tools to monitor and minimize genetic adaptation to captivity. Genomic data can also help determine whether management goals are being met, such as maintaining overall genetic diversity or the integrity of different ancestral population groups, or maintaining variation are specific adaptive loci (Russello and Jensen 2018, this volume). Establishment of captive populations can also have genetic effects on small wild populations from which individuals are taken. For instance, Morrison et al. (2020) used reduced representation sequencing to genotype SNPs in wild and captive populations of the Australian orange-bellied parrot (*Neophema chrysogaster*) and found that removal of half of a juvenile cohort from the wild population to supplement the captive population nearly a decade ago still shows effects on genetic diversity in the wild population. Subsequent release of captive-reared individuals has restored the level of genetic diversity in the wild population (Morrison et al. 2020). Jensen et al. (2018) compared variation at >2,000 SNPs in Pinzón giant tortoise (*Chelonoidis duncanensis*) samples from a single island in the Galápagos Island from before and after a bottleneck that reduced their population size ($N_e$) to just 150–200 in the mid twentieth century. They found that the extent and distribution of genetic variation in the historical and contemporary samples was very similar, which they attributed to a successful ex situ head-start and release program.

With population genomics tools it is possible to identify loci associated with specific phenotypic traits, fitness, or inbreeding depression. It is increasingly possible to design management of captive populations around a specific set of functionally important loci, although there are substantial pitfalls in managing captive wildlife populations for a small number of loci (Kardos and Shafer 2018). However, the possibility of efficient genotyping of individuals with relatively large genetic marker panels means that genetic management of captive populations can target multiple goals at once – for instance, maintaining variation at specific loci or keeping phenotypically distinct populations separate, while still maintaining genome-wide diversity or minimizing genome-wide inbreeding. Another goal of captive population management may include maintaining genetic adaptive potential in the face of specific threats to wild populations, such as disease (Hohenlohe et al. 2019; Storfer et al. 2020, this volume). Genotyping approaches that can be applied across both captive and wild samples (e.g., including non-invasive samples) can help integrate management of captive and natural populations of the same species (Morrison et al. 2020).

## 3.5 *Improving Connections Between Research and Applications*

Despite the potential for wildlife population genomics to address a wide range of issues directly relevant to management actions, there remain gaps between research and application (Holderegger et al. 2019; Taft et al. 2020). It is important for

researchers and practitioners to establish professional connections and to communicate at all stages of wildlife population genomics research. In this case, collaborative partnerships benefit both sides (Taft et al. 2020). Before a research project begins, communication can guide the research toward key metrics or questions needed for management decisions and allow researchers to focus on the types of information and results that would be most informative for management decisions (Holderegger et al. 2019). Conservation practitioners can also learn what types of information are available from population genomics studies, and how to interpret them and apply them to decisions. Managers may be critical in facilitating research, for example by providing samples and providing biological knowledge about wildlife populations. When a study is complete, simply publishing in a scientific journal is often not sufficient for results to be useful for guiding management (Fabian et al. 2019); again, ongoing professional contacts and efforts to communicate results to broader audiences are critical for spreading information between research and practitioner groups.

Population genomics is a challenging science, with high bars to entry particularly given the complexity of laboratory methods, bioinformatics, and data analysis. Training opportunities are critical, and training workshops that involve a mix of researchers and practitioners are most effective at establishing professional connections as described above. However, it is not necessary for everyone involved in population genomics research or using the results to be directly proficient in lab or bioinformatics methods; instead, a major goal of training opportunities should be to teach concepts that allow people to understand what information population genomics studies can provide and to interpret the results in a broader context (Holderegger et al. 2019). Nonetheless, continued efforts to make bioinformatic analysis tools more user-friendly and accessible will facilitate applications of population genomics.

Many of these recommendations for improving connections between wildlife population genomics research and applications are being followed. For instance, Taft et al. (2020) identified a large number of partnerships between researchers and practitioners already established. In part the apparent gap in population genomics results that have actually influenced wildlife management decisions reflects an unavoidable time lag. Many of the case studies highlighted throughout this chapter have not led to direct changes in management of wildlife populations, but they may still contribute to future decisions as understanding of the potential for population genomics to inform wildlife management improves. More broadly, the growing body of population genomics research in wildlife species can contribute to general conclusions about management and conservation actions. For instance, examples of genetic rescue attempts have led to emerging conclusions about the efficacy of this strategy in improving population fitness (Ralls et al. 2018, 2020; Fitzpatrick et al. 2020), which can help provide general guidelines for management decisions. Genomics can contribute to this understanding, for instance by identifying the genetic basis of increased fitness in rescue.

# 4 Approaches and Resources

## *4.1 Options and Challenges for Wildlife*

The wide range of population genomics techniques and approaches, research questions, and applications to wildlife conservation and management questions are illustrated by case studies in particular wildlife taxa (Table 1). These studies demonstrate how the diversity of population genomics techniques can be tailored to a particular study, depending on the resources available, the scientific or management question(s) being addressed, and limitations or challenges for the specific system (Matz 2018). Tools for population genomics are changing rapidly, and this includes advances at all steps in the process: from non-invasive sampling and extraction of DNA from archival or degraded samples, to library preparation protocols and sequencing platforms, to analysis pipelines and software (Luikart et al. 2019; Rajora 2019). At each step, researchers confronting the bewildering array of options should stay grounded in the scientific question(s) being asked and the suitability of any approach for the specific system, as well as how the conclusions might be used to inform a management or conservation action. The resulting choice of approaches may differ widely, and will also be constrained by the time and resources available. In addition to choices of sampling design, library preparation and sequencing, and analysis, there is a growing wealth of resources of genomic information that can be applied across species.

Planning a population genomics study is best done in an integrated way. For example, downstream analyses may require certain numbers of loci or numbers of individuals per population to increase their power to make useful inferences, and these considerations should drive sampling design. Alternatively, the availability of samples or a requirement to use non-invasive or archival samples may drive a study toward particular sequencing and analytical approaches. As an example of an increasingly useful approach, Box 1 discusses these considerations in presenting a general workflow for applying whole-genome sequencing (WGS) in wildlife population genomics studies.

**Box 1 Whole-Genome Sequencing for Wildlife Genomics: A Practical Guide**

The advances in sequencing technology and methods have made whole-genome sequencing (WGS) of multiple individuals a feasible approach for population genomics studies in wildlife. Here, we review a general workflow for WGS data, including library preparation, sequencing, and bioinformatic analysis. Further useful information on designing WGS studies and analysis pipelines is provided by Ekblom and Wolf (2014), Fuentes-Pardo and Ruzzante (2017), Pfeifer (2017), Wong et al. (2019), Bani Baker et al. (2020), and Pereira et al. (2020).

(continued)

**Box 1** (continued)
***Considerations for Library Preparation and Sequencing***

1. *Sampling of individuals* for WGS is an important consideration because often a smaller number of individuals will be sequenced compared to other approaches. For instance, if the goal is to make inferences about a population, such as demographic reconstruction (Fig. 3a), the individuals chosen should be representative of the population in their ancestry. Similarly, inadvertent WGS of an inbred individual would lead to underestimates of population-level heterozygosity or genetic diversity.
2. *The quantity and quality of the DNA* may affect your choice of library preparation and sequencing platform. Most library preparations, which are proprietary for specific sequencing platforms, are optimized for a given range of DNA quantity and quality that are typically easy to achieve using fresh or recently frozen samples. However, often in wildlife studies, samples are degraded due to various factors, such as environmental field conditions or archival storage, making them more challenging to sequence. If sample quality or quantity is lower than specified for a library preparation protocol, it can lead to extensive troubleshooting and limit strong conclusions in downstream analysis. Recent methods have been developed specifically for the use of samples with limited quantity and/or low-quality DNA. For instance, Chiou and Bergey (2018) present a method for enriching target vertebrate DNA and reducing bacterial contamination from fecal samples.
3. *DNA template amplification* with PCR is often used when only low quantities of DNA are available. However, PCR can introduce biases such as potentially removing low-abundance variants from sequenced populations, producing uneven coverage across loci, or introducing mutations into clonally amplified DNA templates that subsequently appear as variants. There are several ways to address this: (1) choose the appropriate library preparation kit given the sample quality, as above, (2) adjust the PCR protocol by minimizing the number of PCR cycles (Aird et al. 2011), and (3) remove duplicates in silico using publicly available bioinformatic tools such as Picard (http://broadinstitute.github.io/picard). Note that removing duplicates reduces overall coverage, so accounting for this filtering step is important to determine how much total sequencing effort is required.
4. *Minimum coverage and insert size* are highly dependent upon the focus of the study and sampling design. The recommended coverage for whole-genome resequencing is >30×/individual when individual-level genotype data will be used (Sims et al. 2014). Recommended coverage for pooled sequencing and ultra-low coverage genome sequencing approaches may be much lower per individual, and inferences are made at the population level

(continued)

**Box 1** (continued)

(Nielsen et al. 2011; Schlötterer et al. 2014; Wang et al. 2016). Further considerations are necessary when addressing questions using structural variants, such as insertion/deletion (indel) and inversion polymorphisms. Standard libraries with short reads (~350–550 bp insert size) are appropriate for detecting small structural variants, such as small indels and copy number variants (CNVs). The detection of large structural variants (>50 kb) such as inversions or translocations may require the use of long-read data (English et al. 2014; Chaisson et al. 2015; Sedlazeck et al. 2018; Mahmoud et al. 2019).

5. *The total sequencing effort* depends on the sequencing platform, accounting for the error rate, initial filtering, and the expected quantity of high-quality sequence data produced, in order to produce sequence data at the required coverage given the species' genome size. For instance, Illumina sequencing has relative low error rates and a multitude of options for models of sequencer, read length, number of reads per sequencing lane, number of lanes that can be run concurrently, and costs. It can be useful to distribute barcoded libraries across multiple lanes to reduce the effect of lane-to-lane variation that can occur with some sequencing platforms (Ross et al. 2013).

***Bioinformatics Workflow for Whole Genomic Sequencing***

1. *Quality filtering of raw sequence data* removes many of the errors produced during sequencing, and is facilitated by the standard *fastq* file format that contains quality scores for each nucleotide. Sequencing platforms differ widely in the error rate at individual nucleotide level, as well as other error types that may be specific to a particular technology. Regardless of the sequencing platform, some level of quality filtering of initial raw data is required (Laehnemann et al. 2016). A quality score is given to each base call by the sequencing platform using Phred scores, which is a logarithmic error probability (Ewing and Green 1998). For example, Q30 indicates that there is a 1 in 1,000 probability of calling an incorrect base (or 99.9% accuracy). Frequently there is an observable trend of decreasing quality with increasing base position, as the quality degrades after many cycles of sequencing (Kircher et al. 2009; Kircher and Kelso 2010), so trimming lower-quality ends of reads can be warranted. Additionally, residual adapter sequences, which are added during the library preparation to bind the DNA template to the sequencing platform, are removed from the ends of each read during initial filtering. Adapters and low-quality base pairs are trimmed using programs such as Trimmomatic (Bolger et al. 2014) and Cutadapt (Martin 2011). Although this trimming step reduces the total number and the length of reads, it raises the quality levels and alignment

(continued)

**Box 1** (continued)

success to a reference which are crucial for the overall success of genomic data analysis.

2. *Read alignment and mapping* typically involves aligning the sequenced fragments to a reference genome or to a *de novo* assembly depending on whether a reference genome is available:

   (a) If a reference genome is available, it can be used to map high-quality reads based on sequence similarity. Burrows–Wheeler Aligner (BWA; Li and Durbin 2009) and Bowtie2 (Langmead and Salzberg 2012) are commonly used programs to perform alignments of short-read data against a reference. It is important to understand how to optimize parameters for each algorithm to minimize alignment artifacts that can arise due to factors such as divergence between the target reads and the reference genome and misalignments around indels. Multiple reviews of alignment and mapping provide further information regarding alignment algorithm and parameter choices (Fonseca et al. 2012; Hatem et al. 2013; Reinert et al. 2015; Ye et al. 2015; Kumar et al. 2019).

   (b) *De novo* assembly involves assembling a new genome without the help of external data. Several recent reviews provide information regarding achieving high-quality *de novo* genome assembly, particularly with non-model systems (Ekblom and Wolf 2014; Koepfli et al. 2015; Phillippy 2017; Liao et al. 2019).

3. *Mapping statistics*, obtained from data provided in the SAM/BAM files that are output from alignment programs, will provide information such as the fraction of reads mapped to the reference genome and mapping quality scores (Phred-scaled), indicating the confidence that the mapping position is likely to be correct based on a combination of sequence similarity to the reference and base quality. Programs such as Qualimap2 (Okonechnikov et al. 2016) and SAMtools (Li et al. 2009a) calculate these summary statistics to help evaluate mapping quality. Further, small targeted regions can be visually assessed for alignment quality using alignment viewers such as the Broad Institute's Integrative Genomics Viewer (IGV; Robinson et al. 2011).
4. *Post-alignment filtering* is recommended to detect and correct spurious alignments and improve the quality of downstream processes such as variant calling. Unpaired reads, reads mapped to multiple positions, and mapped reads with low-quality scores should be removed. Further, local realignment particularly around indels reduces the number of misidentified variants, although newer methods have incorporated this into variant calling algorithms.

(continued)

**Box 1** (continued)

5. *Base quality score recalibration (BQSR)*, implemented in Genome Analysis Toolkit (GATK; McKenna et al. 2010; DePristo et al. 2011) helps to detect systematic errors made by the sequencer when it estimates the quality score of each base call. These non-random errors, caused by the physics or the chemistry of the sequencing reaction or manufacturing flaws of the equipment, can lead to over- or under-estimated base quality scores. These errors are modeled in BQSR by applying machine learning and then quality scores are adjusted accordingly.
6. *Variant calling* identifies sites where at least one individual differs from the reference sequence and estimate individual genotypes at all variant sites. Numerous variant caller methods have been developed, including but not limited to GATK (McKenna et al. 2010), SAMtools (Li et al. 2009a), VarScan (Koboldt et al. 2009), and SOAPsnp (Li et al. 2009b). Variant calling using GATK involves two major steps (Poplin et al. 2018). First, variant genotyping is completed per sample to create intermediate files. Second, another program, HaplotypeCaller, is run on all samples to simultaneously call SNPs and indels. This program reassembles the reads in areas showing signs of variation. HaplotypeCaller tends to be more accurate at calling variants in difficult regions such as regions that contain differing types of variants that are close to each other.
7. *Filtering of variants* with low-quality scores reduces false positives that should be removed from the dataset before downstream analyses. For systems with a large number of validated SNPs, filtering can be completed using variant quality score recalibration (VQSR; van der Auwera et al. 2013). However, often non-model systems do not have these variant databases readily available. In that case, hard filters are applied to remove false positives by detecting variants with characteristics outside their normal distributions. Appropriate choice of threshold values is a function of the data with low-quality scores, imbalanced strand specificity, and skewed allelic imbalance indicators of false positives. Hard filter thresholds can be implemented with programs such as GATK's VariantFiltration (McKenna et al. 2010; DePristo et al. 2011; van der Auwera et al. 2013) and VCFtools (Danecek et al. 2011). It is recommended to test the effects of a range of filtering thresholds particularly when applied to population genetic and demographic inferences, (Mastretta-Yanes et al. 2014; Shafer et al. 2017; Paris et al. 2017).
8. *Variant annotation*, implemented in programs such as Ensembl Variant Effect Predictor (VEP; McLaren et al. 2016) or SnpEff (Cingolani et al. 2012), is the assignment of sequence ontology terms and functional information to variants. This information can include estimates of sequence conservation, computational predictions of putative deleterious effects,

(continued)

**Box 1** (continued)
and predictions about the effect of a variant on protein structure. Variants should be considered putative polymorphisms until validated by PCR amplification and Sanger sequencing or development of a marker panel for additional genotyping. This will ensure that variants discovered are not false positives.

## *4.2 Sampling*

How many samples are required, and how they should be distributed among populations or across a landscape, varies widely depending on the goals of a study (Fig. 5). For instance, studies aiming to understand inbreeding within a population need to sample many individuals, while comparative studies across populations or taxa, such as phylogenetic analysis, may need only a single "representative" individual (Box 1). However, one advance of genomic data is that one or very few individuals can still provide a vast amount of information about a population to the extent those individuals are genetically representative of the population's history. Because each individual's genome derives from an expanding set of ancestors back in time, densely sequencing the whole genome leads to inferences about population history (Fig. 3a). This is particularly important in threatened wildlife species, where the availability of samples may be the greatest constraint on a population genomics study. However, the assumption that focal individuals are representative of a population is critical, and factors such as hidden population structure can strongly affect inferences about demographic history (Mazet et al. 2016; Gaughran et al. 2018).

Population genomics studies in wildlife often aim to use low-quality and/or low-quantity DNA samples, such as archival, environmental, and non-invasive samples collected from scat, hair, or feathers. These samples may have reduced total amounts of genetic material, DNA molecules that are fragmented or degraded, contamination from bacteria or other genetic material, or all of these issues. Andrews et al. (2018, this volume) describe the wide range of genetic and genomic techniques that can be applied in these cases. Many of the library preparation and sequencing approaches below can be optimized for low-quality samples, although others remain challenging. Environmental DNA (eDNA), which is DNA extracted from soil, water, or other environmental samples, has been used primarily for the detection of species presence, such as with species-diagnostic barcode sequences from mitochondrial DNA. Goldberg and Parsley (2020, this volume) describe the potential for eDNA approaches to be extended to population genomics studies in wildlife, in which allelic variation among individuals can be assayed from eDNA samples. This is challenging because eDNA fragments cannot be assigned to individuals, and eDNA samples may contain very few fragments of any particular locus. However, population genomics with eDNA will become more feasible as techniques improve for sequencing single DNA molecules.

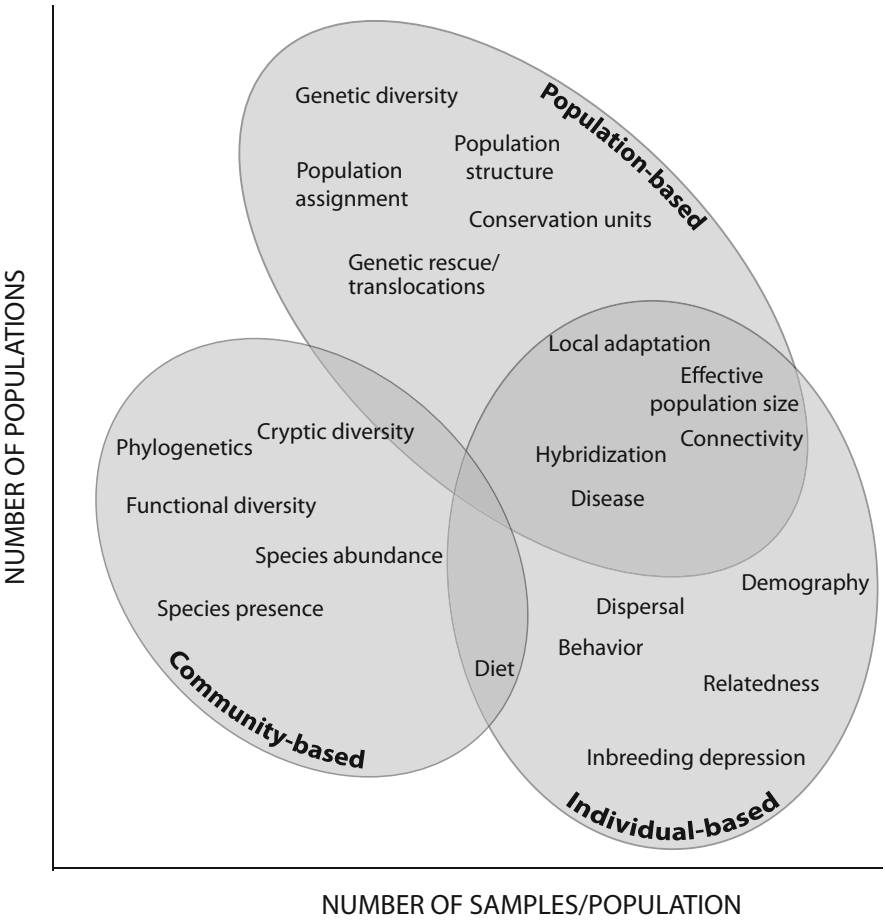


**Fig. 5** Conceptual view of the range of sampling strategies that may be appropriate to address different questions in wildlife population genomics, at different scales as shown in Fig. 1. The number of populations sampled may range from a single focal population with inbreeding or demographic questions, to a large number of populations to address landscape-level questions. Similarly, the number of individuals sampled per population may range from just a single representative of each for comparative or phylogenetic questions, to a large number of individuals to address relatedness or demography within a focal population. Additionally, the total number of individuals sampled presents a trade-off with the amount of genetic information obtained for each individual, given constraints on total sequencing cost. Many population genomics studies in wildlife may be limited by the availability of samples, so that extracting more information per individual is appropriate (e.g., whole-genome sequencing; Box 1)

Many wildlife species are also represented in museum collections, and these samples can provide insights into temporal and spatial variation in many taxa. Often, historical museum samples may represent genetic variants, populations, or even species that are no longer extant in nature (Robinson et al. 2018; van der Valk et al. 2019; Sánchez-Barreiro et al. 2020). Application of genomic methods to

museum samples can reveal how genetic variation has changed in the past and inform understanding of adaptive genetic variation that may have been lost from current populations. For example, Bi et al. (2019) generated genomic sequence data from museum specimens of two species of chipmunk (*Tamias* spp.) spanning 100 years of collection history. They were able to reconstruct demography of the two populations and also identify signatures of positive selection based on rapid shifts in allele frequency.

## *4.3 Library Preparation and Sequencing*

Population genomics in wildlife benefits from a bewildering and growing array of techniques for producing large amounts of genomic DNA sequence data (Fig. 6). Most of these are based on sequencing technologies in which heterogeneous

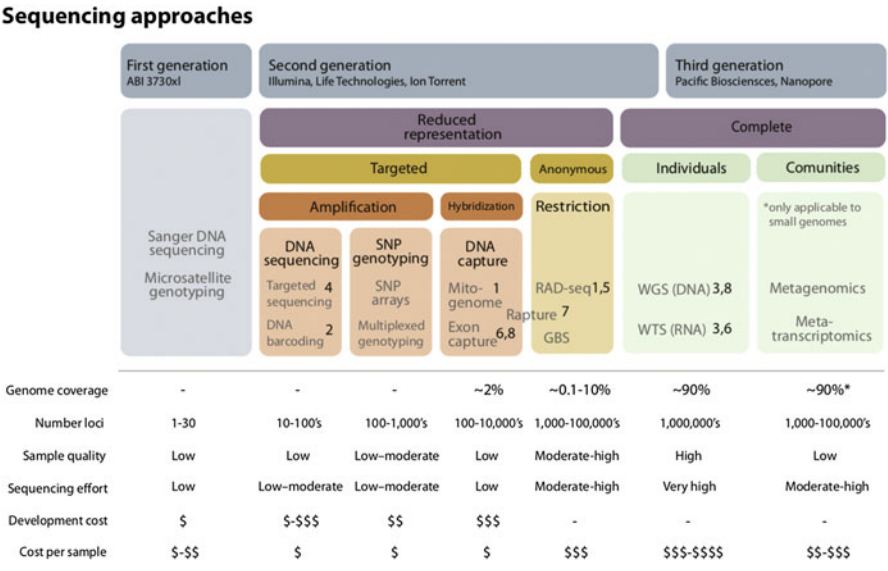


**Fig. 6** Conceptual overview of sequencing approaches for population genomics in wildlife. The top row shows sequencing technologies progressing through methods based on Sanger sequencing (first generation), short-read parallel sequencing (second generation), and long-read sequencing (third generation) (Wong et al. 2019). Genomic sequencing can cover a subset of each genome (reduced representation) or the entire genome (complete). Reduced representation techniques can be either targeted at pre-identified loci, using either amplification with primers or hybridization with probes, or they can be anonymous, for instance using restriction enzymes to survey loci across the genome. Complete genome sequencing may cover individuals or include genomic sequence from multiple individuals or species in a community. Below these groupings are example techniques, with case studies as in Fig. 1: (1) Barbosa et al. (2018); (2) Marshall and Stepien (2019); (3) Hu et al. (2020); (4) Eriksson et al. (2020); (5) Escoda et al. (2019); (6) Rellstab et al. (2019); (7) Peek et al. (2019); (8) Mills et al. (2018). Note that some techniques combine approaches: for instance, Rapture (Ali et al. 2016) combines RADseq with targeted sequence capture. At the bottom are very rough estimates to quantify some features of these techniques, as they might be applied in a wildlife study such as those in Table 1

collections of DNA molecules can be sequenced simultaneously (often called next-generation sequencing, or second- and third-generation sequencing; Heather and Chain 2016; Wong et al. 2019). As a result of these technical advances in recent decades, population genomics techniques may target thousands of loci across the genome. These loci can be either pre-selected based on prior information using capture probes or primers or anonymously distributed across the genome as a result of protocols like RADseq that use restriction enzymes (Hohenlohe et al. 2019; Holliday et al. 2019; Luikart et al. 2019). Data from RADseq are typically used as SNP genotypes, but analyzing them as microhaplotypes can provide higher-resolution data (Baetscher et al. 2018). Alternatively, WGS across a sample of individuals is now feasible even in wildlife species and is particularly well-suited for reconstructing historical demography, estimating inbreeding with runs of homozygosity, or assessing the functional significance of deleterious mutations (Table 1; Box 1). The technology of sequencing continues to provide new platforms for sequencing, including the current transition to third-generation sequencing approaches that provide continuous sequence data for long DNA fragments (Fig. 6). These technologies continue to increase the feasibility and speed of generating reference genome assemblies for wildlife species, adding to the data resources for population genomic studies.

There are important considerations before choosing the most appropriate library preparation and sequencing technique, driven by the limitations of the study system and the scientific question (Benestan et al. 2016; Hohenlohe et al. 2019). One limiting factor may be DNA sample quality. Many genomics techniques require high quality and quantity DNA samples, especially for whole-genome and whole-transcriptome sequencing (Box 1), but also for some reduced representation techniques like some RADseq methods (Andrews et al. 2016). Other techniques, including targeted sequencing with amplification primers or hybridization probes, can be effective with lower-quality DNA samples (Carroll et al. 2018; Bi et al. 2019). Progress continues in optimizing techniques for low-quality samples, including WGS and modified versions of RADseq protocols, so that these approaches are increasingly accessible as well (Russello et al. 2015; Andrews et al. 2018, this volume). New methods for isolation of target DNA prior to library construction can help as well (Chiou and Bergey 2018).

Given a total amount of resources for building libraries and sequencing, the allocation of this budget among numbers of individuals, numbers of populations, and density of sequence data and loci across the genome depends on the scientific question being addressed (Fig. 1). These trade-offs drive the choice of sequencing approach, because sequencing approaches vary widely in cost and sequencing effort per sample (Fig. 6). For example, a study of genetic population structure across a landscape like the one illustrated in Fig. 3b can be applied to identify population units for conservation purposes. This scientific question is best addressed by sampling a relatively large number of individuals distributed geographically, but analyses of population structure require a moderate number of loci. Accordingly, McCartney-Melstad et al. (2018) sampled individuals across nearly the entire species range, and used RADseq to generate data on tens of thousands of SNP loci.

Similarly, Jensen et al. (2020) genotyped 13,488 SNP markers with a RADseq protocol across 358 individuals to identify genetic population clusters in polar bears (*Ursus maritimus*). Alternatively, demographic reconstruction of historical population trends and their consequences, especially in small populations, can be accomplished with high-density WGS on a small number of individuals (Box 1). For instance, this approach was used to identify fine-scale effects of inbreeding in pumas (*Felis concolor*; Saremi et al. 2019) and wolves (*Canis lupus*; Kardos et al. 2018). At the extreme, producing a reference genome assembly for even just a single individual can reveal deep insights into population history and functional genetic variation in wildlife species (Humble et al. 2020; Upadhyay et al. 2020).

Studies focused on adaptive variation may also span these trade-offs depending on the particular question. For example, tests of local adaptation to environmental variables using GEA analysis can benefit from a relatively large number of samples distributed geographically across a wide range of environmental variables, and can still be accomplished with reduced representation approaches (Catchen et al. 2017; Forester et al. 2018, this volume). Alternatively, studies seeking to comprehensively assess the adaptive or functional variation in a wildlife species' genome may require the complete sequence data of WGS, using analyses of gene content and functional inferences about the effects of polymorphisms, rather than analyses that rely on sampling across individuals (Robinson et al. 2018, 2019). For instance, Johnson et al. (2018) produced a high-quality reference genome assembly for koalas (*Phascolarctos cinereus*). They determined that koalas' decline is likely associated with human arrival to Australia, matching the decline of Australian megafauna, and detected decreased genetic diversity in translocated populations originating from a single source population. This study also found adaptations of koalas to the toxicity of eucalyptus foliage and to chlamydia, which has had large impacts on the koala populations over the past century (Polkinghorne et al. 2013).

As costs of sequencing continue to drop, WGS is feasible for larger numbers of samples in wildlife studies (e.g. Lucena-Perez et al. 2020 used WGS on 80 individuals in their study of lynx [*Lynx lynx*] population history), and an increasing number of wildlife population genomics studies apply this technique. However, WGS remains more costly than other techniques, both in the library preparation and sequencing and computational resources to handle WGS datasets, and WGS may not be necessary to answer many questions in wildlife (McMahon et al. 2014; Lewin et al. 2018). In some cases, wildlife taxa present specific challenges for applying population genomics and limit the choice of techniques that can be applied. For instance, some amphibian taxa have remarkably large and complex genomes that may preclude WGS (Funk et al. 2018, this volume), and Weisrock et al. (2018, this volume) provide detailed recommendations for calibrating other methods including RAD sequencing, sequence capture, and amplicon sequencing in this group. More generally, reduced representation approaches will continue to be effective in cases where the scientific question requires large samples of individuals without needing large numbers of loci, and where no prior panel of markers has been developed and an approach like RADseq can be used with no prior data (Andrews et al. 2016). Similarly, traditional genetic techniques like microsatellites will continue to play a

role in wildlife research (e.g., Naude et al. 2020), and even microsatellite genotyping can be accomplished with high-throughput genomic techniques (Bradbury et al. 2018; Tibihika et al. 2019).

In many wildlife applications, it can be useful to use a combination of genomic sequencing approaches. For example, multiple sequencing techniques are commonly combined to produce reference genome assemblies (Humble et al. 2020). Combining WGS of one or a few individuals (at higher depth of coverage), and shallower resequencing of a larger set of (geographically distinct) individuals can provide a greater understanding of the processes governing phylogenetics, population structure, demographics, inbreeding and adaptation, while reducing sequencing effort (Brandies et al. 2019). In many cases it is increasingly feasible for wildlife studies to generate a reference genome assembly concurrently with reduced representation sequencing across a large number of samples, gaining the benefits of a reference genome against which to align the population-level data (Ruegg et al. 2018; Liu et al. 2019). For instance, Oyler-McCance et al. (2020, this volume) describe this approach in sage-grouse (*Centrocercus* spp.), combining WGS to infer demographic history and reduced representation sequence data to detect adaptive differentiation among populations.

As described above for monitoring and other applications, it can be efficient to use an initial dense sequencing approach such as WGS, transcriptome sequencing, or RADseq to develop a smaller panel of markers for genotyping of large numbers of samples over time (Aykanat et al. 2016; Eriksson et al. 2020). These panels can be optimized from genomic data to include adaptive or functionally significant loci (von Thaden et al. 2020). This includes drawing functional genomic information from related species to contribute to wildlife studies, such as the annotated domestic dog (*Canis familiaris*) reference genome that has been used to assess the fitness consequences of mutations in wild canid taxa (Robinson et al. 2018). Genotyping panels can also be optimized for low-quality and non-invasive samples following the initial sequencing of a few higher-quality samples (Natesh et al. 2019; Schmidt et al. 2020). These marker panels can be used to detect species presence (Janecka et al. 2020), to perform individual identification and determine individual distribution (Bourgeois et al. 2019; Giangregorio et al. 2019), to detect and quantify hybridization (Tiesmeyer et al. 2020), or to infer kinship (Escoda et al. 2019).

## *4.4 Resources*

Population genomics has developed a growing foundation of genomic data and resources that can facilitate studies in wildlife species. This includes reference genome assemblies for an increasing number of vertebrates, either wildlife species or their domestic relatives. Well-studied groups like ungulates (Martchenko et al. 2018, this volume) and birds (Toews et al. 2018, this volume) have large numbers of reference genome assemblies. In a more challenging group, Funk et al. (2018, this volume) provide recommendations for building a genome reference set across

amphibians, including at least one reference genome assembly in each amphibian family, and document progress toward this ambitious goal. Increasingly, technological advancements make it feasible to produce a high-quality reference genome assembly for nearly any wildlife species that is of interest for population genomics research (Gopalakrishnan et al. 2017; Armstrong et al. 2019; Rice and Green 2019).

Having a reference genome assembly in a population genomics study provides multiple benefits for all data types, including whole-genome resequencing across a population sample as well as any reduced representation approaches (Box 1; Rochette and Catchen 2017; Shafer et al. 2017). A reference genome allows positioning sequence reads and loci on a map, filtering of duplicate or problematic sequence, higher-confidence identification of loci, statistical analysis such as linkage disequilibrium and sliding-window analyses, identification of candidate genes near markers, and more. More broadly, the increasing number of species with genomic data allows for comparative genomics studies to better understand genomic evolutionary processes, such as changes in chromosome arrangement and recombination across birds (Toews et al. 2018, this volume), as well as phylogenetic analyses to reveal relationships among wildlife species and understand their evolutionary history (Lavretsky 2020, this volume; Ramstad and Dunning 2020, this volume).

Reference genome sequence data are maintained by several institutions that constitute the International Nucleotide Sequence Database Collaboration. They are publicly available online through GenBank of the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/Traces/wgs/), the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browse/genome-assembly-database), and the DNA DataBank of Japan (DDBJ; http://www.ddbj.nig.ac.jp/). GenBank (https://www.ncbi.nlm.nih.gov/bioproject/) and the Genomes Online Database (GOLD; https://gold.jgi.doe.gov/projects; Mukherjee et al. 2017) also provide a list of ongoing projects.

Many reference genome assemblies have also been annotated, meaning that putative genes and functional information have been inferred based on sequence similarity to known genes, transcriptomic data, gene prediction, and other analyses (Dominguez del Angel et al. 2018; Armstrong et al. 2019). As a result, a wildlife population genomics study that identifies loci that are differentiated among populations, subject to selection, or influenced by reduced diversity or hybridization can make functional inferences about these loci. For instance, Grossen et al. (2020) identified deleterious mutations in Alpine ibex (*Capra ibex*) and estimated their effects using genome annotation and functional data from related species, including gene models from the domestic goat (*C. aegagrus*). As a result, the researchers assessed the consequences of severe population bottlenecks and inbreeding on population-level fitness and genetic health of reintroduced ibex populations.

Genome annotation data is available from multiple databases, such as Ensembl (http://www.ensembl.org), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq; Pruitt et al. 2007), and the UCSC genome browser (http://genome.ucsc.edu). Functional information for gene families across species is also available from sources such as the Gene Ontology (GO) database (Ashburner et al. 2000; The Gene Ontology Consortium 2019), the Kyoto Encyclopedia of Genes and Genomes (KEGG;

Kanehisa and Goto 2000; Kanehisa et al. 2012), and the EggNOG database (Huerta-Cepas et al. 2019). Further functional information is available from protein databases such as UniProt Knowledgebase (https://www.uniprot.org; UniProt Consortium 2019) and Pfam (https://pfam.xfam.org). This type of detailed functional information is most useful in wildlife studies when population genomic data has identified a small number of candidate loci that may be important in adaptation, inbreeding depression, or population viability, and understanding the functional mechanisms is important (e.g. Waterhouse et al. 2018).

Reference genome assemblies, annotation, and functional information are also useful in wildlife studies for designing panels of markers that can be used for rapid genotyping, monitoring, or in-depth study of adaptive loci (Meek et al. 2016; Schweizer et al. 2018; Saint-Pé et al. 2019; von Thaden et al. 2020). Increasing publicly available data reduces the cost and investment needed to generate a marker panel for a wildlife species. For example, the large number of domestic ungulate species with genomic resources has translated into marker panel development for wild ungulate taxa (Martchenko et al. 2018, this volume), and genomic resources in dogs have facilitated research in wild canids (Schweizer et al. 2016). In Tasmanian devils, the Rapture approach (RADseq plus sequence capture; Ali et al. 2016) was used to design a panel of nearly 16,000 loci, most of which had putative association with devil facial tumor disease, either based on evidence of selection in response to disease (Epstein et al. 2016) or annotation to cancer or immune-related functions in the reference genome. This panel has been used to assess genetic variants associated with disease-related phenotypes (Margres et al. 2018), the genetic basis of local adaptation (Fraik et al. 2020), and selection in natural populations in response to disease (Stahlke et al. 2020) in a targeted way by genotyping thousands of individuals from natural populations.

## *4.5 Data Analysis*

Population genomic datasets are large, and so bioinformatics and data analysis will be a significant portion of any population genomics study. The bioinformatics and analysis options for genomic data continue to grow. As with the choice of sampling and sequencing approach, the most appropriate analyses depend on the scientific question. The first steps in analyzing a large genomic sequence dataset are typically initial quality filtering, which tend to be fairly similar across data types. Filtering by quality scores, trimming adaptors and low-quality sequence, de-multiplexing samples, and other initial steps are critical (tools for conducting these steps with WGS shown in Box 1 are widely applicable across sequencing types). If a reference is available, sequence reads can be mapped to the reference, and several common software packages are designed for this task (Box 1). If not, *de novo* assembly can be done with multiple tools, depending on whether the data are transcriptome (e.g., Trinity; Grabherr et al. 2011), RADseq (e.g., Stacks; Catchen et al. 2013; PyRAD, Eaton 2014), WGS (see Box 1), or other types. Often the next step will be to identify

loci, such as SNPs, and/or to genotype these loci across a set of individuals. Several software tools, including GATK (McKenna et al. 2010) and SAMtools (Li et al. 2009a), are widely applicable across data types. Others are more specific, such as Stacks (Catchen et al. 2013) written specifically for RADseq data. In other cases, population-level allele frequencies or other statistics will be estimated rather than individual genotypes, using tools such as ANGSD (Korneliussen et al. 2014).

Typically, once a set of genotypes or population-level statistics are produced, further analyses depend on the scientific question and there is a multitude of possibilities. The case studies of wildlife population genomics in Table 1 provide examples of how different analyses are applied and combined. Many of these studies have sampled individuals across populations or a landscape, and a set of basic analyses to examine genetic population structure is common. These include principal components analysis and Bayesian clustering methods, such as STRUCTURE (Pritchard et al. 2000). Phylogenetic analyses are widely used, particularly with samples across divergent populations or species, but they can also be used to separate populations into clusters (e.g., the colors in Fig. 3b represent phylogenetic clusters identified in a maximum-likelihood analysis). Estimates of effective population size can be made in several ways, depending on the sampling (e.g., whether a single or multiple time points were sampled); Fig. 2b illustrates the results from a single time point, using the linkage disequilibrium method implemented in NeEstimator (Do et al. 2014). Historic demographic reconstruction often requires more genomic data; for instance, a few methods based on the sequentially Markovian coalescent (SMC) model are commonly applied to one or a few individuals with WGS data (Fig. 3a). Although these methods require continuous sequence data based on a genome assembly, results are somewhat robust to assembly quality. For instance, reference genomes for wildlife species that remain split into tens of thousands of scaffolds may still be sufficient for inferring demographic history (Patton et al. 2019). With a reduced representation of the genome, demographic inference is still possible with approaches such as approximate Bayesian computation (e.g., Bi et al. 2019).

Several analytical approaches address functional or adaptive variation in genomic data. Multiple software tools have been developed to identify adaptive loci from population genomic datasets, based either on outlier loci or genotype–environment association (Forester et al. 2018, this volume). For example, Tigano et al. (2017) applied outlier analysis to a RADseq dataset in thick-billed murre (*Uria lomvia*) populations, using the software package Bayescan (Foll and Gaggiotti 2008). They found evidence for adaptive divergence among populations, despite no evidence for genome-wide population differentiation. Genotype–environment association can be accomplished with tools such as LFMM (Frichot et al. 2013; see also Forester et al. 2018, this volume), as applied by Ruegg et al. (2018). In cases where samples are available across several time points, adaptive loci can be detected by testing whether shifts in allele frequency at particular loci are consistent with a neutral model of drift or other demographic scenarios. For example, Stahlke et al. (2020) identified signatures of ongoing selection in Tasmanian devils using tools designed for time-series data such as spatpg (Gompert 2016). Genome-wide association studies, for

which several analytical tools have been developed for model systems including humans, can also be applied in wildlife (e.g., Margres et al. 2018), using software such as GEMMA (Zhou and Stephens 2012). With WGS data, the genetics of inbreeding can be investigated by using runs of homozygosity (ROH) (Kardos et al. 2018; Robinson et al. 2019). This method identifies the genomic regions impacted by inbreeding within individuals and can also identify whether genetic rescue from other populations may be successful based on the complementarity of ROH (Saremi et al. 2019). Much can also be learned about adaptive or deleterious loci by inferring the functional consequences of specific alleles with a variety of methods that make use of genome annotations among related taxa (Robinson et al. 2018; Grossen et al. 2020).

The analytical tools described above are a small subset of those available for wildlife population genomics. Studies will often be most successful by combining multiple approaches, drawing multiple conclusions from a genomic dataset. However, specific analyses may not be appropriate in many cases, either because assumptions of the model are violated, the analysis is not designed for a particular data type, or because the amount of data is not sufficient for statistical power. With all steps of the analysis, a critical requirement is to test the effect of parameters and settings on the results (Paris et al. 2017; Shafer et al. 2017).

## 5 Future Prospects in Wildlife Population Genomics

### *5.1 Metagenomics and eDNA*

The studies and techniques described above primarily focus on sequencing of samples from either a single individual or pool of individuals from the same population or species. As genomics tools continue to develop, wildlife population genomics may also make more use of metagenomic sequencing and metabarcoding. Metagenomic sequencing is defined as sequencing genetic material from multiple different taxa within a sample, while metabarcoding specifically refers to identifying the taxa present in a sample using sequence-based signatures, or barcodes (Taberlet et al. 2012; Luikart et al. 2019). These approaches can identify taxa in samples with low DNA quantity and quality like individual non-invasive samples (feces, hair, saliva), bulk samples (multiple individuals), or eDNA samples (Seah et al. 2020). One longstanding application of metagenomic sequencing is assessment of the microbiome – the community of microorganisms – associated with a sample. In wildlife, individual non-invasive samples have provided great insight into the role of the microbiome in adaptation and fitness, diet and diseases of wildlife populations, and even viral communities (Deagle et al. 2019; Hauffe and Barelli 2019; Roth et al. 2019; West et al. 2019; Bergner et al. 2020). Studies of eDNA have mostly focused on detecting species presence and abundance, for instance to detect cryptic or rare species or track invasive species such as Eurasian zebra and quagga mussels (Marshall and Stepien 2019). It is challenging to use eDNA to make population

genetic inferences that depend on data from multiple individuals at a set of loci, but still it has promise for population genomics applications in wildlife (Barnes and Turner 2016; Goldberg and Parsley 2020, this volume). For instance, Sigsgaard et al. (2017) produced estimates of genetic diversity in a whale shark (*Rhincodon typus*) aggregation by detecting mitochondrial DNA in seawater samples. Metagenomic sequencing of the microbial component of eDNA samples, while not directly sequencing wildlife species, can illuminate the environmental conditions in which wildlife populations exist by characterizing the functional genetic diversity of the microbiome (Seeleuthner et al. 2018).

## 5.2 Population Epigenomics

Population epigenomics is a fast-emerging area of research in population genomics (Rajora 2019; Luikart et al. 2019; Moler et al. 2019). It is now well established that epigenomic variation – alterations to genetic material that do not change DNA sequence – can contribute significantly to phenotypic plasticity, abiotic and biotic stress responses, disease conditions, and adaptation to a variety of habitat conditions (reviews in Richards et al. 2017; Moler et al. 2019). Because epigenomic variation may be inherited across generations, it could be of potential evolutionary significance. In wildlife populations, epigenomic variation may be important in the adaptive capacity of populations to respond to environmental pressures, such as climate change (Dawson et al. 2011; Nicotra et al. 2015). Recent advances in high-throughput sequencing technologies to assay genome-wide epigenetic marks, such as bisulfite DNA sequencing, have enabled the field to progress from studying individual epigenomes to investigating epigenomic variation across populations and species (Moler et al. 2019). In many wild animal populations, an abundance of epigenetic (DNA methylation) variation relative to genetic variation has been found (Hu and Barrett 2017).

The role of epigenomic variation in wildlife populations remains poorly understood, although there are some illustrative case studies. Riyahi et al. (2017) studied natural variation in DNA methylation within and among five subspecies of house sparrow (*Passer domesticus*) using the methylation-sensitive amplified polymorphism (MSAP) approach. DNA methylation was not found to be strictly subspecies-specific, but the European subspecies was differentiated from all other Middle East subspecies and the commensal subspecies was differentiated from the non-commensal species by differentially methylated regions. The methylation level was correlated with some morphological traits, such as standardized bill length. Liu et al. (2015) also applied the MSAP approach to three bat species (*Hipposideros armiger*, *Rhinolophus pusillus*, *Miniopterus fuliginosus*). The populations exhibited high epigenetic diversity and significant epigenetic structure within and among populations and individuals. The epigenetic diversity was higher than the corresponding genetic diversity. McNew et al. (2017) studied morphological, genetic, and epigenetic differences between adjacent "urban" and "rural"

populations of each of two species of Darwin's finches (*Geospiza fortis* and *G. fuliginosa*). They did not find differences in large-size copy number variation (CNV) but did find striking epigenetic (methylation) differences between the urban and rural populations of both species. Wenzel and Piertney (2014) examined epigenomic diversity and differentiation among 21 populations of red grouse (*Lagopus lagopus scotica*) in north-east Scotland and tested for association of gastrointestinal parasite load (caecal nematode *Trichostrongylus tenuis*) with hepatic genome-wide and locus-specific methylation states. The populations were found to be significantly epigenetically and genetically differentiated and displayed significant fine-scale epigenetic structure, and parasite load was associated with methylation patterns on a locus-specific, but not genome-wide level. The epigenetic differentiation observed among red grouse populations was considerably higher than genetic differentiation. This study provided an example for epigenetic mechanisms contributing to plasticity and adaptation in the context of host–parasite interactions in natural wildlife populations.

## 5.3 Population Transcriptomics

Population transcriptomics is another fast-emerging research area of population genomics (Rajora 2019; Luikart et al. 2019). Population transcriptomics uses transcriptome-wide data to study variation in gene expression within and among populations to understand mechanisms underlying acclimation and adaptation, phenotypic variation and plasticity, abiotic and biotic stress responses, adaptive evolutionary responses to new environments, and other evolutionary changes (Luikart et al. 2019). Addressing these issues in wildlife populations can be important for understanding population viability and adaptive potential in the face of environmental stressors. As discussed above, whole-transcriptome sequencing (RNASeq) can be applied to identify sequence variation at coding regions of the genome, but it can also be used to assess expression levels across genes, and it does not require prior information to target sequence effort. In animals, most of the population transcriptomics work has been conducted in fish and other aquatic organisms (Alvarez et al. 2015; Connon et al. 2018). Much of this work has addressed three questions: "(1) How much variation in gene expression is there in natural populations and how is it structured? (2) How do environmental stimuli affect gene expression? (3) How does variation in gene expression translate into phenotype?" (Alvarez et al. 2015).

Population transcriptomics research has been limited in terrestrial wildlife populations. In addition to quantifying the role of gene expression in adaptation and population differentiation, transcriptomic studies in wildlife can help understand response to disease. For instance, Campbell et al. (2018) used RNAseq to compare the gene expression profiles of frog (*Rana temporaria*) populations with a history of ranaviral disease and those without disease. They identified over four hundred transcripts that were differentially expressed between populations of different

ranaviral disease history. The differentially expressed transcripts included genes with functions related to immunity, development, protein transport, and olfactory reception. Population transcriptomics has been limited, including in wildlife, because transcriptome sequencing to estimate gene expression levels requires much more sequencing effort than that required to identify sequence differences, and also because RNAseq requires much higher sample quality than DNA sequencing approaches. However, as sequencing costs continue to drop, the feasibility of using RNAseq across individuals sampled in wildlife populations will increase. Technological developments are also likely to improve for preserving RNA from field-collected samples in wildlife.

## 6 Conclusions

The application of population genomics approaches to wildlife continues to expand. It is important for both population genomics researchers and wildlife conservation and management professionals to have an understanding of the range of approaches and questions that can be addressed in wildlife. An ongoing challenge is to improve the connections and communication among these groups. Efforts to provide venues for direct communication and interaction are critical, including cross-disciplinary training and workshops at all career levels. Research studies will benefit from coordination with wildlife professionals at all stages, from design of the questions and approach to interpretation and dissemination of results. New approaches will also emerge in the coming years, such as other "omics" techniques, the use of genetic engineering in wildlife, or approaches for multi-species or community-level genomics. Overall, population genomics provides a critical set of tools to address the biodiversity crisis in wildlife taxa. We hope this chapter provides an overview and framework to advance the field of wildlife population genomics and contribute to improving on-the-ground conservation efforts in urgent times for wildlife species.

**Acknowledgements** This work was supported by funding from a National Science Foundation Rules of Life grant (DEB 1838282) to WCF, NSF Evolutionary Ecology grant (DEB 1754821) to WCF and PAH, NSF Evolutionary Genetics grant (DEB-1655809) to PAH, and Natural Sciences and Engineering Council of Canada Discovery Grant (RGPIN 17-04589) to OPR.

## References

Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011;12:R18. https://doi.org/10.1186/gb-2011-12-2-r18.

Alasaad S, Rozhkovan KV, Atopkin DM, Sánchez A, Romero I, Soriguer RC, et al. RAPD-PCR molecular analysis of the threatened Cabrera's vole populations in the Iberian Peninsula. Afr J Biotechnol. 2013;12:4065–73.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, et al. RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics. 2016;202:389–400. https://doi.org/10.1534/genetics.115.183665.

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30. https://doi.org/10.1111/mec.13948.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709. https://doi.org/10.1038/nrg2844.

Alvarez M, Schrey AW, Richards CL. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol. 2015;24:710–25. https://doi.org/10.1111/mec.13055.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92. https://doi.org/10.1038/nrg.2015.28.

Andrews KS, Nichols KM, Elz A, Tolimieri N, Harvey CJ, Pacunski R, et al. Cooperative research sheds light on population structure and listing status of threatened and endangered rockfish species. Conserv Genet. 2018;19:865–78. https://doi.org/10.1007/s10592-018-1060-0.

Andrews KR, de Barba M, Russello MA, Waits LP. Advances in using non-invasive, archival, and environmental samples for population genomic studies. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_45.

Armstrong EE, Taylor RW, Prost S, Blinston P, van der Meer E, Madzikanda H, et al. Cost-effective assembly of the African wild dog (*Lycaon pictus*) genome using linked reads. Gigascience. 2019;8:giy124. https://doi.org/10.1093/gigascience/giy124.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9. https://doi.org/10.1038/75556.

Auteri GG, Knowles LL. Decimated little brown bats show potential for adaptive change. Sci Rep. 2020;10:3023. https://doi.org/10.1038/s41598-020-59797-4.

Aykanat T, Lindqvist M, Pritchard VL, Primmer CR. From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon *Salmo salar*. J Fish Biol. 2016;89:2658–79. https://doi.org/10.1111/jfb.13149.

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Mol Ecol Resour. 2018;18:296–305. https://doi.org/10.1111/1755-0998.12737.

Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, et al. Landscape genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 261–322.

Bani Baker Q, Hammad M, Al-Rashdan W, Jararweh Y, Al-Smadi M, Al-Zinati M. Comprehensive comparison of cloud-based NGS data analysis and alignment tools. Inform Med Unlocked. 2020;18:100296. https://doi.org/10.1016/j.imu.2020.100296.

Barbosa S, Paupério J, Searle JB, Alves PC. Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. Mol Ecol Resour. 2013;13:43–56. https://doi.org/10.1111/1755-0998.12024.

Barbosa S, Paupério J, Herman JS, Ferreira C, Pita R, Vale-Gonçalves HM, et al. Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. Mol Ecol. 2017;26:951–67. https://doi.org/10.1111/mec.13994.

Barbosa S, Mestre F, White TA, Paupério J, Alves PC, Searle JB. Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors. Mol Ecol. 2018;27:3452–65. https://doi.org/10.1111/mec.14806.

Barnes MA, Turner CR. The ecology of environmental DNA and implications for conservation genetics. Conserv Genet. 2016;17:1–17. https://doi.org/10.1007/s10592-015-0775-4.

Bay RA, Harrigan RJ, Le Underwood V, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science. 2018;359:83–6. https://doi.org/10.1126/science.aan4380.

Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc B. 1996;263:1619–26. https://doi.org/10.1098/rspb.1996.0237.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G. Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Resour. 2009;9:1279–301. https://doi.org/10.1111/j.1755-0998.2009.02699.x.

Bell DA, Robinson ZL, Funk WC, Fitzpatrick SW, Allendorf FW, Tallmon DA, et al. The exciting potential and remaining uncertainties of genetic rescue. Trends Ecol Evol. 2019;34:1070–9. https://doi.org/10.1016/j.tree.2019.06.006.

Benazzo A, Trucchi E, Cahill JA, Delser PM, Mona S, Fumagalli M, et al. Survival and divergence in a small group: the extraordinary genomic history of the endangered Apennine brown bear stragglers. Proc Natl Acad Sci U S A. 2017;114:E9589–97. https://doi.org/10.1073/pnas.1707279114.

Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25:2967–77. https://doi.org/10.1111/mec.13647.

Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, et al. Demographic and environmental drivers of metagenomic viral diversity in vampire bats. Mol Ecol. 2020;29:26–39. https://doi.org/10.1111/mec.15250.

Bi K, Linderoth T, Singhal S, Vanderpool D, Patton JL, Nielsen R, et al. Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. PLoS Genet. 2019;15:e1008119. https://doi.org/10.1371/journal.pgen.1008119.

Blanchong JA, Robinson SJ, Samuel MD, Foster JT. Application of genetics and genomics to wildlife epidemiology. J Wildl Manage. 2016;80:593–608.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bourgeois S, Kaden J, Senn H, Bunnefeld N, Jeffery KJ, Akomo-Okoue EF, et al. Improving cost-efficiency of faecal genotyping: new tools for elephant species. PLoS One. 2019;14:e0210811.

Bowles E, Marin K, Mogensen S, MacLeod P, Fraser DJ. Size reductions and genomic changes within two generations in wild walleye populations: associated with harvest? Evol Appl. 2020;13:1128–44. https://doi.org/10.1111/eva.12987.

Bradbury IR, Wringe BF, Watson B, Paterson I, Horne J, Beiko R, et al. Genotyping-by-sequencing of genome-wide microsatellite loci reveals fine-scale harvest composition in a coastal Atlantic salmon fishery. Evol App. 2018;11:918–1930. https://doi.org/10.1111/eva.12606.

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. Genes. 2019;10:846. https://doi.org/10.3390/genes10110846.

Breed MF, Harrison PA, Blyth C, Byrne M, Gaget V, Gellie NJC, et al. The potential of genomics for restoring ecosystems and biodiversity. Nat Rev Genet. 2019;20:615–28. https://doi.org/10.1038/s41576-019-0152-0.

Campbell LJ, Hammond SA, Price SJ, Sharma MD, Garner TWJ, Birol I, et al. A novel approach to wildlife transcriptomics provides evidence of disease-mediated differential expression and changes to the microbiome of amphibian populations. Mol Ecol. 2018;27:1413–27. https://doi.org/10.1111/mec.14528.

Carroll EL, Bruford MW, DeWoody JA, Leroy G, Strand A, Waits LP, et al. Genetic and genomic monitoring with minimally invasive sampling methods. Evol Appl. 2018;11:1094–119. https://doi.org/10.1111/eva.12600.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis toolset for population genomics. Mol Ecol. 2013;22:3124–40. https://doi.org/10.1111/mec.12354.

Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362–5. https://doi.org/10.1111/1755-0998.12669.

Centeno-Cuadros A, Abbasi I, Nathan R. Sex determination in the wild: a field application of loop-mediated isothermal amplification successfully determines sex across three raptor species. Mol Ecol Resour. 2017;17:153–60. https://doi.org/10.1111/1755-0998.12540.

Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 2015;517:608–11. https://doi.org/10.1038/nature13907.

Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009;10:195–205. https://doi.org/10.1038/nrg2526.

Chiou KL, Bergey CM. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep. 2018;8:1–10. https://doi.org/10.1038/s41598-018-20427-9.

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffman AA. Biological invasions, climate change and genomics. Evol Apps. 2015;8:23–46. https://doi.org/10.1111/eva.12234.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118;iso-2;iso-3. Fly (Austin). 2012;6:80–92. https://doi.org/10.4161/fly.19695.

Connon RE, Jeffries KM, Komoroske LM, Todgham AE, Fangue NA. The utility of transcriptomics in fish conservation. J Exp Biol. 2018;221:jeb148833. https://doi.org/10.1242/jeb.148833.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8. https://doi.org/10.1093/bioinformatics/btr330.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM. Beyond predictions: biodiversity conservation in a changing climate. Science. 2011;332:53–8. https://doi.org/10.1126/science.1200303.

de Barba M, Waits LP, Garton EO, Genovesi P, Randi E, Mustoni A, et al. The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. Mol Ecol. 2010;19:3938–51. https://doi.org/10.1111/j.1365-294X.2010.04791.x.

de Barba M, Miquel C, Lobréaux S, Quenette PY, Swenson JE, Taberlet P. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. Mol Ecol Resour. 2016;17:492–507. https://doi.org/10.1111/1755-0998.12594.

de Villemereuil P, Rutschmann A, Lee KD, Ewen JG, Bekke P, Santure AW. Little adaptive potential in a threatened passerine bird. Curr Biol. 2019;29:889–94. https://doi.org/10.1016/j.cub.2019.01.072.

Deagle BE, Thomas AC, McInnes JC, Clarke LJ, Vesterinen EJ, Clare EL, et al. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? Mol Ecol. 2019;28:391–406. https://doi.org/10.1111/mec.14734.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8. https://doi.org/10.1038/ng.806.

DeYoung RW, Honeycutt RL. The molecular toolbox: genetic techniques in wildlife ecology and management. J Wildl Manage. 2005;69:1362–84. https://doi.org/10.2193/0022-541X(2005)69[1362,TMTGTI]2.0.CO;2.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size ($N_e$) from genetic data. Mol Ecol Resour. 2014;14:209–14. https://doi.org/10.1111/1755-0998.12157.

Dominguez Del Angel V, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, Pettersson OV, et al. Ten steps to get started in genome assembly and annotation. F1000Res. 2018;7:148. https://doi.org/10.12688/f1000research.13598.1.

Eaton DA. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 2014;30:1844–9. https://doi.org/10.1093/bioinformatics/btu121.

Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl. 2014;7:1026–42. https://doi.org/10.1111/eva.12178.

English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. BMC Bioinformatics. 2014;15:180. https://doi.org/10.1186/1471-2105-15-180.

Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Commun. 2016;7:12684. https://doi.org/10.1038/ncomms12684.

Eriksson CE, Ruprecht J, Levi T. More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. Mol Ecol Resour. 2020; https://doi.org/10.1111/1755-0998.13208.

Escoda L, Fernández-González Á, Castresana J. Quantitative analysis of connectivity in populations of a semi-aquatic mammal using kinship categories and network assortativity. Mol Ecol Resour. 2019;19:310–26. https://doi.org/10.1111/1755-0998.12967.

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8:186–94. https://doi.org/10.1101/gr.8.3.186.

Fabian Y, Bollmann K, Brang P, Heiri C, Olschewski R, Rigling A, et al. How to close the science-practice gap in nature conservation? Information sources used by practitioners. Biol Conserv. 2019;235:93–101. https://doi.org/10.1016/j.biocon.2019.04.011.

Ferchaud A-L, Laporte M, Perrier C, Bernatchez L. Impact of supplementation on deleterious mutation distribution in an exploited salmonid. Evol Appl. 2018;11:1053–65. https://doi.org/10.1111/eva.12660.

Ferreira CM, Sabino-Marques H, Barbosa S, Costa P, Encarnação C, Alpizar-Jara R, et al. Genetic non-invasive sampling (gNIS) as a cost-effective tool for monitoring elusive small mammals. Eur J Wildl Res. 2018;64:46. https://doi.org/10.1007/s10344-018-1188-8.

Fitzpatrick SW, Funk WC. Genomics for genetic rescue. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/13836_2019_64.

Fitzpatrick SW, Bradburd GS, Kremer CT, Salerno PE, Angeloni LM. Genomic and fitness consequences of genetic rescue in wild populations. Curr Biol. 2020;30:517–22. https://doi.org/10.1016/j.cub.2019.11.062.

Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol Appl. 2018;11:1035–52. https://doi.org/10.1111/eva.12569.

Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180:977–93.

Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. Bioinformatics. 2012;28:3169–77. https://doi.org/10.1093/bioinformatics/bts605.

Forester BR, Landguth EL, Hand BK, Balkenhol N. Landscape genomics for wildlife research. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_56.

Förster DW, Bull JK, Lenz D, Autenrieth M, Paijmans JLA, Kraus RHS, et al. Targeted resequencing of coding DNA sequences for SNP discovery in nonmodel species. Mol Ecol Resour. 2018;18:1356–73. https://doi.org/10.1111/1755-0998.12924.

Fraik AK, Margres MJ, Epstein B, Barbosa S, Jones M, Hendricks S, et al. Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. Evolution. 2020;74(7):1392–408. https://doi.org/10.1111/evo.14023.

Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30:1687–99. https://doi.org/10.1093/molbev/mst063.

Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol Ecol. 2017;500:207–38. https://doi.org/10.1111/mec.14264.

Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96. https://doi.org/10.1016/j.tree.2012.05.012.

Funk WC, Forester BR, Converse SJ, Darst C, Morey S. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Cons Genet. 2019;20:115–34. https://doi.org/10.1007/s10592-018-1096-1.

Funk WC, Zamudio KR, Crawford AJ. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_61.

Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, et al. Genomics in conservation: case studies and bridging the gap between data and application. Trends Ecol Evol. 2016;31:81–3. https://doi.org/10.1016/j.tree.2015.10.009.

Gaughran SJ, Quinzin MC, Miller JM, Garrick RC, Edwards DL, Russello MA, et al. Theory, practice, and conservation in the age of genomics: the Galápagos giant tortoise as a case study. Evol Appl. 2018;11:1084–93. https://doi.org/10.1111/eva.12551.

Giangregorio P, Norman AJ, Davoli F, Spong G. Testing a new SNP-chip on the Alpine and Apennine brown bear (*Ursus arctos*) populations using non-invasive samples. Conserv Genet Res. 2019;11:355–63. https://doi.org/10.1007/s12686-018-1017-0.

Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K. Genomic quantitative genetics to study evolution in the wild. Trends Ecol Evol. 2017;32:897–908. https://doi.org/10.1016/j.tree.2017.09.004.

Goldberg CS, Parsley MB. Environmental population genomics: challenges and opportunities. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2020_75.

Gompert Z. Bayesian inference of selection in a heterogeneous environment from genetic time-series data. Mol Ecol. 2016;25:121–34. https://doi.org/10.1111/mec.13323.

Gopalakrishnan S, Castruita JA, Sinding MH, Kuderna LF, Räikkönen J, Petersen B, et al. The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. BMC Genomics. 2017;18:1–11. https://doi.org/10.1186/s12864-017-3883-3.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52. https://doi.org/10.1038/nbt.1883.

Grossen C, Biebach I, Angelone-Alasasd S, Keller LF, Croll D. Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. Evol Appl. 2018;11:123–39. https://doi.org/10.1111/eva.12490.

Grossen C, Guillaume F, Keller LF, Croll D. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat Commun. 2020;11:1001. https://doi.org/10.1038/s41467-020-14803-1.

Gupta P, Robin VV, Dharmarajan G. Towards a more healthy conservation paradigm: integrating disease and molecular ecology to aid biological conservation. J Genet. 2020;99:65. https://doi.org/10.1007/s12041-020-01225-7.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 2007;23:167–72. https://doi.org/10.1016/j.tig.2007.02.001.

Hatem A, Bozdağ D, Toland AE, Çatalyürek ÜV. Benchmarking short sequence mapping tools. BMC Bioinformatics. 2013;14:184. https://doi.org/10.1186/1471-2105-14-184.

Hauffe HC, Barelli C. Conserve the germs: the gut microbiota and adaptive potential. Conserv Genet. 2019;20:19–27. https://doi.org/10.1007/s10592-019-01150-y.

Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. Genomics. 2016;107:1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.

Hedges S, Johnson A, Ahlering M, Tyson M, Eggert LS. Accuracy, precision, and cost-effectiveness of conventional dung density and fecal DNA based survey methods to estimate Asian elephant (*Elephas maximus*) population size and structure. Biol Conserv. 2013;159:101–8. https://doi.org/10.1016/j.biocon.2012.12.010.

Hendricks SA, Storfer A, Hohenlohe PA. Population genomics of wildlife cancer. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2020_81.

Heppenheimer E, Harrigan RJ, Rutledge LY, Koepfli K-P, DeCandia AL, Brzeski KE, et al. Population genomic analysis of North American eastern wolves (*Canis lycaon*) supports their conservation priority status. Gene. 2018;9:606. https://doi.org/10.3390/genes9120606.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188:379–97. https://doi.org/10.1086/688018.

Hoban S, Bruford M, Jackson JD, Lopes-Fernandes M, Heuertz M, Hohenlohe PA, et al. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Conserv. 2020;248:108654. https://doi.org/10.1016/j.biocon.2020.108654.

Höglund J, Laurila A, Rödin-Mörch P. Population genomics and wildlife adaptation in the face of climate change. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/13836_2019_69.

Hohenlohe PA, Rutledge LY, Waits LP, Andrews KR, Adams JR, Hinton JW, et al. Comment on "Whole-genome sequence analysis shows two endemic species of North American wolf are admixtures of the coyote and gray wolf". Sci Adv. 2017;3:e1602250. https://doi.org/10.1126/sciadv.1602250.

Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population genomics provides key insights in ecology and evolution. In: Rajora OP, editor. Population genomics: concepts, approaches, and applications. Cham: Springer Nature Switzerland AG; 2019. p. 483–510.

Hohenlohe PA, McCallum HI, Jones ME, Lawrance MF, Hamede RK, Storfer A. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Cons Genet. 2019;20:81–7. https://doi.org/10.1007/s10592-019-01157-5.

Hohenlohe PA, Funk WC, Rajora OP. Population genomics for wildlife conservation and management. Authorea. 2020; https://doi.org/10.22541/au.158480040.06912807.

Holderegger R, Balkenhol N, Bolliger J, Engler JO, Gugerli F, Hochkirch A, et al. Conservation genetics: linking science with practice. Mol Ecol. 2019;28:3848–56. https://doi.org/10.1111/mec.15202.

Holliday JA, Hallerman EM, Haak DC. Genotyping and sequencing technologies in population genetics and genomics. In: Rajora OP, editor. Population genomics: concepts, approaches, and applications. Cham: Springer Nature Switzerland AG; 2019. p. 83–126.

Hu J, Barrett RDH. Epigenetics in natural animal populations. J Evol Biol. 2017;30:1612–32. https://doi.org/10.1111/jeb.13130.

Hu J-Y, Hao Z-Q, Frantz L, Wu S-F, Chen W, Jiang Y-F, et al. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. Natl Sci Rev. 2020;7:798–814. https://doi.org/10.1093/nsr/nwaa031.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14. https://doi.org/10.1093/nar/gky1085.

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM. Inbreeding depression across the lifespan in a wild mammal population. Proc Natl Acad Sci U S A. 2016;113:3585–90. https://doi.org/10.1073/pnas.1518046113.

Humble E, Dobrynin P, Senn H, Chuven J, Scott AF, Mohr DW, et al. Chromosomal-level genome assembly of the scimitar-horned oryx: insights into diversity and demography of a species extinct in the wild. Mol Ecol Resour. 2020; https://doi.org/10.1111/1755-0998.13181.

Hunter ME, Hoban SM, Bruford MW, Segelbacher G, Bernatchez L. Next-generation conservation genetics and biodiversity monitoring. Evol Appl. 2018;11:1029–34. https://doi.org/10.1111/eva.12661.

Janecka JE, Hacker C, Briderick J, Pulugulla S, Auron P, Ringling M, et al. Noninvasive genetics and genomics shed light on the status, phylogeography, and evolution of the elusive snow leopard. In: Ortega J, Maldonado JE, editors. Conservation genetics in mammals. Cham: Springer Nature Switzerland AG; 2020. p. 83–120. https://doi.org/10.1007/978-3-030-33334-8_5.

Jensen EL, Edwards DL, Garrick RC, Miller JM, Gibbs JP, Cayot LJ, et al. Population genomics through time provides insights into the consequences of decline and rapid demographic recovery through head-starting in a Galapagos giant tortoise. Evol Appl. 2018;11:1811–21. https://doi.org/10.1111/eva.12682.

Jensen EL, Tschritter C, de Groot PV, Hayward KM, Branigan M, Dyck M, et al. Canadian polar bear population structure using genome-wide markers. Ecol Evol. 2020;10:3706–14. https://doi.org/10.1002/ece3.6159.

Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation and conservation insights from the koala genome. Nat Genet. 2018;50:1102–11. https://doi.org/10.1038/s41588-018-0153-5.

Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nutile T, et al. Directional dominance on stature and cognition in diverse human populations. Nature. 2015;523:459–62. https://doi.org/10.1038/nature14618.

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40:D109–14. https://doi.org/10.1093/nar/gkr988.

Kardos M, Shafer ABA. The peril of gene-targeted conservation. Trends Ecol Evol. 2018;33:827–39. https://doi.org/10.1016/j.tree.2018.08.011.

Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. 2015;115:63–72. https://doi.org/10.1038/hdy.2015.17.

Kardos M, Åkesson M, Fountain T, Flagstad Ø, Liberg O, Olason P, et al. Genomic consequences of intensive inbreeding in an isolated wolf population. Nat Ecol Evol. 2018;2:124–31. https://doi.org/10.1038/s41559-017-0375-4.

Kircher M, Kelso J. High-throughput DNA sequencing – concepts and limitations. Bioessays. 2010;32:524–36. https://doi.org/10.1002/bies.200900181.

Kircher M, Stenzel U, Kelso J. Improved base calling for the illumina genome analyzer using machine learning strategies. Genome Biol. 2009;10:R83. https://doi.org/10.1186/gb-2009-10-8-r832009.

Kleinman-Ruiz D, Soriano L, Casas-Marce M, Szychta C, Sánchez I, Fernández J, et al. Genetic evaluation of the Iberian lynx ex situ conservation programme. Heredity. 2019;123:647–61. https://doi.org/10.1038/s41437-019-0217-z.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25:2283–5. https://doi.org/10.1093/bioinformatics/btp373.

Koepfli K-P, Paten B, O'Brien SJ. The genome 10K project: a way forward. Annu Rev Anim Biosci. 2015;3:57–111. https://doi.org/10.1146/annurev-animal-090414-014900.

Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 2014;15:356. https://doi.org/10.1186/s12859-014-0356-4.

Kovach RP, Hand BK, Hohenlohe PA, Cosart TF, Boyer MC, Neville HH, et al. Vive la résistance: genome-wide selection against introduced alleles in invasive hybrid zones. Proc B. 2016;283:20161380. https://doi.org/10.1098/rspb.2016.1380.

Kumar S, Agarwal S, Ranvijay. Fast and memory efficient approach for mapping NGS reads to a reference genome. J Bioinform Comput Biol. 2019;17:1950008. https://doi.org/10.1142/S0219720019500082.

Kuparinen A, Festa-Bianchet M. Harvest-induced evolution: insights from aquatic and terrestrial systems. Phil Trans Roy Soc B. 2017;372:20160036. https://doi.org/10.1098/rstb.2016.0036.

Kuparinen A, Hutchings JA, Waples RS. Harvest-induced evolution and effective population size. Evol Appl. 2016;9:658–72. https://doi.org/10.1111/eva.12373.

Kvalnes T, Sæther BE, Haanes H, Røed KH, Engen S, Solberg EJ. Harvest-induced phenotypic selection in an island population of moose, *Alces alces*. Evolution. 2016;70:1486–500. https://doi.org/10.1111/evo.12952.

Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data – high-throughput sequencing errors and their correction. Brief Bioinform. 2016;17:154–79. https://doi.org/10.1093/bib/bbv029.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.

Lavretsky P. Population genomics provides key insights into admixture, speciation, and evolution of closely related ducks of the mallard complex. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2020_76.

Leclerc M, Zedrosser A, Swenson JE, Pelletier F. Hunters select for behavioral traits in a large carnivore. Sci Rep. 2019;9:1–8. https://doi.org/10.1038/s41598-019-48853-3.

Leroy G, Carroll EL, Bruford MW, DeWoody JA, Strand A, Waits L, et al. Next-generation metrics for monitoring genetic erosion within populations of conservation concern. Evol Appl. 2018;11:1066–83. https://doi.org/10.1111/eva.12564.

Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome Project: sequencing life for the future of life. Proc Natl Acad Sci U S A. 2018;115:4325–33. https://doi.org/10.1073/pnas.1720115115.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009a;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, et al. SNP detection for massively parallel whole-genome resequencing. Genome Res. 2009b;19:1124–32. https://doi.org/10.1101/gr.088013.108.

Liao X, Li M, Zou Y, Yi-Pan, Wang J. Current challenges and solutions of de novo assembly. Quant Biol. 2019;7:90–109. https://doi.org/10.1007/s40484-019-0166-9.

Liu S, Sun K, Jiang T, Feng J. Natural epigenetic variation in bats and its role in evolution. J Exp Biol. 2015;218:100–6. https://doi.org/10.1242/jeb.107243.

Liu Y, Liu S, Zhang N, Chen D, Que P, Liu N, et al. Genome assembly of the common pheasant *Phasianus colchicus*: a model for speciation and ecological genomics. Genome Biol Evol. 2019;11:3326–31. https://doi.org/10.1093/gbe/evz249.

Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015;24:1031–46. https://doi.org/10.1111/mec.13100.

Lowe WH, Allendorf FW. What can genetics tell us about population connectivity? Mol Ecol. 2010;19:3038–51. https://doi.org/10.1111/j.1365-294X.2010.04688.x.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2017a;17:142–52. https://doi.org/10.1111/1755-0998.12635.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Responsible RAD: striving for best practices in population genomic studies of adaptation. Mol Ecol Resour. 2017b;17:366–9. https://doi.org/10.1111/1755-0998.12677.

Lucena-Perez M, Marmesat E, Kleinman-Ruiz D, Martínez-Cruz B, Weçek K, Saviljev AP, et al. Genomic patterns in the widespread Eurasian lynx shaped by Late Quaternary climatic fluctuations and anthropogenic impacts. Mol Ecol. 2020;29(4):812–28. https://doi.org/10.1111/mec.15366.

Luikart G, Kardos M, Hand BK, Rajora OP, Aitken SN, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches, and applications. Cham: Springer Nature Switzerland AG; 2019. p. 3–79.

Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome Biol. 2019;20:246. https://doi.org/10.1186/s13059-019-1828-7.

Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 2003;18:189–97. https://doi.org/10.1016/S0169-5347(03)00008-9.

Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24:133–41. https://doi.org/10.1016/j.tig.2007.12.007.

Margres MJ, Jones ME, Epstein B, Kerlin DH, Comte S, Fox S, et al. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. Mol Ecol. 2018;27:4189–99. https://doi.org/10.1111/mec.14853.

Marshall NT, Stepien CA. Invasion genetics from eDNA and thousands of larvae: a targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. Ecol Evol. 2019;9:3515–38. https://doi.org/10.1002/ece3.4985.

Martchenko D, Prewer E, Latch EK, Kyle CJ, Shafer ABA. Population genomics of ungulates. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_30.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2. https://doi.org/10.14806/ej.17.1.200.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. Restriction site-associated DNA sequencing, genotyping error estimation and de novoassembly optimization for population genetic inference. Mol Ecol Resour. 2014;15:28–41. https://doi.org/10.1111/1755-0998.12291.

Matz MV. Fantastic beasts and how to sequence them: ecological genomics for obscure model organisms. Trends Genet. 2018;34:121–32. https://doi.org/10.1016/j.tig.2017.11.002.

Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and a re-evaluation of human evolution. Heredity. 2016;116:362–71. https://doi.org/10.1101/031062.

McCartney-Melstad E, Gidis M, Shaffer HB. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. Heredity. 2018;121:112–25. https://doi.org/10.1038/s41437-018-0097-7.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. https://doi.org/10.1101/gr.107524.110.

McKinney GJ, Larson WA, Seeb LW, Seeb JE. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Mol Ecol Resour. 2017;17:356–61. https://doi.org/10.1111/1755-0998.12649.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17:122. https://doi.org/10.1186/s13059-016-0974-4.

McMahon BJ, Teeling EC, Höglund J. How and why should we implement genomics into conservation? Evol Appl. 2014;7:999–1007. https://doi.org/10.1111/eva.12193.

McNew SM, Beck D, Sadler-Riggleman I, Knutie SA, Koop JAH, Clayton DH, et al. Epigenetic variation between urban and rural populations of Darwin's finches. BMC Evol Biol. 2017;17:183. https://doi.org/10.1186/s12862-017-1025-9.

Meek MH, Baerwald MR, Stephens MR, Goodbla A, Miller MR, Tomalty KMH, et al. Sequencing improves our ability to study threatened migratory species: genetic population assignment in California's Central Valley Chinook salmon. Ecol Evol. 2016;6:7706–16. https://doi.org/10.1002/ece3.2493.

Miller JM, Poissant J, Hogg JT, Coltman DW. Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). Mol Ecol. 2012;21:1583–96. https://doi.org/10.1111/j.1365-294X.2011.05427.x.

Miller JM, Festa-Bianchet M, Coltman DW. Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium HD SNP BeadChip. PeerJ. 2018;6:e4364. https://doi.org/10.7717/peerj.4364.

Mills LS, Bragina EV, Kumar AV, Zimova M, Lafferty DJ, Feltner J, et al. Winter color polymorphisms identify global hot spots for evolutionary rescue from climate change. Science. 2018;359:1033–6. https://doi.org/10.1126/science.aan8097.

Moler E, Abakir A, Eleftheriou M, Johnson JS, Krutovsky KV, Lewis LC, et al. Population epigenomics: advancing understanding of phenotypic plasticity, acclimation, adaptation and diseases. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 179–260.

Morrison CE, Johnson RN, Grueber CE, Hogg CJ. Genetic impacts of conservation management actions in a critically endangered parrot species. Cons Genet. 2020;15:1–9. https://doi.org/10.1007/s10592-020-01292-4.

Muhlfeld CC, Kovach RP, Al-Chokhachy R, Amish SJ, Kershner JL, Leary RF, et al. Legacy introductions and climatic variation explain spatiotemporal patterns of invasive hybridization in a native trout. Glob Chang Biol. 2017;23:4663–74. https://doi.org/10.1111/gcb.13681.

Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. Nucleic Acids Res. 2017;45:D446–56. https://doi.org/10.1093/nar/gkw992.

Natesh M, Taylor RW, Truelove NK, Hadly EA, Palumbi SR, Petrov DA, et al. Empowering conservation practice with efficient and economical genotyping from poor quality samples. Meth Ecol Evol. 2019;10:853–9. https://doi.org/10.1111/2041-210X.13173.

Naude VN, Balme GA, O'Riain J, Hunter LTB, Fattebert J, Dickerson T, et al. Unsustainable anthropogenic mortality disrupts natal dispersal and promotes inbreeding in leopards. Ecol Evol. 2020;10(8):3605–19. https://doi.org/10.1002/ece3.6089.

Nicotra AB, Beever EA, Robertson AL, Hofmann GE, O'Leary J. Assessing the components of adaptive capacity to improve conservation and management efforts under global change. Conserv Biol. 2015;29:1268–78. https://doi.org/10.1111/cobi.12522.

Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12:443–51. https://doi.org/10.1038/nrg2986.

Nunziata SO, Weisrock DW. Estimation of contemporary effective population size and population declines using RAD sequence data. Heredity. 2018;120:196–207. https://doi.org/10.1038/s41437-017-0037-y.

Oh KP, Aldridge CL, Forbey JS, Dadabay CY, Oyler-McCance SJ. Conservation genomics in the Sagebrush Sea: population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). Genome Biol Evol. 2019;11:2023–34. https://doi.org/10.1093/gbe/evz112.

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32:292–4. https://doi.org/10.1093/bioinformatics/btv566.

Oliver MK, Lambin X, Cornulier T, Piertney SB. Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola*

*terrestris*) metapopulations. Mol Ecol. 2009;18:80–92. https://doi.org/10.1111/j.1365-294X.2008.04015.x.

Oyler-McCance SJ, Oh KP, Zimmerman SJ, Aldridge CL. The transformative impact of genomics on Sage-Grouse conservation and management. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply; 2020. https://doi.org/10.1007/13836_2019_65.

Palomares F, Godoy JA, Lopez-Bao JV, Rodríguez A, Roques S, Casas-Marce M, et al. Possible extinction vortex for a population of Iberian lynx on the verge of extirpation. Conserv Biol. 2012;26:689–97.

Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for stacks. Methods Ecol Evol. 2017;188:799–14. https://doi.org/10.1111/2041-210X.12775.

Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK. Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian devils. Mol Biol Evol. 2019;36:2906–21.

Peek RA, Bedwell M, O'Rourke SM, Goldberg C, Wengert GM, Miller MR. Hybridization between two parapatric ranid frog species in the northern Sierra Nevada, California, USA. Mol Ecol. 2019;28:4636–47.

Pereira R, Oliveira J, Sousa M. Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J Clin Med. 2020;9:132. https://doi.org/10.3390/jcm9010132.

Pesavento PA, Agnew D, Keel MK, Woolard KD. Cancer in wildlife: patterns of emergence. Nat Rev Cancer. 2018;18:646–61. https://doi.org/10.1038/s41568-018-0045-0.

Pfeifer SP. From next-generation resequencing reads to a high-quality variant data set. Heredity. 2017;118:111–24. https://doi.org/10.1038/hdy.2016.102.

Phillippy AM. New advances in sequence assembly. Genome Res. 2017;27:xi–xiii. https://doi.org/10.1101/gr.223057.117.

Polkinghorne A, Hanger J, Timms P. Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. Vet Microbiol. 2013;165:214–23. https://doi.org/10.1016/j.vetmic.2013.02.026.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178. https://doi.org/10.1101/201178.

Prates I, Xue AT, Brown JL, Alvarado-Serrano DF, Rodrigues MT, Hickerson MJ, et al. Inferring responses to climate dynamics from historical demography in neotropical forest lizards. Proc Natl Acad Sci U S A. 2016;113:7978–85. https://doi.org/10.1073/pnas.1601063113.

Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, et al. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv. 2017;3:e1603198. https://doi.org/10.1126/sciadv.1603198.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59.

Proctor MF, McLellan BN, Strobeck C, Barclay RM. Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerably small populations. Proc R Soc B Biol Sci. 2005;272:2409–16. https://doi.org/10.1098/rspb.2005.3246.

Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35:D61–5. https://doi.org/10.1093/nar/gkl842.

Rajora OP. Population genomics: concepts, approaches, and applications. Cham: Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/978-3-030-04589-0.

Ralls K, Ballou JD, Dudash MR, Eldridge MD, Fenster CB, Lacy RC, et al. Call for a paradigm shift in the genetic management of fragmented populations. Conserv Lett. 2018;11:e12412. https://doi.org/10.1111/conl.12412.

Ralls K, Sunnucks P, Lacy RC, Frankham R. Genetic rescue: a critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful

genetic variation. Biol Conserv. 2020;251:108784. https://doi.org/10.1016/j.biocon.2020.108784.

Ramstad KM, Dunning LT. Population genomics advances and opportunities in conservation of Kiwi (*Apteryx* spp.). In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2019_71.

Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibanez C, et al. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. Proc Natl Acad Sci U S A. 2019;116:10418–23. https://doi.org/10.1073/pnas.1820663116.

Reed TE, Schindler DE, Hague MJ, Patterson DA, Meir E, Waples RS, et al. Time to evolve? Potential evolutionary responses of Fraser River sockeye salmon to climate change and effects on persistence. PLoS One. 2011;6:e20380. https://doi.org/10.1371/journal.pone.0020380.

Reinert K, Langmead B, Weese D, Evers DJ. Alignment of next-generation sequencing reads. Annu Rev Genomics Hum Genet. 2015;16:133–51. https://doi.org/10.1146/annurev-genom-090413-025358.

Rellstab C, Dauphin B, Zoller S, Brodbeck S, Gugerli F. Using transcriptome sequencing and pooled exome capture to study local adaptation in the giga-genome of *Pinus cembra*. Mol Ecol Resour. 2019;19:536–51. https://doi.org/10.1111/1755-0998.12986.

Rice ES, Green RE. New approaches for genome assembly and scaffolding. Annu Rev Anim Biosci. 2019;7:17–40. https://doi.org/10.1146/annurev-animal-020518-115344.

Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colomé-Tatché M, et al. Ecological plant epigenetics: evidence from model and non-model species, and the way forward. Ecol Lett. 2017;20(12):1576–90. https://doi.org/10.1111/ele.12858.

Riyahi S, Vilatersana R, Schrey AW, Node HG, Aliabadian M, Senar JC. Natural epigenetic variation within and among six subspecies of the house sparrow, *Passer domesticus*. J Exp Biol. 2017;220:4016–23. https://doi.org/10.1242/jeb.169268.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6. https://doi.org/10.1038/nbt.1754.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Curr Biol. 2018;28:1–8. https://doi.org/10.1016/j.cub.2018.08.066.

Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv. 2019;5:eaau0757. https://doi.org/10.1126/sciadv.aau0757.

Rochette NC, Catchen JM. Deriving genotypes from RAD-seq short-read data using stacks. Nat Protoc. 2017;12:2640–59. https://doi.org/10.1038/nprot.2017.123.

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013;14:R51. https://doi.org/10.1186/gb-2013-14-5-r51.

Roth TL, Switzer A, Watanabe-Chailland M, Bik EM, Relman DA, Romick-Rosendale LE, et al. Reduced gut microbiome diversity and metabolome differences in rhinoceros species at risk for iron overload disorder. Front Microbiol. 2019;10:2291. https://doi.org/10.3389/fmicb.2019.02291.

Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, et al. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecol Lett. 2018;21:1085–96. https://doi.org/10.1111/ele.12977.

Russello MA, Jensen EL. Ex situ wildlife conservation in the age of population genomics. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_44.

Russello MA, Waterhouse MD, Etter PD, Johnson EA. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ. 2015;3:e1106. https://doi.org/10.7717/peerj.1106.

Saint-Pé K, Leitwein M, Tissot L, Poulet N, Guinand B, Berrebi P, et al. Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics. BMC Genomics. 2019;20:582. https://doi.org/10.1186/s12864-019-5958-9.

Salmona J, Heller R, Lascoux M, Shafer A. Inferring demographic history using genomic data. In: Rajora OP, editor. Population genomics: concepts, approaches, and applications. Cham: Springer Nature Switzerland AG; 2019. p. 511–38.

Sánchez-Barreiro F, Gopalakrishnan S, Westbury MV, Margaryan A, Vieira FG, Kalthoff D, et al. Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*). bioRxiv. 2020; https://doi.org/10.1101/2020.05.10.086686.

Saremi NF, Supple MA, Byrne A, Cahill JA, Coutinho LL, Dalén L, et al. Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. Nat Commun. 2019;10:4769. https://doi.org/10.1038/s41467-019-12741-1.

Scheele BC, Pasmans F, Skerratt LF, Berger L, Martel A, Beukema W, et al. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. Science. 2019;363:1459–63. https://doi.org/10.1126/science.aav0379.

Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15:749–63. https://doi.org/10.1038/nrg3803.

Schmidt DA, Campbell NR, Govindarajulu P, Larsen KW, Russello MA. Genotyping-in-thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. Mol Ecol Resour. 2020;20:114–24. https://doi.org/10.1111/1755-0998.13090.

Schwartz MK, Luikart G, Waples RS. Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol. 2007;22:25–33. https://doi.org/10.1016/j.tree.2006.08.009.

Schweizer RM, Robinson JA, Harrigan RJ, Silva P, Galverni M, Musiani M, et al. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Mol Ecol. 2016;25:357–79. https://doi.org/10.1111/mec.13467.

Schweizer RM, Durvasula A, Smith J, Vohr SH, Stahler DR, Galaverni M, et al. Natural selection and origin of a melanistic allele in North American gray wolves. Mol Biol Evol. 2018;35:1190–209. https://doi.org/10.1093/molbev/msy031.

Seah A, Lim MCW, McAloose D, Prost S, Seimon TA. MinION-based DNA barcoding of preserved and non-invasively collected wildlife samples. Genes. 2020;11:445. https://doi.org/10.3390/genes11040445.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15:461–8. https://doi.org/10.1038/s41592-018-0001-7.

Seeleuthner Y, Mondy S, Lombard V, Carradec Q, Pelletier E, Wessner M, et al. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nat Commun. 2018;9:1–10. https://doi.org/10.1038/s41467-017-02235-3.

Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett. 2006;9:615–29. https://doi.org/10.1111/j.1461-0248.2006.00889.x.

Shafer ABA, Wolf JBW, Alves PC, Bergstrom L, Bruford MW, Brannstrom I, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30:78–87. https://doi.org/10.1016/j.tree.2014.11.009.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, et al. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol. 2017;8:907–17. https://doi.org/10.1111/2041-210X.12700.

Shah VB, McRae BH. Circuitscape: a tool for landscape ecology. Proceedings of the 7th Python in Science Conference, Vol 7; 2008. p. 62–66.

Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat Ecol Evol. 2017;1:1–5. https://doi.org/10.1038/s41559-016-0004.

Sim Z, Coltman DW. Heritability of horn size in thinhorn sheep. Front Genet. 2019;10:959. https://doi.org/10.3389/fgene.2019.00959.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP, et al. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15:121–32. https://doi.org/10.1038/nrg3642.

Sinclair-Waters M, Ødegård J, Korsvoll SA, Moen T, Lien S, Primmer CR, et al. Beyond large-effect loci: large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. Genet Sel Evol. 2020;52:1–1. https://doi.org/10.1186/s12711-020-0529-8.

Sinding M-HS, Gopalakrishnan S, Vieira FG, Samaniego Castruita JA, Raundrup K, Jørgensen MPH, et al. Population genomics of grey wolves and wolf-like canids in North America. PLoS Genet. 2018;14:e1007745. https://doi.org/10.1371/journal.pgen.1007745.

Stahlke AR, Epstein B, Barbosa S, Patton A, Hendricks SA, Veillet A, et al. Historical and contemporary signatures of selection in response to transmissible cancer in the Tasmanian Devil (*Sarcophilus harrisii*). bioRxiv. 2020; https://doi.org/10.1101/2020.08.07.241885.

Steiner CC, Putnam AS, Hoeck PE, Ryder OA. Conservation genomics of threatened animal species. Ann Rev Anim Biosci. 2013;1:261–81. https://doi.org/10.1146/annurev-animal-031412-103636.

Storfer A, Kozakiewicz CP, Beer MA, Savage AE. Applications of population genomics for understanding and mitigating wildlife disease. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2020_73.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21:2045–50.

Taft HR, McCoskey DN, Miller JM, Pearson SK, Coleman MA, Fletcher NK, et al. Research–management partnerships: an opportunity to integrate genetics in conservation actions. Cons Sci Prac. 2020;2:e218. https://doi.org/10.1111/csp2.218.

Thatte P, Joshi A, Vaidyanathan S, Landguth E, Ramakrishnan U. Maintaining tiger connectivity and minimizing extinction into the next century: insights from landscape genetics and spatially-explicit simulations. Biol Conserv. 2018;218:181–91. https://doi.org/10.1016/j.biocon.2017.12.022.

The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47:D330–8. https://doi.org/10.1093/nar/gky1055.

Tibihika PD, Curto M, Dornstauder-Schrammel E, Winter S, Alemayehu E, Waidbacher H, et al. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. Cons Genet. 2019;20:357–72. https://doi.org/10.1007/s10592-018-1136-x.

Tiesmeyer A, Ramos L, Manuel Lucas J, Steyer K, Alves PC, Astaras C, et al. Range-wide patterns of human-mediated hybridisation in European wildcats. Cons Genet. 2020;21:247–60. https://doi.org/10.1007/s10592-019-01247-4.

Tigano A, Shultz AJ, Edwards SV, Robertson GJ, Friesen VL. Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. Ecol Evol. 2017;7:2370–81. https://doi.org/10.1002/ece3.2819.

Toews DPL, Walsh J, Campagna L. Population genomics of birds: evolutionary history and conservation. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_33.

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15. https://doi.org/10.1093/nar/gky1049.

Upadhyay M, Hauser A, Kunz E, Krebs S, Blum H, Dotsev A, et al. The first draft genome assembly of Snow sheep (*Ovis nivicola*). Genome Biol Evol. 2020;12:1330–6. https://doi.org/10.1093/gbe/evaa124.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1. https://doi.org/10.1002/0471250953.bi1110s43.

van der Valk T, Díez-del-Molino D, Marques-Bonet T, Guschanski K, Dalén L. Historic genomes reveal the genomic consequences of recent population decline in eastern gorillas. Curr Biol. 2019;29:165–70.

vonHoldt BM, Cahill JA, Fan Z, Gronau I, Robinson J, Pollinger JP, et al. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. Sci Adv. 2016;2:e1501714. https://doi.org/10.1126/sciadv.1501714.

vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY. Redefining the role of admixture and genomics in species conservation. Conserv Lett. 2018;11:e12371. https://doi.org/10.1111/conl.12371.

von Thaden A, Nowak C, Tiesmeyer A, Reiners TE, Alves PC, Lyons LA, et al. Applying genomic data in wildlife monitoring: development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. Mol Ecol. 2020;20:662–80. https://doi.org/10.1111/1755-0998.13136.

Waits LP, Paetkau D. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. J Wildl Manage. 2005;69:1419–33. https://doi.org/10.2193/0022-541X(2005)69[1419,NGSTFW]2.0.CO;2.

Waits LP, Cushman SA, Spear SF. Applications of landscape genetics to connectivity research in terrestrial animals. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics concepts, methods, applications. Oxford: Wiley-Blackwell; 2016. p. 199–219. https://doi.org/10.1002/9781118525258.ch12.

Waldvogel AM, Feldmeyer B, Rolshausen G, Exposito-Alonso M, Rellstab C, Kofler R, et al. Evolutionary genomics can improve prediction of species' responses to climate change. Evol Lett. 2020;4:4–18. https://doi.org/10.1002/evl3.154.

Walters AD, Schwartz MK. Population genomics for the management of wild vertebrate populations. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply; 2020. https://doi.org/10.1007/13836_2020_72.

Wang J, Skoog T, Einarsdottir E, Kaartokallio T, Laivuori H, Grauers A, et al. Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. Sci Rep. 2016;6:33256. https://doi.org/10.1038/srep33256.

Waples RS, Lindley ST. Genomics and conservation units: the genetic basis of adult migration timing in Pacific salmonids. Evol Appl. 2018;11:1518–26. https://doi.org/10.1111/eva.12687.

Waples RS, Kays R, Fredrickson RJ, Pacifici K, Mills LS. Is the red wolf a listable unit under the US Endangered Species Act? J Hered. 2018;109:585–97. https://doi.org/10.1093/jhered/esy020.

Waples RS, Naish KA, Primmer CR. Conservation and management of salmon in the age of genomics. Annu Rev Anim Biosci. 2020;8:117–43. https://doi.org/10.1146/annurev-animal-021419-083617.

Waterhouse MD, Erb LP, Beever EA, Russello MA. Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. Mol Ecol. 2018;27:2512–28. https://doi.org/10.1111/mec.14701.

Watsa M, Erkenswick GA, Pomerantz A, Prost S. Portable sequencing as a teaching tool in conservation and biodiversity research. PLoS Biol. 2020;18:e3000667. https://doi.org/10.1371/journal.pbio.3000667.

Weisrock DW, Hime PM, Nunziata SO, Jones KS, Murphy MO, Hotaling S, et al. Surmounting the large-genome "problem" for genomic data generation in salamanders. In: Hohenlohe PA, Rajora

OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_36.

Wellenreuther M, Hansson B. Detecting polygenic evolution: problems, pitfalls, and promises. Trends Genet. 2016;32:155–64. https://doi.org/10.1016/j.tig.2015.12.004.

Wenzel MA, Piertney SB. Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (*Lagopus lagopus scotica*). Mol Ecol. 2014;23:4256–73. https://doi.org/10.1111/mec.12833.

West AG, Waite DW, Deines P, Bourne DG, Digby A, McKenzie VJ, et al. The microbiome in threatened species conservation. Biol Conserv. 2019;229:85–98. https://doi.org/10.1016/j.biocon.2018.11.016.

Williams M, O'Grady J, Ball B, Carlsson J, Eyto E, McGinnity P, et al. The application of CRISPR-Cas for single species identification from environmental DNA. Mol Ecol Resour. 2019;19:1106–14. https://doi.org/10.1111/1755-0998.13045.

Wong K-C, Zhang J, Yan S, Li X, Lin Q, Kwong S, et al. DNA sequencing technologies: sequencing data protocols and bioinformatics tools. ACM Comput Surv. 2019;52(5):1–30. https://doi.org/10.1145/3340286.

Ye H, Meehan J, Tong W, Hong H. Alignment of short reads: a crucial step for application of next-generation sequencing data in precision medicine. Pharmaceutics. 2015;7:523–41. https://doi.org/10.3390/pharmaceutics7040523.

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat Genet. 2013;45:67–71. https://doi.org/10.1038/ng.2494.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4. https://doi.org/10.1038/ng.2310.

Zimmerman SJ, Aldridge CL, Oyler-McCance SJ. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. BMC Genomics. 2020;21:382. https://doi.org/10.1186/s12864-020-06783-9.

# Part II
# Wildlife Sampling and Genomics Technologies

# Advances in Using Non-invasive, Archival, and Environmental Samples for Population Genomic Studies


**Kimberly R. Andrews, Marta De Barba, Michael A. Russello, and Lisette P. Waits**

**Abstract** Recent advances in DNA sequencing and genotyping technologies are rapidly building our capacity to address ecological, evolutionary, and conservation questions for wildlife species. However, a large portion of wildlife genetic research relies on samples containing low quantities and quality of DNA, such as non-invasive, archival, and environmental DNA (eDNA) samples. These samples present unique methodological challenges that are largely responsible for a lag in the adoption of new genetic technologies for many areas of wildlife research. Nonetheless, steady progress is being made as researchers test and refine laboratory protocols and bioinformatic methods tailored to low-quality samples. Here we provide an overview of the progress toward low-quality sample applications for amplicon sequencing, single nucleotide polymorphism (SNP) genotyping, DNA capture, mitogenome sequencing, restriction site-associated DNA sequencing (RADseq), and whole-genome sequencing. We also review methods for generating DNA sequence data from samples comprised of multiple individuals and species, such as eDNA or fecal samples, including metagenome sequencing, metabarcoding, metagenome skimming, and metatranscriptomics. The implementation of these

K. R. Andrews (✉)
Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID, USA

Genetics and Genomics Group, University of Washington JISAO and NOAA Pacific Marine Environmental Lab, Seattle, WA, USA
e-mail: krandrew@uw.edu

M. De Barba
Laboratoire d'Ecologie Alpine (LECA), CNRS, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, Grenoble, France

M. A. Russello
Department of Biology, University of British Columbia, Kelowna, BC, Canada

L. P. Waits
Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_45,

approaches has provided insight into a wide range of questions such as modern and historic population genetic structure and diversity, adaptation, inbreeding, ancient hybridization, occupancy, diet composition, microbiome composition, and many more. As the development of methods tailored for low-quality DNA sources continues to advance over the coming years, we expect these samples to provide unprecedented insight into the ecology, evolution, and conservation of wildlife species.

**Keywords** Ancient DNA · Historical DNA · Metagenomics · Museum samples · Next-generation sequencing

## 1 Introduction

The development of "next-generation" DNA sequencing (NGS) and new genotyping technologies over the last decade has dramatically increased our ability to obtain genetic data for wildlife species. This growing capacity to mine the genomes of non-model organisms has both expanded and refined questions that can be addressed in wildlife ecology, evolution, and conservation. For example, population genomic data provide new opportunities for detecting natural selection and quantifying adaptive genetic variation in natural populations (Schoville et al. 2012) while simultaneously increasing the accuracy and precision of estimates of genome-wide diversity (Väli et al. 2008), population structure, and demographic parameters (Luikart et al. 2003). The application of genomic information has further driven the development of novel approaches for delineating conservation units (Funk et al. 2012), improving wildlife management strategies (Hoffmann et al. 2015), and increasing phylogenetic resolution (Jarvis et al. 2014; Wagner et al. 2013). Furthermore, these new technologies also facilitate the characterization of entire species assemblages from environmental samples (Andersen et al. 2012; Jørgensen et al. 2012).

Despite the increasing availability of new techniques for generating genomic information, the ability to obtain a source of DNA to generate these data can be limited for many wildlife species due to logistical and ethical constraints on tissue sampling. These challenges have been addressed over the past 25 years by optimizing laboratory and data analysis techniques for generating genetic data from non-invasively collected materials, such as feces, hair, and feathers, as well as archival (e.g., museum specimens and subfossils) and environmental samples (Goldberg et al. 2015; Orlando et al. 2015; Waits and Paetkau 2005). However, all of these types of samples often contain DNA of low quantity and poor quality (Pääbo 1989; Shapiro and Hofreiter 2012), as well as DNA from non-target species, all of which can pose challenges for generating genetic and genomic data.

Here, we briefly review the history of non-invasively collected, archival, and environmental sources of DNA for addressing questions in wildlife ecology, evolution, and conservation and highlight the challenges involved in adapting these sampling strategies in the genomics era. We then discuss a variety of NGS-enabled approaches to studying wildlife genetics and genomics with a specific focus on their application to non-invasively collected, archival, and environmental sources of DNA. Throughout, we highlight important considerations for employing the various approaches and explore novel opportunities in light of current and future advances.

## 2 Non-invasive, Archival, and Environmental Sources of DNA for Wildlife Genetics

The field of genetic non-invasive sampling (gNIS) has revolutionized wildlife ecology and management. This field was launched in 1992 when researchers demonstrated that it was possible to obtain and amplify mitochondrial DNA (mtDNA) from hair and fecal samples of brown bears (Höss et al. 1992; Taberlet and Bouvet 1992) and hair samples of chimpanzees (Morin and Woodruff 1992). Over time, researchers expanded the application of gNIS to include feathers (Morin et al. 1994a; Taberlet and Bouvet 1991), egg shells (Pearce et al. 1997), urine (Nota and Takenaka 1999; Valiere and Taberlet 2000), saliva (Wheat et al. 2016; Williams et al. 2003), sloughed skin (Bricker 1996; Valsecchi et al. 1998), insect exuviae (Feinstein 2004), owl pellets (Taberlet and Fumagalli 1996), and other regurgitates (Sugiyama et al. 1993; Valiere et al. 2003). More recently, researchers demonstrated it was possible to obtain bullfrog DNA from water samples (Ficetola et al. 2008). This work launched the new field of environmental DNA (eDNA) analysis which uses samples such as water, soil, and air to target DNA fragments that have been shed by organisms into the surrounding environment (Goldberg et al. 2015; Pedersen et al. 2015; Thomsen and Willerslev 2015). Both gNIS and eDNA analyses have become the genetic sampling and monitoring method of choice for many species because they provide valuable genetic information without catching, handling, or, in some cases, even observing animals (Box 1; Beja-Pereira et al. 2009; Bohmann et al. 2014; Rees et al. 2014; Schwartz et al. 2007; Waits and Paetkau 2005). Also, multiple studies have demonstrated that gNIS and eDNA analyses are more cost-effective than traditional sampling methods (De Barba et al. 2010; Jerde et al. 2011; Solberg et al. 2006; Stenglein et al. 2010).

**Box 1 Genetic non-invasive samples (gNIS), archival samples, and eDNA samples have been used to address a wide range of research questions, including those described below**

*Detection of taxa*

An important application for eDNA samples and gNIS is the detection of target taxa (Waits and Paetkau 2005; Bohmann et al. 2014). These samples can be used to detect the presence of rare, elusive, or invasive species; for example, eDNA approaches have been effective for early detection of invasive species such as Asian carp (Jerde et al. 2011), American bullfrog (Dejean et al. 2012), New Zealand mud snails (Goldberg et al. 2013), and Burmese python (Piaggio et al. 2014). eDNA and gNIS can also be used to conduct biodiversity surveys for both contemporary and ancient environments (Pedersen et al. 2015; Thomsen and Willerslev 2015), detect disease (Kohn and Wayne 1997; Kolby et al. 2015), identify factors influencing occupancy and distribution patterns (Lonsinger et al. 2017; Pansu et al. 2015), and characterize species composition of dietary samples (De Barba et al. 2014; Höss et al. 1992; Taberlet and Fumagalli 1996) and microbial communities (Amato et al. 2013).

*Detection of individuals*

gNIS can be used to identify the presence of individuals across space and time, providing valuable insights into demographic parameters such as individual-level movement patterns (Dixon et al. 2006; Fabbri et al. 2007; Proctor et al. 2005), population size and density (Arandjelovic et al. 2010; Davidson et al. 2014; Kendall et al. 2009; Woods et al. 1999), survival (Rudnick et al. 2005; Woodruff et al. 2016), and sex ratios (Rudnick et al. 2005; Woods et al. 1999). For example, DeMay et al. (2017) used DNA extracted from fecal samples collected from a Columbia Basin pygmy rabbit reintroduction site to estimate dispersal distance and survival rates of reintroduced individuals. Individual identification from gNIS can also be used to identify individual predators responsible for killing domestic animals (Blejwas et al. 2006; Caniglia et al. 2014; Williams et al. 2003) or wildlife species of conservation concern (Ernest et al. 2002; Mumma et al. 2014; Wheat et al. 2016).

*Population genetic analysis*

Population genetic analyses using gNIS and archival samples have been used to characterize modern and historic population structure, diversity, and gene flow (Bi et al. 2013; Epps et al. 2005; Miller and Waits 2003; Quemere et al. 2010; Wultsch et al. 2016). For example, genetic analysis of modern and museum Tasmanian devil specimens indicated genetic diversity has been low for at least 100 years and has not declined as a result of recent disease-related population bottlenecks (Miller et al. 2011). gNIS and archival samples have also been used to identify environmental variables driving population structure

(continued)

**Box 1** (continued)

(Braunisch et al. 2010; Castillo et al. 2014; Cushman et al. 2006; Martinez-Cruz et al. 2007; Wasserman et al. 2010), estimate effective population size (Eggert et al. 2003; Gonzalez et al. 2016; Miller and Kapuscinski 1997), assess recent hybridization and introgression (Adams et al. 2003; Lawson et al. 2017; Steyer et al. 2016), and identify related individuals (Constable et al. 2001; Ford et al. 2011; Morin et al. 1994b; Rudnick et al. 2005). For example, genetic parentage analysis with non-invasive hair and fecal samples collected over 7 years from a reintroduced brown bear population revealed that although population size increased rapidly after reintroduction, reproduction was dominated by a single male, leading to inbreeding and diversity declines (De Barba et al. 2010).

*Phylogenetic analysis*
Phylogenetic analysis using gNIS and archival samples has been used to characterize phylogenetic relationships (Guschanski et al. 2013; Willerslev et al. 2009), assess ancient hybridization (Cahill et al. 2013; Miller et al. 2012; Prufer et al. 2014), and understand transmission pathways of disease (Bos et al. 2011; Schuenemann et al. 2011). For example, phylogenetic analysis of modern, historical, and ancient polar bear and brown bear samples revealed ancient hybridization between these two species (Cahill et al. 2013; Miller et al. 2012). Phylogenetic analysis is also used for eDNA analysis to identify taxa (e.g., Fonseca et al. 2017; Klymus et al. 2017).

Archival specimens represent another important DNA source that does not require traditional sampling and can provide an invaluable resource for reconstructing patterns and processes of evolution across time and space (Wandeler et al. 2007). Such specimens can take the form of preserved hard or soft tissues collected within the last ~200 years for natural history collections (Wandeler et al. 2007), yielding what has been generally termed "historical DNA" (Bouzat et al. 1998). A variety of tissue types have been used as sources of historical DNA including bones (Russello et al. 2005), teeth (Wandeler et al. 2003), skins (Mundy et al. 1997), claw pulp (Casas-Marce et al. 2010), baleen (Rosenbaum et al. 1997), feathers (Ellegren 1991), and fish scales (Nielsen et al. 1999). The quality and quantity of historical DNA can vary by tissue type, preservation method, and storage conditions (Martínkova and Searle 2006; Morin et al. 2007). On a deeper time frame, subfossils, permafrost-preserved specimens, and coprolites can yield "ancient DNA" that is typically $<$100,000 years old (Lindahl 1993) but can extend up to 700,000 years depending upon state of preservation (Orlando et al. 2013). In the pre-NGS era, population-level studies using archival DNA largely relied on conventional markers, such as fragments of the mitochondrial genome and nuclear microsatellite loci, and, later, targeted single nucleotide polymorphisms (SNPs, Morin et al. 2004), to investigate a range of questions in wildlife ecology, evolution, and conservation (Box 1).

The main challenges associated with obtaining genetic data from non-invasively collected, archival, and environmental sources are the low quantity and quality of target DNA present, as well as contamination by "exogenous" DNA, either from non-target species or, in the case of archival specimens, modern DNA. In terms of quality and quantity of recovered DNA, degradation from these sources typically results in fragmentation to about 100–500 bp, with the majority of strand breakage occurring shortly after sampling, shedding, or death as a result of enzymatic and nonenzymatic processes (Hofreiter et al. 2001; Pääbo 1989; Pääbo et al. 2004). The extent of degradation can be further influenced by the environmental conditions at the time of collection (e.g., temperature, pH, humidity, exposure to UV radiation; Barnes et al. 2014; Poinar et al. 1996) and preservation method (Frantzen et al. 1998; Rees et al. 2014). For archival specimens, postmortem DNA damage can be particularly significant, leading to strand breaks, DNA cross-links, and oxidative and hydrolytic lesions that can all impact DNA quality (Pääbo et al. 2004). Co-purification of elements (e.g., aluminum, copper, chromium) or other compounds not thoroughly removed during the DNA extraction process can also bind to active sites and inhibit PCR, the magnitude of which can be influenced by age of specimen and preservation method (Hall et al. 1997). Contamination is another significant concern when working with non-invasively collected, archival, and environmental DNA samples. Ancient samples and contemporary fecal samples are often comprised of more than 95% microbial DNA (Carpenter et al. 2013; Chiou and Bergey 2018; Perry et al. 2010; Snyder-Mackler et al. 2016).

## 3 New Sequencing and SNP Genotyping Technologies

The last decade has seen dramatic advances in DNA sequencing technologies. Whereas Sanger sequencing was the dominant technology for three decades starting in 1977 (Sanger et al. 1977), a diverse array of new technologies started appearing in 2005, each capable of generating orders of magnitude more data than Sanger sequencing. These new technologies have been broadly described as "massively parallel sequencing," "high-throughput sequencing," and "next-generation sequencing;" hereafter we adopt the widely used term "next-generation sequencing" (NGS). Illumina currently dominates the market with short-read "sequencing by synthesis" technology that produces reads up to 300 bp (Metzker 2010; Shendure and Ji 2008), whereas long-read technologies can sequence up to 60 kb with Pacific Biosciences single-molecule real-time sequencing (Eid et al. 2009) and 200 kb with Oxford Nanopore technologies (Goodwin et al. 2015). These technologies have opened up a wide range of new applications for exploring the genomes of both model and non-model organisms (Fig. 1).

A variety of SNP genotyping platforms have also been developed over the last decade. Rather than producing DNA sequence reads, these platforms output genotypes from a preselected set of SNPs for large numbers of samples, with lower cost, effort, and genotyping error rate than NGS (Fig. 2). Single-tube, single-locus SNP

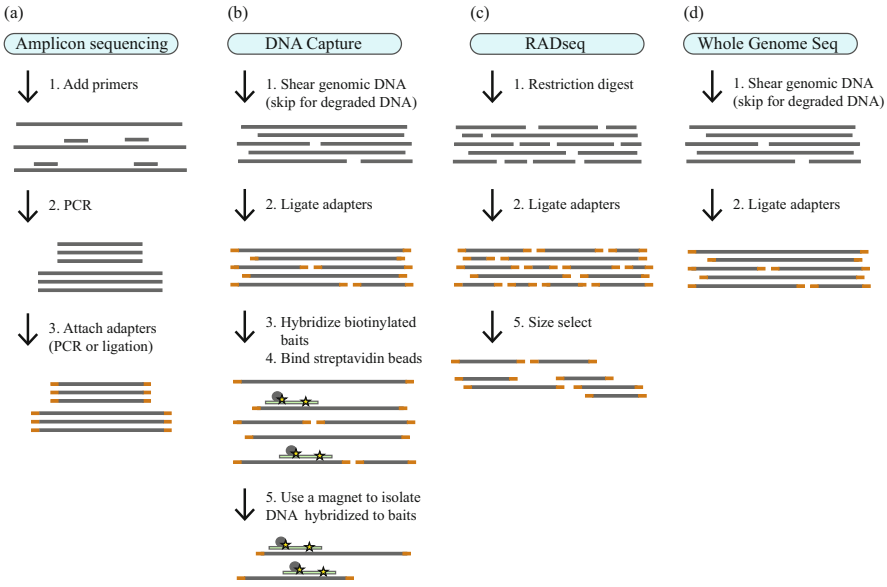


**Fig. 1** Basic steps of four approaches for producing DNA libraries ready for next-generation sequencing: (**a**) Amplicon sequencing, (**b**) DNA capture, (**c**) RADseq, and (**d**) whole-genome sequencing. All methods start with extracted genomic DNA. Orange indicates sequencing adapters. For DNA capture, stars represent biotinylation of baits, and gray circles represent streptavidin-coated magnetic beads bound to biotinylated baits. One RADseq method is illustrated, but many types of RADseq methods have been developed (reviewed in Andrews et al. 2016)

genotyping assays such as TaqMan™ (Higuchi et al. 1993; Holland et al. 1991) and Amplifluor® (Nazarenko et al. 1997) have been used for more than two decades. More recent technologies interrogate hundreds or thousands of loci and individuals simultaneously using a wide range of techniques (Garvin et al. 2010; Ragoussis 2009). SNP genotyping platforms require prior genomic knowledge to design primers and probes, which can be a disadvantage for wildlife species with limited genomic resources available. However, NGS is providing greater access to genomic information in non-model species, thus indirectly contributing to increased use of SNP genotyping platforms in wildlife research.

Non-invasive, archival, and environmental samples present limitations to all methods for generating genetic data but in particular for approaches generating non-targeted data from large numbers of loci. Therefore low-quality samples have limited use under standard protocols for many of the new DNA sequencing and genotyping technologies. However, a number of approaches have been developed to circumvent the limitations of low-quality sequence data for these new technologies, as we describe in the following sections. We start by describing techniques for generating sequence and genotype data for individual target organisms, beginning with the most feasible techniques for low-quality samples. We then describe techniques for generating data from samples comprised of communities of individuals, such as eDNA samples.

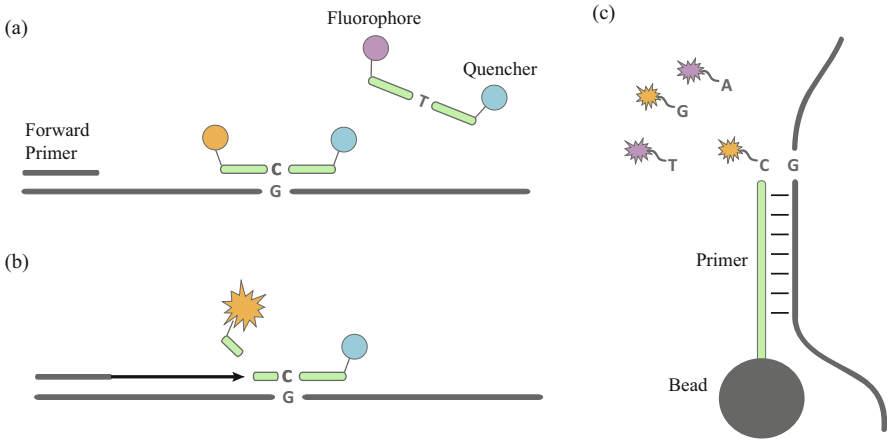


**Fig. 2** Two examples of the many methods employed by SNP genotyping assays. In contrast to next-generation sequencing methods, these assays do not produce sequence reads but instead produce genotypes at a preselected set of SNPs. (**a**, **b**) TaqMan™ assays can be used on Fluidigm Dynamic Arrays. Each assay uses two probes that are complementary to one of the two possible nucleotides at the target SNP. Each probe has a fluorophore and a quencher molecule attached; the quencher inhibits fluorescent signal when in close proximity with the fluorophore. The complementary probe anneals to the sample DNA (**a**). During PCR, Taq polymerase extends the primer and degrades the annealed probe, breaking the association between the fluorophore and the quencher and resulting in a fluorescent signal that is used for genotyping (**b**). (**c**) The primer extension method (also called single-base extension method) is used by iPlex® assays on the MassARRAY and by Illumina GoldenGate and Infinium assays. An oligonucleotide primer is attached to a bead; this primer is complementary to the target site and terminates immediately prior to the SNP site of interest. The primer is enzymatically extended by one fluorescently labeled nucleotide base, and the fluorescent signal of the incorporated base is used for genotyping

## 4 Amplicon Sequencing

Amplicon sequencing (Fig. 1a) is a powerful tool for generating large quantities of genetic data from low-quality samples because it relies on the generation of PCR products using template-specific primers, a method with a well-established success rate for these sample types. However, instead of using traditional Sanger sequencing of PCR products, amplicon sequencing uses NGS to simultaneously sequence millions of DNA fragments, thus generating orders of magnitude more sequence data than prior technologies at a fraction of the cost. Amplicon sequencing is a cost-effective and efficient method for generating sequence data from a relatively small number of loci but for a very large numbers of samples. These methods have the added advantage of being highly flexible in the numbers of loci targeted.

The laboratory methods for amplicon sequencing differ in several ways from those of traditional Sanger sequencing of PCR products. For PCR products to be sequenced on an NGS platform, specific double-stranded oligonucleotides called "sequencing adapters" must be attached to both ends of the PCR products to create a

"DNA library." In addition, PCR products for each individual sample must have a unique barcode identifier so that amplicons from multiple individuals can be sequenced simultaneously, thereby substantially reducing sequencing cost. These requirements are typically accomplished through an initial PCR with target-specific primers, followed by addition of sequencing adapters through a second PCR or a ligation reaction. The barcodes can be incorporated into the primers used for either the first or second PCR. PCRs can be conducted either in singleplex (one primer pair per reaction) or multiplex (multiple primer pairs per reaction). Although singleplex reactions necessarily require many more individual PCRs than do multiplex reactions, Fluidigm Corp (San Francisco, USA) has improved the efficiency of singleplex PCRs by developing chips on which thousands of independent PCRs can be conducted simultaneously in microfluidic chambers. One challenge of multiplex PCR is that amplification performance may be inconsistent across loci within a single reaction. However, Campbell et al. (2015) used a multiplex PCR approach to sequence 192 loci for 2,068 steelhead trout fin tissue samples in a single Illumina HiSeq lane and found that 187 loci were genotyped in >90% of samples, with only three loci genotyped in <70% of samples (Table 1). This study also found 99.9% concordance in genotypes previously generated for the same loci using TaqMan assays.

Amplicon sequencing of microsatellites has been tested for a number of wildlife species and found to have many advantages over traditional microsatellite fragment length analysis including increased accuracy, efficiency, and consistency of genotyping (Table 1; Darby et al. 2016; De Barba et al. 2017; Farrell et al. 2016; Suez et al. 2016; Vartia et al. 2016; Zhan et al. 2017). Specific benefits afforded by amplicon sequencing of microsatellites include unambiguous allele identification, increased information regarding sequence variation not detectable with fragment length analysis, and direct genotype comparability among platforms and laboratories due to automation of the genotyping process. Thus far, one study has used NGS of microsatellite amplicons for non-invasively collected samples and reported significant improvements in genotyping success and error rates from brown bear hair and scats (De Barba et al. 2017).

A very different kind of amplicon sequencing called "Nextera-tagmented, reductively amplified DNA" (NextRAD) genotyping has been tested on low-quality samples (Russello et al. 2015). This method sequences all regions of the genome containing a preselected 9 bp sequence by fragmentation of genomic DNA, ligation of Illumina sequencing adapters, and PCR amplification with a primer containing the 9 bp sequence and the adapter sequence. This method was first used for 96 American pika hair samples and produced data from 3,803 SNPs (Fig. 3a, Table 1; Russello et al. 2015). NextRAD differs from other amplicon sequencing methods in that it indiscriminately generates sequence data for both target and non-target species and therefore will be inefficient for samples containing large quantities of non-target DNA, such as fecal samples, without incorporating a target DNA capture step (see below). NextRAD also targets many more loci than most other amplicon sequencing methods.

**Table 1** Examples of studies using high-throughput DNA sequencing and SNP genotyping for low-quality DNA samples

| Method | Example system | Source material | Number of loci | Application | Citation |
|---|---|---|---|---|---|
| *Amplicon sequencing* | | | | | |
| Microsatellites | Brown bear (*Ursus arctos*) | Feces | 13 microsatellites, 1 sex marker | Proof of method, individual identification, parentage | De Barba et al. (2017) |
| NextRAD | American pika (*Ochotona princeps*) | Hair | 3,803 SNPs | Proof of method, divergent selection and diversity across habitats | Russello et al. (2015) |
| *SNP genotyping platforms* | | | | | |
| Fluidigm | European gray wolf (*Canis lupus*) | Feces, saliva, hair, urine | 96 SNPs | Proof of method, individual identification | Kraus et al. (2015) |
| MassARRAY | Puma (*Puma concolor*) | Feces | 25 SNPs | Proof of method, individual identification | Fitak et al. (2016) |
| Illumina GoldenGate BeadXpress | Coyote (*Canis latrans*) | Feces | 63 SNPs | Hybridization assessment | Monzón et al. (2014) |
| *DNA capture* | | | | | |
| Exons | Alpine chipmunk (*Tamias alpinus*) | Museum toe pads | ~11,000 exons | Temporal shifts in genetic structure and diversity | Bi et al. (2013) |
| Ultraconserved elements (UCEs) | Western scrub jay (*Aphelocoma californica*) | Museum toe pads | 4,460 SNPs | Phylogeny | McCormack et al. (2016) |
| Microsatellites | Lemur (*Propithecus diadema*) | Feces | 5,000 microsatellites | Proof of method, parentage | Kistler et al. (2017) |
| *Mitogenome sequencing* | | | | | |
| Amplicon sequencing | East African white-eyes (*Zosterops* spp.) | Feathers | 13,523–13,596 bp | Phylogeography | Meimberg et al. (2016) |
| Genome skimming | Tasmanian devil (*Sarcophilus harrisii*) | Museum hair shafts | 16,940 bp | Temporal shifts in genetic diversity | Miller et al. (2011) |
| Capture | Western chimpanzee (*Pan troglodytes verus*) | Feces | 16,554 bp | Proof of method | Perry et al. (2010) |

| | | | | | |
|---|---|---|---|---|---|
| *RADseq* | | | | | |
| ddRAD | Lake whitefish (*Coregonus clupeaformis*) | Degraded muscle | Not reported | Proof of method | Graham et al. (2015) |
| Ribo-tailing | Ants (three species) | Museum whole specimen | 16–1,275 SNPs | Proof of method | Tin et al. (2014) |
| Capture | Common ragweed (*Ambrosia artemisiifolia*) | Museum leaf | 22,813 SNPs | Proof of method; historic population structure | Barreiro et al. (2017) |
| *Whole-genome sequencing* | | | | | |
| Shotgun sequencing | Polar bear (*Ursus maritimus*) | Bone | 2.53 Gb | Evolutionary history | Miller et al. (2012) |
| Capture | Przewalski's horse (*Equus ferus* ssp. *przewalskii*) | Bone | 2.40 Gb | Evolutionary history | Der Sarkissian et al. (2015a, b) |
| Capture | Baboon (*Papio papio*) | Feces | 127,654 SNPs | Parentage, relatedness | Snyder-Mackler et al. (2016) |
| *Metagenomics/metabarcoding/metagenome skimming* | | | | | |
| Metagenomics | Giant panda (*Ailuropoda melanoleuca*) | Feces | 37 Mbp | Taxonomic and functional composition of microbiome | Zhu et al. (2011) |
| Metabarcoding | Large herbivores (seven species) | Feces | 1 barcode locus | Dietary niche partitioning | Kartzinel et al. (2015) |
| Metagenome skimming | Banded leaf monkey (*Presbytis femoralis*) | Feces | 3 barcode loci | Diet, parasites, population genetics | Srivathsan et al. (2016) |

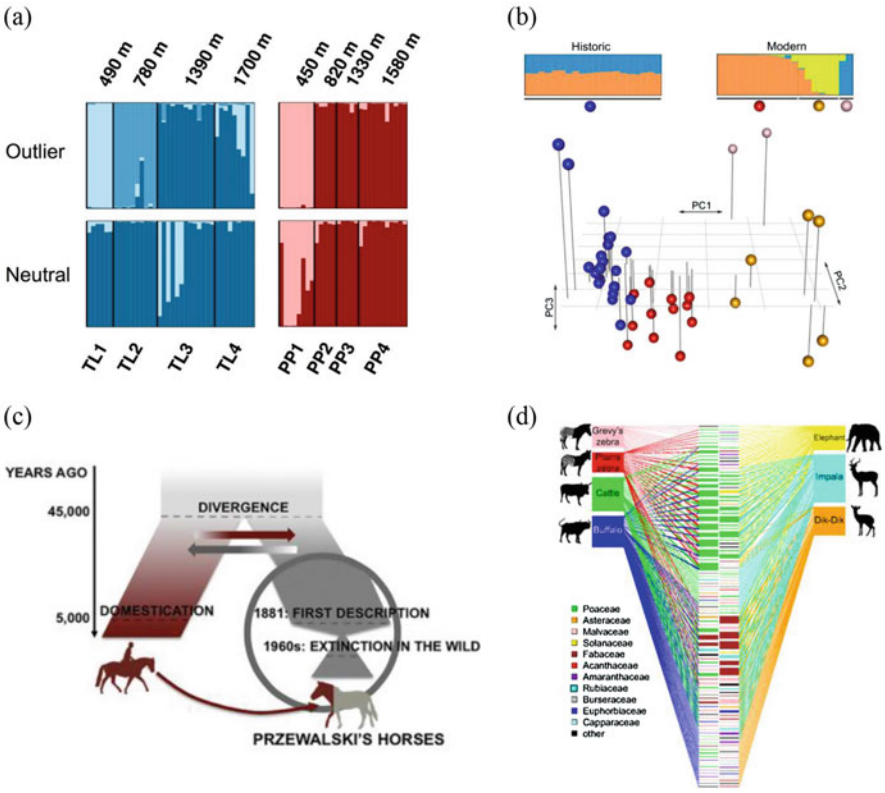


**Fig. 3** Examples of studies using next-generation sequencing of low-quality samples. (**a**) Russello et al. (2015) used NextRAD to evaluate 3,803 SNPs for non-invasively collected hair samples from American pika across two elevational gradients (TL and PP). Outlier analysis and Bayesian clustering analysis revealed 55 loci that are candidates for divergent selection at different elevations. (**b**) Bi et al. (2013) used DNA capture with ~11,000 exons for both contemporary alpine chipmunk samples and museum samples collected in 1915. Bayesian clustering analysis and principal components analysis revealed greater genetic structure for contemporary (non-blue dots) than historic (blue dots) samples. (**c**) Der Sarkissian et al. (2015b) sequenced the genomes of 11 contemporary captive and five historic Przewalski's horses dating 1878–1929. Comparative genomics indicated divergence from domestic horses around 45,000 years ago with ongoing gene flow, variable introgression of domestic alleles, and inbreeding resulting from captivity. (**d**) Kartzinel et al. (2015) conducted fecal sample metabarcoding for seven large mammalian herbivores and discovered unexpectedly strong diet partitioning across taxonomic groups

With the exception of NextRAD, amplicon sequencing requires template-specific primers and therefore requires prior genomic knowledge, which may be unavailable for many wildlife species. However, genomic information is becoming increasingly available for non-model organisms, thus increasing the feasibility of this approach.

## 5 SNP Genotyping Platforms

Numerous SNP genotyping platforms are currently available for analyzing a preselected set of loci (reviewed in Garvin et al. 2010; Ogden 2011; Ragoussis 2009; Slate et al. 2009). These platforms do not use NGS techniques but instead use probes and primers, a wide range of hybridization and enzymatic reactions, and genotype resolution through fluorescent dye or molecular weight detection (Fig. 2). The data output of these platforms is genotype calls rather than sequence reads and therefore requires less bioinformatic processing. Furthermore, the data is generated at a much lower cost and effort than NGS data and usually with a lower genotyping error rate. However, these platforms also require custom-designed taxon-specific probes and primers, which must be developed from prior genomic data. In addition, most platforms require costly, specialized equipment.

The numbers of loci and individuals analyzed at a time, and the level of flexibility in those numbers, vary widely across genotyping platforms. Small-scale platforms analyze tens of individuals and loci at a time and include the Fluidigm Dynamic Array (Fluidigm Corp, San Francisco, USA), MassARRAY (Agena Biosciences, San Digeo, USA), and SNPlex (Applied Biosystems, Foster City, USA). In contrast, platforms developed by Illumina and Affymetrix can genotype thousands of individuals and/or loci simultaneously; for example, the Illumina GoldenGate platform can assay between 96 and 3,072 loci for up to 1,536 samples, and the Illumina Infinium HD iSelect BeadChip arrays can genotype between 3,000 and one million SNPs for up to 24 samples.

The use of SNP genotyping assays is growing in wildlife research, largely due to an increase in accessibility of genomic resources that can be used to design probes and primers (Table 1). Some wildlife studies have taken advantage of genotyping assays developed from closely related domestic species. For example, domestic dog assays have been used for wild canids (von Holdt et al. 2011), cattle assays have been used for deer (Haynes and Latch 2012), and domestic sheep assays have been used for bighorn sheep (Miller et al. 2014). However, the use of SNP assays designed for a different species often results in high proportions of monomorphic loci due to ascertainment bias (Clark et al. 2005; Lachance and Tishkoff 2013). Other wildlife studies have designed species- or population-specific SNP assays using genomic data obtained from transcriptome sequences (Cullingham et al. 2013; Fitak et al. 2016; Hoffman et al. 2012), whole-genome sequences (Nguyen et al. 2014), or other sources of DNA sequence (Holman et al. 2017; Veale and Russello 2016).

Most SNP genotyping assays are not amenable to non-invasive samples due to requirements for high-quantity and/or high-quality input DNA. However, several studies have found success with non-invasive samples when using SNP genotyping assays that target relatively small numbers of loci. For example, a study using $96 \times 96$ Fluidigm Dynamic Array chips, which genotype 96 loci and 96 samples at a time, found low missing data rates ($<10\%$) and low genotyping error rates ($\sim 1\%$) for non-invasive European gray wolf samples including tissue, blood, scat, saliva, hair,

and urine (Kraus et al. 2015). Another study using the same type of assay found even lower rates of missing data and genotyping error (<1%) for fecal samples collected from brown bears in Sweden (Norman and Spong 2015). Thus far, success rates have been lower for studies using MassARRAY for low-quality samples. For example, Fitak et al. (2016) found a relatively high missing data rate (40.2%) for a 25-locus MassARRAY assay for puma fecal samples; however, this was lower than the missing data rate for microsatellites for the same samples (60.1%), indicating sample quality may have been a driving factor.

## 6 DNA Capture

DNA capture is increasingly being used for NGS approaches with low-quality samples (Fig. 1b, Table 1; Gnirke et al. 2009; Jones and Good 2016; Mamanova et al. 2010). This method first involves the creation of genomic DNA libraries for sequencing on a NGS platform. These libraries must contain DNA fragments of the appropriate size for sequencing, which is about 500 bp for the widely used Illumina sequencing platform. For high-quality DNA samples, genomic DNA is usually reduced to the appropriate fragment size range through mechanical or enzymatic shearing, but low-quality samples may already have DNA fragments this size or smaller. Genomic DNA libraries must also have sequencing adapters ligated onto the ends of the DNA fragments. These DNA libraries could be sequenced directly, but for many applications this would generate data from many more loci than are needed. Furthermore, this approach would indiscriminately sequence all DNA present in the sample, which in the case of low-quality samples would often include high proportions of non-target DNA, such as bacteria in many archival samples. DNA capture circumvents these problems by targeting selected loci through hybridization of biotinylated oligonucleotide probes or "baits" to the library and then isolating the hybridized DNA for sequencing. Baits must be designed using prior genomic knowledge, such as a reference genome, transcriptome, restriction site-associated DNA sequencing (RADseq) data (described below), or PCR products. Any number of baits can be synthesized, or baits can be created by directly biotinylating PCR products, RADseq libraries, RNA-transcribed PCR products, or RNA-transcribed DNA libraries. If prior genomic information is not available for the target species, several studies have shown that capture efficiency can be high for baits designed using genomic resources from a moderately divergent species (Bi et al. 2012; Enk et al. 2014; Hedtke et al. 2013; Jin et al. 2012; Vallender 2011).

A number of studies have used DNA capture to target hundreds or thousands of loci for population genetic or phylogenetic analyses with low-quality samples. For example, DNA capture was used to investigate temporal shifts in genetic structure and diversity for alpine chipmunks by sequencing ~11,000 exons from 20 museum samples collected in 1915 and 20 contemporary samples (Fig. 3b, Bi et al. 2013). Similarly, two studies employed DNA capture targeting 9,000 SNPs and the full mitogenome, respectively, to directly investigate the genetic consequences of rapid

population decline and recovery in Pinzon Galapagos giant tortoises sampled pre-bottleneck in 1906 ($n = 78$) and post-bottleneck in 2014 ($n = 150$) (Jensen et al. 2018a, b). Another study used sequence capture of 5,060 ultraconserved elements (UCEs) for phylogenetic analyses of 27 western scrub jay museum specimens up to 120 years old (McCormack et al. 2016). Kistler et al. (2017) described a bioinformatic pipeline for designing probes for a large panel of microsatellite markers for non-model species without a reference genome and developed a capture-based approach for enriching genomic DNA libraries for thousands of these loci. They tested this method on tissue and fecal samples from an endangered lemur species and reported high efficiency in recovering targeted loci, as well as high genotyping accuracy.

DNA capture has also been used to sequence whole mitogenomes, subsets of RADseq loci, and even whole genomes. These approaches are described in more detail in later sections.

## 7 Mitogenome Sequencing

Mitochondrial DNA has been used frequently in analyses of ancient, historical, non-invasive, and forensic samples, because it is present in high copy number in cells relative to nuclear DNA (Table 1). This disparity in copy number helps facilitate recovery of mtDNA sequences through PCR amplification and other approaches (Alacs et al. 2010; Ho and Gilbert 2010; Höss et al. 1992; Paijmans et al. 2013). Mitochondrial markers are linked and maternally inherited and therefore do not provide as much information as multiple nuclear loci. Nonetheless, whole mitogenome sequences can provide greater resolution than single mtDNA markers for phylogenetic and population genetic analysis (Duchêne et al. 2011; Meiklejohn et al. 2014; Paijmans et al. 2013; Rohland et al. 2007) and can enhance the discriminatory power of forensic analysis (Chaitanya et al. 2015).

Before the development of NGS technologies, mitogenomes could be sequenced by PCR amplification and Sanger sequencing of overlapping fragments covering the entire mitogenome, provided sufficient prior knowledge was available for primer design. This approach has been used successfully with a variety of low-quality sample types, including non-invasively collected feces (Bjork et al. 2011; Finch et al. 2014; Matsui et al. 2007) and ancient (Bon et al. 2008; Rogaev et al. 2006) specimens. This process can also be accomplished using two-step multiplex PCRs, which reduce the amount of source biological material required compared to a singleplex PCR approach, and does not require cloning prior to Sanger sequencing (Edwards et al. 2010; Krause et al. 2006; Römpler et al. 2006). However, low-quality samples present a challenge for PCR-based techniques, because DNA degradation can limit the size of fragments available for amplification.

NGS technologies have overcome many of the challenges of Sanger sequencing-based approaches to mitogenome sequencing for low-quality samples. NGS can be used to sequence mtDNA singleplex or multiplex PCR products more efficiently and

inexpensively than Sanger sequencing. Alternatively, mtDNA sequences can be recovered from NGS sequence data generated from entire genomic DNA libraries and assembled into whole mitogenomes, thus eliminating the need for prior knowledge for PCR primer design. This can be accomplished using relatively low quantities of sequence data from genomic DNA libraries, because the high copy number of mtDNA in cells results in a high proportion of the data being of mtDNA origin. The technique of using low quantities of sequence data from genomic DNA libraries to assemble mitogenomes has been called "genome skimming" and can also be used to assemble other high-copy loci like nuclear ribosomal genes, histone genes, and plastomes (Straub et al. 2012). This approach has become widely used for mitogenome sequencing with archival samples, with examples including an ancient polar bear bone sample dating 110,000–130,000 years ago (Lindqvist et al. 2010), a woolly mammoth sample dating 14,000–60,000 years ago (Gilbert et al. 2007), an ice-age woolly rhinoceros sample (Willerslev et al. 2009), and Tasmanian devil museum samples from the early 1900s (Miller et al. 2011).

Mitogenomes can also be sequenced using DNA capture from genomic DNA libraries, and this technique has been used in a number of studies with low-quality samples. For example, this approach was used to sequence mitogenomes for guenon specimens as old as 117 years (Guschanski et al. 2013), extinct sloth species from bone and coprolite samples (13,000–20,000 years old; Slater et al. 2016), and contemporary fecal samples for chimpanzees (Perry et al. 2010). This approach requires prior mitogenome sequence knowledge for bait design, but some studies have found success using baits designed from species 10–20% divergent from the target species (Hawkins et al. 2016; Mason et al. 2011).

## 8 RADseq

Restriction site-associated DNA sequencing is a widely used NGS-based method for sequencing a subset of the genome for non-model organisms (Fig. 1c, Table 1; reviewed in Andrews et al. 2016; Davey et al. 2011). This method sequences regions adjacent to restriction cut sites, which occur across the genome in both coding and noncoding regions. RADseq requires no prior genomic knowledge, uses relatively small quantities of genomic DNA (usually around 100 ng), and is highly flexible in the numbers of loci that can be surveyed. Numerous methods have been developed to generate RADseq data (reviewed in Andrews et al. 2016), but all share some common features: digesting genomic DNA with one or more restriction enzymes, ligating adapters required for sequencing, and sequencing with a high-throughput platform. Most RADseq protocols ligate adapters with unique, sample-specific barcodes early in the library prep; this decreases the time and cost involved in library prep by allowing many samples to be multiplexed early in the protocol and also decreases sequencing costs by allowing many individuals to be sequenced together. Although this approach typically generates data from several thousands to tens of thousands of loci, smaller numbers of loci can be assayed by adding a DNA capture

step to the end of the RADseq library prep using baits designed directly from RADseq data for a selected subset of loci ("Rapture," Ali et al. 2016; "RADcap," Hoffberg et al. 2016).

Few studies have tested the performance of RADseq on low-quality samples. RADseq may be more negatively impacted by degraded DNA than other methods, because digesting degraded DNA with restriction enzymes may result in fragments too small to generate sufficient sequence information. Graham et al. (2015) systematically evaluated the performance of one RADseq method on low-quality samples by allowing lake whitefish muscle samples to degrade by incubation at room temperature for up to 96 h before DNA extraction and library prep. This study found little reduction in RADseq data quantity and quality for moderately degraded DNA but a substantial reduction for severely degraded DNA. Another study used a modified RADseq protocol for six ant specimens collected between 1910 and 1953 and found that a large proportion of sequence reads were not useable because they were too small for genome mapping or were PCR duplicates, although several protocol modifications were proposed that could increase the success of this technique in future studies (Tin et al. 2014). The performance of RADseq for archival and non-invasive samples is likely to vary across the wide range of laboratory protocols that have been developed and also across the variety of sample types, due to the associated variation in quantity and quality of DNA.

Another approach for generating RADseq data from low-quality samples is to use DNA capture of genomic DNA libraries with RADseq locus-specific baits. These approaches differ from Rapture and RADcap (described above), which may perform more poorly for low-quality samples because loci are captured from RADseq libraries rather than genomic DNA libraries. For example, Barreiro et al. (2017) used RADseq data from 190 modern, high-quality common ragweed samples from 37 populations to design and synthesize capture baits for 20,000 RADseq loci. These baits were then used for DNA capture of genomic libraries generated using 38 herbarium samples dating 1835–1913. Although the numbers of sequence reads varied substantially across targeted loci, this approach led to the discovery of 22,813 SNPs in the herbarium samples. Suchan et al. (2016) used a similar approach for museum samples of *Lycaena helle* butterfly but, instead of designing and synthesizing baits, created baits by directly biotinylating RADseq libraries generated from high-quality DNA samples.

## 9 Whole-Genome Sequencing

NGS can also be used to sequence entire genomes (Fig. 1d, Table 1; reviewed in Ekblom and Wolf 2014; Ellegren 2014). Currently, most WGS projects with high-quality samples rely on direct sequencing of genomic DNA libraries ("shotgun sequencing") using the Illumina HiSeq platform. However, the Illumina HiSeq generates sequence reads that are a maximum of 150 bp long, and the short lengths

of these fragments can present a limitation when assembling a genome de novo. Thus, many whole-genome sequencing projects with high-quality samples use a combination of both Illumina sequencing and long-read sequencing technologies such as those developed by Pacific Biosciences or Oxford Nanopore. Even when combining short- and long-read technologies, however, the process of de novo genome assembly from raw sequence data requires a large quantity of sequence data and complex computational algorithms (Ekblom and Wolf 2014; Ellegren 2014). Once a genome has been assembled for a species, "genome resequencing" of additional samples from the same species can be conducted with less sequencing and computational effort, because the raw sequence data can be aligned to the reference genome (Ellegren 2014; Fuentes-Pardo and Ruzzante 2017). Furthermore, a reference genome can provide a valuable resource for sequencing and genotyping approaches that target a subset of the genome, such as those described above.

WGS and genome resequencing are challenging for low-quality samples due to low quantities of DNA, degradation of DNA, and high proportions of non-target DNA which is sequenced indiscriminately with standard WGS protocols. However, the high-throughput sequencing capability of NGS, along with the ability of NGS to sequence short DNA fragments, dramatically increases the feasibility of WGS for low-quality samples when compared to Sanger sequencing (Hofreiter et al. 2015; Rizzi et al. 2012). A growing number of studies are using NGS to sequence whole genomes from archival samples, including samples from extinct and endangered species (Table 1; reviewed in Der Sarkissian et al. 2015a; Leonardi et al. 2017). For example, Mikheyev et al. (2015) sequenced and compared the genomes of honeybees collected from 32 colonies in 1977 and 2010, and Der Sarkissian et al. (2015b) sequenced the genomes of five historic Przewalski's horses dating to 1878–1929 (Fig. 3c). A number of genomes have been sequenced from ancient bone, tissue, and hair specimens including human, Neanderthal, mammoth, polar bear, horse, common rat, and pigeon (e.g., Cahill et al. 2013; Green et al. 2006; Hung et al. 2014; Miller et al. 2008, 2012; Orlando et al. 2013; Poinar et al. 2006; Rowe et al. 2011).

Although NGS has dramatically increased the feasibility of WGS with low-quality samples, this approach is still more costly than WGS of high-quality samples and requires special protocols for DNA extraction, library prep, and bioinformatic analyses to accommodate degraded DNA and the high prevalence of non-target DNA in the samples (reviewed in Orlando et al. 2015). One of the primary laboratory methods used for WGS of low-quality samples is DNA capture with probes designed from genome sequence data generated using high-quality modern samples, and this approach was used for most of the examples listed above. Some recent DNA capture studies have reduced the time and cost involved in bait development by generating a genomic DNA library using a high-quality sample from the study species and then transcribing the library into biotinylated RNA fragments which can be directly used as baits. For example, Carpenter et al. (2013) used this approach for genome-wide enrichment of ancient human teeth, bone, and hair samples, with baits generated by transcribing a DNA library created using a blood sample from a contemporary human. For this study, DNA libraries sequenced without the capture step yielded an average of 1.2% reads that mapped to

the human genome, but libraries sequenced with the capture step yielded up to 59% mapped reads. Notably, this technique requires a large quantity of genomic DNA from a high-quality sample, which can be a limitation for many wildlife studies.

The use of WGS for contemporary non-invasive samples has lagged behind that for ancient samples and has focused on primate fecal samples with a DNA capture approach. Perry et al. (2010) demonstrated the feasibility of WGS from fecal samples by sequencing more than 1.5 Mb of the genome (including the whole mitogenome and parts of two chromosomes) from six western chimpanzee fecal samples, using 55,000 120 bp capture baits designed from a chimpanzee reference genome. Snyder-Mackler et al. (2016) used a capture approach similar to that of Carpenter et al. (2013) described above for genome-wide enrichment of 62 wild baboon fecal samples, with baits generated by RNA transcription of a genomic library from a high-quality blood sample from the same species (Table 1). This study found up to 40-fold target enrichment as a result of the capture step.

Chiou and Bergey (2018) introduced another method called "FecalSeq" for enriching whole genomic DNA from low-quality vertebrate samples that requires no prior genomic knowledge or high-quality samples. Instead of using baits, this method uses methyl-CpG-binding domain (MBD) proteins to selectively bind DNA having high levels of methylation, taking advantage of the fact that vertebrate DNA has a higher frequency of methylation than does bacterial DNA. This approach resulted in a 195-fold increase in the amount of target DNA for baboon fecal samples that started with a mean of 0.34% host DNA (Chiou and Bergey 2018).

## 10 Metagenome Sequencing, Metabarcoding, and Metatranscriptomics

Whereas the techniques we have described thus far generate sequence and genotype data for single target organisms, other techniques called metagenome sequencing, metabarcoding, and metatranscriptomics generate sequence data for samples comprised of multiple individuals and species, such as eDNA samples, insect traps, plankton tows, fecal samples, oral swabs, and many other sample types (reviewed in Aguiar-Pulido et al. 2016; Mendoza et al. 2015; Taberlet et al. 2018; Wooley et al. 2010). "Metagenomics" refers to studies generating data from a large portion of the genomes present in these samples (the "metagenome"); these studies typically use a shotgun sequencing approach and focus on both taxonomic-informative and functional genes present in species that have relatively small genomes, such as microbial species. For example, Zhu et al. (2011) used wild giant panda fecal samples for metagenomic analysis of the gut microbiome and identified putative bacterial genes coding for cellulose-digesting enzymes, thus clarifying how giant pandas can partially digest bamboo despite the absence of genes coding for cellulose-degrading enzymes within the panda genome (Table 1). Similarly, fecal metagenomics revealed functional genes related to metabolism in the microbiome of Asian

elephants (Ilmberger et al. 2014), snub-nosed monkeys (Xu et al. 2015), and Galapagos iguanas (Hong et al. 2015).

"Metabarcoding" differs from metagenomics because it does not focus on a large portion of the genome but instead typically uses amplicon sequencing of PCR products for a small number of taxonomic-informative loci to delineate the species composition of a sample (reviewed in Taberlet et al. 2018). Metabarcoding is similar to "DNA barcoding," but starts with samples comprised of communities of individuals rather than single individuals, and simultaneously generates sequence data from these different individuals and species. As in DNA barcoding, metabarcoding usually focuses on a certain taxonomic group, such as bacteria, plants, or plankton, and usually uses high-copy genetic markers such as mtDNA and chloroplast loci. However, shorter fragments are amplified for non-invasive and eDNA samples to accommodate DNA degradation, and highly conserved and versatile primers are preferred for robust and efficient amplification of many taxa (Deagle et al. 2014; Valentini et al. 2009). The development of NGS technologies has led to a dramatic rise in metabarcoding studies because these technologies eliminate the need for a time-consuming cloning step that would be required for Sanger sequencing of samples containing amplicons from multiple individuals and species.

Metabarcoding is becoming widely used for wildlife studies with non-invasive and eDNA samples. For example, fecal sample metabarcoding has been used to characterize diet in herbivores, carnivores, insectivores, and omnivores (e.g., Fig. 3d, Table 1; Bohmann et al. 2011; De Barba et al. 2014; Deagle et al. 2009; Kartzinel et al. 2015; Shehzad et al. 2012; Valentini et al. 2009). Fecal metabarcoding has also been used to characterize microbial communities in a number of wildlife species, primarily using 16S ribosomal RNA markers (reviewed in Escobar-Zepeda et al. 2015). For example, Amato et al. (2013) used fecal metabarcoding to characterize microbial communities of howler monkeys across different habitats and found lower microbiome diversity in suboptimal habitats. Metabarcoding of eDNA water samples has been used to characterize fish, amphibian, and invertebrate communities, detect invasive species, and even infer population structure and diversity (e.g., Klymus et al. 2017; Sigsgaard et al. 2016; Thomsen et al. 2012). This approach has also been used to characterize ancient and contemporary species assemblages of flora and fauna using eDNA extracted from soil samples (reviewed in Pedersen et al. 2015).

Another approach for generating sequence data from barcode loci is "metagenome skimming" (Crampton-Platt et al. 2016; Papadopoulou et al. 2015). This method is conceptually similar to genome skimming (described above) and starts with sequencing of metagenomic DNA libraries, followed by bioinformatic identification of the portion of the sequence data that is the barcode loci. This technique is feasible even with relatively low quantities of metagenomic sequence data, because barcode loci are typically high copy number in cells, and therefore a large proportion of the sequence data generated by metagenome sequencing will be from the barcode loci. This technique is more expensive than metabarcoding, but overcomes several of its limitations, including PCR bias, lack of taxonomic resolution from using small numbers of barcode loci or short barcode loci, and detection of species for which

primers anneal poorly. Furthermore, metagenome skimming can also be used to assemble mitogenomes and identify non-target taxa. For example, Srivathsan et al. (2016) used both metagenome skimming and metabarcoding of fecal samples to assess the species composition of banded leaf monkey diet and found that metagenome skimming had higher taxonomic resolution (Table 1). In addition, the fecal sample metagenome sequence data confirmed the presence of parasites and allowed assembly of whole mitogenomes of the monkeys. Metagenome skimming has been described using a variety of terms in the literature, including "metagenomics," "targeted metagenomics," "metabarcoding," and "PCR-free single/multiple loci metabarcoding," potentially leading to confusion (Mendoza et al. 2015).

Metatranscriptomics involves sequencing RNA extracted from a sample comprised of a community of individuals and species (Aguiar-Pulido et al. 2016; Carvalhais et al. 2012; Moran 2009). This method provides information regarding the genes expressed by the community and has primarily been used to investigate microbial communities. For example, metatranscriptomics may help identify bacterial genes with functions making them effective probiotics for the fungus responsible for chytridiomycosis in amphibians (Rebollar et al. 2016). However, metatranscriptomics has a number of limitations, including the instability of RNA, the requirement for reference data regarding gene sequences and functions, and others (Moran 2009).

## 11 Choosing a Method

The optimal choice of sequencing or genotyping method for a project involving low-quality samples will depend on the research question and budget, as well as the quantity and quality of DNA available. Different research questions vary in the required numbers and specificity of targeted loci, and the numbers of loci generated by different methods vary substantially (Fig. 4). Currently, DNA capture is the most widely used technique for generating sequence data from low-quality samples because it targets specific loci of interest and has high flexibility in the number of loci targeted. Amplicon sequencing and SNP genotyping also target loci of interest and are less time-consuming and expensive than DNA capture for large sample sizes; these methods are likely to become widely used for research projects with low-quality samples requiring fewer than several hundred loci. For studies targeting the mitogenome, the choice between amplicon sequencing, genome skimming, and DNA capture will largely rest on project budget and the availability of reference sequences for designing primers or baits.

In contrast to DNA capture and amplicon sequencing, non-targeted approaches like RADseq and WGS are advantageous because they require no prior genomic information (Fig. 4), thus eliminating the time, cost, and bioinformatic expertise required to design primers and probes. Furthermore, non-targeted approaches are minimally affected by the many disadvantages associated with ascertainment bias

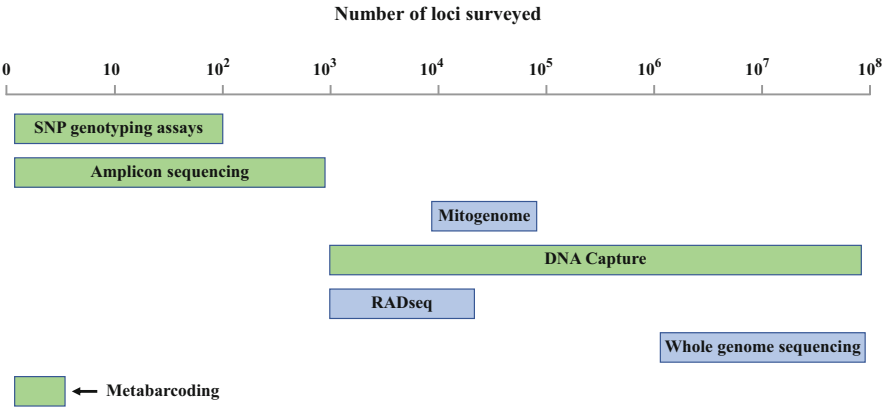


**Fig. 4** Numbers of loci typically interrogated for next-generation sequencing and SNP genotyping approaches with wildlife species. Green indicates methods that require prior genomic knowledge to design primers and probes, and blue indicates methods that require no prior genomic knowledge. Note the scale is logarithmic

(Clark et al. 2005; Lachance and Tishkoff 2013). However, generating RADseq and WGS data is more technically challenging for non-invasive samples than targeted approaches, due to greater sensitivity to DNA quantity and quality. In addition, these methods non-discriminately sequence all DNA present in the sample and therefore will waste sequencing effort for samples containing large quantities of exogenous DNA, unless a DNA capture step is incorporated. In addition, RADseq and WGS generate data from much larger numbers of loci than are required for many research questions and thus will not be time- and cost-efficient for many studies. However, WGS could be highly valuable for species or populations for which few samples are available, such as endangered or extinct species, because a number of ecological and evolutionary insights can be gained from whole-genome sequences from even a single individual. In addition, both RADseq and WGS data can be used as a resource for designing primers and probes for targeted approaches.

For studies of environmental samples, metabarcoding is currently the most efficient, affordable, and widely used approach for taxonomic composition analysis. Metagenome skimming can also be used for this type of analysis and can overcome some of the limitations of metabarcoding but requires more sequencing and bioinformatic effort. Metagenomic and metatranscriptomic approaches can provide insight into the functional genomic composition of communities but require even greater sequencing and bioinformatic effort; currently, these approaches are typically restricted to species with relatively small genomes, such as microbes, and are often limited by the availability of functional genomic information in reference databases.

In summary, recent advances in DNA sequencing and genotyping technologies have dramatically increased the amount of genetic data available for low-quality samples, thereby expanding the types of questions we can address using non-invasively collected, archival, and environmental samples. These technologies

are rapidly evolving to become cheaper, faster, and easier and produce longer and more accurate sequence reads. Although applications of these technologies for low-quality samples will continue to lag behind those for high-quality samples, nonetheless we expect steady advances in techniques tailored for these sample types in the coming years. Continued efforts should also be made to maximize initial amount and quality of DNA for low-quality samples by optimizing sample collection, preservation, and DNA extraction to minimize DNA degradation and loss, and improve recovery of endogenous DNA. Furthermore, development should continue for effective bioinformatic tools and standards tailored for data generated from low-quality samples to ensure accurate data collection. In the coming decade, we expect these advances to provide unprecedented insights into wildlife ecology, evolution, and conservation.

**Acknowledgments** We thank Brendan Epstein and Paul Hohenlohe for helpful comments on the manuscript. K.R.A. and L.P.W. were funded by the College of Natural Resources at the University of Idaho. M.D.B. acknowledges the support of the Laboratoire d'Ecologie Alpine, Grenoble, France. M.A.R acknowledges the support of the NSERC Discovery program (grant # 2014-04736). This is PMEL contribution number 4805 and Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063 contribution number 2018-0160.

## Glossary

**Amplicon sequencing** High-throughput sequencing of PCR products from multiple individuals simultaneously

**Archival sample** Historic (collected within the last ~200 years) or ancient (usually up to ~100,000 years old) tissue sample

**Ascertainment bias** Inference bias resulting from the process by which genetic loci were selected

**Bait** Biotinylated oligonucleotide probe used to isolate or "capture" target DNA for sequencing

**DNA capture** Method relying on baits to isolate or "capture" target DNA prior to high-throughput sequencing

**DNA library** DNA that has been prepared for high-throughput sequencing, with DNA fragments the appropriate length, and with sequencing adapters ligated to ends of fragments

**Environmental DNA (eDNA)** DNA present in environmental samples such as water, soil, and air

**Genetic non-invasive sample (gNIS)** DNA sample collected without handling the study organism, e.g., shed hair or fecal sample

**Genome skimming** Using low quantities of sequence data from genomic DNA libraries to study high-copy loci like mitogenomes, nuclear ribosomal genes, histone genes, and plastomes

**Metabarcoding** High-throughput sequencing of PCR products generated from taxonomic-informative markers for an environmental sample

**Metagenome skimming** Using shotgun sequencing data from samples comprised of multiple individuals and/or species (e.g., environmental samples) to study high-copy loci like mitogenomes, nuclear ribosomal genes, histone genes, and plastomes. Often focuses on taxonomic-informative markers to identify community composition

**Metatranscriptomics** Sequencing RNA extracted from an environmental sample

**Next-generation sequencing** A variety of high-throughput DNA sequencing technologies developed over the last decade

**Nextera-tagmented, reductively amplified DNA (NextRAD)** High-throughput sequencing of PCR products from genomic regions containing a preselected 9 bp sequence

**Restriction site-associated DNA sequencing (RADseq)** High-throughput sequencing of genomic regions adjacent to restriction cut sites

**Sanger sequencing** "Traditional" low-throughput DNA sequencing technology developed in 1977

**Sequencing by synthesis** High-throughput sequencing technology used by Illumina that detects the incorporation of single bases into replicating DNA strands

**Shotgun sequencing** Direct sequencing of genomic DNA libraries

**Single nucleotide polymorphism (SNP)** Variant at a single nucleotide site

**SNP genotyping platforms** A variety of technologies that generate SNP genotype data rather than sequence reads

## References

Adams JR, Kelly BT, Waits LP. Using faecal DNA sampling and GIS to monitor hybridization between red wolves (*Canis rufus*) and coyotes (*Canis latrans*). Mol Ecol. 2003;12:2175–86. https://doi.org/10.1046/j.1365-294X.2003.01895.x.

Aguiar-Pulido V, Huang WR, Suarez-Ulloa V, Cickovski T, Mathee K, Narasimhan G. Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. Evol Bioinforma. 2016;12:5–16. https://doi.org/10.4137/ebo.s36436.

Alacs EA, Georges A, FitzSimmons NN, Robertson J. DNA detective: a review of molecular approaches to wildlife forensics. Forensic Sci Med Pathol. 2010;6:180–94. https://doi.org/10.1007/s12024-009-9131-7.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, et al. RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics. 2016;202:389–400. https://doi.org/10.1534/genetics.115.183665.

Amato KR, Yeoman CJ, Kent A, Righini N, Carbonero F, Estrada A, et al. Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. ISME J. 2013;7:1344–53. https://doi.org/10.1038/ismej.2013.16.

Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, et al. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Mol Ecol. 2012;21:1966–79. https://doi.org/10.1111/j.1365-294X.2011.05261.x.

Andrews K, Good J, Miller M, Luikart G, Hohenlohe P. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92. https://doi.org/10.1038/nrg.2015.28.

Arandjelovic M, Head J, Kuhl H, Boesch C, Robbins MM, Maisels F, et al. Effective non-invasive genetic monitoring of multiple wild western gorilla groups. Biol Conserv. 2010;143:1780–91. https://doi.org/10.1016/j.biocon.2010.04.030.

Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. Environmental conditions influence eDNA persistence in aquatic systems. Environ Sci Technol. 2014;48:1819–27. https://doi.org/10.1021/es404734p.

Barreiro FS, Vieira FG, Martin MD, Haile J, Gilbert MTP, Wales N. Characterizing restriction enzyme-associated loci in historic ragweed (*Ambrosia artemisiifolia*) voucher specimens using custom-designed RNA probes. Mol Ecol Res. 2017;17:209–20. https://doi.org/10.1111/1755-0998.12610.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G. Advancing ecological understandings through technological transformations in noninvasive genetics. Mol Ecol Resour. 2009;9:1279–301. https://doi.org/10.1111/j.1755-0998.2009.02699.x.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics. 2012;13:1–14. https://doi.org/10.1186/1471-2164-13-403.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. Unlocking the vault: next-generation museum population genomics. Mol Ecol. 2013;22:6018–32. https://doi.org/10.1111/mec.12516.

Bjork A, Liu WM, Wertheim JO, Hahn BH, Worobey M. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. Mol Biol Evol. 2011;28:615–23. https://doi.org/10.1093/molbev/msq227.

Blejwas KM, Williams CL, Shin GT, McCullough DR, Jaeger MM. Salivary DNA evidence convicts breeding male coyotes of killing sheep. J Wildl Manag. 2006;70:1087–93. https://doi.org/10.2193/0022-541x(2006)70[1087:sdecbm]2.0.co;2.

Bohmann K, Monadjem A, Noer CL, Rasmussen M, Zeale MRK, Clare E, et al. Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. PLoS One 2011;6. https://doi.org/10.1371/journal.pone.0021441.

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, et al. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol Evol. 2014;29:358–67. https://doi.org/10.1016/j.tree.2014.04.003.

Bon C, Caudy N, de Dieuleveult M, Fosse P, Philippe M, Maksud F, et al. Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. Proc Natl Acad Sci U S A. 2008;105:17447–52. https://doi.org/10.1073/pnas.0806143105.

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of *Yersinia pestis* from victims of the Black Death. Nature. 2011;478:506–10. https://doi.org/10.1038/nature10549.

Bouzat JL, Lewin HA, Paige KN. The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. Am Nat. 1998;152:1–6. https://doi.org/10.1086/286145.

Braunisch V, Segelbacher G, Hirzel AH. Modelling functional landscape connectivity from genetic population structure: a new spatially explicit approach. Mol Ecol. 2010;19:3664–78. https://doi.org/10.1111/j.1365-294X.2010.04703.x.

Bricker J. Purification of high quality DNA from shed skin. Herpetol Rev. 1996;27.

Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsyanikov N, et al. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. PLoS Genet. 2013;9. https://doi.org/10.1371/journal.pgen.1003345.

Campbell NR, Harmon SA, Narum SR. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Res. 2015;15:855–67. https://doi.org/10.1111/1755-0998.12357.

Caniglia R, Fabbri E, Galaverni M, Milanesi P, Randi E. Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. J Mammal. 2014;95:41–59. https://doi.org/10.1644/13-mamm-a-039.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am J Hum Genet. 2013;93:852–64. https://doi.org/10.1016/j.ajhg.2013.10.002.

Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. Application of metatranscriptomics to soil environments. J Microbiol Methods. 2012;91:246–51. https://doi.org/10.1016/j.mimet.2012.08.011.

Casas-Marce M, Revilla E, Godoy JA. Searching for DNA in museum specimens: a comparison of sources in a mammal species. Mol Ecol Res. 2010;10:502–7. https://doi.org/10.1111/j.1755-0998.2009.02784.x.

Castillo JA, Epps CW, Davis AR, Cushman SA. Landscape effects on gene flow for a climate-sensitive montane species, the American pika. Mol Ecol. 2014;23:843–56. https://doi.org/10.1111/mec.12650.

Chaitanya L, Ralf A, van Oven M, Kupiec T, Chang J, Lagace R, et al. Simultaneous whole mitochondrial genome sequencing with short overlapping amplicons suitable for degraded DNA using the ion torrent personal genome machine. Hum Mutat. 2015;36:1236–47. https://doi.org/10.1002/humu.22905.

Chiou KL, Bergey CM. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep. 2018;8:1975. https://doi.org/10.1038/s41598-018-20427-9.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 2005;15:1496–502. https://doi.org/10.1101/gr.4107905.

Constable JL, Ashley MV, Goodall J, Pusey AE. Noninvasive paternity assignment in Gombe chimpanzees. Mol Ecol. 2001;10:1279–300. https://doi.org/10.1046/j.1365-294X.2001.01262.x.

Crampton-Platt A, Yu DW, Zhou X, Vogler AP. Mitochondrial metagenomics: letting the genes out of the bottle. Gigascience 2016;5. https://doi.org/10.1186/s13742-016-0120-y.

Cullingham CI, Cooke JEK, Dang S, Coltman DW. A species-diagnostic SNP panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. Tree Genet Genomes. 2013;9:1119–27. https://doi.org/10.1007/s11295-013-0608-x.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Nat. 2006;168:486–99. https://doi.org/10.1086/506976.

Darby BJ, Erickson SF, Hervey SD, Ellis-Felege SN. Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. Ecol Evol. 2016;6:4502–12. https://doi.org/10.1002/ece3.2221.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510. https://doi.org/10.1038/nrg3012.

Davidson GA, Clark DA, Johnson BK, Waits LP, Adams JR. Estimating cougar densities in northeast Oregon using conservation detection dogs. J Wildl Manag. 2014;78:1104–14. https://doi.org/10.1002/jwmg.758.

De Barba M, Waits LP, Garton EO, Genovesi P, Randi E, Mustoni A, et al. The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. Mol Ecol. 2010;19:3938–51. https://doi.org/10.1111/j.1365-294X.2010.04791.x.

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, et al. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol Ecol Resour. 2014;14:306–23. https://doi.org/10.1111/1755-0998.12188.

De Barba M, Miquel C, Lobreaux S, Quenette PY, Swenson JE, Taberlet P. High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. Mol Ecol Res. 2017;17:492–507. https://doi.org/10.1111/1755-0998.12594.

Deagle BE, Kirkwood R, Jarman SN. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Mol Ecol. 2009;18:2022–38. https://doi.org/10.1111/j.1365-294X.2009.04158.x.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match Biol Lett 2014;10(9). https://doi.org/10.1098/rsbl.2014.0562.

Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. J Appl Ecol. 2012;49:953–9. https://doi.org/10.1111/j.1365-2664.2012.02171.x.

DeMay SM, Becker PA, Rachlow JL, Waits LP. Genetic monitoring of an endangered species recovery: demographic and genetic trends for reintroduced pygmy rabbits (*Brachylagus idahoensis*). J Mammal. 2017;98:350–64. https://doi.org/10.1093/jmammal/gyw197.

Der Sarkissian C, Allentoft ME, Avila-Arcos MC, Barnett R, Campos PF, Cappellini E, et al. Ancient genomics. Phil Trans Roy Soc B Biol Sci. 2015a;370. https://doi.org/10.1098/rstb.2013.0387.

Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol. 2015b;25:2577–83. https://doi.org/10.1016/j.cub.2015.08.032.

Dixon JD, Oli MK, Wooten MC, Eason TH, McCown JW, Paetkau D. Effectiveness of a regional corridor in connecting two Florida black bear populations. Conserv Biol. 2006;20:155–62. https://doi.org/10.1111/j.1523-1739.2006.00292.x.

Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One 2011;6. https://doi.org/10.1371/journal.pone.0027138.

Edwards CJ, Magee DA, Park SDE, McGettigan PA, Lohan AJ, Murphy A, et al. A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). PLoS One 2010;5. https://doi.org/10.1371/journal.pone.0009255.

Eggert LS, Eggert JA, Woodruff DS. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park. Ghana Mol Ecol. 2003;12:1389–402. https://doi.org/10.1046/j.1365-294X.2003.01822.x.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8. https://doi.org/10.1126/science.1162986.

Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl. 2014;7:1026–42. https://doi.org/10.1111/eva.12178.

Ellegren H. DNA typing of museum birds. Nature. 1991;354:113. https://doi.org/10.1038/354113a0.

Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol. 2014;29:51–63. https://doi.org/10.1016/j.tree.2013.09.008.

Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard JM, Poinar HN. Ancient whole genome enrichment using baits built from modern DNA. Mol Biol Evol. 2014;31:1292–4. https://doi.org/10.1093/molbev/msu074.

Epps CW, Palsboll PJ, Wehausen JD, Roderick GK, Ramey RR, McCullough DR. Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. Ecol Lett. 2005;8:1029–38. https://doi.org/10.1111/j.1461-0248.2005.00804.x.

Ernest HB, Rubin ES, Boyce WM. Fecal DNA analysis and risk assessment of mountain lion predation of bighorn sheep. J Wildl Manag. 2002;66:75–85. https://doi.org/10.2307/3802873.

Escobar-Zepeda A, de Leon AVP, Sanchez-Flores A. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Front Genet. 2015;6. https://doi.org/10.3389/fgene.2015.00348.

Fabbri E, Miquel C, Lucchini V, Santini A, Caniglia R, Duchamp C, et al. From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Mol Ecol. 2007;16:1661–71. https://doi.org/10.1111/j.1365-294X.2007.03262.x.

Farrell ED, Carlsson JEL, Carlsson J. Next Gen Pop Gen: implementing a high-throughput approach to population genetics in boarfish (*Capros aper*). R Soc Open Sci. 2016;3. https://doi.org/10.1098/rsos.160651.

Feinstein J. DNA sequence from butterfly frass and exuviae. Conserv Genet. 2004;5:103–4. https://doi.org/10.1023/b:coge.0000014058.34840.94.

Ficetola GF, Miaud C, Pompanon F, Taberlet P. Species detection using environmental DNA from water samples. Biol Lett. 2008;4:423–5. https://doi.org/10.1098/rsbl.2008.0118.

Finch TM, Zhao N, Korkin D, Frederick KH, Eggert LS. Evidence of positive selection in mitochondrial complexes I and V of the African elephant. PLoS One 2014;9. https://doi.org/10.1371/journal.pone.0092587.

Fitak RR, Naidu A, Thompson RW, Culver M. A new panel of SNP markers for the individual identification of North American pumas. J Fish Wildl Manag. 2016;7:13–27. https://doi.org/10.3996/112014-jfwm-080.

Fonseca VG, Sinniger F, Gaspar JM, Quince C, Creer S, Power DM, et al. Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. Sci Rep. 2017;7. https://doi.org/10.1038/s41598-017-06687-x.

Ford MJ, Hanson MB, Hempelmann JA, Ayres KL, Emmons CK, Schorr GS, et al. Inferred paternity and male reproductive success in a killer whale (*Orcinus orca*) population. J Hered. 2011;102:537–53. https://doi.org/10.1093/jhered/esr067.

Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK, Kohn MH. Empirical evaluation of preservation methods for faecal DNA. Mol Ecol. 1998;7:1423–8. https://doi.org/10.1046/j.1365-294x.1998.00449.x.

Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations, and practical recommendations. Mol Ecol. 2017; https://doi.org/10.1111/mec.14264.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96. https://doi.org/10.1016/j.tree.2012.05.012.

Garvin MR, Saitoh K, Gharrett AJ. Application of single nucleotide polymorphisms to non-model species: a technical review. Mol Ecol Res. 2010;10:915–34. https://doi.org/10.1111/j.1755-0998.2010.02891.x.

Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, et al. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science. 2007;317:1927–30. https://doi.org/10.1126/science.1146971.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27:182–9. https://doi.org/10.1038/nbt.1523.

Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). Freshwat Sci. 2013;32:792–800. https://doi.org/10.1899/13-046.1.

Goldberg CS, Strickler KM, Pilliod DS. Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biol Conserv. 2015;183:1–3. https://doi.org/10.1016/j.biocon.2014.11.040.

Gonzalez EG, Blanco JC, Ballesteros F, Alcaraz L, Palomero G, Doadrio I. Genetic and demographic recovery of an isolated population of brown bear *Ursus arctos* L., 1758. PeerJ. 2016;4. https://doi.org/10.7717/peerj.1928.

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 2015;25:1750–6. https://doi.org/10.1101/gr.191395.115.

Graham C, Glenn T, McArthur A, Boreham D, Kieran T, Lance S, et al. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Res. 2015;15:1304–15. https://doi.org/10.1111/755-0998.12404.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of neanderthal DNA. Nature. 2006;444:330–6. https://doi.org/10.1038/nature05336.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, et al. Next-generation museomics disentangles one of the largest primate radiations. Syst Biol. 2013;62:539–54. https://doi.org/10.1093/sysbio/syt018.

Hall LM, Willcox MS, Jones DS. Association of enzyme inhibition with methods of museum skin preparation. Biotechniques 1997;22:928–30, 932–4.

Hawkins MTR, Hofman CA, Callicrate T, McDonough MM, Tsuchiya MTN, Gutierrez EE, et al. In-solution hybridization for mammalian mitogenome enrichment: pros, cons and challenges associated with multiplexing degraded DNA. Mol Ecol Res. 2016;16:1173–88. https://doi.org/10.1111/1755-0998.12448.

Haynes GD, Latch EK. Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus spp*.) using the BovineSNP50 BeadChip. PLoS One 2012;7. https://doi.org/10.1371/journal.pone.0036536.

Hedtke SM, Morgan MJ, Cannatella DC, Hillis DM. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS One 2013;8. https://doi.org/10.1371/journal.pone.0067908.

Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology. 1993;11:1026–30. https://doi.org/10.1038/nbt0993-1026.

Ho SYW, Gilbert MTP. Ancient mitogenomics. Mitochondrion. 2010;10:1–11. https://doi.org/10.1016/j.mito.2009.09.005.

Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, et al. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Res. 2016;16:1264–78. https://doi.org/10.1111/1755-0998.12566.

Hoffman JI, Tucker R, Bridgett SJ, Clark MS, Forcada J, Slate J. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the antarctic fur seal. Mol Ecol Res. 2012;12:861–72. https://doi.org/10.1111/j.1755-0998.2012.03158.x.

Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management. Clim Change Responses 2015;2. https://doi.org/10.1186/s40665-014-0009-x.

Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S. Ancient DNA. Nat Rev Genet. 2001;2:353–9. https://doi.org/10.1038/35072071.

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, et al. The future of ancient DNA: technical advances and conceptual shifts. Bioessays. 2015;37:284–93. https://doi.org/10.1002/bies.201400160.

Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain-reaction product by utilizing the 5′-3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc Natl Acad Sci U S A 1991;88:7276–80. https://doi.org/10.1073/pnas.88.16.7276.

Holman LE, de la Serrana DG, Onoufriou A, Hillestad B, Johnston IA. A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. Aquaculture. 2017;476:59–64. https://doi.org/10.1016/j.aquaculture.2017.04.001.

Hong PY, Mao YJ, Ortiz-Kofoed S, Shah R, Cann I, Mackie RI. Metagenomic-based study of the phylogenetic and functional gene diversity in Galapagos land and marine iguanas. Microb Ecol. 2015;69:444–56. https://doi.org/10.1007/s00248-014-0547-6.

Höss M, Kohn M, Pääbo S, Knauer F, Schröder W. Excrement analysis by PCR. Nature. 1992;359:199. https://doi.org/10.1038/359199a0.

Hung CM, Shaner PJL, Zink RM, Liu WC, Chu TC, Huang WS, et al. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc Natl Acad Sci U S A. 2014;111:10636–41. https://doi.org/10.1073/pnas.1401526111.

Ilmberger N, Gullert S, Dannenberg J, Rabausch U, Torres J, Wemheuer B, et al. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. PLoS One 2014;9. https://doi.org/10.1371/journal.pone.0106707.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014;346:1320–31. https://doi.org/10.1126/science.1253451.

Jensen EL, Edwards DL, Garrick RC, Miller JM, Gibbs JP, Cayot LJ, et al. Population genomics through time provides insights into the consequences of decline and rapid demographic recovery through head-starting in a Galapagos giant tortoise. Evol Appl. 2018a;In press. https://doi.org/10.1111/eva.12682.

Jensen EL, Miller JM, Edwards DL, Garrick RC, Tapia W, Caccone A, Russello MA. Temporal mitogenomics of the Galapagos giant tortoise from Pinzón reveals potential biases in population genetic inference. J Hered. 2018b;In press. https://doi.org/10.1093/jhered/esy016/4963692.

Jerde CL, Mahon AR, Chadderton WL, Lodge DM. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv Lett. 2011;4:150–7. https://doi.org/10.1111/j.1755-263X.2010.00158.x.

Jin X, He MZ, Ferguson B, Meng YH, Ouyang LM, Ren JJ, et al. An effort to use human-based exome capture methods to analyze chimpanzee and macaque exomes. PLoS One 2012;7. https://doi.org/10.1371/journal.pone.0040637.

Jones M, Good J. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25:185–202. https://doi.org/10.1111/mec.13304.

Jørgensen T, Kjær KH, Haile J, Rasmussen M, Boessenkool S, Andersen K, et al. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. Mol Ecol. 2012;21:1980–8. https://doi.org/10.1111/j.1365-294X.2011.05278.x.

Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, et al. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proc Natl Acad Sci U S A. 2015;112:8019–24. https://doi.org/10.1073/pnas.1503283112.

Kendall KC, Stetz JB, Boulanger J, Macleod AC, Paetkau D, White GC. Demography and genetic structure of a recovering grizzly bear population. J Wildl Manag. 2009;73:3–17. https://doi.org/10.2193/2008-330.

Kistler L, Johnson SM, Irwin MT, Louis EE, Ratan A, Perry GH. A massively parallel strategy for STR marker development, capture, and genotyping. Nucleic Acids Res. 2017. https://doi.org/10.1093/nar/gkx574.

Klymus KE, Marshall NT, Stepien CA. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. PLoS One 2017;12. https://doi.org/10.1371/journal.pone.0177643.

Kohn MH, Wayne RK. Facts from feces revisited. Trends Ecol Evol. 1997;12:223–7. https://doi.org/10.1016/s0169-5347(97)01050-1.

Kolby JE, Smith KM, Ramirez SD, Rabemananjara F, Pessier AP, Brunner JL, et al. Rapid response to evaluate the presence of amphibian chytrid fungus (*Batrachochytrium dendrobatidis*) and ranavirus in wild amphibian populations in Madagascar. PLoS One 2015;10. https://doi.org/10.1371/journal.pone.0125330.

Kraus RHS, Vonholdt B, Cocchiararo B, Harms V, Bayerl H, Kuhn R, et al. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. Mol Ecol Res. 2015;15:295–305. https://doi.org/10.1111/1755-0998.12307.

Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, et al. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature. 2006;439:724–7. https://doi.org/10.1038/nature04432.

Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. Bioessays. 2013;35:780–6. https://doi.org/10.1002/bies.201300014.

Lawson LP, Fessl B, Vargas FH, Farrington HL, Cunninghame HF, Mueller JC, et al. Slow motion extinction: inbreeding, introgression, and loss in the critically endangered mangrove finch (*Camarhynchus heliobates*). Conserv Genet. 2017;18:159–70. https://doi.org/10.1007/s10592-016-0890-x.

Leonardi M, Librado P, Sarkissian CD, Schubert M, Alfarhan AH, Alquraishi SA, et al. Evolutionary patterns and processes: lessons from ancient DNA. Syst Biol. 2017;66:E1–E29. https://doi.org/10.1093/sysbio/syw059.

Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362:709–15. https://doi.org/10.1038/362709a0.

Lindqvist C, Schuster SC, Sun YZ, Talbot SL, Qi J, Ratan A, et al. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. Proc Natl Acad Sci U S A. 2010;107:5053–7. https://doi.org/10.1073/pnas.0914266107.

Lonsinger RC, Gese EM, Bailey LL, Waits LP. The roles of habitat and intraguild predation by coyotes on the spatial dynamics of kit foxes. Ecosphere 2017;8. https://doi.org/10.1002/ecs2.1749.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94. https://doi.org/10.1038/nrg1226.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7:111–8. https://doi.org/10.1038/nmeth.1419.

Martinez-Cruz B, Godoy JA, Negro JJ. Population fragmentation leads to spatial and temporal genetic structure in the endangered Spanish imperial eagle. Mol Ecol. 2007;16:477–86. https://doi.org/10.1111/j.1365-294X.2007.03147.x.

Martínkova N, Searle JB. Amplification success rate of DNA from museum skin collections: a case study of stoats from 18 museums. Mol Ecol Notes. 2006;6:1014–7. https://doi.org/10.1111/j.1471-8286.2006.01482.x.

Mason VC, Li G, Helgen KM, Murphy WJ. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome Res. 2011;21:1695–704. https://doi.org/10.1101/gr.120196.111.

Matsui A, Rakotondraparany F, Hasegawa M, Horai S. Determination of a complete lemur mitochondrial genome from feces. Mamm Stud. 2007;32:7–16. https://doi.org/10.3106/1348-6160(2007)32[7:DOACLM]2.0.CO;2.

McCormack JE, Tsai WLE, Faircloth BC. Sequence capture of ultraconserved elements from bird museum specimens. Mol Ecol Res. 2016;16:1189–203. https://doi.org/10.1111/1755-0998.12466.

Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. Mol Phylogenet Evol. 2014;78:314–23. https://doi.org/10.1016/j.ympev.2014.06.003.

Meimberg H, Schachtler C, Curto M, Husemann M, Habel JC. A new amplicon based approach of whole mitogenome sequencing for phylogenetic and phylogeographic analysis: an example of East African white-eyes (Aves, Zosteropidae). Mol Phylogenet Evol. 2016;102:74–85. https://doi.org/10.1016/j.ympev.2016.05.023.

Mendoza MLZ, Sicheritz-Ponten T, Gilbert MTP. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Brief Bioinform. 2015;16:745–58. https://doi.org/10.1093/bib/bbv001.

Metzker ML. Applications of next-generation sequencing sequencing technologies – the next generation. Nat Rev Genet. 2010;11:31–46. https://doi.org/10.1038/nrg2626.

Mikheyev AS, Tin MMY, Arora J, Seeley TD. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. Nat Commun. 2015;6. https://doi.org/10.1038/ncomms8991.

Miller LM, Kapuscinski AR. Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. Genetics. 1997;147:1249–58.

Miller CR, Waits LP. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. Proc Natl Acad Sci U S A. 2003;100:4334–9. https://doi.org/10.1073/pnas.0735531100.

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, et al. Sequencing the nuclear genome of the extinct woolly mammoth. Nature. 2008;456:387–U51. https://doi.org/10.1038/nature07446.

Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, et al. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). Proc Natl Acad Sci U S A. 2011;108:12348–53. https://doi.org/10.1073/pnas.1102838108.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao FQ, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A. 2012;109:E2382–E90. https://doi.org/10.1073/pnas.1210506109.

Miller JM, Malenfant RM, David P, Davis CS, Poissant J, Hogg JT, et al. Estimating genome-wide heterozygosity: effects of demographic history and marker type. Heredity. 2014;112:240–7. https://doi.org/10.1038/hdy.2013.99.

Monzón J, Kays R, Dykhuizen DE. Assessment of coyote-wolf-dog admixture using ancestry-informative diagnostic SNPs. Mol Ecol. 2014;23:182–97. https://doi.org/10.1111/mec.12570.

Moran MA. Metatranscriptomics: eavesdropping on complex microbial communities. Microbiome. 2009;4:329–34.

Morin PA, Woodruff DS. Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In: Martin RD, Dixson AF, Wickings EJ, editors. Paternity in primates: genetic tests and theories. Basel, Switzerland: Karger; 1992. p. 63–81.

Morin PA, Messier J, Woodruff DS. DNA extraction, amplification, and direct sequencing from Hornbill feathers. J Sci Soc Thail. 1994a;20:31–41.

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS. Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science. 1994b;265:1193–201.

Morin PA, Luikart G, Wayne RK, Grp SNPW. SNPs in ecology, evolution and conservation. Trends Ecol Evol. 2004;19:208–16. https://doi.org/10.1016/j.tree.2004.01.009.

Morin PA, Hedrick NM, Robertson KM, Leduc CA. Comparative mitochondrial and nuclear quantitative PCR of historical marine mammal tissue, bone, baleen, and tooth samples. Mol Ecol Notes. 2007;7:404–11. https://doi.org/10.1111/j.1471-8286.2007.01699.x.

Mumma MA, Soulliere CE, Mahoney SP, Waits LP. Enhanced understanding of predator-prey relationships using molecular methods to identify predator species, individual and sex. Mol Ecol Resour. 2014;14:100–8. https://doi.org/10.1111/1755-0998.12153.

Mundy NI, Unitt P, Woodruff DS. Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. Auk. 1997;114:126–9.

Nazarenko IA, Bhatnagar SK, Hohman RJ. A closed tube format for amplification and detection of DNA based on energy transfer. Nucleic Acids Res. 1997;25:2516–21. https://doi.org/10.1093/nar/25.12.2516.

Nguyen TTT, Hayes BJ, Ingram BA. Genetic parameters and response to selection in blue mussel (Mytilus galloprovincialis) using a SNP-based pedigree. Aquaculture. 2014;420:295–301. https://doi.org/10.1016/j.aquaculture.2013.11.021.

Nielsen EE, Hansen MM, Loeschcke V. Analysis of DNA from old scale samples: technical aspects, applications and perspectives for conservation. Hereditas. 1999;130:265–76. https://doi.org/10.1111/j.1601-5223.1999.00265.x.

Norman AJ, Spong G. Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. Ecol Evol. 2015;5:3056–65. https://doi.org/10.1002/ece3.1588.

Nota Y, Takenaka O. DNA extraction from urine and sex identification of birds. Mol Ecol. 1999;8:1237–8. https://doi.org/10.1046/j.1365-294X.1999.00682_2.x.

Ogden R. Unlocking the potential of genomic technologies for wildlife forensics. Mol Ecol Resour. 2011;11:109–16. https://doi.org/10.1111/j.1755-0998.2010.02954.x.

Orlando L, Ginolhac A, Zhang GJ, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013;499:74–8. https://doi.org/10.1038/nature12323.

Orlando L, Gilbert MTP, Willerslev E. Applications of next generation sequencing: reconstructing ancient genomes and epigenomes. Nat Rev Genet. 2015;16:395–408. https://doi.org/10.1038/nrg3935.

Pääbo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A. 1989;86:1939–43. https://doi.org/10.1073/pnas.86.6.1939.

Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, et al. Genetic analyses from ancient DNA. Annu Rev Genet. 2004;38:645–79. https://doi.org/10.1146/annurev.genet.37.110801.143214.

Paijmans JLA, Gilbert MTP, Hofreiter M. Mitogenomic analyses from ancient DNA. Mol Phylogenet Evol. 2013;69:404–16. https://doi.org/10.1016/j.ympev.2012.06.002.

Pansu J, Giguet-Covex C, Ficetola GF, Gielly L, Boyer F, Zinger L, et al. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Mol Ecol. 2015;24:1485–98. https://doi.org/10.1111/mec.13136.

Papadopoulou A, Taberlet P, Zinger L. Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. Mol Ecol. 2015;24:3515–7. https://doi.org/10.1111/mec.13263.

Pearce JM, Fields RL, Scribner KT. Nest materials as a source of genetic data for avian ecological studies. J Field Ornithol. 1997;68:471–81.

Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, et al. Ancient and modern environmental DNA. Phil Trans Roy Soc B Biol Sci. 2015;370. https://doi.org/10.1098/rstb.2013.0383.

Perry GH, Marioni JC, Melsted P, Gilad Y. Genomic-scale capture and sequencing of endogenous DNA from feces. Mol Ecol. 2010;19:5332–44. https://doi.org/10.1111/j.1365-294X.2010.04888.x.

Piaggio AJ, Engeman RM, Hopken MW, Humphrey JS, Keacher KL, Bruce WE, et al. Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. Mol Ecol Res. 2014;14:374–80. https://doi.org/10.1111/1755-0998.12180.

Poinar HN, Hoss M, Bada JL, Paabo S. Amino acid racemization and the preservation of ancient DNA. Science. 1996;272:864–6. https://doi.org/10.1126/science.272.5263.864.

Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science. 2006;311:392–4. https://doi.org/10.1126/science.1123360.

Proctor MF, McLellan BN, Strobeck C, Barclay RMR. Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerably small populations. Proc R Soc B Biol Sci. 2005;272:2409–16. https://doi.org/10.1098/rspb.2005.3246.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505:43–9. https://doi.org/10.1038/nature12886.

Quemere E, Crouau-Roy B, Rabarivola C, Louis EE, Chikhi L. Landscape genetics of an endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. Mol Ecol. 2010;19:1606–21. https://doi.org/10.1111/j.1365-294X.2010.04581.x.

Ragoussis J. Genotyping technologies for genetic research. Annu Rev Genomics Hum Genet. 2009;10:117–33. https://doi.org/10.1146/annurev-genom-082908-150116.

Rebollar EA, Antwis RE, Becker MH, Belden LK, Bletz MC, Brucker RM, et al. Using "omics" and integrated multi-omics approaches to guide probiotic selection to mitigate chytridiomycosis and other emerging infectious diseases. Front Microbiol. 2016;7. https://doi.org/10.3389/fmicb.2016.00063.

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC. The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. J Appl Ecol. 2014;51:1450–9. https://doi.org/10.1111/1365-2664.12306.

Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D. Ancient DNA studies: new perspectives on old samples. Genet Sel Evol. 2012;44. https://doi.org/10.1186/1297-9686-44-21.

Rogaev EI, Moliaka YK, Malyarchuk BA, Kondrashov FA, Derenko MV, Chumakov I, et al. Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. PLoS Biol. 2006;4:403–10. https://doi.org/10.1371/journal.pbio.0040073.

Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M. Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biol. 2007;5:1663–71. https://doi.org/10.1371/journal.pbio.0050207.

Römpler H, Dear PH, Krause J, Meyer M, Rohland N, Schöneberg T, et al. Multiplex amplification of ancient DNA. Nat Protoc. 2006;1:720–8. https://doi.org/10.1038/nprot.2006.84.

Rosenbaum HC, Egan MG, Clapham PJ, Brownell RL, Desalle R. An effective method for isolating DNA from historical specimens of baleen. Mol Ecol. 1997;6:677–81. https://doi.org/10.1046/j.1365-294X.1997.00230.x.

Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, et al. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. Mol Ecol Res. 2011;11:1082–92. https://doi.org/10.1111/j.1755-0998.2011.03052.x.

Rudnick JA, Katzner TE, Bragin EA, Rhodes OE, Dewoody JA. Using naturally shed feathers for individual identification, genetic parentage analyses, and population monitoring in an endangered Eastern imperial eagle (*Aquila heliaca*) population from Kazakhstan. Mol Ecol. 2005;14:2959–67. https://doi.org/10.1111/j.1365-294X.2005.02641.x.

Russello MA, Glaberman S, Gibbs JP, Marquez C, Powell JR, Caccone A. A cryptic taxon of Galapagos tortoise in conservation peril. Biol Lett. 2005;1:287–90. https://doi.org/10.1098/rsbl.2005.0317.

Russello MA, Waterhouse MD, Etter PD, Johnson EA. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ. 2015;3. https://doi.org/10.7717/peerj.1106.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74:5463–7. https://doi.org/10.1073/pnas.74.12.5463.

Schoville SD, Bonin A, Francois O, Lobreaux S, Melodelima C, Manel S. Adaptive genetic variation on the landscape: methods and cases. In: Futuyma DJ, editor. Annual review of ecology, evolution, and systematics. Annual review of ecology evolution and systematics, vol 43. 2012; p. 23–43. https://doi.org/10.1146/annurev-ecolsys-110411-160248.

Schuenemann VJ, Bos K, DeWitte S, Schmedes S, Jamieson J, Mittnik A, et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. Proc Natl Acad Sci U S A. 2011;108:E746–E52. https://doi.org/10.1073/pnas.1105107108.

Schwartz MK, Luikart G, Waples RS. Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol. 2007;22:25–33. https://doi.org/10.1016/j.tree.2006.08.009.

Shapiro B, Hofreiter M. Ancient DNA. Methods and protocols. New York: Springer; 2012. https://doi.org/10.1007/978-1-61779-516-9_1.

Shehzad W, Riaz T, Nawaz MA, Miquel C, Poillot C, Shah SA, et al. Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. Mol Ecol. 2012;21:1951–65. https://doi.org/10.1111/j.1365-294X.2011.05424.x.

Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135–45. https://doi.org/10.1038/nbt1486.

Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat Ecol Evol. 2016;1. https://doi.org/10.1038/s41559-016-0004.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM. Gene mapping in the wild with SNPs: guidelines and future directions. Genetica. 2009;136:97–107. https://doi.org/10.1007/s10709-008-9317-z.

Slater GJ, Cui P, Forasiepi AM, Lenz D, Tsangaras K, Voirin B, et al. Evolutionary relationships among extinct and extant sloths: the evidence of mitogenomes and retroviruses. Genome Biol Evol. 2016;8:607–21. https://doi.org/10.1093/gbe/evw023.

Snyder-Mackler N, Majoros WH, Yuan ML, Shaver AO, Gordon JB, Kopp GH, et al. Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. Genetics. 2016;203:699–714. https://doi.org/10.1534/genetics.116.187492.

Solberg KH, Bellemain E, Drageset OM, Taberlet P, Swenson JE. An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. Biol Conserv. 2006;128:158–68. https://doi.org/10.1016/j.biocon.2005.09.025.

Srivathsan A, Ang A, Vogler AP, Meier R. Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. Front Zool. 2016;13. https://doi.org/10.1186/s12983-016-0150-4.

Stenglein JL, Waits LP, Ausband DE, Zager P, Mack CM. Efficient, noninvasive genetic sampling for monitoring reintroduced wolves. J Wildl Manag. 2010;74:1050–8. https://doi.org/10.2193/2009-305.

Steyer K, Kraus RHS, Molich T, Anders O, Cocchiararo B, Frosch C, et al. Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). Conserv Genet. 2016;17:1183–99. https://doi.org/10.1007/s10592-016-0853-2.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next generation sequencing for plant systematics. Am J Bot. 2012;99:349–64. https://doi.org/10.3732/ajb.1100335.

Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, et al. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0151651.

Suez M, Behdenna A, Brouillet S, Graca P, Higuet D, Achaz G. MicNeSs: genotyping microsatellite loci from a collection of (NGS) reads. Mol Ecol Res. 2016;16:524–33. https://doi.org/10.1111/1755-0998.12467.

Sugiyama Y, Kawamoto S, Takenaka O, Kumazaki K, Miwa N. Paternity discrimination and intergroup relationships of chimpanzees at Bossou. Primates. 1993;34:545–52. https://doi.org/10.1007/bf02382665.

Taberlet P, Bouvet J. A single plucked feather as a source of DNA for bird genetic studies. Auk. 1991;108:959–60.

Taberlet P, Bouvet J. Bear conservation genetics. Nature. 1992;358:197.

Taberlet P, Fumagalli L. Owl pellets as a source of DNA for genetic studies of small mammals. Mol Ecol. 1996;5:301–5. https://doi.org/10.1111/j.1365-294X.1996.tb00318.x.

Taberlet P, Bonin A, Zinger L, Coissac E. Environmental DNA – biodiversity research and monitoring. Oxford: Oxford University Press; 2018. https://doi.org/10.1093/oso/9780198767220.001.0001.

Thomsen PF, Willerslev E. Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv. 2015;183:4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

Thomsen PF, Kielgast J, Iversen LL, Moller PR, Rasmussen M, Willerslev E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS One 2012;7. https://doi.org/10.1371/journal.pone.0041732.

Tin MMY, Economo EP, Mikheyev AS. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. PLoS One 2014;9. https://doi.org/10.1371/journal.pone.0096793.

Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol. 2009;24:110–7. https://doi.org/10.1016/j.tree.2008.09.011.

Väli U, Einarsson A, Waits L, Ellegren H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol. 2008;17:3808–17. https://doi.org/10.1111/j.1365-294X.2008.03876.x.

Valiere N, Taberlet P. Urine collected in the field as a source of DNA for species and individual identification. Mol Ecol. 2000;9:2150–2. https://doi.org/10.1046/j.1365-294x.2000.01114-2.x.

Valiere N, Fumagalli L, Gielly L, Miquel C, Lequette B, Poulle ML, et al. Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. Anim Conserv. 2003;6:83–92. https://doi.org/10.1017/s1367943003003111.

Vallender EJ. Expanding whole exome resequencing into non-human primates. Genome Biol. 2011;12:1–10. https://doi.org/10.1186/gb-2011-12-9-r87.

Valsecchi E, Glockner-Ferrari D, Ferrari M, Amos W. Molecular analysis of the efficiency of sloughed skin sampling in whale population genetics. Mol Ecol. 1998;7:1419–22. https://doi.org/10.1046/j.1365-294x.1998.00446.x.

Vartia S, Villanueva-Canas JL, Finarelli J, Farrell ED, Collins PC, Hughes GM, et al. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. R Soc Open Sci. 2016;3. https://doi.org/10.1098/rsos.150565.

Veale AJ, Russello MA. Sockeye salmon repatriation leads to population re-establishment and rapid introgression with native kokanee. Evol Appl. 2016;9:1301–11. https://doi.org/10.1111/eva.12430.

von Holdt BM, Pollinger JP, Earl DA, Knowles JC, Boyko AR, Parker H, et al. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. 2011;21:1294–305. https://doi.org/10.1101/gr.116301.110.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, et al. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol Ecol. 2013;22:787–98. https://doi.org/10.1111/mec.12023.

Waits LP, Paetkau D. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. J Wildl Manag. 2005;69:1419–33. https://doi.org/10.2193/0022-541x(2005)69[1419:ngstfw]2.0.co;2.

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM. Patterns of nuclear DNA degeneration over time – a case study in historic teeth samples. Mol Ecol. 2003;12:1087–93. https://doi.org/10.1046/j.1365-294X.2003.01807.x.

Wandeler P, Hoeck PEA, Keller LF. Back to the future: museum specimens in population genetics. Trends Ecol Evol. 2007;22:634–42. https://doi.org/10.1016/j.tree.2007.08.017.

Wasserman TN, Cushman SA, Schwartz MK, Wallin DO. Spatial scaling and multi-model inference in landscape genetics: *Martes americana* in northern Idaho. Landsc Ecol. 2010;25:1601–12. https://doi.org/10.1007/s10980-010-9525-7.

Wheat RE, Allen JM, Miller SDL, Wilmers CC, Levi T. Environmental DNA from residual saliva for efficient noninvasive genetic monitoring of brown bears (*Ursus arctos*). PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0165259.

Willerslev E, Gilbert MTP, Binladen J, Ho SYW, Campos PF, Ratan A, et al. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. BMC Evol Biol. 2009;9. https://doi.org/10.1186/1471-2148-9-95.

Williams CL, Blejwas K, Johnston JJ, Jaeger MM. A coyote in sheep's clothing: predator identification from saliva. Wildl Soc Bull. 2003;31:926–32.

Woodruff SP, Lukacs PM, Christianson D, Waits LP. Estimating Sonoran pronghorn abundance and survival with fecal DNA and capture-recapture methods. Conserv Biol. 2016;30:1102–11. https://doi.org/10.1111/cobi.12710.

Woods JG, Paetkau D, Lewis D, McLellan BN, Proctor M, Strobeck C. Genetic tagging of free-ranging black and brown bears. Wildl Soc Bull. 1999;27:616–27.

Wooley JC, Godzik A, Friedberg I. A primer on metagenomics. PLoS Comput Biol 2010;6. https://doi.org/10.1371/journal.pcbi.1000667.

Wultsch C, Waits LP, Kelly MJ. A Comparative analysis of genetic diversity and structure in jaguars (*Panthera onca*), pumas (*Puma concolor*), and ocelots (*Leopardus pardalis*) in fragmented landscapes of a critical Mesoamerican linkage zone. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0151043.

Xu B, Xu WJ, Li JJ, Dai LM, Xiong CY, Tang XH, et al. Metagenomic analysis of the *Rhinopithecus bieti* fecal microbiome reveals a broad diversity of bacterial and glycoside hydrolase profiles related to lignocellulose degradation. BMC Genomics 2015;16. https://doi.org/10.1186/s12864-015-1378-7.

Zhan LY, Paterson IG, Fraser BA, Watson B, Bradbury IR, Ravindran PN, et al. MEGASAT: automated inference of microsatellite genotypes from sequence data. Mol Ecol Res. 2017;17:247–56. https://doi.org/10.1111/1755-0998.12561.

Zhu LF, Wu Q, Dai JY, Zhang SN, Wei FW. Evidence of cellulose metabolism by the giant panda gut microbiome. Proc Natl Acad Sci U S A. 2011;108:17714–9. https://doi.org/10.1073/pnas.1017956108.

# Environmental Population Genomics: Challenges and Opportunities


**Caren S. Goldberg and Meghan B. Parsley**

**Abstract** Genomic material originating from macroorganisms and collected in environmental samples has the potential to be used for population and community genomic analyses, yielding insights into metrics such as population diversity, functional connectivity, adaptive variation, and age structure. Fractionation studies indicate that the size of environmental DNA (eDNA) from fishes is large enough to contain nuclei or cell fragments and pilot work indicates that nuclear single nucleotide polymorphisms can be amplified from eDNA samples, providing evidence that genomic analysis of eDNA samples is possible. Additionally, environmental RNA (eRNA) may be more persistent in environmental samples than previously thought. To use eDNA or eRNA for insights into population and community genomics or transcriptomics, collection methods can be designed to focus on the individual genotype or on collecting a representative sample of the population or community. In addition to the technical challenges of collecting, preserving, and analyzing these materials, differences in genomic production among individuals and validation of marker sets specific to the target species or community are required. In this chapter we focus on what genomic information may be able to be harvested from environmental samples and how this material may be distributed in the environment, as well as explore approaches for how sampling design could be used to gain insights into populations and communities.

**Keywords** eDNA · Environmental DNA · eRNA · Noninvasive sampling · Population genomics

C. S. Goldberg (✉) · M. B. Parsley
School of the Environment, Washington State University, Pullman, WA, USA
e-mail: caren.goldberg@wsu.edu

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_75, 

## 1 What is eDNA?

Environmental DNA (eDNA) is genomic material captured by sampling soil, water, air, or other aspects of the environment to detect macroorganisms (Rees et al. 2014). The field of eDNA research has focused so far on detecting species based on their mitochondrial DNA (mtDNA) sequences using quantitative PCR or metabarcoding (Thomsen and Willerslev 2015). However, there is evidence that additional genomic information (e.g., nuclear DNA, environmental RNA (eRNA)) is available in environmental samples that could yield insights into population diversity, functional connectivity, adaptive variation, and age structure. Environmental DNA analysis methods are similar to those for ancient and noninvasive genetic samples, where discrete materials from individuals are analyzed (bone, hair, scat; Thomsen and Willerslev 2015). However, inference from eDNA beyond species detection is additionally challenged by the mixing of genomic materials from individuals of the same species (Andrews et al. 2018). Using eDNA for population genetics has been mentioned in the literature (Barnes and Turner 2016) with many technical challenges detailed recently by Adams et al. (2019). Here we focus on what genomic information may be able to be harvested from environmental samples, how this material may be distributed in the environment, and how we can design sampling to gain information about populations and communities.

The physical properties (i.e. "state" sensu Barnes and Turner 2016) of eDNA are of primary concern when inferring the amount, quality, and type of genomic material that can be captured from environmental samples. Fractionation studies have demonstrated that most aquatic eDNA shed from fish is contained in particles between 1.2 and 10 μm (Turner et al. 2014; Wilcox et al. 2015), consistent with the size of a nucleus or cell fragments, and that particles $>10$ μm decay quickly after shedding (Jo et al. 2019). Additional evidence that eDNA is made up of more than small fragments comes from the sequencing of whole mitochondria from eDNA samples collected using filtration and 1.2 μm polycarbonate track etch filters (Deiner et al. 2017). Further work indicates that genotyping single nucleotide polymorphisms (SNPs) from eDNA is possible if concentrations are high enough (Box 1). However, it is also possible that eDNA may be free DNA that is bound to particles in the water column, with occasional whole mitochondria.

Mitochondrial DNA has largely been the focus of eDNA work so far as it is the most abundant genomic material in animal cells and is well-characterized for many species. Sequences from mtDNA are often used as barcodes for species because of the lack of within-species variation in some sections (Hajibabaei et al. 2007). For some species, variation within mtDNA can be informative; mtDNA haplotypes for whale sharks have been recovered from seawater and were similar in frequency to

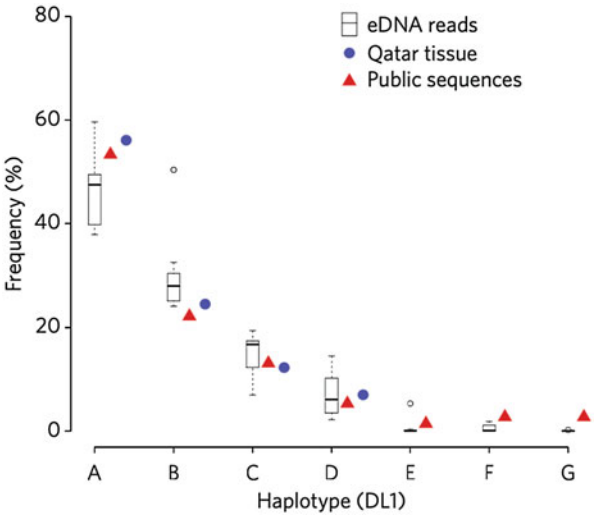


**Fig. 1** eDNA has the potential to provide estimates of allele frequencies in natural populations: frequencies of mitochondrial DNA haplotypes in eDNA samples match estimates from tissue samples and from public sequence data in whale sharks (*Rhincodon typus*). Reproduced with permission from Sigsgaard et al. (2016)

that characterized for the population from tissue samples (Fig. 1; Sigsgaard et al. 2016). Additionally, mtDNA preserved in sediments can be used to study biodiversity through time (Bálint et al. 2018). Methods for mtDNA capture and analysis from eDNA samples are now well-characterized (Goldberg et al. 2016), methods for characterization of RNA are nascent (Pochon et al. 2017; Ammon et al. 2019), and published nuclear DNA analysis from natural systems has so far been limited to repetitive regions (18S and ITS; Dysthe et al. 2018; Stat et al. 2017). In aquaria, degradation rates of nuclear DNA did not differ from that of mtDNA (Bylemans et al. 2018), indicating that differences in production and increased ability to detect rare particles may be the key to genotyping nuclear DNA from aquatic eDNA samples. In contrast, nuclear DNA may be more difficult to recover from soil samples (Emmons et al. 2017). If a small number of eDNA samples could represent the diversity and allele frequencies of nuclear DNA in populations, this would open a new avenue for sampling large numbers of populations noninvasively and inexpensively.

**Box 1 Genotyping SNPs from eDNA for Chinook**

Data from collaboration between Caren S. Goldberg (Washington State University) and Shawn Narum (Columbia River Inter-Tribal Fish Commission)

To determine whether we could distinguish environmental DNA (eDNA) signals of the federally endangered Upper Columbia Spring Chinook salmon (*Oncorhynchus tshawytscha*) evolutionarily significant unit (ESU) from those of the non-listed Upper Columbia Summer-/Fall-run Chinook Salmon ESU, we applied single nucleotide polymorphism (SNP) analysis to a set of eDNA samples testing positive for Chinook (Laramie et al. 2015). To determine the panel of SNPs that would only amplify for Chinook, we analyzed a set of tissue samples from Chinook and nontarget species (brown trout, *Salmo trutta*; bull trout, *S. confluentus*; brook trout, *Salvelinus fontinalis*; cutthroat trout, *O. clarkii*; coho salmon, *O. kisutch*; rainbow trout, *O. mykiss*; sockeye salmon, *O. nerka*; mountain whitefish, *Prosopium williamsoni*) with the 192-SNP panel described in Warheit et al. (2013). From this set, we identified those that were found to be the most distinct between these ESUs (i.e., allele frequencies within both lineages furthest from 0.50 and in different directions from each other) and tested them against a panel of co-occurring salmonids. From these results, we identified 5 SNPs that were unique to Chinook to analyze eDNA samples, with fragment lengths up to 1,118 base pairs.

We genotyped 27 eDNA samples for 5 SNPs using the Type-it Fast SNP Probe PCR Kit (Qiagen, Inc.). Reactions were 10 µl in volume with 1X Type-it Fast SNP Probe PCR Master Mix, 0.2 µM of each primer, and 0.2 µM of each probe and each included 2 µl of sample. Cycling began with 15 min at 95 °C followed by 50 cycles of 95 °C for 15 s and 60 °C for 30 s on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories). Samples were run in triplicate with an allele only counted if it amplified in two or more of the reactions. We used overall genotype frequencies from Hess et al. (2011) to calculate the probability of each genotype being produced by each run of salmon and considered the run with the higher frequency to be the one producing the genotype. This assumes that only one run was dominant enough to produce the genotype detected. In a true mixed stock run, additional analytical processes would need to be developed to calculate the probability that each run was contributing to the sample but may not account for all alleles detected.

We obtained enough genotype information ($\geq$4 markers) to calculate probabilities of spring-run Chinook presence at eight sampled locations (Fig. 2; some locations had multiple samples collected). Within this set, we detected spring-run Chinook salmon in all but one location where they were thought to be located (Fig. 2). For both sites where one marker dropped out, the marker was 313 base pairs. This indicates it is possible to genotype nuclear

(continued)

**Box 1** (continued)
DNA from eDNA samples if concentrations are high, but collection, preservation, and/or analysis methods would have to be improved before this method could be applied.

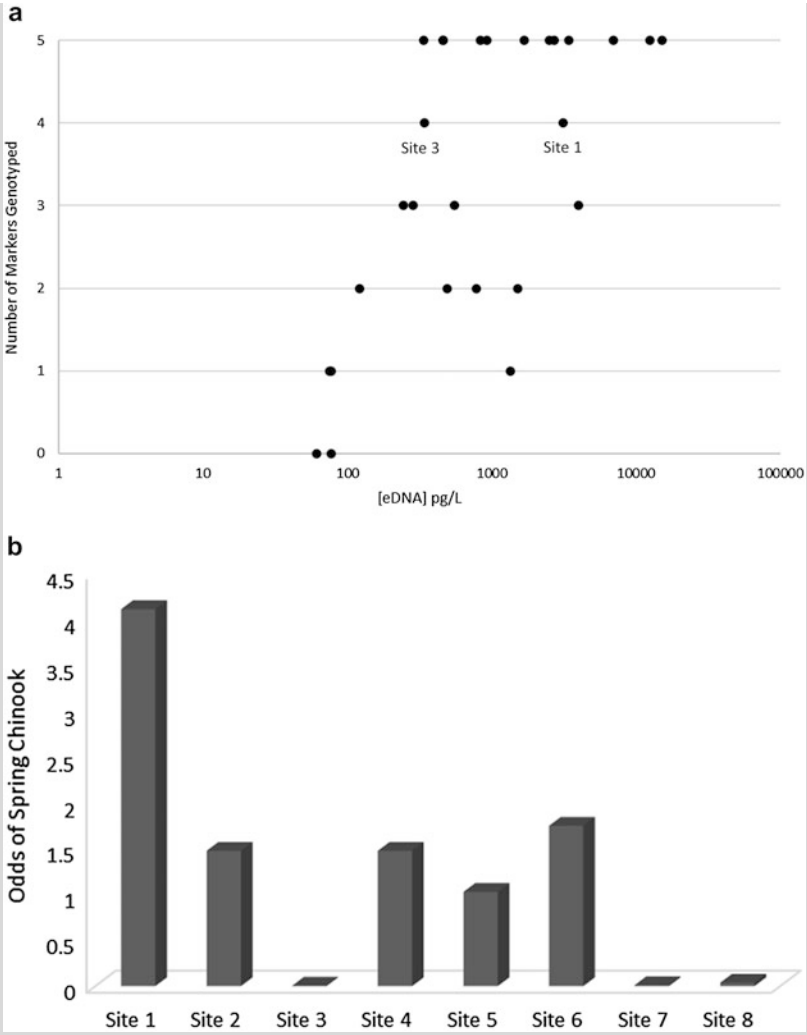


**Fig. 2** (**a**) Relationship between Chinook eDNA concentration in sample replicate and the number of single nucleotide polymorphism markers that generated data for run-specific analysis. A minimum of four markers was required for run identification; sites at this level are identified. (**b**) Odds of genotype from eDNA sample of Chinook being sourced from spring rather than fall-run Chinook. A value of 1 indicates no information on whether the sample is from spring or fall Chinook; a value of 4 indicates that the eDNA from that site was four times more likely to be sourced from spring Chinook. Sites with known spring and fall Chinook (Sites 1–6 and Sites 7–8, respectively) were correctly identified with eDNA, with the exception of Site 3

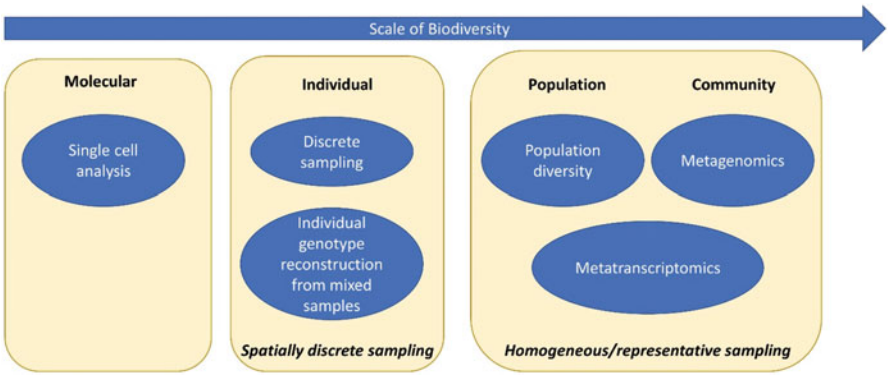


**Fig. 3** Potential approaches to sampling for population and community genomic analysis of environmental DNA

## 2 Environmental Genomic Information for Insights Across Scales of Biodiversity

The analysis of environmental genomic material has the potential to provide insights at multiple scales of biodiversity (Fig. 3). At the molecular level, methods used for the analysis of a single cell could potentially be applied to environmental samples (detailed in Adams et al. 2019), if single cells can be harvested from large volumes of water. At the individual level, it may be possible to reconstruct individual genotypes using single samples in soil or water for species with discrete space use (e.g., burrows). Another option for species where space use of individuals is distinct but overlapping (i.e., not schooling behavior) is to use the spatial information to reconstruct genotypes from mixed samples. Alternatively, a homogeneous or representative sample of the area could provide information at the population or community level. In theory, population diversity estimates could be obtained from a representative sample (similar to pool-seq; Schlötterer et al. 2014), without the complexity of reconstructing individual genotypes. Finally, a community diversity approach can be taken, looking at gene expression from eRNA or genotypes across species (metagenomics, metatranscriptomics). Below we describe the potential avenues for conducting these analyses with environmental samples as well as their associated challenges.

## 3 Individual Genotypes from eDNA Samples

Individual genotypes may potentially be inferred at the molecular or individual level using single-cell analysis or spatially informed eDNA sampling. If eDNA exists as whole cells or nuclei, one individual-based approach would be to harvest whole cells

from water samples for genotyping. This would require a change in sample collection protocols as filtering likely causes cells to burst (Thomas et al. 2018) thus mixing the sample. If whole cells can be harvested, this would provide individual genomes and avoid the issue of inter- and intra-specific mixing of samples that is a major challenge in genomic analysis of eDNA. However, methods have not yet been developed to harvest these rare individual cells from large water samples. The current state of the field for this application is detailed in Adams et al. (2019).

Another approach to population genomic analysis of environmental samples would be to use the spatial location of samples to collect or reconstruct individual genotypes. In the simplest case, non-overlapping space use by animals may present an opportunity for individual-level sampling. For example, if a single crayfish occupies a burrow, crayfish eDNA from the soil or water around that burrow should be highly dominated by the genotype from that individual. Similarly, mtDNA has successfully been obtained from snow tracks of mammalian species (Franklin et al. 2019; Kinoshita et al. 2019), though these methods have not yet been optimized for the collection of nuclear eDNA for population genomic analysis. If a single sample from those substrates could be used to obtain an individual genotype, this method could be used in standard population genomic analysis, similar to hair or scat (Andrews et al. 2018). However, this would only work for species with discrete space use and requires a detailed understanding of the spatial ecology and habitat of the target species.

A third potential approach would be to use the unique spatial-temporal signature of an individual to probabilistically recreate genotypes from a uniform distribution of spatial samples. This assumes that the DNA of any individual has the greatest concentration closest to the location of the individual when sampled and decreases with distance. With many individual spatial samples, it may be possible to probabilistically estimate the number of individuals contributing DNA to the environment and reconstruct their genotypes, providing the basis for genomic analyses at the individual level. In contrast to conditions for collecting mixed samples, this approach would likely work better in systems with very limited hydrological movement (i.e., wetlands) and with more territorial or sedentary species. Additionally, DNA copy number of target genes would have to be accurately estimated, and the number of samples necessary to collect from each location may be cost-prohibitive. For this approach, developing validated marker sets that do not cross-amplify with co-occurring species would be required. For example, if a SNP for a target species was fixed in a co-occurring congener, results using that marker would be highly biased. This may require additional marker development, even for species where informative markers for within-species analysis are well-characterized (Box 1). This challenge will increase in more diverse systems and would need to be addressed using tissue-based validations, requiring well-curated tissue libraries without the low levels of cross-contamination common in tissue collections and extractions.

## 4 Population and Community Genomics from eDNA Samples

An alternative approach to focusing on individual genotypes is to take advantage of the mixing of DNA in aquatic systems to collect samples representing a whole population or community (i.e., metagenomics). Allele frequencies could then be estimated using read depth as a surrogate for copy number (Deagle et al. 2019). Collecting a representative sample is the challenge for this approach, as eDNA typically does not travel far from its source (Dunker et al. 2016; Port et al. 2016; Fremier et al. 2019) and sampling at any one location will yield a biased estimate of population diversity. Therefore, an integrated sample through space is necessary, either by combining samples collected from many points or (for aquatic systems) through continuous sampling (e.g., Thomas et al. 2019). Lotic systems are likely to already be better mixed than lentic systems, and a representative sample for a reach may be able to be collected from the thalweg of a stream (the length of inference would need to be estimated using hydrological models; Fremier et al. 2019; Song et al. 2017). Additionally, species that regularly travel across the environment being sampled will already provide for a more mixed sample than more sedentary or territorial species. Inference is further complicated by the variability of eDNA production rates across individuals, with some "super-shedders" producing 105× more eDNA than same-aged conspecifics (Barnes et al. 2014; Klymus et al. 2015; Strickler et al. 2015; Fig. 4). The genotype of these individuals may be overestimated in a sample, providing biased allele frequency results.

Some of the challenges of spatial inference and sampling have already been encountered in the characterization of communities through metabarcoding. For example, samples taken in marine systems reflected communities within 60 m (Port et al. 2016), indicating a challenge for inferring the spatial extent of community estimates. Community diversity metrics that use read count to estimate abundance have found significant but noisy correlations, potentially due to different levels of

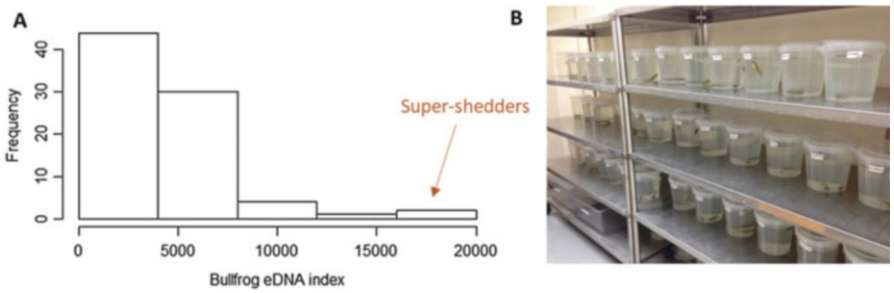


**Fig. 4** (**a**) Histogram of production of environmental DNA (eDNA) by American bullfrog tadpoles housed in microcosms. (**b**) Animals were sourced from the same location; water source, feeding, and care were identical among tanks. Tanks were then exposed to different treatments in an experiment of eDNA degradation (Strickler et al. 2015)

production among individuals or species, as well as the effects of primer bias (Kelly et al. 2014; Hänfling et al. 2016). At the population level, analyzing pooled extracts of many individuals to estimate population allele frequencies is an approach taken to reduce costs (pool-seq) but works best with large numbers of individuals (50–100; Schlötterer et al. 2014), and sequences of target species would have to be enriched prior to analysis, for example, through methylation-based enrichment (Chiou and Bergey 2018) or targeted baits (Andrews et al. 2018). Additionally, finding markers that do not amplify across species would still be required for population-level analysis. At the community level, markers would have to be unbiased across the group of target organisms.

If we can generate allele frequencies for populations from representative eDNA samples, it opens the toolbox of population genomic analyses, including studies of gene flow, diversity, adaptation, phylogeography, and population history. The ability to visit a system and collect water or soil rather than capturing and sampling 20–30 individuals would greatly increase the number of populations that can be sampled and analyzed for the same costs. Additionally, collecting environmental samples could reduce impacts to species and habitat compared to traditional sampling methods. However, the challenges of sampling in relation to space use and accounting for differences in eDNA production rates would have to be addressed in order to use mixed environmental samples for population or community genomics. In addition, the time required to develop and validate suites of species-specific markers suitable for mixed environmental samples may delay the application of these methods.

## 5 Environmental RNA for Increased Insights into Population and Community Genomics

Genomic material that can be analyzed in environmental samples to provide insights into population or community ecology and genomics may extend beyond DNA. For example, RNA decays more rapidly than DNA (Stoeck et al. 2007) and could alleviate the problem of false positives from deceased organisms or allochthonous DNA input in systems (Goldberg et al. 2016). Environmental RNA may provide a more accurate indicator of population size from environmental samples due to the absence of exogenous input from dead individuals or outside sources that can bias eDNA-based estimates (Tillotson et al. 2018). The detection of RNA from organisms in environmental samples has been applied in other fields such as the detection of pathogens (Girones et al. 2010) as well as characterizing changes in transcript profiles of marine microbial communities (Ottesen et al. 2011). Though several studies have been conducted using eRNA in other contexts, methodology and applications are lacking for macroorganisms. Preliminary work metabarcoding and targeting mitochondrial RNA from marine samples indicates that eRNA of

metazoans may be recoverable from environmental samples (Pochon et al. 2017; Ammon et al. 2019), which could be used to provide valuable information at the population or community level.

Environmental RNA could be used to gain population-level information by measuring variation in gene expression between populations, within species, and across communities. RNA products produced during specific life stages or eRNA: eDNA ratios at specific times during development could be indicative of age structure or reproductive activity of populations. Environmental RNA signatures of organisms in specific physiological conditions, such as reproductive activity, could be used to differentiate between breeding and nonbreeding sites, a major goal of endangered species monitoring programs. Finally, the detection of specific eRNA products could be used to estimate gene expression at the population or community level. This could help inform population or community responses to common stressors such as pesticides or other changes in environmental conditions such as global climate change (Adams et al. 2019; Cristescu 2019).

Despite its promise, there are many challenges for using eRNA to investigate populations. The high degradation rates that could make RNA useful for distinguishing between current and legacy positive signals also mean that detection of RNA could be very challenging. However, recent evidence from other fields such as paleogenetics, cellular biology, and functional genomics provide support for abundant excretion of RNA into the environment and more stability in the environment than was previously thought (Cristescu 2019). Environmental RNA analysis will also require further transcriptomic and genomic resources to target-specific gene products. Methods for sample collection, preservation, and extraction will need to be developed and optimized to deal with eRNA degradation and specificity of gene products from non-model species in order to move the field of eRNA forward.

## 6 Conclusion

The explosion of studies in the field of environmental genetic analysis in the last decade is highly promising for increasing our understanding of ecology and evolution. It appears that detection of species using mitochondrial sequences in water or soil is just the beginning for the information that we may be able to obtain from environmental samples. By combining improved molecular techniques for collection, preservation, and analysis of samples with knowledge of the spatial ecology of target species, we have the potential to revolutionize how we learn about populations and communities.

## References


Adams CIM, Knapp M, Gemmell NJ, Jeunen G-J, Bunce M, Lamare MD, et al. Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool? Genes. 2019;10:192. https://doi.org/10.3390/genes10030192.

Ammon U, Wood SA, Laroche O, Zaiko A, Lavery SD, Inglis GJ, et al. Linking environmental DNA and RNA for improved detection of the marine invasive fanworm *Sabella spallanzanii*. Front Mar Sci. 2019;6:621. https://doi.org/10.3389/fmars.2019.00621.

Andrews KR, De Barba M, Russello MA, Waits LP. Advances in using non-invasive, archival, and environmental samples for population genomic studies. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_45.

Bálint M, Pfenninger M, Grossart HP, Taberlet P, Vellend M, Leibold MA, et al. Environmental DNA time series in ecology. Trends Ecol Evol. 2018;33:945–57. https://doi.org/10.1016/j.tree.2018.09.003.

Barnes MA, Turner CR. The ecology of environmental DNA and implications for conservation genetics. Conserv Genet. 2016;17:1–17. https://doi.org/10.1007/s10592-015-0775-4.

Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. Environmental conditions influence eDNA persistence in aquatic systems. Environ Sci Technol. 2014;48:1819–27. https://doi.org/10.1021/es404734p.

Bylemans J, Furlan EM, Gleeson DM, Hardy CM, Duncan RP. Does size matter? An experimental evaluation of the relative abundance and decay rates of aquatic environmental DNA. Environ Sci Technol. 2018;52:6408–16. https://doi.org/10.1021/acs.est.8b01071.

Chiou KL, Bergey CM. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci Rep. 2018;8:1–10. https://doi.org/10.1038/s41598-018-20427-9.

Cristescu ME. Can environmental RNA revolutionize biodiversity science? Trends Ecol Evol. 2019;34:694–7. https://doi.org/10.1016/j.tree.2019.05.003.

Deagle BE, Thomas AC, McInnes JC, Clarke LJ, Vesterinen EJ, Clare EL, et al. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? Mol Ecol. 2019;28:391–406. https://doi.org/10.1111/mec.14734.

Deiner K, Renshaw MA, Li Y, Olds BP, Lodge DM, Pfrender ME. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. Methods Ecol Evol. 2017;8:1888–98. https://doi.org/10.1111/2041-210X.12836.

Dunker KJ, Sepulveda AJ, Massengill RL, Olsen JB, Russ OL, Wenburg JK, et al. Potential of environmental DNA to evaluate northern pike (*Esox lucius*) eradication efforts: an experimental test and case study. PLoS One. 2016;11:1–21. https://doi.org/10.1371/journal.pone.0162277.

Dysthe JC, Franklin TW, McKelvey KS, Young MK, Schwartz MK. An improved environmental DNA assay for bull trout (*Salvelinus confluentus*) based on the ribosomal internal transcribed spacer. PLoS One. 2018;13:1–11. https://doi.org/10.1371/journal.pone.0206851.

Emmons AL, DeBruyn JM, Mundorff AZ, Cobaugh KL, Cabana GS. The persistence of human DNA in soil following surface decomposition. Sci Justice. 2017;57:341–8. https://doi.org/10.1016/j.scijus.2017.05.002.

Franklin TW, McKelvey KS, Golding JD, Mason DH, Dysthe C, Pilgrim KL, et al. Using environmental DNA methods to improve winter surveys for rare carnivores: DNA from snow and improved noninvasive techniques. Biol Conserv. 2019;229:50–8. https://doi.org/10.1016/j.biocon.2018.11.006.

Fremier AK, Strickler KM, Parzych J, Powers S, Goldberg CS. Stream transport and retention of environmental DNA pulse releases in relation to hydrogeomorphic scaling factors. Environ Sci Technol. 2019;53:6640–6649. https://doi.org/10.1021/acs.est.8b06829.

Girones R, Ferrus MA, Alonso JL, Rodriguez-Manzano J, Calgua B, Corrêa AA, et al. Molecular detection of pathogens in water – the pros and cons of molecular techniques. Water Res. 2010;44:4325–39. https://doi.org/10.1016/j.watres.2010.06.030.

Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, et al. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol. 2016;7:1299–1307. https://doi.org/10.1111/2041-210X.12595.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 2007;23:167–72. https://doi.org/10.1016/j.tig.2007.02.001.

Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, et al. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Mol Ecol. 2016;25:3101–19. https://doi.org/10.1111/mec.13660.

Hess J, Matala A, Zendt J, Fredericksen C, Sharp W, Narum S. Introgressive hybridization between major Columbia River Chinook salmon lineages within the Klickitat River due to hatchery practices. Can J Fish Aquat Sci. 2011;68:1876–1891. https://doi.org/10.1139/f2011-107.

Jo T, Murakami H, Yamamoto S, Masuda R, Minamoto T. Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. Ecol Evol. 2019;9:1135–46. https://doi.org/10.1002/ece3.4802.

Kelly RP, Port JA, Yamahara KM, Crowder LB. Using environmental DNA to census marine fishes in a large mesocosm. PLoS One. 2014;9:e86175. https://doi.org/10.1371/journal.pone.0086175.

Kinoshita G, Yonezawa S, Murakami S, Isagi Y. Environmental DNA collected from snow tracks is useful for identification of mammalian species. Zoolog Sci. 2019;36:198–207. https://doi.org/10.2108/zs180172.

Klymus KE, Richter CA, Chapman DC, Paukert C. Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. Biol Conserv. 2015;183:77–84. https://doi.org/10.1016/j.biocon.2014.11.020.

Ottesen EA, Marin R III, Preston CM, Young CR, Ryan JP, Scholin CA, et al. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. ISME J. 2011;5:1881–95. https://doi.org/10.1038/ismej.2011.70.

Pochon X, Zalko A, Fletcher LM, Laroche O, Wood SA. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. PLoS One. 2017;12:e0187636. https://doi.org/10.1371/journal.pone.0187636.

Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, et al. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. Mol Ecol. 2016;25:527–41. https://doi.org/10.1111/mec.13481.

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC. The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. J Appl Ecol. 2014;51:1450–9. https://doi.org/10.1111/1365-2664.12306.

Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15:749–63. https://doi.org/10.1038/nrg3803.

Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat Ecol Evol. 2016;1:0004. https://doi.org/10.1038/s41559-016-0004.

Song JW, Small MJ, Casman EA. Making sense of the noise: the effect of hydrology on silver carp eDNA detection in the Chicago area waterway system. Sci Total Environ. 2017;605–606:713–20. https://doi.org/10.1016/j.scitotenv.2017.06.255.

Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, et al. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Sci Rep. 2017;7:1–11. https://doi.org/10.1038/s41598-017-12501-5.

Stoeck T, Zuendorf A, Breiner HS, Behnke A. A molecular approach to identify active microbes in environmental eukaryote clone libraries. Microb Ecol. 2007;53:328–39. https://doi.org/10.1007/s00248-006-9166-1.

Strickler KM, Fremier AK, Goldberg CS. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biol Conserv. 2015;183:85–92. https://doi.org/10.1016/j.biocon.2014.11.038.

Thomas AC, Howard J, Nguyen PL, Seimon TA, Goldberg CS. ANDe™: a fully integrated environmental DNA sampling system. Methods Ecol Evol. 2018;9:1379–85. https://doi.org/10.1111/2041-210X.12994.

Thomas AC, Tank S, Nguyen PL, Ponce J, Sinnesael M, Goldberg CS. A system for rapid eDNA detection of aquatic invasive species. Environ DNA. 2019. https://doi.org/10.1002/edn3.25.

Thomsen PF, Willerslev E. Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv. 2015;183:4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

Tillotson MD, Kelly KP, Duda JJ, Hoy M, Kralj J, Quinn TP. Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. Biol Conserv. 2018;220:1–11. https://doi.org/10.1016/j.biocon.2018.01.030.

Turner CR, Barnes MA, Xu CCY, Jones SE, Jerde CL, Lodge DM. Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods Ecol Evol. 2014;5:676–84. https://doi.org/10.1111/2041-210X.12206.

Warheit KI, Seeb L, Templin WD, Seeb J Moving GSI into the next decade: SNP coordination for Pacific Salmon Treaty Fisheries. 2013. Publication number FPT-13-09. https://wdfw.wa.gov/sites/default/files/publications/01629/wdfw01629.pdf

Wilcox TM, McKelvey KS, Young MK, Lowe WH, Schwartz MK. Environmental DNA particle size distribution from Brook Trout (*Salvelinus fontinalis*). Conserv Genet Resour. 2015;7:639–41. https://doi.org/10.1007/s12686-015-0465-z.

# Surmounting the Large-Genome "Problem" for Genomic Data Generation in Salamanders


**David W. Weisrock, Paul M. Hime, Schyler O. Nunziata, Kara S. Jones, Mason O. Murphy, Scott Hotaling, and Justin D. Kratovil**

**Abstract** Salamanders have some of the largest genomes among all extant organisms, due in large part to the proliferation of repetitive elements and the expansion of intron size. This increased complexity and size has limited the application of genomic tools to the population genetic and phylogenetic study of salamanders, even as these methods have become common for most other organisms. However, the generation of genomic data in salamanders is not out of reach for most researchers. High-quality and informative data sets can be acquired for salamander-centric research projects with careful consideration of the genomic tool(s) most appropriate for the question at hand and how best to apply these to a salamander genome. Here, we review a range of genomic tools representing the current best options for use in the study of genome-wide variation within and between salamander species. This includes the use of

D. W. Weisrock (✉) · K. S. Jones
Department of Biology, University of Kentucky, Lexington, KY, USA
e-mail: david.weisrock@uky.edu

P. M. Hime
Department of Biology, University of Kentucky, Lexington, KY, USA

Biodiversity Institute, University of Kansas, Lawrence, KS, USA

S. O. Nunziata
Department of Biology, University of Kentucky, Lexington, KY, USA

Department of Entomology, University of Maryland, College Park, MD, USA

M. O. Murphy
Department of Biology, University of Kentucky, Lexington, KY, USA

Department of Biology, Miami University, Oxford, OH, USA

S. Hotaling
Department of Biology, University of Kentucky, Lexington, KY, USA

School of Biological Sciences, Washington State University, Pullman, WA, USA

J. D. Kratovil
Department of Biology, University of Kentucky, Lexington, KY, USA

Department of Entomology, University of Kentucky, Lexington, KY, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_36,

transcriptomics (RNAseq), restriction site-associated DNA sequencing (RADseq), sequence capture enrichment methods, and PCR-based parallel tagged amplicon sequencing. Each of these methods has a particular set of benefits, as well as limitations in the study of salamander genomics. We highlight their trade-offs and the factors that should be considered when choosing among them, and we provide descriptions of exemplar studies that illustrate their empirical applications. By making informed decisions about the choice and implementation of these subgenomic methods, we believe that they can be broadly and effectively applied as important resources for the study of salamander evolution and conservation.

**Keywords** Amphibian · Conservation genetics · Parallel tagged amplicon sequencing · Phylogenetics · Population genetics · RADseq · RNAseq · Targeted sequence capture

## 1 Introduction

As next-generation sequencing (NGS) and the genomic revolution have swept forward the population genetic and phylogenetic study of non-model species, the use of new genomic tools for these pursuits in salamanders has lagged behind. The principal reason for this lag is their ridiculously large genomes – larger than almost all other vertebrate species (Sessions 2008). At their smallest, salamander genome sizes in the range of ~15 gigabases (Gb) can be found in many species of the genus *Desmognathus*, roughly five times the size of the human genome. At their largest, genomes have expanded to an astounding ~120 Gb in the Neuse River waterdog, *Necturus lewisi* (Gregory 2018). The approximately 700 salamander species fall somewhere in this range, typically around ~30–50 Gb. These absurdly large salamander genome sizes are particularly evident when put in the context of other major clades. For example, most mammals have ~3–4 Gb genomes, with a range of 1.6–6.3 Gb (Kapusta et al. 2017). Salamanders are exceptional even among other amphibians, with maximum haploid genome sizes within frogs of ~12 Gb (Olmo 1973) and within caecilians of ~14 Gb (Beçak et al. 1970). It is also worth noting that massive genome sizes in salamanders are not the result of polyploidization, as nearly all salamanders are diploid, with the exception of the unisexual members of the genus *Ambystoma* (Gibbs and Denton 2016).

Unsurprisingly, extremely high sequencing costs and the lack of availability for computational resources that can handle the inordinately large amount of data needed to produce a salamander genome have proved prohibitive in sequencing and assembling reference-quality salamander genomes. This has begun to change, as recent studies have produced genomic constructs for the axolotl (Ambystomatidae: *Ambystoma mexicanum*; Keinath et al. 2015; Nowoshilow et al. 2018) and the Iberian ribbed newt (Salamandridae: *Pleurodeles waltl*; Elewa et al. 2017). However, even these efforts have yielded highly fragmentary assemblies, highlighting some of the broader limitations and challenges in salamanders, ranging from the use

of finite sequencing resources in an immense genome to the difficulty of placing subgenomic sequence data in the context of a whole-genome assembly.

It is likely that no single mechanism explains the evolution of large genome size in salamanders. Transposable elements (TEs) are common components of the genomes of most eukaryotes. However, studies of salamanders from the families Ambystomatidae, Cryptobranchidae, and Plethodontidae have revealed a disproportionately larger number of long terminal repeat retrotransposons, relative to other vertebrates, suggesting that the proliferation of these elements may be a driving factor in salamander genome gigantism (Sun et al. 2012; Sun and Mueller 2014; Nowoshilow et al. 2018). Introns are also substantially longer in salamanders relative to other vertebrate genomes and may contain greater numbers of regulatory regions (Smith et al. 2009; Nowoshilow et al. 2018). Salamanders also have very low metabolic rates relative to other vertebrates, and correlations between metabolic rate, cell volume size, and genome size have been proposed (Licht and Lowcock 1991). Given their vast size, it is likely that other notable aspects of salamander genomes will be discovered which set them apart from other vertebrates (e.g., Madison-Villar et al. 2016; Mohlhenrich and Mueller 2016; Elewa et al. 2017; Nowoshilow et al. 2018).

The challenge posed by large genomes varies across NGS tools, with each method posing its own suite of challenges. The use of PCR and capture-based approaches is constrained by the lack of baseline genome sequence information for most salamanders, limiting the generation of effective primers or capture baits. When these resources are available, large genome size does not seem to have a negative effect on PCR amplification of loci, but it does have an effect on capture-based enrichment methods, where capture baits are searching for "needles" in an extremely large "haystack." For anonymous locus methods, such as restriction site-associated DNA sequencing (or RADseq), challenges arise from the fact that larger genomes contain higher numbers of restriction enzyme recognition sites, and close attention is required to optimize the number of anonymous fragments for sequencing. RNA sequencing-based (RNAseq) approaches may be less hampered by large genome size, but certain analyses of the resulting data may be limited by the current lack of whole-genome resources for salamanders.

While these constraints have hindered the application of genomic data in the study of salamanders at the micro- and macroevolutionary levels, they are not insurmountable. Improvements in sequencing technologies continue to increase the amount of sequence data that can be generated while also decreasing costs. In addition, as researchers begin to take the plunge into the pool of available genomic tools and apply these to population genetic and phylogenetic questions in salamanders (Fig. 1), many of the kinks are beginning to be worked out of the data generation protocols, and a set of "best-practice" guidelines are emerging. This developing access to genome-wide data in salamanders brings with it a large genome upside: bigger genomes also harbor greater information about evolutionary history. For example, increased access to variable sites across the genome increases the probability of detecting recent coalescent events that can be informative of very recent population history. In addition, salamander genomes may contain a larger number of

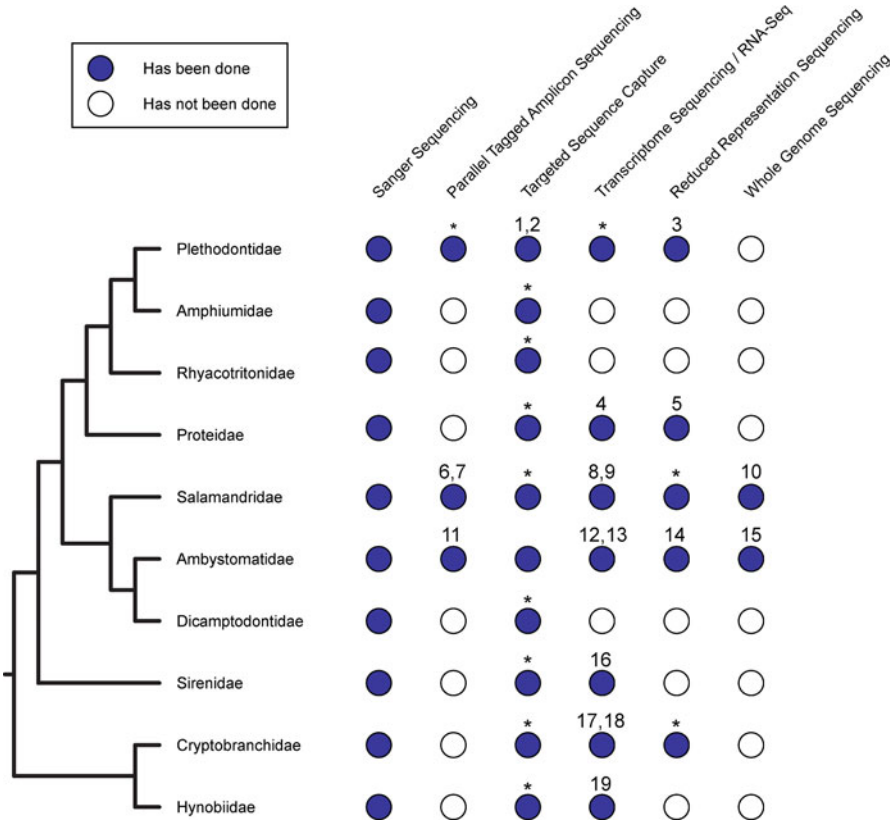


**Fig. 1** A phylogenetic perspective of the variety of subgenomic and genomic methods that have been applied across the ten extant salamander families. Filled circles indicate that a particular method has been used in a given family. Open circles denote cases where a particular method has not been used for a family. Numbers above filled circles identify empirical examples for a particular method applied to a salamander family (1: Newman and Austin 2016; 2: Bryson et al. 2018; 3: Lucas et al. 2016; 4: Irisarri et al. 2017; 5: Murphy et al. 2018; 6: Zieliński et al. 2014a; 7: Wielstra et al. 2014a; 8: Czypionka et al. 2015; 9: Looso et al. 2013; 10: Elewa et al. 2017; 11: O'Neill et al. 2013; 12: Putta et al. 2004; 13: Nowoshilow et al. 2018; 14: Nunziata et al. 2017; 15: Nowoshilow et al. 2018; 16, 17: Irisarri et al. 2017; 18: Qi et al. 2016; 19: Matsunami et al. 2015). Filled circles marked with asterisks represent unpublished applications by the authors. Full references for these examples can be found in the literature cited

truly independent markers of species tree history, owing to the greater amount of recombinatorial decoupling of genetic variation over large chromosomal stretches. From these perspectives, salamanders may serve as unique systems for the study of evolutionary history.

Here, our primary goal is to review the many methods available for generating genomic data for studies in natural populations of non-model organisms and provide insight and guidance into their application in salamander genomes. While the spirit of this chapter lies within the context of conservation and wildlife genomics, many

of the methods commonly used to study population-level genetic variation can be similarly applied at the phylogenetic level, and, when appropriate, we identify the strengths and weaknesses of each method at different scales of evolutionary divergence. We have written this review with the expectation that the reader will have a general familiarity with basic laboratory and sequencing methods, and we refer the reader to a number of reviews covering the new era of NGS in population and phylogenomics for more detail on sequencing methods (e.g., Davey et al. 2011; Lemmon and Lemmon 2013). Finally, we note that salamanders are not the only organisms with expanded genome sizes and the lessons learned in the application of genomic tools in salamanders can be leveraged in the study of other large-genome species.

## 2 Genomic Data Generation in Salamanders

Researchers interested in the study of genomic variation in natural populations now have a wide range of methods available for generating data that is appropriately targeted at their particular question (Fig. 2). While whole-genome sequencing is beginning to be a tractable approach for studying genetic variation within and among species with "normal" genome sizes, it is unlikely that this will become a reality for salamanders any time soon. However, other methods are available to comprehensively survey aspects of the genome. Deciding which method to use requires the consideration of factors that would apply to any taxonomic group, which largely revolve around the scale of divergence and what levels of genetic variation will be most informative for the questions at hand (Fig. 2). Salamanders, however, bring an extra set of genome-specific considerations. For example, the targeting of specific loci in the genome will require prior knowledge of genome sequence information and will likely require the availability of a relatively closely related genomic

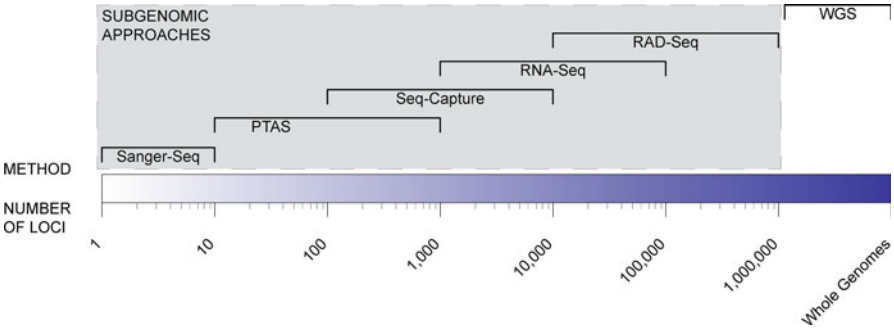


**Fig. 2** A variety of subgenomic methods are available for data generation in salamanders. Different approaches will yield different numbers of loci, and researchers may select a method of data generation suited to the target numbers of loci for their study. Ranges of numbers of loci are approximate. WGS: whole-genome sequencing

resource from which to draw this information. In addition, the number of individuals that can be sequenced in parallel on an NGS platform will scale proportionally with the size of the genome under study. Each sequencing method can be affected differently by these different factors and, when coupled with the time and resources available to a project, will mean that different researchers may make different choices about the methods best suited to their project. As a note, all salamander genome size estimates presented here are taken from Gregory (2018).

## 3 Restriction Site-Associated DNA Sequencing (and Related Approaches)

### *3.1 General Overview*

RAD sequencing (e.g., Miller et al. 2007), in its many varieties, provides one of the most straightforward ways to narrow down the number of genomic regions for sequencing. Through fragmentation of the genome with restriction enzymes and the subsequent reduction of this fragment pool to a particular size range, a substantially reduced subset of the genome can be created for sequencing. The use of the same restriction enzymes and size selection across multiple individuals provides the opportunity to recover a shared set of loci amenable to evolutionary analysis. These methods have been a boon to the ecological and evolutionary study of wild populations (Andrews et al. 2016), as well as for functional genomics (Baird et al. 2008), because they allow for the generation of large genome-wide data sets without the need for substantial prior information about the genomes under study. RAD sequencing is commonly used to uncover genetic variation, typically in the form of single-nucleotide polymorphisms (SNPs), and is most frequently used in studies at the population level or at the population-species interface. It can also be applied in a phylogenetic context across multiple species; however, increased evolutionary divergence reduces the shared overlap in orthologous loci among species. In this chapter, we do not review RAD-based protocols in detail and instead refer the reader to several original papers and reviews detailing their use (Miller et al. 2007; Elshire et al. 2011; Peterson et al. 2012; Andrews et al. 2016). In addition, we encourage readers to familiarize themselves with other complexities of these data that are not salamander-genome specific (e.g., allele and locus dropout; Gautier et al. 2013).

### *3.2 Salamander Genome Limitations*

Larger genomes have a greater number of potential restriction enzyme cut sites and, as a result, more potential fragments to be sequenced. In addition, most researchers

have a finite amount of sequencing effort that can be applied to a project. The sequencing of a single fragment (or locus) from an individual and the confident determination of its nucleotide composition and variation (i.e., homo- or heterozygous) require multiple independent sequence reads from the same fragment (the depth of sequencing coverage). Consequently, salamander researchers will need to consider how a RAD sequencing protocol can be optimized to reduce the overall set of fragments that can be sequenced and how this will intersect with their limited sequencing resources to permit the recovery of useful genomic data across multiple individuals.

The most important consideration for reducing the number of genomic fragments when working with large genomes is the choice of restriction enzymes, which has a large influence on the number of fragments that are produced. Restriction enzymes with longer, and rarer, recognition sites (e.g., 6 or 8 bp) will yield fewer fragments than those with smaller recognition sites. As a further step in reducing the number of fragments to be sequenced, researchers can use two restriction enzymes instead of one (i.e., a double digest, or ddRAD) and sequence only those fragments containing both cut sites (Peterson et al. 2012). It is likely that all salamander RAD sequencing studies will require a ddRAD-like approach to produce a library reduced enough to optimize sequencing efforts. Finally, the selection of a particular size range from the resulting fragment distribution provides yet another mechanism for reducing the number of fragments for sequencing.

Based on the study-specific requirements for numbers of loci, numbers of individuals, and per-locus depth of sequencing coverage, as well as the available sequencing resources, practitioners may optimize the restriction enzyme(s) and/or size selection window accordingly. Optimizing ddRAD approaches involves performing single- and double-digests of genomic DNA for multiple pairs of restriction enzymes and empirically estimating the number of sequenceable fragments within different size selection windows (as in Peterson et al. 2012, supplemental materials). While genomic resources are not required for this estimation, a best guess of genome size for the species under study can be used, which are available for all salamander families (Gregory 2018). Software is also available to perform in silico predictions of fragment numbers when a genome assembly is available (Lepais and Weir 2014). This can also be done using randomly generated sequence data as a proxy for an unknown genome, although we are unaware of any attempt to use this as a preparatory step for ddRAD in something as large as a salamander genome. Ultimately, after considering the constraints of the possible numbers of loci per individual, researchers should then select a restriction enzyme combination and size selection window best suited to their particular question and resources.

During the planning stage of a RAD sequencing project, it may be useful to quantify the interactions of important factors that will influence data generation. Based on the estimated number of loci per individual, the desired mean depth of sequencing coverage per locus, the number of individuals to be included in the study, and the estimated proportion of raw sequencing reads that can be assembled into

loci, one can estimate the total amount of sequencing effort needed for the study according to Eq. 1.

$$\text{SeqEffort} = (L \times C \times I)/R \quad (1)$$

SeqEffort is the number of total reads to be sequenced, $L$ is the average number of loci (fragments) per individual, $C$ is the mean per-locus depth of sequencing coverage, $I$ is the number of individuals, and $R$ is the proportion of sequencing reads passing all quality filters (e.g., sequencing quality scores, removal of PCR duplicates) and assembled into loci (on-target rate).

For example, in order to sequence 100,000 loci per individual to 30× mean coverage for 100 individuals, and an 85% on-target rate, 352,941,176 reads (or read pairs, for paired-end sequencing) would be required. This is, of course, idealized, and other factors will come into play. In our experiences with ddRAD in salamanders, empirical on-target rate was ~85% when we aimed for ~30× coverage, but numbers of individuals and loci varied by species and project.

When dealing with salamander genomes, restriction enzyme combinations may still yield an exceedingly high number of loci, and it may be difficult to optimize ddRAD protocols to produce fewer than tens of thousands to hundreds of thousands of loci per individual. These expectations are based upon our experience across five families of salamanders (Table 1). More recent modifications of the general RAD method may provide the means to further winnow down the numbers of loci produced per individual, either by performing an additional restriction enzyme digestion step (e.g., Graham et al. 2015) or by subsequently performing targeted enrichment on a subset of loci generated in an initial round of RAD sequencing [e.g., Rapture (Ali et al. 2016) or RADcap (Hoffberg et al. 2016)].

Once data are in hand, there are inherent complications with assembling and analyzing large numbers of loci that will be generated by a RAD sequencing protocol in a large genome. Perhaps the most obvious is that the computation time required to assemble sequencing reads into loci, and to compare loci across multiple individuals, scales with numbers of loci. Standard software packages (Catchen et al. 2011, 2013; Eaton 2014) work well for RAD locus assembly, but access to high-performance parallel computing resources is highly desirable. One special consideration in the assembly of salamander RAD sequencing data is the detection and filtering of potential paralogous loci. The large proportion of repetitive elements in salamander genomes greatly increases the probability that paralogs will be sequenced and that they might masquerade as orthologous loci. Paralogous loci should exhibit some characteristic patterns if assembled as a single locus, including extremely high sequence coverage and/or biologically implausible numbers of alleles (i.e., >2 for diploid species). Many assembly programs include functions that can filter according to these factors. Finally we note that while genome assemblies are now available for two salamander families, high rates of divergence across salamander families – and even between genera – are likely to limit the usefulness of

**Table 1** Use of RAD-based sequencing in different salamander species

| Family | Species | Genome size (Gb) | Target selection size (bp) | Est. # of sequence fragments | Maximum # loci recovered for an individual | Reference |
|---|---|---|---|---|---|---|
| Ambystomatidae | *Ambystoma opacum* | ~30 | 300 | 273,120 | 300,869 | Nunziata et al. (2017) |
| | *Ambystoma talpoideum* | | 300 | 133,450 | 131,201 | |
| Cryptobranchidae | *Andrias* | ~55 | 500 | 332,445 | 398,087 | Hime et al. unpublished |
| | *Cryptobranchus alleganiensis* | ~55 | 500 | – | 445,982 | Hime et al. unpublished |
| Plethodontidae | *Desmognathus fuscus* | ~15 | 300 | 259,376 | 67,724 | Kratovil et al. unpublished |
| | *Desmognathus marmoratus/ quadramaculatus* | ~15 | 500 | 166,085 | 85,520 | Jones and Weisrock unpublished |
| | *Eurycea* sp. | ~25 | 250–350 | – | – | Lucas et al. (2016) |
| Proteidae | *Necturus maculosus* | ~85 | 300 | 981,452 | 387,292 | Murphy et al. (2018) |
| Salamandridae | *Laotriton laoensis* | ~50 | 300 | – | 101,901 | Jones et al. unpublished |

All studies used a double-digest method with the enzymes *SphI* and *EcoRI* for library preparation except for Lucas et al. (2016), which used *MseI* and *EcoR1*

salamander genome assemblies to a narrow range of closely related species. Hence RAD-based locus assembly will continue to be almost exclusively de novo for the foreseeable future.

### 3.3 Examples in Salamanders

RAD-based sequencing has been successfully implemented in multiple families of salamanders spanning a wide range of genome sizes (Table 1). Nunziata et al. (2017) provide a useful illustration of the application of ddRAD sequencing in salamanders, with a study of fine-scale population demographics in the ambystomatids *Ambystoma opacum* and *A. talpoideum* (genome sizes estimated between 24 and 36 Gb). After performing a series of test digests, *EcoRI* and *SphI* were identified as the best restriction enzyme pair, and a size selection window of 270 to 330 bp was estimated to contain ~130,000 and ~270,000 unique fragments in *A. talpoideum* and *A. opacum*, respectively. To sequence this fragment pool to a read depth of 10×, a maximum of 24 individuals were multiplexed per lane of an Illumina HiSeq 2500. After assembly, individuals of *A. talpoideum* had as many as 131,000 reconstructed loci, while *A. opacum* had ~300,000 loci, close to predictions based on their fragment distributions. A more important perspective is the number of loci recovered across multiple individuals, and here the number of shared loci is expected to drop. In *A. opacum*, when restricting the data to only include loci recovered from 95% of individuals, just 15,740 loci were retained. Increasing the allowed level of missing data to 15% bumped this up to 40,326 shared loci. While this is a substantially smaller number of loci than predicted for each individual, this reduction is typical of RAD sequencing studies. Furthermore, the data generated in Nunziata et al. (2017) were, nonetheless, highly informative, yielding important insights into the population demographics of rapidly changing salamander populations.

Lucas et al. (2016) used a genotyping-by-sequencing approach to estimate genetic diversity and gene flow in a wetland metapopulation of an undescribed species of *Eurycea*, which has a best-guess genome size estimate of ~25 Gb. A more limited sequencing effort was used in this study, but this still resulted in the assembly of ~6,200 unique loci and the identification of ~7,000 shared SNPs. While the complete details covering levels of missing data are not provided, this study still serves as an example of successful RAD sequencing in a salamander.

Near the upper end of the genome-size spectrum, ddRAD has been successfully used in a population structure study of the common mudpuppy, *Necturus maculosus* (~80–95 Gb genome; Murphy et al. 2018). Despite this truly massive genome size and corresponding massive number of fragments estimated per individual (~1 million, Table 1), this still resulted in a final data set of ~10,000 shared loci – with no missing data – across all sampled individuals (distributed across three river basins in Kentucky). Our lab has had similar success in phylogeographic studies of the hellbender salamander, *Cryptobranchus alleganiensis* (~55 Gb genome) and its

Asian sister genus, *Andrias* (~46–50 Gb genome; Hime et al. unpublished). While hundreds of thousands of loci were successfully assembled per individual, ~74,000 loci were still shared among ~100 individuals from across the Eastern United States. Maybe more surprisingly, we recovered ~43,000 shared loci between *Andrias* and *Cryptobranchus*, despite a divergence between these two clades of at least 15 million years (Kumar et al. 2017). The level of shared loci recovered between these genera is encouraging given the expectation of locus dropout due to the accumulation of substitutions in restriction enzyme recognition sites over time; however, we note that this level of shared recovery is not always found in interspecies comparisons. For example, our research in different plethodontid radiations has yielded low levels of shared ddRAD loci across species. Consequently, we emphasize that more empirical studies will be needed to know how generalizable patterns and levels of shared loci will be across interspecific salamander studies. While the increased divergence between species will lower the number of shared loci recovered, these loci will have higher levels of variability, relative to their patterns within species, and thus should still provide a large amount of information for interspecific questions (e.g., Lemmon and Lemmon 2012).

## *3.4 Guidelines for RAD-Based Sequencing in Salamanders*

Given the wide range of genome sizes and compositions across salamanders, no single RAD-based sequencing protocol is expected to work for all species. Different species may require different restriction enzyme combinations, fragment size selection windows, and varying numbers of individuals that can be sequenced in parallel. It is likely that a ddRAD protocol will be required, as the fragment pool resulting from a single restriction enzyme digest will be too large and would dilute sequencing effort too much to produce useful results across multiple individuals. Beyond this one blanket recommendation, we recommend that the implementation of ddRAD-based studies in salamander species consider the following:

1. Following the protocol outlined in Peterson et al. (2012), researchers should evaluate the fragment distributions resulting from both single and double digests using an Agilent Bioanalyzer (or similar equipment). When combined with ballpark metrics of genome size, this allows for an estimation of the number of fragments within particular fragment-size windows. As a note, our lab has consistently found the enzyme combination of *SphI* and *EcoRI* to generate appropriate numbers of fragments, but this does not necessarily mean that these will be the best for all salamander RAD sequencing studies. Furthermore, not all studies will use the same fragment size selection window as applied in our studies in different families, and a thorough assessment of potential fragment numbers is encouraged to the identify particular fragment sizes to use and to be avoided.
2. Increase the amount of input genomic DNA above the ~50 ng range typically used in RAD studies. In our experiences, starting genomic DNA amounts in the

range of 1 μg has worked well for species with genomes in the range of 15–20 Gb (e.g., *Desmognathus*), and as much as 2.5 μg of genomic DNA was used for larger genomes (e.g., *Necturus maculosus*). These higher amounts of starting DNA ensure that sufficient quantities of genomic material remain after double digestion and size selection. Starting with larger quantities of DNA can also limit the number of PCR cycles required to reach desired final library concentrations (thus reducing the potential for PCR-induced errors).

3. Be cautious with the urge to increase the number of multiplexed individuals that are sequenced. A threshold exists that when passed will result in most sequenced loci being recovered at unacceptably low coverage to distinguish genuine SNP variation from sequencing error. What this coverage threshold is will depend on the study. For example, higher coverage will be necessary for population genetic questions and analyses where diploid genotypes for all individuals are important, and lower levels may suffice in studies where population-level estimates of allele frequencies are of interest. Here, we refrain from providing guidelines for levels of multiplexing on a "lane" of sequencing, as sequencing technologies – and RAD library protocols – continue to increase in efficiency and output. If the depth of coverage of initial rounds of sequencing is too low, be prepared to increase sequencing effort accordingly.

## 4 Transcriptomics and RNAseq

### *4.1 General Overview*

Transcriptome sequencing (including RNAseq) involves the purification of transcribed RNA from a tissue or set of tissues, conversion to complementary DNA, and subsequent high-throughput sequencing. By focusing sequencing effort on transcribed regions of the genome, researchers are able to target coding regions, to the exclusion of other genomic content. Consequently, transcriptomics can be easily applied in salamanders, providing access to a large amount of genomic content with no more difficulty than its use in organisms with smaller genomes. These transcriptomic resources can then have multiple applications in salamander wildlife genomics. As perhaps its main application, transcriptomics is used to study differences in gene expression across tissues, individuals, or populations, to better understand the effects of spatial, environmental, or temporal factors on cellular processes (e.g., Trapnell et al. 2013). Transcriptomics also represents an effective method for directly targeting SNPs in coding regions, either as those segregating in a population or as fixed diagnostic markers between groups of interest (Zieliński et al. 2014a). As discussed above, it can be a direct and effective way to identify candidate loci to be developed into sequence capture-based or PTAS-based markers. Transcriptomics can also provide important context for many of the previously discussed anonymous loci generated through a RAD-based approach (Amores et al. 2011).

## 4.2 Salamander Genome Limitations

From a data generation perspective, the large salamander genome poses no significant challenge, relative to other taxa with smaller genomes. In this method, the RNA polymerase machinery does the important enrichment work for you. Researchers should be aware of the general challenges in employing transcriptomics in natural settings. This includes acquiring similar enough tissues from individuals under study to increase the probability of recovering the same set of expressed orthologous loci. Studies of speciation and local adaptation should also consider the particular tissue and developmental stage being sampled and whether their expressed genes will include the loci relevant to the study at hand. The need to rely on nondestructive sampling or challenges in field collecting (e.g., acquiring necessary permitting, or finding rare species) can all pose limitations to properly implementing a transcriptomic approach for population and evolutionary studies. In addition, with a lack of a whole-genome assembly, there are also likely to be many things that are unknown going into the study, including the number of potential loci to be expected. Finally, we point out that the computational overhead of transcriptome assembly for large salamander genomes is also not expected to be more burdensome than in other taxa.

## 4.3 Examples in Salamanders

The works of Putta et al. (2004) and Habermann et al. (2004) represent the earliest efforts in generating large-scale transcriptomic data from salamanders, with studies in the Mexican axolotl (*A. mexicanum*) and eastern tiger salamander (*A. t. tigrinum*). These studies predated current NGS technologies and were generated as ESTs that provided sequence data from one end of a transcribed exonic region. However, this still resulted in the identification of ~35,000 ESTs and >10,000 contigs with high sequence similarity to known human coding sequences. While many of the goals of this work were aimed at generating resources for the study of salamander regenerative developmental biology, these resources have also had substantial downstream applications in the generation of a genome-wide linkage map (Smith et al. 2005), the generation of PCR-based nuclear markers for the study of species boundaries in related Mexican species (Weisrock et al. 2006), and the study of hybridization and admixture between native and introduced species (Fitzpatrick et al. 2010).

In a more recent example, Keinath et al. (2017) provide an example for the use of transcriptomics in the generation of a high-quality linkage map for *Notophthalmus viridescens*. This work is particularly exciting in that it demonstrated a relatively simple and fast process for developing linkage maps from large-genome species without the requirement for tremendous sequencing resources (only one HiSeq2000 lane was used) or >F1 generations (a single mother and her 28 offspring were used). Given that whole-genome sequence assemblies for most salamanders are likely to be

unavailable in the near future, transcriptome-based linkage maps will continue to serve as our best resources for studying genome structure and the placement of ecologically and functionally relevant loci (e.g., Voss and Smith 2005).

Transcriptomic data sets have also been recently used in phylogenomic studies at both shallow and deep evolutionary histories of salamanders. Rodríguez et al. (2017) used RNAseq data (along with ddRAD data) to resolve the recent history of divergence among species of the salamandrid genus *Salamandra*. Using rather modest sequencing effort on an Illumina MiSeq, the authors were still able to assemble a data set of 3,170 orthologous loci sampled from seven *Salamandra* species and two *Lyciasalamandra* outgroup species. In a study of deep phylogenetic relationships across jawed vertebrates, Irisarri et al. (2017) included sequence data generated using RNAseq from representatives of a number of salamander families. This study provides a good perspective on the sequencing effort required to recover known orthologous protein-coding genes. Using RNA sourced from multiple tissue types from the species *Andrias davidianus* (Cryptobranchidae), *Calotriton asper* (Salamandridae), *Proteus anguinus* (Proteidae), and *Siren lacertina* (Sirenidae), and a half of an Illumina MiSeq flow cell per species, they recovered between 59 and 81% of 233 core vertebrate genes (CVGs), a reference collection of one-to-one vertebrate orthologs that can be used to benchmark transcriptomic studies (Hara et al. 2015). Using a substantially greater sequencing effort in the salamandrid *Pleurodeles waltl* (381 million reads, or over 27× the number of sequence reads than in the above discussed species), recovery of CVGs approached 98%. Collectively, this demonstrated that standard transcriptomic sequencing approaches applied to diverse RNA pools in salamanders can lead to nearly complete recovery of the standard set of orthologous vertebrate genes but also that rather small sequencing efforts can still recover large sets of expressed genes.

A number of additional transcriptomic projects have been completed in salamanders to understand cellular responses in gene expression in an environmental context. Qi et al. (2016) used an RNAseq approach to study the immune response of the Chinese giant salamander, *Andrias davidianus*, when infected by a bacterial pathogen. This work yielded ~19,000 annotated coding genes and demonstrated the utility of RNAseq-based approaches in salamanders for identifying genes that potentially underlie functionally relevant pathways for immunity. Czypionka et al. (2015) used an initial round of transcriptome sequencing in *Salamandra salamandra*, coupled with the subsequent use of microarrays containing probes matching a set of ~22,000 assembled contigs identified as having open reading frames (ORFs), to study shared versus differential patterns of gene expression between nonlethally sampled tail clips and lethally sampled whole larvae. Interestingly, this work showed that a large proportion of genes (51%) had similar changes in expression among tail and whole-body tissues across different temperature treatments, suggesting that nonlethally sampled tail tissues may serve as a good proxy for environmentally influenced gene expression. Matsunami et al. (2015) used RNAseq to examine gene expression changes underlying phenotypic plasticity in *Hynobius retardatus* in response to different predators. This work generated ~740,000 assembled contigs, among which ~175,000 could be identified as protein coding based on the presence of

ORFs. Based on this large genomic resource, dozens of genes were identified that had differential expression under different predator regimes, and ultimately this led to new insights into the understanding of the evolution of phenotypic plasticity.

### *4.4 Guidelines for Applications in Salamanders*

There is little salamander-specific advice that we can offer for the use of transcriptome sequencing, as there are inherent limitations to the use of this tool in species with large genomes. Standard laboratory and computational methods will apply. Perhaps the one relevant point to make is that the design and implementation of these projects can be done according to researchers' downstream goals. If the goal is marker development for subsequent use in sequence capture and PTAS studies, a single tissue source from an animal (e.g., a tail tip) can be sufficient to generate enough candidate loci. This may be ideal when nondestructive sampling is preferred, or when tissue sources are rare. Alternatively, when projects are aimed at identifying as many coding genes as possible, either in an attempt to uncover orthologous loci identified in other species or to study their expression differences across different treatments, multiple tissue sources from an animal are required.

## 5 Sequence Capture and Enrichment

### *5.1 General Overview*

Sequence capture methods use synthetic oligonucleotide probes to target and enrich for genomic regions identified a priori. Biotinylated probes are annealed with fragmented and barcoded genomic DNA of a target species, with the probes finding their complementary match to target loci. These "captured" fragments are then sequestered by hybridization to streptavidin-linked beads and clonally amplified by high-fidelity PCR. The resulting enrichment products for multiple individuals are then sequenced in parallel on a NGS platform. By using the same set of probes across all individuals in a study, sequence capture methods provide an effective approach for generating data from shared orthologous loci. Sequence capture methods were kick-started with the generation of protocols to perform probe hybridization in solution (Gnirke et al. 2009) and since have been dominated by two general approaches, anchored hybrid enrichment (AHE; Lemmon et al. 2012) and ultraconserved elements (UCEs; Faircloth et al. 2012). They have also been implemented in a custom fashion in numerous taxonomic groups, typically in the form of exon and candidate locus capture (e.g., Bi et al. 2012; Linnen et al. 2013; Portik et al. 2016).

A sequence capture approach for genomic data generation can have many benefits over other genomic methods. First, it provides a methodologically efficient approach

for sequencing known regions of the genome, as opposed to anonymous loci that are typically sequenced using RAD-based methods. Second, it can lead to the generation of highly complete data sets across individuals and species, as it does not suffer from allele dropout due to a single mutation or substitution in a restriction enzyme recognition site. Third, capture probes can be quite forgiving to mismatches with genomic templates, much more than can be tolerated in the annealing of PCR primers. Consequently, probes based on one species can be effectively used across a relatively wide range of divergent taxa; however, there are limitations to the level of divergence between probe taxa and capture taxa that we discuss below.

## *5.2 Salamander Genome Limitations*

In any sequence capture reaction, capture probes must sift through a pool of genomic DNA to find complementary matches with their target loci. While this is an efficient method for enriching a sample with a desired set of loci for sequencing, the vast nature of a genome also leads to a large amount of "off-target" capture (Guo et al. 2012), or the enrichment of additional genomic regions that are not part of the specific set of targeted loci. This can occur for a number of reasons, including the promiscuous annealing of probes to nontargeted DNA under different reaction conditions and the carry through of high-copy regions of the genome (e.g., mitochondrial DNA). In salamander genomes, the expansion of many aspects of genomic content (e.g., larger introns and greater number of TEs) is expected to increase the amount of off-target enrichment. This leads to at least two complications: (1) capture probes are diluted across the genome in proportion to genome size, yielding a lower level of enrichment of targeted loci, and as a result, (2) greater sequencing effort will be required to recover targeted loci at a sufficient depth of read coverage.

The evolutionary divergence between probe taxa and target species also exerts a strong influence on capture success. For instance, in capture reactions applied to frogs, Hedtke et al. (2013) found that the numbers of recovered loci dropped precipitously with increasing divergence time between probe species and target species (also see Lemmon et al. 2012). Although a straightforward work-around is to design probes specifically from the taxon or taxa under study, the current scarcity of genomic resources for salamanders means that this is not likely to be a simple fix for many researchers.

There are at least two possible remedies that can be applied to mitigate off-target enrichment in salamanders. The first is to use capture probes with high specificity to the taxa under study. While targeting conserved stretches of DNA provides one mechanism for increasing the probability that probe sequences will have high complementarity to the template DNA, this can still lead to an exceptionally high level of off-target enrichment in salamanders. Even with high conservation between the probe taxon and the study taxon (e.g., when they are the same species), "on-target" sequence reads may at best only account for 20–30% of the total sequence reads (Bi et al. 2012; Faircloth et al. 2012), and this is expected to be substantially

lower in salamanders. We have explored this approach in both hellbender salamanders (*Cryptobranchus*) and dusky salamanders (*Desmognathus*). Here we compared locus recovery using the original Lemmon et al. (2012) AHE method and probe set (in which the closest probe taxon was the frog *Xenopus* [*Silurana*] *tropicalis*) to a custom probe set that included capture probes designed specifically from genomic resources for *Cryptobranchus* and *Desmognathus* which we developed de novo. In the first set of tests, the evolutionary distance between the probe taxon and target species was ~300 million years, and we recovered 54.1% (277/512) and 72.8% (373/512) of loci from *Cryptobranchus* and *Desmognathus*, respectively. In contrast, use of the taxon-specific probe set in these two taxa increased locus recovery to 93.0% (319/343) in *Cryptobranchus* and 99.7% (342/343) in *Desmognathus*.

The other possible remedy to increase capture efficiency in salamanders is through decreasing the negative effect of the highly repetitive fraction of the genome (McCartney-Melstad et al. 2016). Using $c_0$t-1 DNA (Kallioniemi et al. 1992) developed from the species of interest, a large portion of the repetitive DNA in the genome can be "blocked," increasing the probability that individual capture probes will find their complementary match and reducing the amount of off-target capture. $C_0$t-1 DNA itself is the repetitive fraction of the genome and is created by fragmenting the genome, denaturing the DNA into single-strand form, and then slowly reannealing into double-strand form. Because high-copy, single-strand fragments should find their complementary match sooner than single-copy fragments, the collection of the early stages of fragment reannealing (i.e., $c_0$t-1) yields a collection of mostly repetitive DNA. Adding DNA from this $c_0$t-1 fraction to sequence capture reactions can be effective in blocking repetitive genomic DNA and increasing the probability that capture probes will find their on-target matches in the genome. In a test study applying exon-based capture probes based on transcriptomic resources for the Mexican axolotl, *Ambystoma mexicanum*, to the closely related tiger salamanders *A. californiense* and *A. mavortium*, McCartney-Melstad et al. (2016) demonstrated that the inclusion of $c_0$t-1 had a positive effect on on-target sequence reads. Furthermore, they found even more increased recovery of on-target sequence reads when using both $c_0$t-1 and increased concentrations of input DNA. While these effects did not push capture efficiency to very high levels of on-target sequencing, it did nearly double the rate of on-target sequencing, from ~10% without $c_0$t-1 and using standard DNA input concentrations to nearly 20% when using high amounts of both. The generation of twice as many on-target reads can substantially improve the number of recovered loci and provide the necessary read coverage needed to confidently identify SNP variation.

While the use of $c_0$t-1 blocker DNA can be an effective improvement to sequence capture in salamanders, it also has its limitations. First, it is not clear what level of divergence will be tolerated between the genomes of the $c_0$t-1 species and the targeted study species. Genomic repeat landscapes can be expected to change with increased divergence from a common ancestor, and it is likely that $c_0$t-1 may need to be derived from the study taxon or closely related species to be most effective. The generation of $c_0$t-1 DNA itself may pose a limitation in many taxa, particularly when

tissue resources are small (e.g., when using nondestructive sampling) or rare (e.g., endangered or difficult to obtain species). Much remains to be explored with this method, including the potential for whole-genome amplification of the $c_0$t-1 fraction (as per the suggestion in McCartney-Melstad et al. 2016) to create large amounts of material from limited sources.

One additional caveat to consider in sequence capture is that the use of conserved genomic regions for probe development may prove limiting in the recovery of genetic variation for use in population-level studies. While targeted sequence capture based on conserved regions can perform well at shallow scales in terms of the proportion of target loci that are successfully captured, this performance may come at the cost of variation in loci. Conserved core regions of loci are less likely to contain informative sites at shallow scales, and it may be the more variable flanking regions of loci (e.g., introns or 3′ untranslated ends of coding regions) that are of greatest utility for questions at the population genetic level. Salamander researchers targeting population-level questions may consider developing more species-specific capture markers for their projects (e.g., Lemmon et al. 2012), but given the lack of genomic resources for most families, these will probably need to be generated for the species or clade of interest.

## 5.3 Examples in Salamanders

To date, works of Newman and Austin (2016) and Bryson et al. (2018) represent the only published studies applying sequence capture – both in the form of UCEs – to address evolutionary questions in salamanders. Newman and Austin (2016) took steps to make the standard UCE approach more amphibian specific by restricting their UCE capture probe set to 2,064 probes covering 1,745 loci that had >85% sequence similarity to the *Xenopus* [*Silurana*] *tropicalis* genome. Excluding more divergent probes presumably has the effect of increasing the amount of on-target sequencing. A total of ~600 million sequence reads were generated from capture libraries for a set of 94 *Plethodon serratus* individuals sampled from across their range, along with two outgroup *P. cinereus* individuals. This resulted in the recovery of a large number of loci across the majority of individuals, but the exact number of loci retained for analysis varied across missing data filtering strategies. At the most individual inclusive level, a total of 321 loci (out of 1,745, or ~18%) were recovered for a set of 85 *P. serratus* and the two *P. cinereus*, where each locus was missing from no more than 20% of individuals. The number of retained loci increased to 1,327 (76%) when allowing for up to 40% missing individuals per locus. Newman and Austin (2016) also explored additional filtering strategies aiming at preserving the number of retained loci by removing individuals with greater numbers of missing loci.

Bryson et al. (2018) used a more standard tetrapod UCE probe set targeting 5,060 loci to capture loci for a range-wide study of the Mexican plethodontid, *Isthmura belli*. In an effort to increase capture efficiency, $c_0$t-1 blocker derived from chicken

was used as part of the capture protocol, although it is unclear what effect this had given the tremendous evolutionary divergence between chicken and salamanders. While total sequencing effort was not described in this paper, it is clear that this approach yielded informative and useful data. Using a 50% missing data threshold, 1,094 loci were recovered across sequenced individuals. Increasing the missing data threshold to 30% reduced this to 796 loci.

Overall, the genomic takeaway from these studies is that sequence capture can work quite successfully in salamanders without major augmentations to standard protocols. However, it also provides another example of the effect of divergence between the probe taxon and target taxon. If highly complete data sets are desired, researchers will either have to decide between accepting a dropout of a high percentage of their loci, or individuals, when using divergent probes, or they may have to invest the effort in developing new probes for the taxa under study.

### *5.4 Guidelines for Applications in Salamanders*

The primary piece of advice we can provide for the use of sequence capture-based methods in salamanders is to use probes with high sequence specificity to the taxa under study. The use of standard methods and probe sets such as AHE and UCE will likely result in successful data generation. But, the use of probes generated from genomic resources derived from the taxon or clade under study will more likely result in increases in capture efficiency, decreases in missing data, and reduction in the sequencing effort required. When projects are limited in their taxonomic scope, and when sufficient tissue samples are available, the use of $c_0t$-1 blocking DNA derived from the taxa under study will be greatly beneficial in improving capture efficiency and reducing the amount of sequencing effort (and corresponding fundage) required.

## 6 Parallel Tagged Amplicon Sequencing

### *6.1 General Overview*

Parallel tagged amplicon sequencing (PTAS) emerged early in the transition from studies using one or a few loci to those using large numbers of loci. PTAS couples more traditional methods of sequence enrichment (i.e., PCR) with high-throughput sequencing to generate moderate- to large-scale data sets. The overall methodology is straightforward; given available PCR primers for multiple loci, amplicons from an individual are generated and then pooled and indexed either prior to or as part of library preparation. Multiple individuals are then sequenced in parallel on an NGS platform. Despite its simplicity, PTAS has not been as widely adopted for population genomic and phylogenomic studies, relative to sequence capture and RAD-based

methods. This can be attributed to a number of reasons, including the much greater genomic sampling offered by the other methods, the lack of primer pairs for large suites of nuclear loci for most species, and the relatively larger laboratory effort required for PCR enrichment, although this latter issue can potentially be mitigated by merging multiplex PCR methods with PTAS (Campbell et al. 2015). Nonetheless, PTAS can still serve as an optimal data generation method for some labs and projects where the desired number of loci is modest and less than what would be more efficiently, and more cheaply, generated using a sequence capture approach (e.g., <100 loci; Shen et al. 2013). For many questions, data from a modest number of loci may be sufficient to produce well-supported results (e.g., Hime et al. 2016), and labs entertaining the idea of a genomic approach to their research may find the simple segue from PCR to NGS appealing, especially given the cost and equipment needs for other genomic methods. Generally, PTAS yields a low level of missing data, and because PCR is typically followed by confirmation with gel electrophoresis, researchers can have confidence in generating data across all loci that show positive amplification.

## 6.2 Salamander Genome Limitations

Genome size is not expected to have a direct negative impact on the use of PTAS. There is no evidence that PCR performance is influenced by genome size, and once amplicons have been generated, NGS is expected to perform just as well as with other enrichment strategies. PTAS may actually provide one of the most efficient sequencing strategies: as long as PCR does not amplify many secondary loci, amplicon pools should be highly enriched for the target loci, with little sequencing effort being squandered on off-target loci. As an example of this efficiency, we have included amplicon pools of different sets of loci amplified from two individuals (one *Ambystoma* and one *Desmognathus*) in the same indexed library and recovered all loci with high sequence coverage using a single multiplexed MiSeq run.

Perhaps the biggest limitation to the use of a PTAS approach in salamanders is the lack of developed PCR primer pairs for large numbers of loci. Few genomic resources are available, and primer development based solely on the currently available amphibian model genome (*Xenopus* [*Silurana*] *tropicalis*) is unlikely to yield primers that can be applied across salamanders with high levels of success. Developing primers from exons conserved across multiple vertebrate genomes has led to a toolkit of ~100 highly successful PCR primers that have been applied across salamander families (Shen et al. 2013) and across major clades within the Plethodontidae (Shen et al. 2016). One potential downside to relying on conserved exonic genomic regions for PCR primer development is that genetic variation may be limited when working at the population level. This is unlikely to be solved by anchoring primers in adjacent exons and amplifying across introns given the large intron size in salamanders (Smith et al. 2009; Nowoshilow et al. 2018). As discussed below, one of the best options for developing variable PCR-based markers is through the generation of transcriptomic

data for the taxa under study, the identification of candidate orthologous and variable loci, and the development of species- or clade-specific PCR primers.

## 6.3 Examples in Salamanders

PTAS has been used most successfully in phylogeographic and population-level studies of two diverse salamander radiations. It was first applied in a systematic study of North American tiger salamanders (the *A. tigrinum* complex) aimed at understanding population structure and species boundaries across the range of the species complex (O'Neill et al. 2013). Using transcriptomic resources – generated as expressed sequence tags (ESTs) – for *A. t. tigrinum* and *A. mexicanum* (Putta et al. 2004), PCR primers were developed for a large suite of nuclear markers and then tested for successful PCR amplification, leading to a suite of primer pairs for 95 nuclear loci that amplified across the entire species complex (~5 million years of divergence). These loci were amplified from a total of 95 individuals, pooled and indexed by individual, and sequenced on a Roche 454 platform. Despite the use of this older NGS platform and a low number of sequence reads (~344,000) relative to current technology, this still resulted in a high proportion (~81%) of on-target sequences that could be successfully assigned to an indexed individual. This also resulted in a relatively high level of data completeness across individuals, with an average of just 11% missing data per individuals. The library preparation and NGS methods from this study represent older versions of genomic technology, and improvements in both will further improve the efficiency and recovery of PTAS data (Feng et al. 2016).

In a similarly designed set of independent studies in newts of the genera *Lissotriton* (Zieliński et al. 2014a) and *Triturus* (Wielstra et al. 2014a), PTAS markers were developed to study evolutionary patterns within and between species. Both studies used newer NGS platforms (Ion Torrent and Illumina), which yielded higher levels of read coverage per locus versus the 454-based approach employed by O'Neill et al. (2013). They also used a multiplexed approach toward their PCR work, bundling PCR primers for as many as 11 loci and greatly reducing the amount of laboratory effort required for amplicon generation. It should be clarified, however, that not all loci can be co-amplified in the same PCR reaction, as some primers can dimerize based on their nucleotide composition. Consequently, researchers interested in performing a similar multiplexed PCR strategy will need to invest some initial effort in determining locus pooling compatibility, and it is possible that given a limited set of PCR markers, not all loci will be able to be multiplexed.

While the data from these PTAS studies are on a much smaller scale than those typically generated with RAD sequencing or sequence capture, they have nonetheless been very informative about evolutionary patterns at shallow scales of divergence. The broadscale work of O'Neill et al. (2013) was transitioned into an informative study exploring the numbers of loci and information content required

for species delimitation at shallow scales of divergence (Hime et al. 2016). The marker set and methods of Wielstra et al. (2014a) were used to gain important insights into the genetically admixed history of a recently described species (Wielstra and Arntzen 2014) and reconstruct the broader phylogeny of multiple species of *Triturus* (Wielstra et al. 2014b). Similarly, the work of Zieliński et al. (2014a) has been expanded to study the demographics of speciation between species of *Lissotriton* (Zieliński et al. 2016).

### *6.4 Guidelines for Applications in Salamanders*

A PTAS strategy may be ideal for many researchers who (a) desire a multilocus approach to their work, (b) are addressing questions that don't require densely sampled genomic markers, and (c) prefer to not be deluged with massive data sets. This may be particularly true for many salamander researchers given the complexities discussed above for the use of RAD and capture-based sequencing protocols. Our best advice for the use of PTAS in salamander studies is to work from a pool of candidate markers developed from a closely related species. In the immediate future, this will probably require the generation of transcriptome resources to provide this pool of candidates, but given the modest number of loci that will go into a PTAS study, a large transcriptomic sequencing effort will not be required. We also recommend the selection of sets of primer pairs that generate amplicons of roughly the same length (within a few hundred base pairs of each other), which will maximize sequencing efficiency on NGS platforms, and we advocate the use of qPCR to quantify and normalize pooled amplicon concentrations between individuals in order to produce more even coverage across individuals.

## 7 Other Approaches

We have discussed four major genomic techniques that are commonly used in the study of population and interspecific genetic variation in non-model species. However, these do not represent the only genomic approaches to be used for these sorts of data. Many additional approaches have been devised which either augment one of these sets of tools to meet a particular need or which hybridize two of these tools together into a new method that overcomes some of the individual shortcomings. For example, the ddRAD method has been extended to include a third restriction enzyme digestion step, which has the effect of further reducing the fragment pool that goes into sequencing (Graham et al. 2015). RAD-based sequencing has also been paired with sequence capture methods to yield the perfectly named Rapture (Ali et al. 2016) and RADcap (Hoffberg et al. 2016) methods. In both cases, an initial round of RAD sequencing is used to develop candidate markers, which are then subsequently sequenced using a capture-based approach. This has the benefit of negating the

effect of locus dropout due to recognition site mutations and decreasing the amount of missing data in a study. This may be particularly useful for RAD-based studies that aim to extend across species or previous projects.

One combination approach that has been used in salamanders is the identification of diagnostic SNPs between species of *Lissotriton montandoni* and *L. vulgaris* using comparative transcriptomics, followed by SNP-based genotyping of 192 loci sampled from >400 individuals using an Illumina GoldenGate assay to study admixture dynamics in contact zones between these species (Zieliński et al. 2014b). These assays do not actually yield sequence data and instead provide genotypes for each SNP based on patterns of fluorescence generated through PCR using allele-specific genotyping primers (Fan et al. 2006). Consequently, this component of the work may not be as "genomic" as other methods, but for salamander researchers interested in screening large numbers of individuals for population genetic study, it may actually be ideal to limit the genomic work (and its corresponding effort) to transcriptomic study of a relatively small amount of individuals, followed by mass genotyping of diagnostic or variable loci, especially given that these SNP genotype data can be highly informative about fine-scale population processes (Zieliński et al. 2014b). Alternatively, if genomic tools are not a limitation, this type of an approach can also be accomplished through the development of probes to the loci of interest and the use of sequence capture for locus recovery and genotyping.

## 8 Conclusions and Future Perspectives

Even with the large-genome constraints on the generation of genomic data for population genetic and phylogenetic analysis in salamanders, most current subgenomic methods can be effectively used when implemented properly. RADseq can be an effective tool for population-level and shallow-scale phylogenetic projects provided that researchers have enough sequencing depth for the numbers of fragments and individuals under study. Transcriptomics can serve multiple roles, from gene/locus discovery to studies of gene expression. Here, there is little genome-specific limitation in salamanders, with the greatest hurdle potentially being access to tissues for the extraction of relevant RNA pools. Sequence capture methods are a powerful source of phylogenetic markers from targeted regions of the genome but will work best in salamanders when capture probes are designed from relatively closely related taxa and when, again, enough sequencing depth is generated to adequately sequence the numbers of targeted fragments and individuals. Finally, parallel tagged amplicon sequencing has served as an effective method to generate moderate to large data sets in salamanders, without requiring the large sequencing effort needed with other genomic methods. However, it does require known and effective primers and more substantial laboratory effort to implement.

Other possibilities exist and we have not covered all of them here. Instead we emphasize that we foresee continued development of new methods and proposed variations on new and existing methods. Whole-genome sequencing and assembly in

salamanders is just starting to take root, and the prospects of their broader application in salamander research loom large on the horizon. But, the era of and utility of subgenomic approaches in salamanders are here now. Successful implementations of these methods continue to be demonstrated, and it is likely that more studies and papers are in the works as we write this chapter. Our overarching piece of advice to those interested in adding a genomic perspective to their salamander research is to consider these methods in detail and choose the one that is best suited to your questions, your resources, and what you are personally willing to tackle.

**Acknowledgments** Our explorations of the use of genomic tools in salamanders have been generously funded by a range of organizations. We wish to thank the National Science Foundation for much of this support through awards DEB-0949532 and DEB-1355000 (to DWW), DEB-1601586 (to DWW and PMH), DEB-1601470 (to DWW and SON), DEB-1406876 (to DWW and JDK), a National Science Foundation Graduate Research Fellowship under Grant No. 3048109801 (to PMH), and a Blue Waters Graduate Research Fellowship from the National Center for Supercomputing Applications (to PMH). Our work in *Cryptobranchus* and *Necturus* has also been funded through the Kentucky Science and Engineering Foundation through awards KSEF-2913-RDE-016 and KSEF-3266-RDE-018 (both to DWW and SJ Price, UK Forestry). We also thank the many societies that have provided graduate student research grants, including the Society of Systematic Biologists, the Society for the Study of Evolution, the Society for the Study of Amphibians and Reptiles, the Society for Freshwater Science, the Kentucky Society of Natural History, and the Kentucky Academy of Sciences. Our research would not have been possible without the exceptional computational resources provided by the University of Kentucky Center for Computational Sciences and Lipscomb High Performance Computing Cluster. We also wish to thank the many people who have been instrumental in laying down the groundwork for much of research and providing invaluable consulting and assistance with lab work and computation, including S. Randal Voss, Jeramiah Smith, Catherine Linnen, Mary Foley, Alan Lemmon, and Emily Moriarty Lemmon.

## References

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, et al. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics. 2016;202:389–400.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics. 2011;188:799–808.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3:e3376.

Beçak W, Beçak ML, Schreiber G, Lavalle D, Amorim FO. Interspecific variability of DNA content in Amphibia. Cell Mol Life Sci. 1970;26:204–6.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics. 2012;13:403.

Bryson RW Jr, Zarza E, Grummer JA, Parra-Olea G, Flores-Villela O, Klicka J, et al. Phylogenomic insights into the diversification of salamanders in the *Isthmura bellii* group across the Mexican highlands. Mol Phylogenet Evol. 2018;125:78–84.

Campbell NR, Harmon SA, Narum SR. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Res. 2015;15:855–67.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. G3. 2011;1:171–82.
Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22:3124–40.
Czypionka T, Krugman T, Altmüller J, Blaustein L, Steinfartz S, Templeton AR. Ecological transcriptomics – a non-lethal sampling approach for endangered fire salamanders. Methods Ecol Evol. 2015;6:1417–25.
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510.
Eaton DA. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 2014;30:1844–9.
Elewa A, Wang H, Talavera-Lopez C, Joven A, Brito G, Kumar A, et al. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. Nat Commun. 2017;8:2286.
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6:e19379.
Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol. 2012;61:717–26.
Fan JB, Chee MS, Gunderson KL. Highly parallel genomic assays. Nat Rev Genet. 2006;7:632–44.
Feng YJ, Liu QF, Chen MY, Liang D, Zhang P. Parallel tagged amplicon sequencing of relatively long PCR products using the Illumina HiSeq platform and transcriptome assembly. Mol Ecol Res. 2016;16:91–102.
Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB. Rapid spread of invasive genes into a threatened native species. Proc Natl Acad Sci U S A. 2010;107:3606–10.
Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhue C, Pudlo P, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 2013;22:3165–78.
Gibbs HL, Denton RD. Cryptic sex? Estimates of genome exchange in unisexual mole salamanders (*Ambystoma* sp.). Mol Ecol. 2016;25:2805–15.
Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27:182–9.
Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, et al. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Res. 2015;15:1304–15.
Gregory TR. Animal genome size database. 2018. http://www.genomesize.com. Accessed 14 Apr 2018.
Guo Y, Long J, He J, Li CI, Cai Q, Shu XO, et al. Exome sequencing generates high quality data in non-target regions. BMC Genomics. 2012;13:194.
Habermann B, Bebin AG, Herklotz S, Volkmer M, Eckelt K, Pehlke K, et al. An *Ambystoma mexicanum* EST sequencing project: analysis of 17,352 expressed sequence tags from embryonic and regenerating blastema cDNA libraries. Genome Biol. 2004;5:R67.
Hara Y, Tatsumi K, Yoshida M, Kajikawa E, Kiyonari H, Kuraku S. Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. BMC Genomics. 2015;16:977.
Hedtke SM, Morgan MJ, Cannatella DC, Hillis DM. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS One. 2013;8:e67908.

Hime PM, Hotaling S, Grewelle RE, O'Neill EM, Voss SR, Shaffer HB, et al. The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. Mol Ecol. 2016;25:5959–74.

Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, et al. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Res. 2016;16:1264–78.

Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire J-Y, Kupfer A, et al. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol. 2017;1:1370–8.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992;258:818–21.

Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proc Natl Acad Sci U S A. 2017;114:E1460–9.

Keinath MC, Timoshevskiy VA, Timoshevskaya NY, Tsonis PA, Voss SR, Smith JJ. Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. Sci Rep. 2015;5:16413.

Keinath MC, Voss SR, Tsonis PA, Smith JJ. A linkage map for the newt *Notophthalmus viridescens*: insights in vertebrate genome and chromosome evolution. Dev Biol. 2017;426:211–8.

Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34:1812–9.

Lemmon AR, Lemmon EM. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Syst Biol. 2012;61:745–61.

Lemmon EM, Lemmon AR. High-throughput genomic data in systematics and phylogenetics. Annu Rev Ecol Evol Syst. 2013;44:99–121.

Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol. 2012;61:727–44.

Lepais O, Weir JT. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Res. 2014;14:1314–21.

Licht LE, Lowcock LA. Genome size and metabolic rate in salamanders. Comp Biochem Physiol B Biochem Mol Biol. 1991;100:83–92.

Linnen CR, Poh YP, Peterson BK, Barrett RD, Larson JG, Jensen JD, et al. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science. 2013;339:1312–6.

Looso M, Preussner J, Sousounis K, Bruckskotten M, Michel CS, Lignelli E, et al. A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. Genome Biol. 2013;14:R16.

Lucas LK, Gompert Z, Gibson JR, Bell KL, Buerkle CA, Nice CC. Pervasive gene flow across critical habitat for four narrowly endemic, sympatric taxa. Freshw Biol. 2016;61:933–46.

Madison-Villar MJ, Sun C, Lau NC, Settles ML, Mueller RL. Small RNAs from a big genome: the piRNA pathway and transposable elements in the salamander species *Desmognathus fuscus*. J Mol Evol. 2016;83:126–36.

Matsunami M, Kitano J, Kishida O, Michimae H, Miura T, Nishimura K. Transcriptome analysis of predator- and prey-induced phenotypic plasticity in the Hokkaido salamander (*Hynobius retardatus*). Mol Ecol. 2015;24:3064–76.

McCartney-Melstad E, Mount GG, Shaffer HB. Exon capture optimization in amphibians with large genomes. Mol Ecol Res. 2016;16:1084–94.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 2007;17:240–8.

Mohlhenrich ER, Mueller RL. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. Evolution. 2016;70:2865–78.

Murphy MO, Jones KS, Price SJ, Weisrock DW. A genomic assessment of population structure and gene flow in an aquatic salamander identifies the roles of spatial scale, barriers, and river architecture. Freshw Biol. 2018;63:407–19.

Newman CE, Austin CC. Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Mol Ecol. 2016;25:6162–74.
Nowoshilow S, Schloissnig S, Fei JF, Dahl A, Pang AW, Pippel M, et al. The axolotl genome and the evolution of key tissue formation regulators. Nature. 2018;554:50–5.
Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. Mol Ecol. 2017;26:1060–74.
Olmo E. Quantitative variations in the nuclear DNA and phylogenesis of the Amphibia. Caryologia. 1973;26:43–68.
O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, Aguilar-Miguel X, et al. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Mol Ecol. 2013;22:111–29.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7:e37135.
Portik DM, Smith LL, Bi K. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol Ecol Res. 2016;16:1069–83.
Putta S, Smith JJ, Walker JA, Rondet M, Weisrock DW, Monaghan J, et al. From biomedicine to natural history research: EST resources for ambystomatid salamanders. BMC Genomics. 2004;5:54.
Qi Z, Zhang Q, Wang Z, Ma T, Zhou J, Holland JW, et al. Transcriptome analysis of the endangered Chinese giant salamander (*Andrias davidianus*): immune modulation in response to *Aeromonas hydrophila* infection. Vet Immunol Immunopathol. 2016;169:85–95.
Rodríguez A, Burgon JD, Lyra M, Irisarri I, Baurain D, Blaustein L, et al. Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. Mol Phylogenet Evol. 2017;115:16–26.
Sessions SK. Evolutionary cytogenetics in salamanders. Chromosom Res. 2008;16:183–201.
Shen XX, Liang D, Feng YJ, Chen MY, Zhang P. A versatile and highly efficient toolkit including 102 nuclear markers for vertebrate phylogenomics, tested by resolving the higher level relationships of the Caudata. Mol Biol Evol. 2013;30:2235–48.
Shen XX, Liang D, Chen MY, Mao RL, Wake DB, Zhang P. Enlarged multilocus data set provides surprisingly younger time of origin for the Plethodontidae, the largest family of salamanders. Syst Biol. 2016;65:66–81.
Smith JJ, Kump DK, Walker JA, Parichy DM, Voss SR. A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in *Ambystoma*. Genetics. 2005;171:1161–71.
Smith JJ, Putta S, Zhu W, Pao GM, Verma IM, Hunter T, et al. Genic regions of a large salamander genome contain long introns and novel genes. BMC Genomics. 2009;10:19.
Sun C, Mueller RL. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. Genome Biol Evol. 2014;6:1818–29.
Sun C, Shepard DB, Chong RA, Lopez Arriaza J, Hall K, Castoe TA, et al. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol. 2012;4:168–83.
Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31:46–53.
Voss SR, Smith JJ. Evolution of salamander life cycles: a major-effect quantitative trait locus contributes to discrete and continuous variation for metamorphic timing. Genetics. 2005;170:275–81.
Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR. Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. Mol Ecol. 2006;15:2489–503.
Wielstra B, Arntzen JW. Kicking *Triturus arntzeni* when it's down: large-scale nuclear genetic data confirm that newts from the type locality are genetically admixed. Zootaxa. 2014;3802:381–8.

Wielstra B, Duijm E, Lagler P, Lammers Y, Meilink WR, Ziermann JM, et al. Parallel tagged amplicon sequencing of transcriptome-based genetic markers for *Triturus* newts with the ion torrent next-generation sequencing platform. Mol Ecol Res. 2014a;14:1080–9.

Wielstra B, Arntzen JW, van der Gaag KJ, Pabijan M, Babik W. Data concatenation, Bayesian concordance and coalescent-based analyses of the species tree for the rapid radiation of *Triturus* newts. PLoS One. 2014b;9:e111011.

Zieliński P, Stuglik MT, Dudek K, Konczal M, Babik W. Development, validation and high-throughput analysis of sequence markers in nonmodel species. Mol Ecol Res. 2014a;14:352–60.

Zieliński P, Dudek K, Stuglik MT, Liana M, Babik W. Single nucleotide polymorphisms reveal genetic structuring of the carpathian newt and provide evidence of interspecific gene flow in the nuclear genome. PLoS One. 2014b;9:e97431.

Zieliński P, Nadachowska-Brzyska K, Dudek K, Babik W. Divergence history of the Carpathian and smooth newts modelled in space and time. Mol Ecol. 2016;25:3912–28.

# Part III
# Wildlife Populations: Ecology, Evolution and Adaptation

# Landscape Genomics for Wildlife Research


**Brenna R. Forester, Erin L. Landguth, Brian K. Hand, and Niko Balkenhol**

**Abstract** Landscape genomics investigates how spatial and environmental factors influence geographic patterns of genome-wide genetic variation. Adaptive landscape genomics focuses on how these spatial and environmental processes structure the amount and distribution of selection-driven genetic variation among populations, which ultimately determines how phenotypic variation is arrayed across landscapes. This adaptive landscape genomics approach can be used to identify the causal factors underlying local adaptation and has great potential to guide decision-making in applied wildlife research, especially in light of anthropogenic climate and land use change. Conservation and management applications include delineating conservation units, designing conservation monitoring programs, and predicting changes in the spatial distribution and potential loss of adaptive genomic variation under environmental change. However, there remains great untapped potential for the application of adaptive landscape genomics to wildlife research, including moving beyond correlative genotype-environment association tests. In this chapter, we explore and discuss the potential of adaptive landscape genomics for improving wildlife research, including case studies that illustrate its application in wildlife management and conservation. We also present a comprehensive workflow for adaptive landscape genomics studies in wildlife, including sampling design, genomic and environmental data production, and data analysis. We conclude with avenues and perspectives for future work and ongoing challenges in adaptive landscape genomics.

B. R. Forester (✉)
Department of Biology, Colorado State University, Fort Collins, CO, USA
e-mail: brenna.forester@colostate.edu; brenna.forester@gmail.com

E. L. Landguth
School of Public and Community Health Sciences, University of Montana, Missoula, MT, USA

B. K. Hand
Flathead Lake Biological Station, University of Montana, Polson, MT, USA

N. Balkenhol
Faculty of Forest Sciences, Wildlife Sciences, University of Goettingen, Göttingen, Germany

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_56,

**Keywords** Adaptive capacity · Adaptive genetic variation · Conservation genomics · Genome-wide association studies · Genotype-environment associations · Natural selection

## 1 Adaptive Landscape Genomics and Wildlife Research

For over three decades, the study of genetic variation in wildlife species has been used to investigate ecological and evolutionary questions, especially those related to management and conservation (e.g., Frankham 1995; Frankham et al. 2017). More recently, wildlife species have been a focus of landscape genetics research, including the investigation of functional connectivity in heterogeneous landscapes, the identification of source-sink dynamics and barriers to gene flow, and linkage of processes, such as land use change and degradation, to patterns of genetic variation (Storfer et al. 2010; Selkoe et al. 2015; Waits et al. 2015).

Prior to the development of next-generation sequencing (NGS), however, the examination of adaptive genetic variation in wildlife was relatively limited (Bensch and Åkesson 2005). In the past decade, NGS technologies, with their falling costs, improvements in genotyping yield and quality, and increasing accessibility (Goodwin et al. 2016), have created an unprecedented opportunity to study local adaptation and natural selection in wildlife populations. In an adaptive landscape genomics framework, these data allow for the explicit investigation of how environmental and spatial processes structure the amount and distribution of selection-driven genetic variation among wildlife populations. This, in turn, informs our understanding of the ecological and evolutionary processes at work, as well as how best to manage and conserve the adaptive capacity of wildlife populations in the face of complex and interacting environmental changes (Sgro et al. 2011; Harrisson et al. 2014; McMahon et al. 2014; Hoffmann et al. 2015; Funk et al. 2018).

Despite these advances, there are many practical challenges of working with wildlife that limit the use of adaptive landscape genomics, as reflected in the small number of published empirical studies to date (see below). For many wildlife species, sampling can be a major obstacle, including obtaining adequate sample sizes, appropriate sample stratification across environmental gradients, and DNA of sufficient quality and quantity. Study inference can often be limited by a lack of basic ecological information including the proximal environmental drivers of natural selection. Finally, a lack of genomic resources and an inability to use experimental approaches in controlled environments (e.g., common gardens) for the majority of wildlife species limits validation of identified candidate adaptive variants. These issues are not insurmountable, however, and the integration of complementary data and analyses can be used to improve inference in many cases.

In the following sections, we present a workflow for adaptive landscape genomics studies in wildlife, including sampling design, genomic and environmental data production, and data analysis. We follow up with a discussion of the potential of

adaptive landscape genomics for wildlife studies, including case studies that illustrate the application of adaptive landscape genomics in wildlife management and conservation. We conclude with avenues and perspectives for future work and ongoing challenges in adaptive landscape genomics.

## 2 Data Production for Adaptive Landscape Genomics Research in Wildlife

Adaptive landscape genomics includes a range of techniques for identifying and analyzing spatially structured, selection-driven genetic variation, including correlative genotype-environment associations (GEA), phenotypic approaches like genome-wide association studies (GWAS) and quantitative trait locus mapping (QTL), candidate-gene methods, and exome and transcriptomic approaches (Storfer et al. 2015; Fig. 1). In this chapter, we focus on the identification of adaptive variants through GEA, because this is the analytical framework that explicitly incorporates environmental variation into the identification of selection-driven variation and is the most widely used landscape genomics approach in wildlife to date (Balkenhol et al. 2017). The reasons for this are practical and include (1) no requirement for

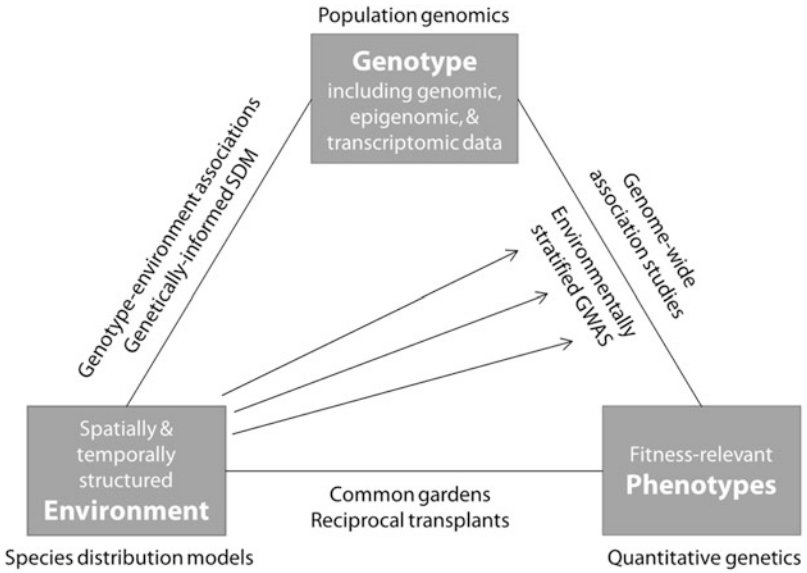


**Fig. 1** Expanded view of the adaptive landscape genomics framework, illustrating how multiple data types (boxes) can inform the relationships among spatially and temporally structured environmental heterogeneity, genotypes (including genomic, epigenomic, and transcriptomic data), and fitness-relevant phenotypes. Analytical approaches (black text) can be integrated to improve our understanding of adaptation in wildlife species. Figure adapted from Sork et al. (2013)

phenotypic data, which can be difficult to collect in the field on large numbers of individuals; (2) no requirement for manipulative experiments such as crosses, common gardens, and reciprocal transplants, which are impossible for many wildlife species; and (3) no requirement for prior genomic resources. While any landscape genomic study would be improved by the inclusion of complementary data sets and resources (Sect. 4), they are not essential to addressing many of the questions relevant to adaptive genetic variation in wildlife.

## 2.1 Landscape Sampling Designs

Sampling design is a fundamental component of robust research, but it is often unclear how different sampling strategies affect interpretation of landscape genomic results, including those obtained using GEA. So far, we know that the power of GEA methods can depend strongly on sampling design, which includes how samples are arrayed across the landscape, the total number of samples, and whether sampling is individual- or population-based. Generally, it is best to array samples across the maximum range of environmental variation that is thought to drive selection, for example, collecting samples from the lowest and highest elevation populations for a montane species. Though intermediate samples (i.e., along a transect) can be useful in addressing some questions (e.g., the spatial scale of local adaptation), for most GEA studies, intermediate samples will contribute little in terms of power (Lotterhos and Whitlock 2015). Paired sampling that maximizes environmental distance while minimizing geographic distance has shown high power and low false-positive rates in simulations, since this design maximizes environmental signal while minimizing the confounding effects of population structure (Lotterhos and Whitlock 2015). This sampling design may not be practical for many real species inhabiting complex landscapes, however.

Perhaps unsurprisingly, it is clear from simulation studies that increasing total sample size increases power to detect signatures of selection (De Mita et al. 2013; Lotterhos and Whitlock 2015). However, how those samples are arrayed, either within populations or as individuals, has complex trade-offs in power and false-positive rates. Population-based analyses combine individual genotypes into allele frequency estimates for the population, whereas individual-based analyses use individual genotypes as the response. Population-based sampling involves a compromise between the number of populations sampled and the number of individuals used within each population to estimate allele frequencies. Generally, fewer individuals sampled across more populations is the most effective approach for population-based GEA analysis (De Mita et al. 2013), though the impact of sample allocation can be method- and demography-dependent (Lotterhos and Whitlock 2015). Individual-based sampling and analysis tends to improve power due to the increase in the number of observations; however, this can also lead to increased false positives for univariate (though not multivariate) GEAs (de Villemereuil et al. 2014; Forester et al. 2018). Several studies have also revealed important impacts of

sampling design, including sample size, on characterizing the signal of IBD and other processes that influence spatial genetic patterns (Landguth et al. 2012; Oyler-McCance et al. 2013; Prunier et al. 2013). For example, a simulation study by Landguth et al. (2012) found that although the strength of environmental correlation values was not affected by sample size, the variance increased as sample size decreased. This suggests that an increase in noise of spatial genetic data could play a role in the ability to correctly identify loci under selection.

When possible, replicating sampling across multiple environmental gradients can improve the strength of inference (i.e., more evidence for true adaptive detection, Table 1), if the same candidate loci are identified (e.g., Hohenlohe et al. 2010; Swaegers et al. 2015; Brauer et al. 2016). While a lack of replicated detection can be indicative of false positives (Buehler et al. 2014), parallel adaptation through different genes and genetic architectures (i.e., the underlying genetic basis of a phenotypic trait) is also a possible explanation and one that can occur for a variety of reasons, including metapopulation dynamics, limited dispersal, and habitat heterogeneity (Ralph and Coop 2015a, b; Bernatchez 2016). Disentangling false positives from parallel adaptation involving different genes and architectures is not a simple task in species with limited genomic resources, so a lack of replicated detection should be interpreted cautiously. This complexity illustrates why GEA studies conducted in a small part of a species range should not be naively extrapolated to the entirety of the range, especially for species with strong geographic population structure (Hand et al. 2016).

There are still many avenues for research with regard to sampling protocol for GEA methods. Future work should explore simulations that evaluate sample allocation and effort for GEA method performance in both discrete populations and continuously distributed individuals (e.g., Prunier et al. 2013; Landguth and Schwartz 2014) and across spatially complex environments that control for both gene flow and selection. Most studies thus far have used a spatially random sample drawn from the population(s) to test GEA performance, but in reality, a truly random sampling design is not only difficult to achieve in the field but can also have reduced power relative to more strategic sampling strategies (Lotterhos and Whitlock 2015). Future simulations should consider different sampling designs as well as the number of loci sampled and the effect of missing data.

## *2.2 Genomic Data Production*

Genomic data production begins with determining which molecular approach will be best suited to the questions under consideration, to the amount of data that will be needed to answer those questions, and to the budget available (Benestan et al. 2016; Hohenlohe et al. 2017). Because most wildlife species lack genomic resources, questions related to adaptive variation in wildlife will usually be best served with an anonymous reduced representation sequencing approach (though see Sect. 4 for a discussion of whole-genome resequencing in wildlife species).

**Table 1** Approaches to improve inference of local adaptation beyond anonymous NGS studies of genotype-environment associations

| | Improvement in understanding of local adaptation | Data, logistical, and analytical requirements | General considerations for wildlife species | Examples |
|---|---|---|---|---|
| Replication of GEA | Identification of the same candidate adaptive loci across replicated gradients improves confidence; lack of replication can indicate parallel adaptation through alternate genetic architectures or false positives | Additional field work, sequencing, and bioinformatics required for additional samples | If possible, should be included in wildlife GEA studies due to relatively low cost and potentially high benefit | Hohenlohe et al. (2010), Buehler et al. (2014), Swaegers et al. (2015), Brauer et al. (2016), and Hand et al. (2016) |
| Simulation | Many uses including optimizing sampling design to maximize potential for identifying local adaptation; corroborating empirical findings; forecasting future distributions of adaptive variation; developing new theory | No additional data required; analytical efforts depend on the complexity of the simulation framework and scenarios | Given the logistical complexities and limitations of working with many wildlife species, there is considerable room for the creative application of simulation in adaptive studies of wildlife | Cooke et al. (2014), Landguth et al. (2015), Lotterhos and Whitlock (2015), and Creech et al. (2017) |
| Annotated genome | Can reduce data loss in anonymous NGS studies and improve data quality (e.g., genotyping accuracy) and inference (e.g., candidate validation); annotation can help with biological/ecological interpretation of candidate markers; can be used for targeted exon capture | Reference genome of focal or related species; quality of the reference and level of annotation will determine how useful it is; if producing a new genome, cost and effort may be high, depending on the size and complexity of the genome | Main uses to date have been in wildlife with genome-enabled domesticated relatives. There are few high-quality genomes for wildlife species, though that is changing as costs fall. Caution is needed if using genome of a distantly related species due to ascertainment bias | Roffler et al. (2016) and Ruegg et al. (2018) |
| Whole-genome resequencing | Identifies all genetic variation across sequenced individuals including structural variants (e.g., copy number variants) and mutations in regulatory elements. Increases power for detection of local adaptation and genetic basis of phenotypes | Requires a reference genome and large computational and storage capacity; greater expense than reduced representation sequencing but produces maximum information content | Not accessible for most wildlife species currently, especially those with large genome sizes (such as many amphibians); more information than is needed to address most conservation and management-related questions | Kardos et al. (2015) and Toews et al. (2016) |

| | | | | |
|---|---|---|---|---|
| Phenotypic data/GWAS | Identifying overlap in loci detected in environmentally stratified GWAS and loci detected by GEA improves strength of inference since independent detections are more likely to reflect true adaptive processes | GWAS requires dense genotyping to have sufficient power. Requires measurement of fitness-related phenotypic traits in many individuals | Unlike QTL mapping, GWAS does not require pedigree data or crosses, making it much more feasible for phenotypic studies in wildlife species | Wenzel et al. (2016) and Lasky et al. (2015) |
| Common gardens and reciprocal transplants | Provide direct evidence for local adaptation: common gardens establish the genetic basis of fitness (or related traits) while controlling for plasticity, while reciprocal transplants are a direct test of local adaptation | Logistically difficult, time-consuming, and labor-intensive, especially if experiments are replicated across many environments | Can only be used in species that can be reared under controlled conditions or in field-based experiments. Multiple generations raised in controlled conditions provides strongest inference. Have mainly been used in plants | De Kort et al. (2014) and Lasky et al. (2018) |
| Epigenetics | Extends our understanding of how the environment shapes phenotypic plasticity and local adaptation through non-genetic variation in gene expression. The extent of transgenerational transmission, especially in environmentally induced epigenetic responses, is very poorly understood | Reference-free data production methods are available. Many studies use a transcriptome or annotated genome for data production and/or interpretation | Feasible for studies of associations between environmental and epigenetic variation; establishing transgenerational inheritance requires experimental approaches that are not possible in many wildlife species | Dimond and Roberts (2016), Lea et al. (2016), and Trucchi et al. (2016) |
| Transcriptomics | Can be used to identify how environmental variation affects gene expression and functionally adaptive variation; can provide insight into the capacity of species to adapt to changing environmental conditions | Tissue requirements often involve sacrifice of animals. Especially in field-based studies, replicate sampling across treatments is required. RNA-specific lab spaces are required. Data can be produced de novo or by mapping to a reference genome | Requires careful experimental design and (for best inference) controlled, experimental treatments. Field-based investigations are possible, but inference of environmental processes driving differential gene expression is limited | Barshis et al. (2013), O'Neil et al. (2014), and Narum and Campbell (2015) |

Reduced representation methods require no prior genomic information and randomly subsample the genome to identify single nucleotide polymorphisms (SNPs). The most common anonymous sequencing family is restriction-site associated DNA sequencing (RADseq), which targets DNA adjacent to restriction enzyme cut sites (Andrews et al. 2016). While there are limitations to RADseq approaches (reviewed in Andrews et al. 2016), these methods are relatively inexpensive and produce data from both neutral and adaptive genomic regions which can be used to address a variety of management questions. Depending on genome size and levels of linkage disequilibrium, varying proportions of the genome will be sampled by RADseq methods (Lowry et al. 2016; McKinney et al. 2017). Most management questions related to adaptive variation, however, do not require a complete assessment of adaptive variation across the entire genome, nor the functional validation of candidate loci (Sect. 3). In many cases, characterizing the geographic and environmental patterns of potentially adaptive variation across populations will be sufficient (Catchen et al. 2017; Hohenlohe et al. 2017).

For species without genomic resources, downstream bioinformatics analyses of RADseq data (i.e., locus assembly, genotype, and SNP calling) are done de novo (e.g., Catchen et al. 2013). In cases where a high-quality reference genome is available for the focal or a closely related species (e.g., ~tens of millions of years divergent), RAD loci can be aligned to the reference, which can reduce data loss when compared to the stringent filtering required in a de novo framework (Table 1; Cosart et al. 2011). However, aligning to a poor quality or divergent reference genome can result in informative loci being lost. Current best practices for RADseq bioinformatics advise first building loci de novo, then aligning consensus loci to the reference (Paris et al. 2017).

When prior genomic resources are available (such as a previous RADseq assessment), targeted capture can be used to sequence data from a subset of previously identified regions (Jones and Good 2016). Recent developments that combine RADseq and capture methods (Ali et al. 2016; Hoffberg et al. 2016) have expanded the accessibility of targeted capture and can be used to optimize genotyping across individuals by ensuring consistent coverage of the same loci. In cases where a modest number of SNPs (50–500) are needed, genotyping in thousands by sequencing (GT-seq) can provide a means for genotyping SNPs in thousands of individuals in a single lane of sequencing (Campbell et al. 2015). These methods are especially useful for developing genetic monitoring panels (Flanagan et al. 2018; Schwartz et al. 2007) since they maximize coverage of targeted markers while minimizing cost per individual genotype. See Flanagan et al. (2018) and Benestan et al. (2016) for more detailed overviews of analytical pipelines for reduced representation and targeted sequencing methods, including data filtering, locus assembly and alignment, and genotype and SNP calling.

Finally, when a high-quality, well-annotated reference genome is available and the goal is to capture SNPs in genes of known function, exon capture can be used (e.g., Roffler et al. 2016) (Table 1), though it is often more costly than other approaches (Harvey et al. 2016; Manel et al. 2016). Exon capture can also target ultra-conserved elements and expressed sequence tags (e.g., McCartney-Melstad et al. 2016). Transcriptome sequencing is another option to identify transcribed

portions of the genome; however, rigorous transcriptomic studies will be challenging to implement in most wildlife species (Sect. 4, Table 1).

DNA requirements vary for these methods, though most anonymous and targeted NGS sequencing protocols require 50–100 ng of high-quality DNA, which can usually be satisfied with nonlethal blood or tissue samples. However, for very small animals, such as invertebrates, whole individuals may have to be sacrificed (e.g., Lozier 2014). Noninvasive genetic samples, including hair snags and fecal pellets, have been widely used in microsatellite studies of wildlife species that are difficult and/or expensive to capture (e.g., American black bear, Cushman et al. 2006). Unfortunately, the low levels of DNA present in these samples have posed a challenge for NGS methods, though new techniques are being developed to facilitate the use of these samples in genomic analyses (Andrews et al. 2018). For example, a recent study successfully used NGS on hair samples from American pika (*Ochotona princeps*), indicating that, with appropriate precautions and supporting information (a genomic scaffold in this case), even small amounts of DNA (as little as 1 ng) can be used for adaptive landscape genomics (Russello et al. 2015).

## *2.3 Environmental Data Production*

Ideally, the environmental data component of GEA analysis should be developed based on the physiology and ecology of the focal species and known or hypothesized drivers of selection. When these factors are not well known, an exploratory approach will be required using available biotic and abiotic predictors. Because of its focus on local environmental conditions, adaptive landscape genomics does not necessarily require continuous environmental surfaces (e.g., interpolated raster layers created using a geographic information system, or GIS). However, these layers are often the default environmental data sources since they cover broad geographic areas (ensuring spatial coverage of sampled individuals and populations) and, for climate data, include long-term temporal resolution (Leempoel et al. 2017). Especially for long-lived wildlife species, sufficient temporal coverage in climate variables is important to capture interannual variability that matches long-term selective pressures. Most publicly available gridded data sets include data at these coarser annual and decadal scales (e.g., CHELSA; Karger et al. 2017). By contrast, high temporal resolution (e.g., daily data; TOPOFIRE; Holden et al. 2013) may be necessary for some species and questions, since extremes, cumulative exposure, and threshold events may be important selective pressures (e.g., Welbergen et al. 2008; Vasseur et al. 2014; Buckley and Huey 2016; Gardner et al. 2016). These data are available, but usually on reduced spatial scales. The relatively coarse spatial scale of these interpolated data sets (usually ~1 km resolution) can also be problematic for smaller species, since these data cannot capture relevant microclimates and can underestimate habitat and climatic heterogeneity (Nadeau et al. 2017). In these cases, field-based sensors for site-specific data collection (e.g., HOBOs, Onset Computer Corp.) and modeling may provide the highest spatial resolution (e.g., Peterman and Semlitsch 2013), but these data will be limited by their reduced

temporal component (i.e., limited to the time period during which data were collected). Exceptions to this temporal data issue include major catastrophic or selective events, such as severe drought or virulent disease (e.g., Grant and Grant 1993; Epstein et al. 2016). Finally, another option for site-specific climate data with potentially long temporal resolution is weather station data (e.g., RAWS; https://raws.dri.edu/); however, these data are typically sparse in space.

There are a large number of free climate and land cover GIS data sets available. Caution and skepticism are warranted, however, since these data sets can vary widely in their quality (Daly 2006). High-quality global data are available from CHELSA (Karger et al. 2017) and the Consortium for Spatial Information (www.cgiar-csi.org/data) for climate and water balance and from EarthEnv (www.earthenv.org) for land cover and vegetation. Other regional climate resources (mostly North America) include TopoWx (Oyler et al. 2015), ClimateNA (Wang et al. 2016), and water balance data through AdaptWest (Dobrowski et al. 2013) for North America and Copernicus for global- to European-scale data (http://climate.copernicus.eu and http://land.copernicus.eu). In addition, the growing availability of environmental data from remotely sensed data products (e.g., http://earthdata.nasa.gov), including climate and land use time series datasets, offers unprecedented opportunities in landscape genomics to account for environmental variables (and changes in those variables) influencing local adaptation.

Once available data are identified, the next step is to extract and/or calculate relevant predictors. Whenever possible, it is best to use proximal (e.g., temperature, precipitation) as opposed to distal (e.g., elevation, latitude) predictors, since proximal variables are more ecologically relevant and may decouple from their distal proxies over time, for example, under climate change (Lookingbill and Urban 2005). Many environmental predictors will be highly correlated with each other (e.g., $|r| > 0.7$, Dormann et al. 2013), which is statistically problematic for most GEA methods and confounds interpretation. Methods for dealing with collinearity include pruning predictors using a pairwise correlation matrix or variance inflation factor (retaining the variable most relevant or ecologically interpretable) or using PCA to reduce a large number of predictors into orthogonal synthetic predictors, though this can limit ecological interpretation. Finally, because GEAs are correlative models, environmental predictors identified as important by a GEA analysis may not be the driving selective pressures. Collinear predictors that were removed during pruning and other correlated but unmeasured biotic and abiotic conditions may also be important (Rellstab et al. 2015).

## *2.4 Data Analysis with Genotype-Environment Associations*

Genotype-environment association (GEA) analysis (also called environmental association analysis) refers to a group of statistical analyses used in adaptive landscape genomics to partition neutral from potentially adaptive genetic variation. GEA methods identify candidate adaptive loci based on associations between allele

distributions and environmental variables hypothesized to drive selection, reflecting a pattern of selected alleles at higher frequency in certain environments (reviewed in Rellstab et al. 2015). This is in contrast to a population genomics approach, where adaptive loci are identified by differentiating locus-specific patterns (caused by locus-specific processes including selection) from the genome-wide pattern (caused by genome-wide processes, such as genetic drift, demographic processes, and gene flow; Luikart et al. 2003). These differentiation-based methods are useful for detecting strong divergent selection (Storz 2005) and are especially valuable when environmental predictors are not available or when the number of sampled populations is small. Limitations of differentiation-based approaches include a requirement for population-based sampling and a basis in theoretical population genetic models that are violated by many empirical systems. By contrast, many GEA methods can be used with either individual- or population-based sampling, which is advantageous when environmental gradients are continuous or when populations are not clearly distinguishable (Jones et al. 2013). Additionally, most GEA methods do not use an underlying population genetic model and so are not sensitive to the deviations common in empirical systems. GEA methods are not limited to detecting divergent selection, but can also detect weaker selective pressures, such as selection on standing genetic variation (Forester et al. 2018; Fig. 2). Finally, the inclusion of environmental predictors improves power when compared with differentiation-based methods (De Mita et al. 2013; de Villemereuil et al. 2014; Rellstab et al. 2015).

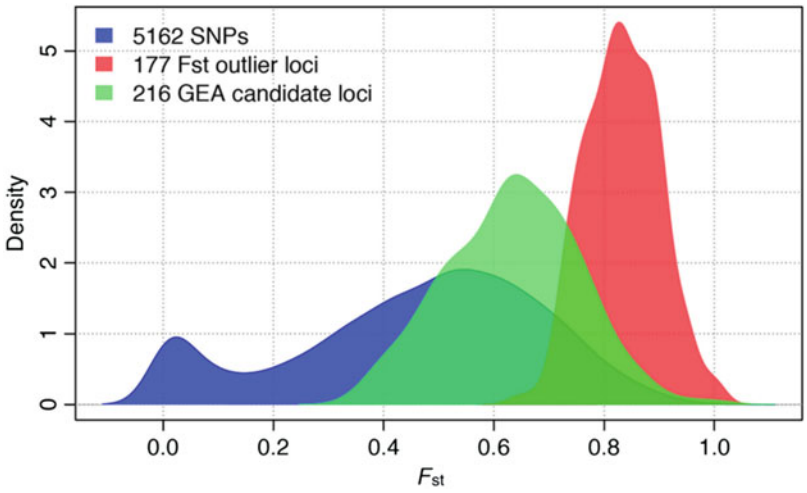


**Fig. 2** Density distribution of $F_{ST}$ values for SNPs identified in southern pygmy perch (*Nannoperca australis*): 5,162 neutral and candidate adaptive SNPs (blue), 177 candidate loci identified using $F_{ST}$ outlier/differentiation-based methods (red), and 216 candidate loci identified using GEA methods (green). The narrow $F_{ST}$ distribution of outlier candidates with a high mean $F_{ST}$ (~0.8) is indicative of divergent selection and allelic fixation, while the broader $F_{ST}$ distribution of GEA candidates with a lower mean $F_{ST}$ (~0.6) is indicative of polygenic selection from standing genetic variation. Figure from Brauer et al. (2016)

However, differentiation-based tests can provide complementary information to GEA results, since they can detect the impact of selective pressures that may not be captured by the selected environmental predictors.

Both sets of methods are sensitive to neutral genetic structure, which can produce spatial patterns that resemble selection, resulting in elevated false-positive rates in the absence of correction. Most GEA methods incorporate an approach for controlling for neutral genetic structure, including covariance matrices, the probability of membership from clustering or ordination analyses, spatial predictors, and other spatial analyses such as trend surface analysis (Rellstab et al. 2015). These corrections are applied in a diversity of GEA methods, most of which are univariate, meaning they test one locus and one predictor variable at a time. These methods include generalized linear models (e.g., Joost et al. 2007; Stucki et al. 2017), linear mixed effects models (e.g., Coop et al. 2010; Frichot et al. 2013; Lasky et al. 2014; Yoder et al. 2014), and nonparametric models (e.g., partial Mantel test, Hancock et al. 2011). Univariate GEAs can produce elevated false-positive rates due to the multiple comparisons required to test individual SNPs and predictors. Multiple test corrections, such as Bonferroni, can be overly conservative (potentially removing true-positive detections), while alternative correction methods, such as false discovery rate (Benjamini and Hochberg 1995), rely on an assumption of a null distribution of *p*-values, which may often be violated in empirical data sets. While these issues should not discourage the use of univariate methods (though corrections should be chosen carefully, see François et al. (2016) for an overview), other statistical approaches may be better suited to the high dimensionality of genomic data sets.

In particular, multivariate GEAs, which can analyze all loci and predictor variables simultaneously, are well suited to data sets comprising hundreds of individuals sampled at many thousands of genetic markers. These methods can more effectively detect multilocus selection since they consider how groups of markers covary in response to environmental predictors (Rellstab et al. 2015; Forester et al. 2018). This is important because many adaptive processes are expected to result in weak, polygenic molecular signatures. These include selection on standing genetic variation, recent or contemporary selection that has not yet led to allele fixation, and conditional neutrality (Yeaman and Whitlock 2011; Le Corre and Kremer 2012; Savolainen et al. 2013; Tiffin and Ross-Ibarra 2014). Multivariate GEAs include redundancy analysis (RDA), which was recently tested in a simulation framework and showed a superior combination of high true-positive and low false-positive rates while being robust to different demographic histories, sampling designs, and sample sizes (Forester et al. 2018). While additional testing is needed, multivariate GEAs show promise as a powerful complement to univariate detection approaches.

Finally, our understanding of the performance of GEA methods under realistic genomic architectures and sampling effects has been limited by the relatively simplistic simulation frameworks used to date. Additional testing of GEA methods on more complex genetic architectures (e.g., conditional neutrality vs. antagonistic pleiotropy, multilocus vs. polygenic selection) and realistic genomic sampling conditions is needed (e.g., Yoder and Tiffin 2017). In sum, it is important to realize that different approaches for detecting loci influenced by selection can yield different

conclusions, and a common standard for reporting the analysis of adaptive genomic data has yet to be developed (see Ahrens et al. 2018).

## 3 Applications and Potential of Adaptive Landscape Genomics in Wildlife Research

Adaptive landscape genomics has been used in many studies to identify candidate adaptive variation, its environmental drivers, and spatial distribution. While these studies are important in that they provide a baseline for additional research, there are still a limited number of wildlife studies that move beyond data generation to directly address ecological and evolutionary questions and/or management issues. In this section, we discuss recent applications of adaptive landscape genomics in wildlife research and highlight notable case studies. We then follow up with areas for advancement that have not yet been implemented in wildlife genomics.

### *3.1 Current Applications of Adaptive Landscape Genomics in Wildlife Research*

#### 3.1.1 What Are the Ecological and Evolutionary Processes Underlying Spatial Patterns of Neutral and Adaptive Genetic Variation?

Explicit investigation of how neutral and candidate adaptive genetic variation relate to spatial and environmental variation can provide insight into the ecological and evolutionary processes that generate observed genetic patterns. These patterns include isolation by distance (IBD) and isolation by environment (IBE), which can be explained by the processes of isolation by dispersal limitation and isolation by adaptation, respectively (Orsini et al. 2013; Wang and Bradburd 2014). These analyses can provide information on species biology, ecology, and evolutionary history, including estimates of dispersal distance, biased dispersal (e.g., due to fitness advantage or natal habitat preference), colonization history, natural or sexual selection against immigrants, and reduced hybrid fitness. In a recent application, Manthey and Moyle (2015) tested for patterns of IBD and IBE in white-breasted nuthatches (*Sitta carolinensis*) inhabiting the sky islands of the southeastern United States. By investigating both neutral and candidate adaptive markers, they identified IBE as the significant pattern structuring both neutral and adaptive markers, with an absence of IBD. Extremes of temperature and precipitation structured environmental adaptation due to nonrandom gene flow among populations, pointing to a generative process of isolation by adaptation as a result of biased dispersal (i.e., birds selecting more suitable environments).

### 3.1.2 What Are the Relative Roles of Genetic Drift and Natural Selection in Structuring Genetic Variation in Small Populations?

Genetic drift is expected to be the dominant genetic process in populations with small effective sizes, constraining selection and lowering adaptive capacity (e.g., Lande 1988; Willi et al. 2006). However, recent studies have suggested that, while small populations show a decline in genetic variation due to drift, these declines may not overwhelm selection or necessarily lower adaptive capacity (Brauer et al. 2016; Funk et al. 2016; Wood et al. 2016). This could have important implications for conservation and management of these populations, including informing genetic rescue and assisted migration efforts (Sect. 3.2.1). Brauer et al. (2016) provide a compelling example of retained adaptive divergence in spite of strong genetic drift and geographic isolation in the threatened southern pygmy perch (*Nannoperca australis*). Using univariate and multivariate GEAs, they found signatures of parallel polygenic adaptation to environmental and physical gradients that were replicated across demographically independent populations (Fig. 2). The smaller and more isolated headwater populations had less standing genetic variation at candidate adaptive loci in comparison to larger downstream populations, pointing to these latter populations as sources for genetic rescue or assisted migration efforts into recently and anthropogenically isolated populations.

### 3.1.3 How Can Knowledge of Adaptive Differentiation Inform the Delineation of Conservation Units?

Genomic data can improve the delineation of conservation units through increased resolution into neutral differentiation (e.g., Lah et al. 2016; Peters et al. 2016) and characterization of adaptive differentiation (Funk et al. 2012; Prince et al. 2017; Ruegg et al. 2018) (Fig. 3). For example, genomic data were used to identify neutral and adaptive differentiation in Baltic Sea herring (*Clupea harengus*) where previous studies using smaller genetic datasets had found little evidence for differentiation (Guo et al. 2016). This study provided additional evidence that current herring management units may have negative impacts on fisheries yields since they are poorly aligned with biological units based on local adaptation to salinity and temperature. As this case illustrates, adaptive differentiation can inform the delineation of ecotypes: populations (or subspecies) that are adapted to local environmental conditions. Defining ecotypes can be especially important in conservation efforts where the emphasis is not only on maintaining neutral genetic diversity but also overall evolutionary potential (Harrisson et al. 2014). While adaptive differentiation can be characterized using differentiation-based approaches (e.g., Cooke et al. 2014; Moura et al. 2014), landscape genomics provides additional insight into the environmental drivers of local adaptation, which can better inform conservation efforts (e.g., Pavey et al. 2015). For example, a recent study of the willow flycatcher

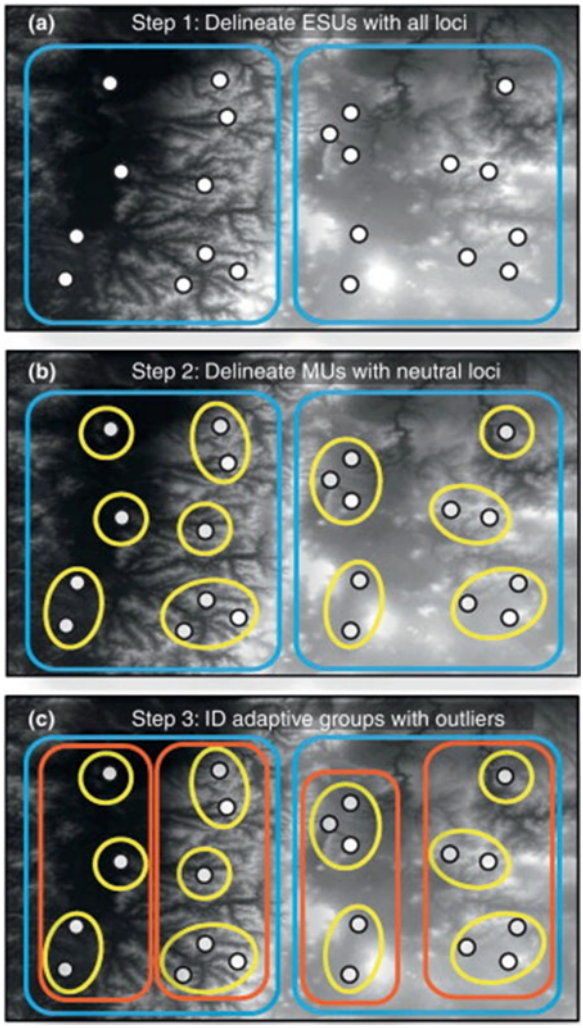


**Fig. 3** Workflow for using genomic data to delineate and test for adaptive differentiation among conservation units. White circles are sampling locations, blue outlines are evolutionarily significant units (ESUs), yellow outlines are management units (MUs), and orange outlines are adaptively similar groups of MUs. The grayscale background is an elevation layer (low to high represented by black to white). (**a**) Step 1: Delineate ESUs with all loci. (**b**) Step 2: Delineate MUs with neutral loci. (**c**) Step 3: Identify adaptive groups with outliers. Figure from Funk et al. (2012)

complex, including the endangered Southwestern willow flycatcher (*Empidonax traillii extimus*), supported the validity of this subspecies classification based on local adaptation related to temperature extremes (Ruegg et al. 2018). This study also found that the Southwestern subspecies was at the greatest risk for climate-mediated extinction due to high genomic vulnerability (a measure of the mismatch between adaptive genotypes and future environmental conditions).

### 3.1.4 How Can Adaptive Landscape Genomics Inform the Design of Conservation Monitoring Programs?

Depending on the conservation needs of a species, a genomic monitoring plan may be essential to effective management; however, there are few examples of genomic monitoring in the published literature. The best monitoring plans identify criteria for biologically significant change and develop a strategy for management intervention given detection of this change *prior* to initiating monitoring (Flanagan et al. 2018; Schwartz et al. 2007). This approach best ensures that monitoring will trigger timely management interventions, rather than just documenting decline and possibly extinction (Lindenmayer et al. 2013). An initial genomic assessment (e.g., RADseq study) can be used to identify a subset of neutral and candidate adaptive markers to be targeted for a monitoring panel using sequence capture or SNP arrays (e.g., Hohenlohe et al. 2011; Amish et al. 2012; Houston et al. 2014; Wright et al. 2015; Aykanat et al. 2016). While monitoring of neutral genetic variation can inform important parameters such as changes in genetic diversity and population size, monitoring of candidate adaptive variation can provide information on the status of adaptively divergent populations (Sect. 3.1.3) and management interventions such as assisted gene flow (Sect. 3.2.2).

For example, Hess et al. (2015) transitioned an initial NGS assessment (Hess et al. 2013) into a robust and multifaceted monitoring program for declining Pacific lamprey (*Entosphenus tridentatus*) (Fig. 4). They identified 96 neutral and candidate adaptive markers that were diagnostic for parentage analysis, cryptic species identification, and characterization of neutral and candidate adaptive genetic variation. These markers were incorporated into a SNP array and are currently being used to monitor the effectiveness of a diverse set of management actions including translocation, artificial propagation, and habitat restoration, as well as track population size, facilitate species identification at early life stages, and link adaptive markers to lamprey phenotypes (body size and migration timing). Appropriate sampling design for temporal monitoring of genetic change is still not well understood and will

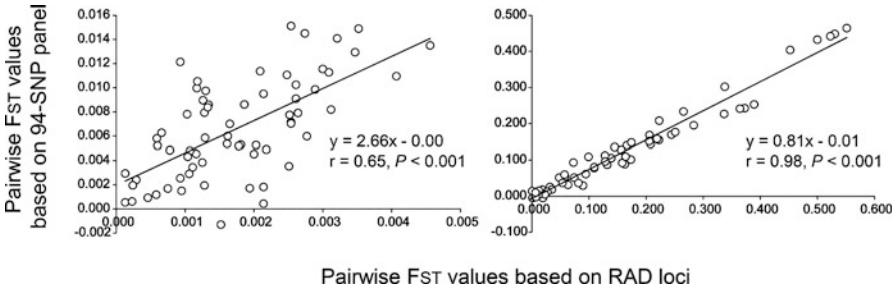


**Fig. 4** Correlation between pairwise $F_{ST}$ values from the full RADseq marker set and the 96-SNP monitoring panel developed for Pacific lamprey (*Entosphenus tridentatus*), showing neutral (left) and adaptive (right) SNPs for Columbia River Basin samples. Mantel tests indicate good representation of genomic-scale $F_{ST}$ values in the monitoring panel. Figure from Hess et al. (2015)

depend on the biology and demography of the species, the study objectives, and the power of the markers and sampled individuals to detect change (Schwartz et al. 2007; Allendorf et al. 2008; Hoban et al. 2014). Simulations, an underutilized tool in conservation management, will continue to play an important role in optimizing sampling design for genomic monitoring (Balkenhol and Landguth 2011; Hoban et al. 2013).

### 3.1.5 What Are the Genomic Implications of Range Expansion Under Climate Change?

Range shifts are some of the best-documented responses to climate change, with species across many taxa showing (sometimes idiosyncratic) changes in their distribution in response to changing climatic conditions (Hickling et al. 2006; Chen et al. 2011). Range expansions should leave predictable signals of founder effects and allele surfing at neutral loci (Excoffier et al. 2009), accompanied by changes in traits to facilitate adaptation and the rate of spread (Phillips et al. 2010). Linking genomic signals of evolution to these phenotypic changes has been rare, and it is not well understood how rapid trait changes on the expanding front are mediated by allele frequencies within populations. Swaegers et al. (2015) addressed these questions using a carefully planned adaptive landscape genomics study in a range-expanding damselfly (*Coenagrion scitulum*), accompanied by existing phenotypic data. By evaluating five different core-edge sets of populations, these authors demonstrated replicated neutral changes predicted by theory in independently established edge populations: founder effects, reduced gene flow, and higher levels of genetic drift. Using candidate adaptive markers, they identified parallel evolution for increased flight endurance in edge populations across four of the five populations, indicating convergent evolution from a locus that was polymorphic in the shared ancestral population (Fig. 5). Finally, using a multivariate GEA, the authors identified a genomic signal of adaptation to changing thermal regimes. This is one of only a few studies that has demonstrated a genetic basis to phenotypic changes during range expansion in response to climate change.

### 3.1.6 Can We Predict the Spatial Distribution of Adaptive Genetic Variation Under Changing Climates?

Species distribution models are commonly used in the ecological and conservation literature to predict changes in species distributions in response to climate change (e.g., Early and Sax 2011; Forester et al. 2013; Guisan et al. 2013; Hazen et al. 2013). These models generally use a static species-climate relationship for prediction and do not consider intraspecific variation in climate responses due to plasticity or local adaptation. However, with the increasing availability of genetic data for species of conservation concern, the incorporation of intraspecific variability into these models is now possible (e.g., Ikeda et al. 2017). These genetically informed models have

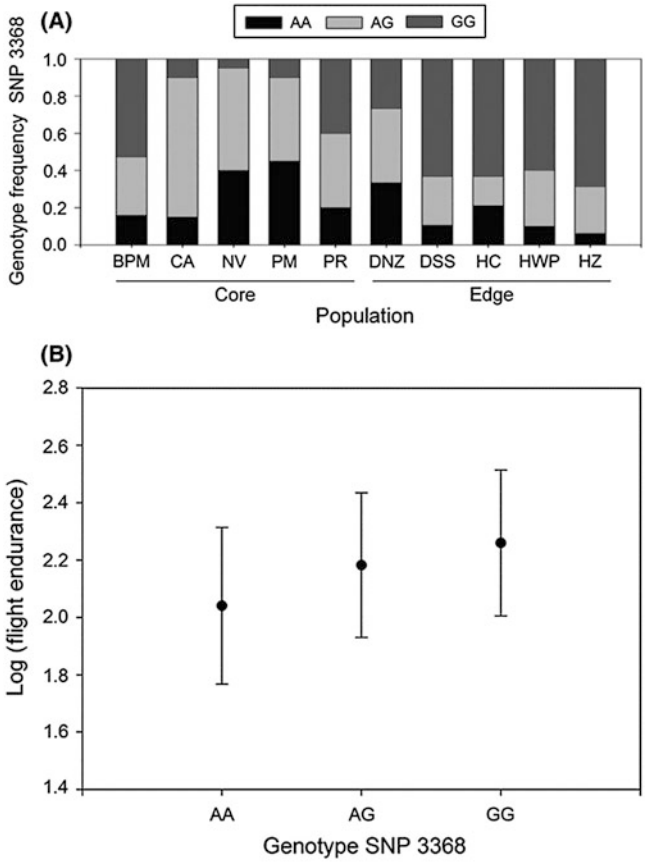


**Fig. 5** (**a**) Genotype frequencies of candidate adaptive SNP 3368 in core and edge damselfly (*Coenagrion scitulum*) populations. (**b**) Flight endurance for the three genotypes of SNP 3368 (log-transformed flight endurance in seconds with 95% confidence intervals). The GG genotype was found at higher frequency in edge (expanding) populations and was associated with the highest mean average flight endurance. Figure from Swaegers et al. (2015)

been used to predict where and when future climates may disrupt patterns of local adaptation (Jay et al. 2012; Fitzpatrick and Keller 2015) and inform assessments of adaptive capacity under future climate change (Razgour et al. 2018; Bay et al. 2018). In a recent study, Creech et al. (2017) used a simulation approach to investigate the spread of adaptive genotypes in desert bighorn sheep, a habitat specialist expected to be threatened by habitat loss and further fragmentation due to climate change. In this novel approach, landscape resistance models were developed for desert bighorn sheep (*Ovis canadensis nelsoni*) in three different regions that varied in habitat quantity and configuration, using data from neutral genetic markers (mostly noninvasively collected). Simulations based on these resistance models were used to investigate how the spread of an adaptive allele varied based on selection strength and whether the adaptive variant was derived from standing genetic variation or a

new mutation. Adaptation from standing genetic variation had a much higher incidence of spread and likelihood of persistence than a novel mutation, especially when landscapes were more highly connected (Fig. 6). These results are in line with

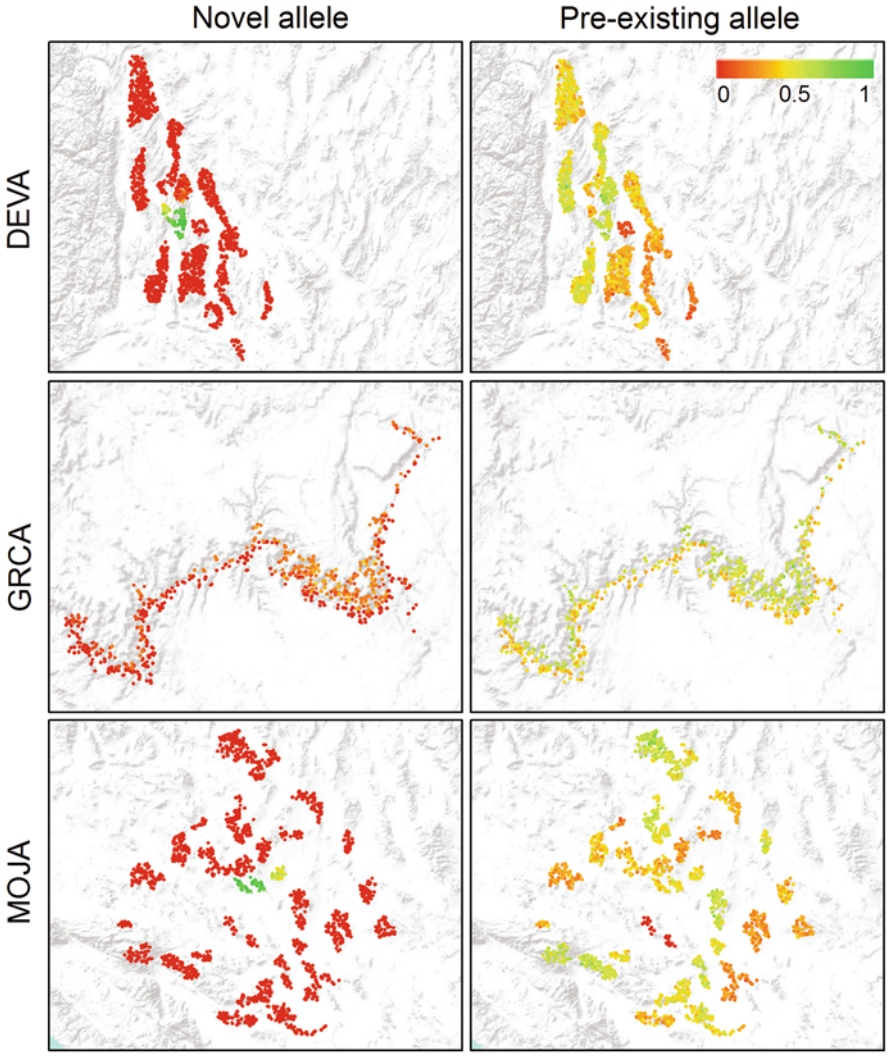


**Fig. 6** Simulated spread of an adaptive allele in populations of desert bighorn sheep (*Ovis canadensis nelsoni*) in different regions of the United States over 100 years. Regions are Death Valley in the northern Mojave Desert (DEVA), the Grand Canyon in northern Arizona (GRCA), and the southern Mojave Desert (MOJA). Colored dots are individual locations with color gradient reflecting the proportion of simulation Monte Carlo replicates in which the adaptive allele is present ($\geq$1 copy) in each individual at year 100, assuming strong selection and a medium dispersal threshold. Left and right columns show presence of the adaptive allele after novel mutation and selection on standing genetic variation, respectively. Figure from Creech et al. (2017)

empirical and conceptual work (reviewed in Hendry 2013) and highlight the importance of maintaining standing genetic variation in desert bighorn sheep populations, as well as the potential need for assisted gene flow (see below) targeting multiple locations in isolated populations.

## 3.2 *Underutilized Applications of Adaptive Landscape Genomics in Wildlife Research*

There are a variety of questions and applications informed by adaptive landscape genomics that have not yet been implemented in wildlife research, but which have significant potential to improve our understanding of ecological and evolutionary processes and management applications. Below we highlight the potential of adaptive landscape genomics for questions related to wildlife management.

### 3.2.1 Using Adaptive Landscape Genomics to Inform Genetic Rescue

The purpose of genetic rescue is to improve the viability of small, isolated, and declining populations by increasing neutral genetic diversity through the movement of individuals between populations (Whiteley et al. 2015). Genetic rescue can be beneficial for populations that are at risk of or are currently experiencing deleterious effects from inbreeding depression, and has been used successfully in a number of high-profile conservation efforts (e.g., Florida panthers (Johnson et al. 2010); wolves (Vilà et al. 2003; Adams et al. 2011); and bighorn sheep (Miller et al. 2012)). The main concern with genetic rescue is outbreeding depression, a reduction in fitness due to mixing divergently adapted genotypes (Edmands 2007); however, recent reviews have highlighted the potentially large benefits and limited risks when genetic rescue is carefully implemented (Hedrick and Fredrickson 2010; Weeks et al. 2011; Frankham 2015; Whiteley et al. 2015). Adaptive landscape genomics can play an important role in minimizing the risks of outbreeding depression by providing an assessment of the environmental and spatial factors that structure adaptive genetic variation across populations. By identifying source populations that minimize adaptive (and also environmental) divergence from the target population, the risk of adaptive incompatibilities and outbreeding depression can be reduced. While assessment of adaptive genetic differentiation is not a requirement for a successful genetic rescue program, it provides additional insight into the characteristics of target and potential source populations that can be used to maximize the effectiveness of this management intervention (Whiteley et al. 2015; Fitzpatrick and Funk 2018). Since genetics studies in wildlife will increasingly use NGS methods to develop molecular markers, incorporating adaptive variation in plans for genetic rescue should become a more common approach.

### 3.2.2 Using Adaptive Landscape Genomics to Inform Assisted Gene Flow

In contrast to genetic rescue, assisted gene flow is a proactive management technique that involves the directional movement of "preadapted" individuals between populations (and within a species range) to facilitate adaptation to changing conditions. It has been advocated for long-lived, sessile species such as trees (e.g., Steane et al. 2014) and species that have a limited ability to track climate conditions that they are currently adapted to (Sgro et al. 2011; Aitken and Whitlock 2013). Unlike genetic rescue, the target population for assisted gene flow should have a large effective population size (to maximize the effectiveness of selection and minimize the impact of genetic drift), and the source and target populations should be divergent based on their adaptive genotypes, where the source population has adaptive variation expected to be advantageous under future conditions in the target population. The concerns about assisted gene flow include the disruption of local adaptation; the loss of distinct, locally adapted lineages; and outbreeding depression. Additionally, because NGS methods that sample the genome do not provide a complete assessment of adaptation, it is possible to reduce fitness with introduced individuals due to maladaptation to an unsampled adaptive parameter, in spite of beneficial adaptation to changing climatic conditions. One option for addressing this issue is to maximize available adaptive variation in source individuals by using a "portfolio effect" or "composite provenancing" that covers a broader range of future climate conditions as well as other, potentially unknown but important selective parameters (Schindler et al. 2010; Sgro et al. 2011; Weeks et al. 2011; Aitken and Whitlock 2013). However, for many populations and species that either lack the capacity for long-distance movement or have no available suitable habitats to disperse into, adaptation in place will be the only alternative to maladaptation, extirpation, and extinction. In these cases, consideration of the potentially far-reaching benefits and careful evaluation to minimize the risks of assisted gene flow can provide an important option for the management of vulnerable populations (Weeks et al. 2011; Aitken and Whitlock 2013).

### 3.2.3 Using Adaptive Landscape Genomics to Inform Site Prioritization to Maximize Evolutionary Potential

Conservation plans are generally focused on protecting the maximum amount of diversity (e.g., the number of different species) in the fewest number of sites and/or at the lowest cost. By selecting sites with complementary sets of species (i.e., sites that are most dissimilar), biodiversity protection can be maximized while minimizing the number of sites in the network (Margules and Pressey 2000). Site prioritization can also be extended to intraspecific diversity to ensure

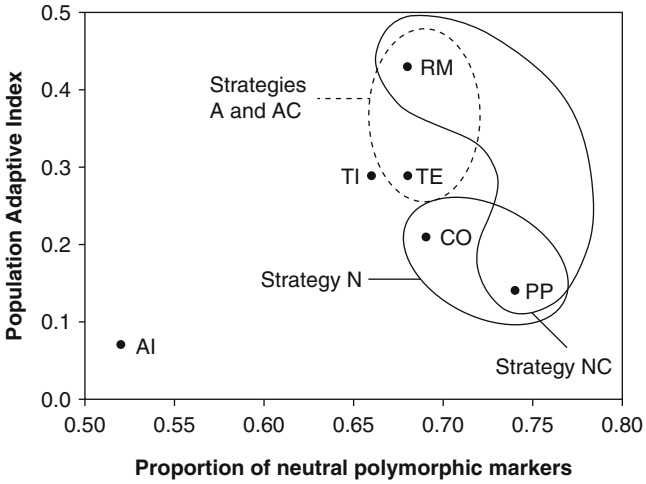


**Fig. 7** Pairs of populations of the common frog (*Rana temporaria*) chosen for conservation based on different strategies: strategy N, protection of the two populations with the two highest neutral diversities; strategy NC, protection of the two populations with the highest pairwise neutral diversity; strategy A, protection of the two populations with the two highest population adaptive indexes; and strategy AC, protection of the two populations with the highest pairwise population adaptive indexes. NC and AC use the principle of complementarity applied to the neutral and adaptive data, respectively, to maximize the breadth of conserved genetic variation. Figure from Bonin et al. (2007)

sufficient protection of genetic diversity within species conservation plans (Bonin et al. 2007) (Fig. 7). When neutral genetic data are available, the goal is most often to prioritize populations that maximize the within and between group variability of the species (e.g., Ottewell et al. 2016). However, with increasing pressures on species to adapt to rapidly changing environmental conditions, conserving the maximum amount of adaptive genetic diversity is also essential to ensure the evolutionary potential of threatened species (Nicotra et al. 2015). The inclusion of adaptive genetic variation into the site prioritization framework was first advocated a decade ago by Bonin et al. (2007). However, few empirical examples have been published in academic journals, which likely reflects the lack of genomic data for species of conservation concern, though some studies may have been published in the gray literature due to the applied nature of the work (Garner et al. 2016). The decreasing costs of genomics should ensure a revival of this framework for informing conservation plans.

### 3.2.4 Using Museum Collections to Better Understand Changes in Adaptive Variation over Time

Museum collections represent a rich source of historical genetic variation that can be invaluable in understanding the evolutionary consequences of recent environmental

change in wildlife species (Holmes et al. 2016). For example, Miller and Waits (2003) amplified eight neutral microsatellite markers in 110 museum samples of Yellowstone grizzly bear to assess the impact of past anthropogenic isolation and culling on future genetic viability of this population. The transition from these genetic-scale museum studies to genomic-scale data that can be used to investigate selection and adaptation has been hampered by the challenges of working with highly degraded DNA. Fortunately, several recent approaches have been used to develop genomic-scale marker sets, including adaptive markers, from museum samples (Andrews et al. 2018), including targeted sequencing of immune response loci in the Pale-headed Brushfinch (Hartmann et al. 2014), exon capture in alpine chipmunks (Bi et al. 2013), and whole-genome sequencing in honey bees (Mikheyev et al. 2015). These methods currently require more genomic information than can be provided by a reduced representation NGS study, but increasing amounts of genomic data (including reference genomes) for non-model species and falling sequencing costs indicate that historical data may soon be an option for many wildlife species. Techniques are even being developed for formalin-fixed samples (characteristic of, e.g., amphibian, reptile, and fish specimens), though these approaches are currently limited to neutral markers developed for phylogenomic studies (Hykin et al. 2014; Ruane and Austin 2017).

### 3.2.5 Using Adaptive Landscape Genomics to Inform the Management of Hybridization

Hybridization is a fundamental management problem addressed in conservation genetics (Bohling 2016), and genomic methods are being increasingly used to identify and manage hybridization at high resolution in a diversity of wildlife species (e.g., Fitzpatrick et al. 2009; vonHoldt et al. 2011; Hohenlohe et al. 2013; Kovach et al. 2016; Wayne and Shaffer 2016). However, natural and human-mediated hybridization may also be an effective tool for improving the adaptive capacity of threatened species in response to rapid anthropogenic change, such as climate change (Hoffmann et al. 2015; Hamilton and Miller 2016). To our knowledge there are no examples of managed hybridization in wildlife populations with the specific goal of facilitating adaptive introgression, though experimental (e.g., salt tolerance in yeast, Stelkens et al. 2014) and accidental (e.g., insecticide resistance in mosquitos, Norris et al. 2015) cases demonstrate the power of hybridization to improve adaptive capacity in response to strong selective pressures. As with assisted gene flow, adaptive landscape genomics could play an integral role in determining appropriate source populations for managed adaptive introgression to facilitate evolutionary resilience in the face of rapid environmental change.

## 4 Future Research Avenues in Wildlife Landscape Genomics: Improving and Moving Beyond Genotype-Environment Associations

Adaptive landscape genomics has led to valuable insights in wildlife studies and can be used to address important ecological and evolutionary questions and management issues. Future studies will be able to choose from an even larger number of statistical methods for conducting adaptive landscape genomics, and this choice will hopefully be guided by emerging recommendations regarding their relative suitability for addressing specific research questions (Rellstab et al. 2015; Balkenhol et al. 2017). Here, we have largely focused on one type of analytical approach used in adaptive landscape genomics, the analysis of genotype-environment associations. As discussed above, GEA is a main component of most current landscape genomic studies aiming to find relationships between selection-driven loci and environmental heterogeneity. Nevertheless, we emphasize that using additional analytical approaches and incorporating complementary data in wildlife research can improve our understanding of adaptation (Table 1, Figs. 1 and 8). Some of these are already in use in wildlife studies, such as replication of GEA sampling across gradients (Sect. 2.1), the incorporation of a well-annotated reference genome (Sect. 2.2), and

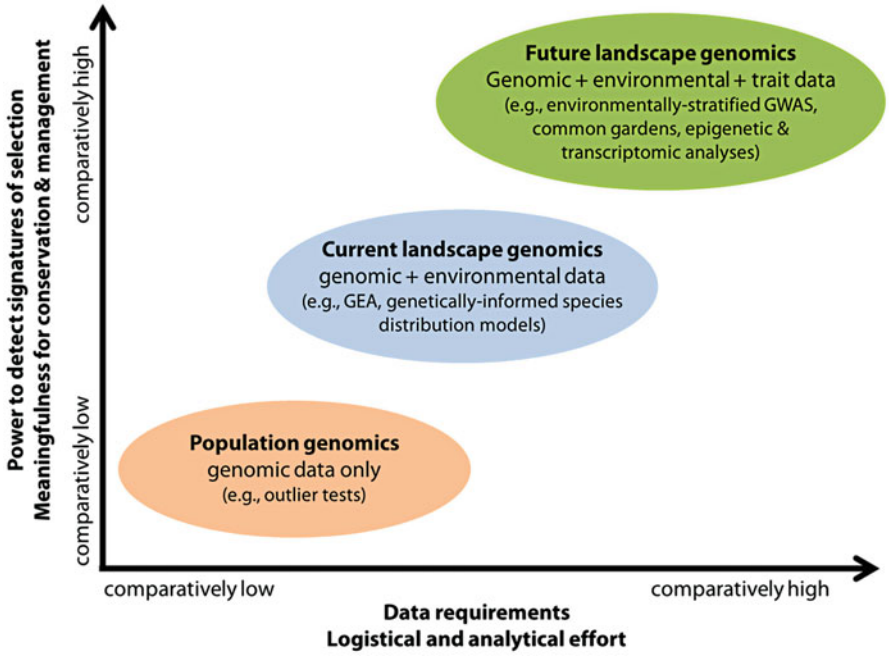


**Fig. 8** Analytical approaches to understanding local adaptation (ellipses). Their relative data and analytical requirements (*x*-axis) and power and utility for conservation and management (*y*-axis) range from comparatively low to comparatively high

the use of simulations (Sect. 3.1.6). Simulations in particular are underutilized in landscape genomics research, and recent applications illustrate their value in corroborating empirical findings (e.g., Cooke et al. 2014) and in developing new theories (reviewed in Landguth et al. 2015) (Table 1). Additionally, as sequencing costs fall, whole-genome resequencing is becoming an option for some wildlife species (e.g., Kardos et al. 2015; Toews et al. 2016; Therkildsen and Palumbi 2017), providing increased marker density when compared to reduced representation methods, in addition to identifying other genetic variation such as structural variants (Fuentes-Pardo and Ruzzante 2017) (Table 1). However, the limited accessibility and affordability of this approach over reduced representation methods makes whole-genome resequencing less feasible for conservation and management applications in wildlife (Fuentes-Pardo and Ruzzante 2017).

## *4.1 Integrating Phenotypic Data Through Environmentally Stratified GWAS and GEA*

One of the major advantages of the GEA-based approach to adaptive landscape genomics is that no phenotypic data are required. Nevertheless, natural selection acts on the phenotype, not the genotype. Thus, our understanding of the processes that shape patterns of adaptive genetic variation in heterogeneous environments will likely not be complete without considering phenotypic variability. Because of this, particularly interesting complements to GEA are analytical approaches that account for phenotypic variability when assessing landscape influences on genomic variation. There are a few examples of adaptive landscape genomics studies in wildlife that have used a post hoc approach to incorporating phenotypes, by correlating trait data with candidate loci identified through GEA analysis (Swaegers et al. 2015; Funk et al. 2016).

A more promising approach is to combine detections from GEA analysis with those from genome-wide association studies (GWAS) that are conducted for many individuals across environmental gradients. Whereas GEA methods look for relationships between genotypes and environment, GWAS uses statistical approaches to test for relationships between phenotypes and genotypes. For example, Lasky et al. (2015) used GEA associations with climate to predict GWAS-derived phenotypic variation in adaptive traits in the important food crop *Sorghum bicolor*. In another case, Berg and Coop (2014) combined GWAS with quantitative genetics and GEA to detect signals of local adaptation in several human traits. Unless a reference genome is available, the genes detected by GWAS are anonymous, i.e., their exact locations on the genome and their functions are usually unknown. However, one could argue that for the fate of individuals and populations, and for wildlife conservation management, it is not usually crucial to identify the location and function of adaptive genes, but rather to understand their effects on fitness. In either case, linking the phenotype-genotype results of environmentally stratified GWAS

with genotype-environment-derived GEA detections can improve the strength of inference in adaptive landscape genomic studies, since independently detected and overlapping loci are more likely to reflect true adaptive processes.

GWAS studies in wildlife species are currently somewhat limited, and (similar to GWAS in other groups) their success in identifying genetic variation underlying fitness-relevant phenotypes is mixed (e.g., Johnston et al. 2011; Santure et al. 2013; Wenzel et al. 2016). However, in wildlife species where data on fitness-relevant phenotypic traits can be collected, the integration of environmentally stratified GWAS and GEA will provide greatly improved inference for adaptation that can inform ecological, evolutionary, and management questions.

## *4.2 The Value of Experimental Manipulations in Informing Assessments of Adaptation*

Experimental manipulations provide the most direct evidence for the genetic basis of a fitness-related trait and/or local adaptation (Savolainen et al. 2013; de Villemereuil et al. 2016). Common gardens are used to rear individuals from different habitats/environmental conditions under common controlled or field conditions. While common gardens are designed to study the genetic basis of traits while controlling for phenotypic plasticity, they can be confounded by genotype-by-environment interactions (Kawecki and Ebert 2004; Merilä and Hendry 2014), though replication across environments can alleviate this problem (de Villemereuil et al. 2016). Reciprocal transplants are a type of common garden where individuals from different environments are reared in both their native and non-native (introduced) environment. Reciprocal transplants measure the contribution of both genetic and environmental variation to fitness and can be used to identify local adaptation. The most robust inference for these methods comes from rearing multiple generations under common conditions to reduce maternal effects (Kawecki and Ebert 2004), and using individuals of known pedigree to facilitate quantitative genetic study (though estimates of relatedness can be made using molecular data).

These stringent conditions will make experimental manipulations inaccessible for most wildlife species due to challenges associated with rearing in controlled conditions, limitations associated with listed species status, and other logistical complications. Experimental studies do exist for wildlife species, including amphibians (e.g., Berven 1982; Bernardo 1994) and salmonids (Fraser et al. 2011; Christie et al. 2016), and it is these studies that provide the highest inferential strength for understanding the functionality of genomic and phenotypic variation in an evolutionary context.

When experimental methods are possible, there are multiple alternatives for integrating these data with GEA results. For example, environmentally stratified common gardens could be used as a follow-up to GEA to compare the fitness (or related trait) of individuals from environmentally divergent habitats who carry

or lack a candidate adaptive allele identified by GEA (Holderegger et al. 2008). Replicating individuals within habitats that carry/do not carry the allele of interest is important to control for divergent genetic backgrounds across populations, so this approach would be most useful for candidate loci that have not diverged to fixation. Another approach is to assess the relationship between candidate SNPs identified with GEA and traits measured in a common garden (De Kort et al. 2014). Recently, Lasky et al. (2018) developed a synthetic approach that integrates data from multiple common gardens stratified across environments with GEA in a genome-wide association model of genotype-by-environment interactions. This novel approach uses imputed fitness values for GEA data to coherently synthesize evidence from common gardens and GEA tests, increased power to detect signatures of local adaptation.

## *4.3 Epigenetics as a Mechanism for Rapid Adaptive Responses*

Epigenetic modifications (i.e., phenotypic changes that are mediated by the regulation of gene expression, rather than alterations in the DNA sequence) are influenced by genome-environment interactions and can therefore shape patterns of adaptive genomic variation in heterogeneous environments (Verhoeven et al. 2016; Whipple and Holeski 2016). Epigenetic variation may also be a mechanism allowing rapid adaptation to changing environmental conditions (via plasticity), even in the face of small population sizes and low genetic diversity (Massicotte et al. 2011; Bernatchez 2016). However, our understanding of epigenetic processes in natural settings is currently quite limited, and most population and quantitative genetic theory does not include epigenetic effects. Method development in the field is proceeding quickly, with newer approaches based on reduced representation sequencing possible in species without a reference genome (e.g., Trucchi et al. 2016; van Gurp et al. 2016). These advances should make studies of DNA methylation (the most widely studied epigenetic mechanism) more accessible in wildlife species, providing insight into the role of environmentally induced epigenetic modifications in plastic responses to environmental change. For example, a recent study identified a role for epigenetic modifications in plastic responses of three reef-building corals to ocean acidification and thermal stress, demonstrating a previously unknown adaptive response of these species to climate change (Dimond and Roberts 2016). However, experimental designs such as multigenerational common gardens, which are not feasible for most wildlife species, are ultimately needed to establish transgenerational inheritance of epigenetic modifications (Whipple and Holeski 2016). This will constrain our understanding of the role of epigenetics in enhancing the evolutionary potential of wildlife species in a management context.

### 4.4 The Importance of Differential Gene Expression Across Landscapes

Detailed landscape genomic inference can be derived from approaches that analyze the functionality and expression of genes, such as transcriptome analyses (i.e., RNA-seq), across landscapes (Storfer et al. 2015). Because these approaches require experimental work and high-quality gene annotation to establish robust relationships between functional genomic variation and gene expression, they have been most frequently applied to model species. While de novo implementations of transcriptomics are available for non-model species (e.g., Haas et al. 2013), the quality of inference from these studies is limited (Alvarez et al. 2015; Todd et al. 2016). The most robust experimental design for transcriptomic studies involves controlled, hypothesis-driven, experimental treatments to identify the processes underlying differential gene expression across relevant tissue types. This imposes limitations for field-based studies of many wildlife species, though progress is being made in investigating alternative, less destructive tissue sampling approaches (Czypionka et al. 2015). While detecting differences in gene expression under field conditions in wild populations is possible, inference is often confounded due to the sensitivity of gene expression to environmental conditions. Field-based studies therefore require very careful design and execution, (relatively expensive) biological replication, and careful documentation of potentially confounding biological, environmental, and laboratory effects (Todd et al. 2016). Even in these cases, field-based RNA-seq studies are mostly confined to generating hypotheses for future research (Todd et al. 2016). Despite these constraints, there are a few examples of common garden approaches to documenting differential gene expression in wild species in a climate change context (e.g., Barshis et al. 2013; O'Neil et al. 2014; Narum and Campbell 2015; Thomas and Palumbi 2017), which illustrate the potential for transcriptomic studies to inform adaptive capacity in response to climate change (Hoffmann et al. 2015). Additionally, when there are existing transcriptomic resources available, they can be used to help validate and suggest functional relevance of loci detected with GEA (Szulkin et al. 2016).

Finally, a related sequencing method, whole-exome sequencing (WES), uses transcriptomic resources to develop probes targeting exons (protein coding genes) as well as functional but noncoding regions (Warr et al. 2015). WES, not to be confused with exon capture (which targets subsets of the exome, usually based on an annotated genome, e.g., Roffler et al. 2016), is currently restricted to humans and other model species, as well as a few domesticated species and crops. WES is unlikely to be applied in non-model wildlife species in the near future, though exome resources developed for model and domesticated species could be leveraged in studies of closely related species.

## 5 Conclusions

Overall, GEA has already led to several exciting findings in wildlife landscape genomics, and the rapid development of new and improved GEA methods and software for their implementation will likely ensure their continuing role as a workhorse for adaptive landscape genomics. However, we need to be careful not to equate "landscape genomics" with "genotype-environment associations." The first is a scientific field, while the latter is a mere set of analytical tools. Several authors have already argued that landscape genomics focusing on neutral processes (i.e., "landscape genetics") is not a distinct scientific field, but rather a collection of new and more powerful methods to test old ideas that mainly stem from other disciplines (e.g., Dyer 2015; Rissler 2016). Even though some of the important "classic" population genetic concepts and theories are certainly still valid for landscape genomics, we agree that the theoretical and conceptual development in both neutral and adaptive landscape genomics lags far behind the methodological progress we have made.

As highlighted by Bernatchez (2016), theories of adaptive capacity and evolutionary potential in nature have generally not been able to keep up with the fast developments for gathering and analyzing large amounts of genomic data. Specifically, most of the theories and models used by geneticists and evolutionary biologists do not account for spatiotemporal impacts of environmental heterogeneity in a realistic manner and also seldom include the complex interrelationships of processes impacting genomic variation (e.g., polygenic and balancing selection, genomic architecture, epigenetics). Because of this, predictions from classical theories and models often do not match the genomic patterns that we see in nature. Hence, it is crucial to compare findings from empirical genomic studies to predictions derived from existing theory and simulation studies. Contrasting expected genomic patterns with those observed in field and experimental settings can shed light on the missing pieces in our understanding of adaptive processes in heterogeneous and changing environments.

This will also be vital for developing new theories and hypotheses that refine our thinking about the links between environmental complexity, genomic variation, and evolutionary processes. Such a theory-focused approach is not only important for the future development of landscape genomics (Balkenhol et al. 2015; Dyer 2015) but is also necessary to better inform conservation managers and policy-makers about the most important challenges to expect under changing environmental conditions. Our predictions of future evolutionary trajectories of populations and species require a much better theoretical and conceptual understanding of how adaptive capacity and evolutionary potential vary across groups with different life histories and ecological niches. As stated before, future studies should evaluate whether and when such detailed information is indeed needed to successfully manage wildlife populations.

Thus, the future of landscape genomics will hopefully move beyond the statistical detection of associations between environmental and genetic data and mature toward a field with a solid theoretical and conceptual foundation. Such progress has to go

hand-in-hand with more holistic research approaches that combine (quasi-) experimental study designs with simulations and empirical analyses that use the full range of available tools for assessing environmental impacts on selection-driven genetic and phenotypic variation and underlying processes. We particularly advocate the integration of GWAS across environmental gradients, because understanding the impacts of environmental heterogeneity on genomic variation that is relevant for phenotypic variation can provide us with information more closely related to fitness and population dynamics. Additionally, GWAS studies, though potentially labor-intensive for the collection of phenotypic trait data, are more feasible in wildlife species compared to other options for improving inference of local adaptation (Table 1). Finally, we need to begin to move beyond the single-species studies that are typical of adaptive landscape genomics to date. Considering multiple species or entire communities in landscape genomics (i.e., community landscape genomics, Hand et al. 2015) is clearly challenging but necessary because neutral and adaptive genomic patterns are not only shaped by the physical characteristics of an area ("landscape" in a strict sense) but also by the interactions among species. Landscape community genomics could be facilitated by eDNA (environmental DNA) approaches that rely on samples of, e.g., water or soil that contain genetic material from wildlife that have been in contact with these environmental samples (Ficetola et al. 2008). While current eDNA approaches do not yet allow population and landscape genomic studies, future technological and analytical developments will likely enable such applications (Bohmann et al. 2014).

Considering all of the abovementioned research approaches in adaptive landscape genomics of wildlife will ultimately help us to not only quantify and predict genetic patterns in changing environments but also to understand the function of these patterns and their relevance for individual fitness, population dynamics, and species persistence.

**Acknowledgments** Thanks to Kim Andrews and Paul Hohenlohe for helpful comments that improved the chapter. This work was supported in part by funds provided by National Science Foundation grants EF-1442597 and DEB-1340852 to ELL and BRF, NASA grant NNX14AC91G to ELL, and National Science Foundation grant DEB-1639014 and NASA grant NNX14AB84G to BKH.

## References

Adams JR, Vucetich LM, Hedrick PW, Peterson RO, Vucetich JA. Genomic sweep and potential genetic rescue during limiting environmental conditions in an isolated wolf population. Proc R Soc B Biol Sci. 2011;278:3336–44.

Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, Dadaniec RY. The search for loci under selection: trends, biases and progress. Mol Ecol. 2018;27:1342–56.

Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. Annu Rev Ecol Evol Syst. 2013;44:367–88.

Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics. 2016;202:389–400.

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N. Genetic effects of harvest on wild animal populations. Trends Ecol Evol. 2008;23:327–37.
Alvarez M, Schrey AW, Richards CL. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol. 2015;24:710–25.
Amish SJ, Hohenlohe PA, Painter S, Leary RF, Muhlfeld C, Allendorf FW, Luikart G. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. Mol Ecol Resour. 2012;12:653–60.
Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.
Andrews KR, DeBarba M, Russello MA, Waits LP. Techniques for using non-invasive, archival, and environmental samples in population genomic studies. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer; 2018.
Aykanat T, Lindqvist M, Pritchard VL, Primmer CR. From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon *Salmo salar*. J Fish Biol. 2016;89:2658–79.
Balkenhol N, Landguth EL. Simulation modelling in landscape genetics: on the need to go further. Mol Ecol. 2011;20:667–70.
Balkenhol N, Cushman SA, Waits LP, Storfer A. Current status, future opportunities, and remaining challenges in landscape genetics. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics. Hoboken: Wiley; 2015.
Balkenhol N, Dudaniec RY, Krutovsky KV, et al. Landscape genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP, editor. Population genomics concepts, strategies and approaches. Cham: Springer; 2017.
Barshis DJ, Ladner JT, Oliver TA, Seneca FO, Traylor-Knowles N, Palumbi SR. Genomic basis for coral resilience to climate change. Proc Natl Acad Sci U S A. 2013;110:1387–92.
Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science. 2018;359:83–6.
Benestan LM, Ferchaud A-L, Hohenlohe PA, Garner BA, Naylor GJ, Baums IB, Schwartz MK, Kelley JL, Luikart G. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25:2967–77.
Benjamini Y, Hochberg Y. Controlling the false discovery rate – a practical and powerful approach. J R Stat Soc B Methodol. 1995;57:289–300.
Bensch S, Åkesson M. Ten years of AFLP in ecology and evolution: why so few animals? Mol Ecol. 2005;14:2899–914.
Berg JJ, Coop G. A population genetic signal of polygenic adaptation. PLoS Genet. 2014;10:e1004412.
Bernardo J. Experimental analysis of allocation in two divergent, natural salamander populations. Am Nat. 1994;143:14–38.
Bernatchez L. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. J Fish Biol. 2016;89:2519–56.
Berven KA. The genetic basis of altitudinal variation in the wood frog *Rana sylvatica*. I. An experimental analysis of life history traits. Evolution. 1982;36:962–83.
Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. Unlocking the vault: next-generation museum population genomics. Mol Ecol. 2013;22:6018–32.
Bohling JH. Strategies to address the conservation threats posed by hybridization and genetic introgression. Biol Conserv. 2016;203:321–7.
Bohmann K, Evans A, Gilbert MTP, et al. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol Evol. 2014;29:358–67.
Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. Conserv Biol. 2007;21:697–708.

Brauer CJ, Hammer MP, Beheregaray LB. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Mol Ecol. 2016;25:5093–113.
Buckley LB, Huey RB. Temperature extremes: geographic patterns, recent changes, and implications for organismal vulnerabilities. Glob Chang Biol. 2016;22:3829–42.
Buehler D, Holderegger R, Brodbeck S, Schnyder E, Gugerli F. Validation of outlier loci through replication in independent data sets: a test on *Arabis alpina*. Ecol Evol. 2014;4:4296–306.
Campbell NR, Harmon SA, Narum SR. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Resour. 2015;15:855–67.
Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22:3124–40.
Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362–5.
Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD. Rapid range shifts of species associated with high levels of climate warming. Science. 2011;333:1024–6.
Christie MR, Marine ML, Fox SE, French RA, Blouin MS. A single generation of domestication heritably alters the expression of hundreds of genes. Nat Commun. 2016;7:10676.
Cooke GM, Landguth EL, Beheregaray LB. Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. Evolution. 2014;68:1947–60.
Coop G, Witonsky D, Rienzo AD, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185:1411–23.
Cosart T, Beja-Pereira A, Chen S, Ng SB, Shendure J, Luikart G. Exome-wide DNA capture and next generation sequencing in domestic and wild species. BMC Genomics. 2011;12:347.
Creech TG, Epps CW, Landguth EL, Wehausen JD, Crowhurst RS, Holton B, Monello RJ. Simulating the spread of selection-driven genotypes using landscape resistance models for desert bighorn sheep. PLoS One. 2017;12:e0176960.
Cushman SA, McKelvey KS, Hayden J, Schwartz MK. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Nat. 2006;168:486–99.
Czypionka T, Krugman T, Altmüller J, Blaustein L, Steinfartz S, Templeton AR, Nolte AW. Ecological transcriptomics – a non-lethal sampling approach for endangered fire salamanders. Methods Ecol Evol. 2015;6:1417–25.
Daly C. Guidelines for assessing the suitability of spatial climate data sets. Int J Climatol. 2006;26:707–21.
De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus glutinosa. Mol Ecol. 2014;23:4709–21.
De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol Ecol. 2013;22:1383–99.
de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. Genome scan methods against more complex models: when and how much should we trust them? Mol Ecol. 2014;23:2006–19.
de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I. Common garden experiments in the genomic era: new perspectives and opportunities. Heredity. 2016;116:249–54.
Dimond JL, Roberts SB. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Mol Ecol. 2016;25:1895–904.
Dobrowski SZ, Abatzoglou J, Swanson AK, Greenberg JA, Mynsberge AR, Holden ZA, Schwartz MK. The climate velocity of the contiguous United States during the 20th century. Glob Chang Biol. 2013;19:241–51.
Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 2013;36:27–46.
Dyer RJ. Is there such a thing as landscape genetics? Mol Ecol. 2015;24:3518–28.

Early R, Sax DF. Analysis of climate paths reveals potential limitations on species range shifts. Ecol Lett. 2011;14:1125–33.
Edmands S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol. 2007;16:463–75.
Epstein B, Jones M, Hamede R, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Commun. 2016;7:12684.
Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. Annu Rev Ecol Evol Syst. 2009;40:481–501.
Ficetola GF, Miaud C, Pompanon F, Taberlet P. Species detection using environmental DNA from water samples. Biol Lett. 2008;4:423–5.
Fitzpatrick SW, Funk WC. Population genomics for genetic rescue. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer; 2018.
Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol Lett. 2015;18:1–16.
Fitzpatrick BM, Johnson JR, Kump DK, Shaffer HB, Smith JJ, Voss SR. Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. BMC Evol Biol. 2009;9:176.
Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban SM. Guidelines for using genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol Appl. 2018;11:1035–52.
Forester BR, DeChaine EG, Bunn AG. Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. Divers Distrib. 2013;19:1480–95.
Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. Mol Ecol. 2018;27:2215–33.
François O, Martins H, Caye K, Schoville SD. Controlling false discoveries in genome scans for selection. Mol Ecol. 2016;25:454–69.
Frankham R. Conservation genetics. Annu Rev Genet. 1995;29:305–27.
Frankham R. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Mol Ecol. 2015;24:2610–8.
Frankham R, Ballou JD, Ralls K, et al. Genetic management of fragmented animal and plant populations. Oxford: Oxford University Press; 2017.
Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity. 2011;106:404–20.
Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30:1687–99.
Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations, and practical recommendations. Mol Ecol. 2017;26:5369–406.
Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.
Funk WC, Lovich RE, Hohenlohe PA, et al. Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol Ecol. 2016;25:2176–94.
Funk WC, Forester BR, Converse SJ, Darst C, Morey S. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet. 2018; https://doi.org/10.1007/s10592-018-1096-1.
Gardner JL, Amano T, Sutherland WJ, Clayton M, Peters A. Individual and demographic consequences of reduced body condition following repeated exposure to high temperatures. Ecology. 2016;97:786–95.
Garner BA, Hand BK, Amish SJ, et al. Genomics in conservation: case studies and bridging the gap between data and application. Trends Ecol Evol. 2016;31:81–3.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17:333–51.
Grant BR, Grant PR. Evolution of Darwin's finches caused by a rare climatic event. Proc R Soc B Biol. 1993;251:111–7.
Guisan A, Tingley R, Baumgartner JB, et al. Predicting species distributions for conservation decisions. Ecol Lett. 2013;16:1424–35.
Guo B, Li Z, Merilä J. Population genomic evidence for adaptive differentiation in the Baltic Sea herring. Mol Ecol. 2016;25:2833–52.
Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016;30:33–41.
Hancock AM, Brachi B, Faure N, et al. Adaptation to climate across the Arabidopsis thaliana genome. Science. 2011;334:83–6.
Hand BK, Lowe WH, Kovach RP, Muhlfeld CC, Luikart G. Landscape community genomics: understanding eco-evolutionary processes in complex environments. Trends Ecol Evol. 2015;30:161–8.
Hand BK, Muhlfeld CC, Wade AA, et al. Climate variables explain neutral and adaptive variation within salmonid metapopulations: the importance of replication in landscape genetics. Mol Ecol. 2016;25:689–705.
Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. Using genomics to characterize evolutionary potential for conservation of wild populations. Evol Appl. 2014;7:1008–25.
Hartmann SA, Schaefer HM, Segelbacher G. Genetic depletion at adaptive but not neutral loci in an endangered bird species. Mol Ecol. 2014;23:5712–25.
Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst Biol. 2016;65:910–24.
Hazen EL, Jorgensen S, Rykaczewski RR, et al. Predicted habitat shifts of Pacific top predators in a changing climate. Nat Clim Chang. 2013;3:234–8.
Hedrick PW, Fredrickson R. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. Conserv Genet. 2010;11:615–26.
Hendry AP. Key questions in the genetics and genomics of eco-evolutionary dynamics. Heredity. 2013;111:456–66.
Hess JE, Campbell NR, Close DA, Docker MF, Narum SR. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Mol Ecol. 2013;22:2898–916.
Hess JE, Campbell NR, Docker MF, et al. Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. Mol Ecol Resour. 2015;15:187–202.
Hickling R, Roy DB, Hill JK, Fox R, Thomas CD. The distributions of a wide range of taxonomic groups are expanding polewards. Glob Chang Biol. 2006;12:450–5.
Hoban S, Gaggiotti O, Bertorelle G. Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): a software for choosing the appropriate number of markers and samples. Methods Ecol Evol. 2013;4:299–303.
Hoban S, Arntzen JA, Bruford MW, et al. Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. Evol Appl. 2014;7:984–98.
Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, Glenn TC. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour. 2016;16:1264–78.
Hoffmann A, Griffin P, Dillon S, et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management. Clim Chang Res. 2015;2:1–23.
Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010;6:e1000862.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour. 2011;11:117–22.
Hohenlohe PA, Day MD, Amish SJ, et al. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol Ecol. 2013;22:3002–13.
Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population genomics provides key insights in ecology and evolution. In: Rajora OP, editor. Population genomics concepts, approaches and applications. Cham: Springer; 2017.
Holden ZA, Jolly W, Parson R, Warren A, Landguth EL, Abatzoglou J. TOPOFIRE: a system for monitoring insect and climate induced impacts on fire danger in complex terrain. Consortium for Integrated Climate Research in Western Mountains. 2013. http://topofire.dbs.umt.edu/topofire_v2/topofire_portal_v2.php.
Holderegger R, Herrmann D, Poncet B, et al. Land ahead: using genome scans to identify molecular markers of adaptive relevance. Plant Ecol Divers. 2008;1:273–83.
Holmes MW, Hammond TT, Wogan GOU, et al. Natural history collections as windows on evolutionary processes. Mol Ecol. 2016;25:864–81.
Houston RD, Taggart JB, Cézard T, et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics. 2014;15:90.
Hykin SM, Bi K, McGuire JA. Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. PLoS One. 2014;10:e0141579.
Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG. Genetically informed ecological niche models improve climate change predictions. Glob Chang Biol. 2017;23:164–76.
Jay F, Manel S, Alvarez N, et al. Forecasting changes in population genetic structure of alpine plants in response to global warming. Mol Ecol. 2012;21:2354–68.
Johnson WE, Onorato DP, Roelke ME, et al. Genetic restoration of the Florida panther. Science. 2010;329:1641–5.
Johnston SE, McEwan JC, Pickering NK, et al. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol Ecol. 2011;20:2555–66.
Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25:185–202.
Jones MR, Forester BR, Teufel AI, et al. Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. Evolution. 2013;67:3455–68.
Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, Taberlet P. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol. 2007;16:3955–69.
Kardos M, Luikart G, Bunch R, et al. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol Ecol. 2015;24:5616–32.
Karger DN, Conrad O, Böhner J, et al. Climatologies at high resolution for the earth's land surface areas. Sci Data. 2017;4:170122.
Kawecki TJ, Ebert D. Conceptual issues in local adaptation. Ecol Lett. 2004;7:1225–41.
Kovach RP, Hand BK, Hohenlohe PA, et al. Vive la résistance: genome-wide selection against introduced alleles in invasive hybrid zones. Proc R Soc B. 2016;283:20161380.
Lah L, Trense D, Benke H, et al. Spatially explicit analysis of genome-wide SNPs detects subtle population structure in a mobile marine mammal, the harbor porpoise. PLoS One. 2016;11:e0162792.
Lande R. Genetics and demography in biological conservation. Science. 1988;241:1455–60.
Landguth EL, Schwartz MK. Evaluating sample allocation and effort in detecting population differentiation for discrete and continuously distributed individuals. Conserv Genet. 2014;15:981–92.

Landguth EL, Fedy BC, Oyler-McCance SJ, et al. Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Mol Ecol Resour. 2012;12:276–84.
Landguth E, Cushman SA, Balkenhol N. Simulation modeling in landscape genetics. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics. Hoboken: Wiley; 2015. p. 99–113.
Lasky JR, Marais D, L D, et al. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. Mol Biol Evol. 2014;31:2283–96.
Lasky JR, Upadhyaya HD, Ramu P, et al. Genome-environment associations in sorghum landraces predict adaptive traits. Sci Adv. 2015;1:e1400218.
Lasky JR, Forester BR, Reimherr M. Coherent synthesis of genomic associations with phenotypes and home environments. Mol Ecol Resour. 2018;18:91–106.
Le Corre V, Kremer A. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 2012;21:1548–66.
Lea A, Altmann J, Alberts SC, Tung J. Resource base influences genome-wide DNA methylation levels in wild baboons (Papio cynocephalus). Mol Ecol. 2016;25:1681–96.
Leempoel K, Duruz S, Rochat E, Widmer I, Orozco-terWengel P, Joost S. Simple rules for an efficient use of geographic information systems in molecular ecology. Front Ecol Evol. 2017;5:1–33.
Lindenmayer DB, Piggott MP, Wintle BA. Counting the books while the library burns: why conservation monitoring programs need a plan for action. Front Ecol Environ. 2013;11:549–55.
Lookingbill T, Urban D. Gradient analysis, the next generation: towards more plant-relevant explanatory variables. Can J For Res. 2005;35:1744–53.
Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015;24:1031–46.
Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2016;17:142–52.
Lozier JD. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. Mol Ecol. 2014;23:788–801.
Luikart G, England P, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94.
Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, Aurelle D. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. Mol Ecol. 2016;25:170–84.
Manthey JD, Moyle RG. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. Mol Ecol. 2015;24:3628–38.
Margules CR, Pressey RL. Systematic conservation planning. Nature. 2000;405:243–53.
Massicotte R, Whitelaw E, Angers B. DNA methylation: a source of random variation in natural populations. Epigenetics. 2011;6:421–7.
McCartney-Melstad E, Mount GG, Shaffer HB. Exon capture optimization in amphibians with large genomes. Mol Ecol Resour. 2016;16:1084–94.
McKinney GJ, Larson WA, Seeb LW, Seeb JE. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on breaking RAD by Lowry et al., (2016). Mol Ecol Resour. 2017;17:356–61.
McMahon BJ, Teeling EC, Höglund J. How and why should we implement genomics into conservation? Evol Appl. 2014;7:999–1007.
Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evol Appl. 2014;7:1–14.
Mikheyev AS, Tin MMY, Arora J, Seeley TD. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. Nat Commun. 2015;6:7991.

Miller CR, Waits LP. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implications for conservation. Proc Natl Acad Sci U S A. 2003;100:4334–9.

Miller JM, Poissant J, Hogg JT, Coltman DW. Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). Mol Ecol. 2012;21:1583–96.

Moura AE, Kenny JG, Chaudhuri R, et al. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. Mol Ecol. 2014;23:5179–92.

Nadeau CP, Urban MC, Bridle JR. Coarse climate change projections for species living in a fine-scaled world. Glob Chang Biol. 2017;23:12–24.

Narum SR, Campbell NR. Transcriptomic response to heat stress among ecologically divergent populations of redband trout. BMC Genomics. 2015;16:103.

Nicotra AB, Beever EA, Robertson AL, Hofmann GE, O'Leary J. Assessing the components of adaptive capacity to improve conservation and management efforts under global change. Conserv Biol. 2015;29:1268–78.

Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc Natl Acad Sci U S A. 2015;112:815–20.

O'Neil ST, Dzurisin JDK, Williams CM, et al. Gene expression in closely related species mirrors local adaptation: consequences for responses to a warming world. Mol Ecol. 2014;23:2686–98.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol. 2013;22:5983–99.

Ottewell KM, Bickerton DC, Byrne M, Lowe AJ. Bridging the gap: a genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. Divers Distrib. 2016;22:174–88.

Oyler JW, Ballantyne A, Jencso K, Sweet M, Running SW. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. Int J Climatol. 2015;35:2258–79.

Oyler-McCance SJ, Fedy BC, Landguth EL. Sample design effects in landscape genetics. Conserv Genet. 2013;14:275–85.

Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for stacks. Methods Ecol Evol. 2017;8:1360–73.

Pavey SA, Gaudin J, Normandeau E, Dionne M, Castonguay M, Audet C, Bernatchez L. RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American Eel. Curr Biol. 2015;25:1666–71.

Peterman WE, Semlitsch RD. Fine-scale habitat associations of a terrestrial salamander: the role of environmental gradients and implications for population dynamics. PLoS One. 2013;8:e62184.

Peters JL, Lavretsky P, DaCosta JM, Bielefeld RR, Feddersen JC, Sorenson MD. Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). Biol Conserv. 2016;203:272–81.

Phillips BL, Brown GP, Shine R. Life-history evolution in range-shifting populations. Ecology. 2010;91:1617–27.

Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller MR. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv. 2017;3:e1603198.

Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P, Lena JP. Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. Mol Ecol. 2013;22:5516–30.

Ralph PL, Coop G. The role of standing variation in geographic convergent adaptation. Am Nat. 2015a;186:S5–S23.

Ralph PL, Coop G. Convergent evolution during local adaptation to patchy landscapes. PLoS Genet. 2015b;11:e1005630.

Razgour O, Taggart JB, Manel S, et al. An integrated framework to identify wildlife populations under threat from climate change. Mol Ecol Resour. 2018;18:18–31.
Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24:4348–70.
Rissler LJ. Union of phylogeography and landscape genetics. Proc Natl Acad Sci U S A. 2016;113:8079–86.
Roffler GH, Amish SJ, Smith S, Cosart T, Kardos M, Schwartz MK, Luikart G. SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. Mol Ecol Resour. 2016;16:1147–64.
Ruane S, Austin CC. Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. Mol Ecol Resour. 2017;17:1003–8.
Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, Paxton EH, Smith TB, Coulson T. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecol Lett. 2018;21:1085–96.
Russello MA, Waterhouse MD, Etter PD, Johnson EA. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ. 2015;3:e1106.
Santure AW, De Cauwer I, Robinson MR, Poissant J, Sheldon BC, Slate J. Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. Mol Ecol. 2013;22:3949–62.
Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20.
Schindler DE, Hilborn R, Chasco B, Boatright CP, Quinn TP, Rogers LA, Webster MS. Population diversity and the portfolio effect in an exploited species. Nature. 2010;465:609–12.
Schwartz MK, Luikart G, Waples RS. Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol. 2007;22:25–33.
Selkoe KA, Scribner KT, Galindo HM. Waterscape genetics – applications of landscape genetics to rivers, lakes, and seas. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics. Hoboken: Wiley; 2015. p. 220–46.
Sgro C, Lowe A, Hoffmann A. Building evolutionary resilience for conserving biodiversity under climate change. Evol Appl. 2011;4:326–37.
Sork V, Aitken S, Dyer R, Eckert AJ, Legendre P, Neale D. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet Genomes. 2013;9:901.
Steane DA, Potts BM, McLean E, Prober SM, Stock WD, Vaillancourt RE, Byrne M. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. Mol Ecol. 2014;23:2500–13.
Stelkens RB, Brockhurst MA, Hurst GDD, Greig D. Hybridization facilitates evolutionary rescue. Evol Appl. 2014;7:1209–17.
Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? Mol Ecol. 2010;19:3496–514.
Storfer A, Antolin MF, Manel S, Epperson BK, Scribner KT. Genomic approaches in landscape genetics. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics. Hoboken: Wiley; 2015. p. 149–64.
Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 2005;14:671–88.
Stucki S, Orozco-terWengel P, Forester BR, et al. High performance computation of landscape genomic models including local indicators of spatial association. Mol Ecol Resour. 2017;17:1072–89.
Swaegers J, Mergeay J, Van Geystelen A, Therry L, Larmuseau MHD, Stoks R. Neutral and adaptive genomic signatures of rapid poleward range expansion. Mol Ecol. 2015;24:6163–76.
Szulkin M, Gagnaire P-A, Bierne N, Charmantier A. Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. Mol Ecol. 2016;25:542–58.

Therkildsen NO, Palumbi SR. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Mol Ecol Resour. 2017;17:194–208.
Thomas L, Palumbi SR. The genomics of recovery from coral bleaching. Proc R Soc B. 2017;284:20171790.
Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol. 2014;29:673–80.
Todd EV, Black MA, Gemmell NJ. The power and promise of RNA-seq in ecology and evolution. Mol Ecol. 2016;25:1224–41.
Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr Biol. 2016;26:2313–8.
Trucchi E, Mazzarella AB, Gilfillan GD, Lorenzo MT, Schönswetter P, Paun O. BsRADseq: screening DNA methylation in natural populations of non-model species. Mol Ecol. 2016;25:1697–713.
van Gurp TP, Wagemaker NCAM, Wouters B, Vergeer P, Ouborg JNJ, Verhoeven KJF. epiGBS: reference-free reduced representation bisulfite sequencing. Nat Methods. 2016;13:322–4.
Vasseur DA, DeLong JP, Gilbert B, et al. Increased temperature variation poses a greater risk to species than climate warming. Proc R Soc Lond B Biol. 2014;281:20132612.
Verhoeven KJF, vonHoldt BM, Sork VL. Epigenetics in ecology and evolution: what we know and what we need to know. Mol Ecol. 2016;25:1631–8.
Vilà C, Sundqvist A-K, Flagstad Ø, et al. Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. Proc R Soc Lond B Biol Sci. 2003;270:91–7.
vonHoldt BM, Pollinger JP, Earl DA, et al. A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. 2011;21:1294–305.
Waits LP, Cushman SA, Spear SF. Applications of landscape genetics to connectivity research in terrestrial animals. In: Balkenhol N, Cushman SA, Storfer AT, Waits LP, editors. Landscape genetics. Hoboken: Wiley; 2015. p. 199–219.
Wang IJ, Bradburd GS. Isolation by environment. Mol Ecol. 2014;23:5649–62.
Wang T, Hamann A, Spittlehouse D, Carroll C. Locally downscaled and spatially customizable climate data for historical and future periods for North America. PLoS One. 2016;11:e0156720.
Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome sequencing: current and future perspectives. G3: Genes Genom Genet. 2015;5:1543–50.
Wayne RK, Shaffer HB. Hybridization and endangered species protection in the molecular era. Mol Ecol. 2016;25:2680–9.
Weeks AR, Sgro CM, Young AG, et al. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evol Appl. 2011;4:709–25.
Welbergen JA, Klose SM, Markus N, Eby P. Climate change and the effects of temperature extremes on Australian flying-foxes. Proc R Soc Lond B Biol. 2008;275:419–25.
Wenzel MA, Douglas A, James MC, Redpath SM, Piertney SB. The role of parasite-driven selection in shaping landscape genomic structure in red grouse (*Lagopus lagopus scotica*). Mol Ecol. 2016;25:324–41.
Whipple AV, Holeski LM. Epigenetic inheritance across the landscape. Front Genet. 2016;7:189.
Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA. Genetic rescue to the rescue. Trends Ecol Evol. 2015;30:42–9.
Willi Y, Van Buskirk J, Hoffmann AA. Limits to the adaptive potential of small populations. Annu Rev Ecol Evol Syst. 2006;37:433–58.
Wood JLA, Yates MC, Fraser DJ. Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. Evol Appl. 2016;9:640–57.
Wright B, Morris K, Grueber CE, et al. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. BMC Genomics. 2015;16:791.

Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. Evolution. 2011;65:1897–911.

Yoder JB, Tiffin P. Effects of gene action, marker density, and time since selection on the performance of landscape genomic scans of local adaptation. J Hered. 2017;109:16–28.

Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic signature of adaptation to climate in *Medicago truncatula*. Genetics. 2014;196:1263–75.

# Population Genomics of Ungulates


**Daria Martchenko, Erin Prewer, Emily K. Latch, Christopher J. Kyle, and Aaron B. A. Shafer**

**Abstract** Humans have long relied on ungulates for food, clothing, manual labor, and transportation. Ungulates were among the first species to be domesticated and managed in the wild, but more than one-third of species are currently of conservation concern. Starting in the late twentieth century, ungulate research and management began employing genetic tools to assess attributes like the degree of population structure, inbreeding, and variation in functionally important genes. As sequencing technology advanced, research on ungulates shifted to now assay variation across the entire genome. More than 20 ungulates have had their genome assembled with a mean length of 2.6 Gb and N50 of 26 Mb. Genomic studies have provided deeper insights into the evolutionary relationships among giraffes and bovids, while camelids and horses have had their entire species demographic histories reconstructed using novel Markovian coalescent models. Moreover, artificial and natural selection has left clear signatures on ungulate genomes with high-throughput sequencing techniques being used to identify the genetic basis to important phenotypic traits. Novel assembly strategies and genomic assays are regularly being employed on ungulates, and research on this ecological and economically valuable group will help chart the course of the emerging field of wildlife genomics.

**Keywords** Adaptation · Cross-species · Demography · Domestication · Genome assembly · SNP chip

---

Daria Martchenko and Erin Prewer contributed equally to this work.

D. Martchenko · E. Prewer
Environmental and Life Sciences, Trent University, Peterborough, ON, Canada

E. K. Latch
Behavioral and Molecular Ecology Research Group, Department of Biological Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

C. J. Kyle · A. B. A. Shafer (✉)
Environmental and Life Sciences, Trent University, Peterborough, ON, Canada

Department of Forensics, Trent University, Peterborough, ON, Canada
e-mail: aaronshafer@trentu.ca

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_30,

## 1 Introduction to Ungulates

Terrestrial ungulates are a diverse assemblage of species inhabiting ecosystems from the arctic to the desert (Fig. 1a). A defining characteristic of ungulates are their terminal phalanges enclosed in thick hoofs (Fig. 1b), thereby excluding cetaceans that are genetically embedded within the clade (Shimamura et al. 1997). Taxonomically, ungulates are represented by two orders, Cetartiodactyla or even-toed ungulates, such as pigs, goats, and camels, and Perissodactyla or odd-toed ungulates such as horses, tapirs, and rhinoceroses (Groves and Grubb 2011). Ungulates are under threat from stressors such as climate change, habitat loss, and overharvest (Ripple et al. 2015). In addition, most species exhibit some degree of migration or require large home ranges, with many of these territories and migration routes impacted by environmental perturbations (Dou et al. 2013; Post et al. 2008) and landscape modification (Gedir et al. 2015). Currently, 95 out of 235 ungulate species (numbers excluding the Cetacea order) are listed as critically endangered, endangered, or vulnerable by the International Union for the Conservation of Nature (IUCN 2017).

Humans have had a relationship with ungulates for thousands of years as they are critical sources of food, clothing, and general tools that assist in manual labor and transportation. Ungulates were among the first species to be domesticated starting with sheep and goats approximately 10,000 years ago (Chessa et al. 2009). Wild ungulates are an important subsistence source in indigenous communities (Vors and Boyce 2009) while also generating large revenues through sport hunting and ecotourism (Gordon et al. 2004). The first known efforts of game management

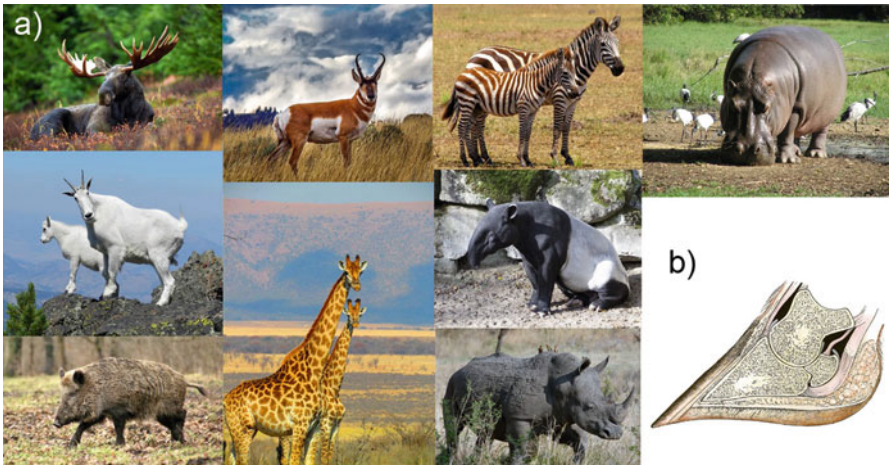


**Fig. 1** Example of ungulates (**a**) and their characteristic enclosed phalanges, or hoofs (**b**). All images fall under Creative Commons attribution

involving ungulates can be traced to the Mongol Empire in the thirteenth century (Leopold 1987). Some semblance of game management appeared in Europe a century later where there were defined hunting seasons (Leopold 1987), while efforts at management in North America can be traced back to the late seventeenth century when a hunting ban on deer was enacted in Rhode Island (Brown 2013). To put their economic importance in perspective, in 2011 alone, 14 billion USD was spent hunting deer and elk in the United States (U.S. Fish and Wildlife Service 2012). Ungulates are also a keystone species in many ecosystems, with grazing enhancing ecosystem productivity by increasing the nutritional quality of forage and supporting higher plant biodiversity (Truett et al. 2001). While anthropogenic impacts have contributed to the aforementioned conservation listing of many ungulates, some species have benefited from human interventions in the form of habitat modification and the removal of predators, notably cervids (i.e., the deer family), where populations have required controlled culling to reduce numbers (Gordon et al. 2004).

Ungulate research and management changed in the late twentieth century, and a large number of individual-based, long-term ungulate research projects collecting samples for DNA analysis were established during this time (Table 1). Here, genetic tools have been vital for inferring parentage within the study populations, and more broadly, documenting population structure and quantifying patterns of gene flow and isolation (Scribner 1993). For example, reintroductions of species are often done with a limited number of individuals (Rhodes and Latch 2010; Scribner 1993); therefore monitoring genetic processes in populations has become a critical element of gauging the success of management efforts (Hogg et al. 2006; Miller et al. 2012). Genetic tools have also been instrumental for improving our understanding of life history characteristics, particularly mating structure. Targeting functional and immune genes like the major histocompatibility complex (MHC) has been commonplace in ungulate research. This is because variability in vertebrate MHC is associated with higher pathogen resistance (Janeway 2001; Quéméré et al. 2015) and often linked to individual fitness characteristics, such as body size, weight and parasite load, and secondary sexual traits such as antlers (e.g., Ditchkoff et al. 2001; Paterson et al. 1998; Winternitz et al. 2013). These genotype-phenotype relationships (see Box 1), however, are nuanced, and, for example, in red deer some MHC variants were associated with higher resistance to some parasites but not others, and there was no relationship with antler size (Buczek et al. 2016). These genetic queries of ungulates have benefitted from recent advances in sequencing technology that have facilitated the transition from genetic to genome-wide datasets, resulting in numerous genome assemblies and resources for ungulates.

**Table 1** Ongoing and long-term ungulate research projects involving genetic sampling

| Species name | Common name | Location | Year |
|---|---|---|---|
| *Ovis canadensis* | Bighorn sheep | Ram Mountain, Alberta, Canada | 1972 |
| *Ovis canadensis* | Bighorn sheep | Sheep River, Alberta, Canada | 1981 |
| *Ovis canadensis* | Bighorn sheep | Bison Range, Montana, USA | 1979 |
| *Ovis aries* | Soay sheep | St. Kilda, Scotland | 1985 |
| *Oreamnos americanus* | Mountain goat | Caw Ridge, Alberta, Canada | 1988 |
| *Cervus elaphus* | Red deer | Isle of Rum, Scotland | 1972 |
| *Cervus elaphus* | Red deer | Trondelag, Norway | 1987 |
| *Equus ferus caballus* | Feral horses | Sable Island, Nova Scotia, Canada | 2007 |
| *Odocoileus virginianus* | White-tailed deer | Anticosti Island, Quebec, Canada | 2002 |
| *Capreolus capreolus* | Roe deer | Trois Fontaines, France | 1975 |
| *Capreolus capreolus* | Roe deer | Chize, France | 1977 |
| *Rangifer tarandus* | Svalbard reindeer | Svalbard archipelago, Norway | 1994, 2002 |
| *Bison bison* | Bison | Konza Prairie, Kansas, USA | 1987 |
| *Bison bison* | Bison | Wind Cave, South Dakota, USA | 1960s |
| *Ovis gmelini* | European mouflon | Caroux-Espinouse Massif, France | 1985 |
| *Rupicapra rupicapra* | Alpine chamois | Les Bauges, France | 1985 |
| *Capra ibex* | Alpine ibex | Belledonne, France | 1985 |
| *Dama dama* | Fallow deer | Phoenix Park, Ireland | 1985 |
| *Dama dama* | Fallow deer | San Rossore, Italy | 1984 |
| *Tragelaphus strepsiceros* | Greater kudu | Kruger Park, South Africa | 1984–1993 |
| *Alces alces* | Moose | Vega, Norway | 1992 |
| *Antilocapra americana* | Pronghorn antelope | Bison Range, Montana, USA | 1981 |
| *Rupicapra pyrenaica* | Pyrenean chamois | Orlu, France | 1985 |
| *Rupicapra pyrenaica* | Pyrenean chamois | Parc National des Pyrenees, France | 1993 |

## 2 A Brief History of Whole-Genome Assemblies in Ungulates

The first assemblies of ungulate genomes started with domestic species of high interest and economic importance. The first completed ungulate draft genomes were of the cow (*Bos taurus*) in 2004 and domestic horse (*Equus caballus*) in 2007, both of which used early forms of whole-genome shotgun sequencing (Elsik et al. 2009; Wade et al. 2009). This method randomly fragments genomic DNA and inserts it into either fosmid or bacterial artificial chromosomes (BAC) vectors. The DNA is then clonally amplified, and random clones are selected and sequenced using Sanger sequencing. With technological improvements, subsequent ungulate genomes applied hybrid approaches that combined BAC-end or fosmid-end sequencing with newer high-throughput sequencing (HTS) techniques. In this approach, library

preparation occurred in a cell-free system, rather than cloned in vectors, and produced significantly more data in less time. Both domestic goat (*Capra hircus*) and domestic pig (*Sus scrofa*) used BAC and fosmid sequencing to create long read pairs that facilitated the placement of the newer high-throughput but short-read data into contigs and scaffolds (Dong et al. 2015; Groenen et al. 2012). Improvements in assembly continued with the addition of mate-pair library preparations that circularized long fragments of DNA, thereby creating long insert reads without the use of vectors. This approach has been used to assemble both the domestic yak (*Bos mutus*) and sheep genome (*Ovis aries*; Jiang et al. 2014; Qiu et al. 2012).

Most genome sequencing projects to date have been concentrated on model or domestic species. With the improvement and lowered cost in sequencing technologies, that focus has shifted to include more wildlife species (Ellegren 2014). Current genome assemblies rely on whole-genome sequencing of short-read and mate-pair fragments (or alternative long-read sequencing strategies) with HTS technology. For ungulate assemblies, the most popular sequencing technology is Illumina sequencing by synthesis that includes, for example, the wild goat (*Capra aegagrus*; Dong et al. 2015), giraffe (*Giraffa camelopardalis*; Agaba et al. 2016), and Przewalski's horse genomes (*Equus przewalskii*; Huang et al. 2014). A trend has emerged toward using multiple sequencing technologies to improve genome assembly; for example, the Bactrian camel (*Camelus ferus*; Jirimutu et al. 2012) used both Illumina and SOLiD sequencing, while the donkey (*Equus asinus*; Huang et al. 2015) assembly used both Illumina and 454 pyrosequencing technologies. The bioinformatics combination of high-coverage sequencing of a wild species and use of a closely related reference are also commonplace. Both the okapi (*Okapia johnstoni*; Agaba et al. 2016) and bighorn sheep (*Ovis canadensis*; Miller et al. 2015) assemblies relied on closely related [domestic] genomes. Overall we collected data on 25 ungulate genomes; these showed high variation in N50 scaffold size (range from 10,458 to 100,310,653) with a mean N50 of 27,074,698. Genome size estimates were more consistent ranging from 1.99 to 3.31 Gb, with a mean length of 2.61 Gb (Table 2).

## 3 The Importance of Domestic Genomes

Ungulate research has benefited from genome projects on domestic animals more so than most other taxonomic groups. Resources like primers, SNP chips, and annotated genomes generated in domesticated species including cow, sheep, goat, pig, and horse can be applied to related ungulates, resulting in so-called "genome-enabled" taxa (Kohn et al. 2006). While there are limits to how distantly related a species one can use (Cosart 2013; Miller et al. 2012), and concerns related to potential biases (Powell et al. 2016; Shafer et al. 2016), the genome-enabled nature of ungulates has permitted addressing genome-scale questions in ungulates well before other taxonomic groups. The application of cross-amplified markers has been used to generate relatively dense linkage maps (Poissant et al. 2010; Slate et al. 2002), assess population structure (Miller et al. 2011; Haynes

**Table 2** Whole-genome assemblies in ungulates

| Species name | Common name | Assembly length (Mb) | N50 (scaffold) | Coverage | Sequencing technology |
|---|---|---|---|---|---|
| *Camelus ferus* | Wild Bactrian camel | 2,009 | 2,005,940 | 30× | Illumina GAIIx; 454 GS-FLX Titanium; SOLid 3 |
| *Camelus dromedarius* | Arabian camel | 2,004 | 4,188,677 | 65× | Illumina HiSeq2000 |
| *Camelus bactrianus* | Bactrian camel | 1,993 | 8,812,066 | 79× | Illumina HiSeq2000 |
| *Bison bison* | American bison | 2,828 | 7,192,658 | 60× | 454; Illumina HiSeq |
| *Bubalus bubalis* | Water buffalo | 2,836 | 1,412,388 | 70× | Illumina GAIIx; Illumina HiSeq; 454 |
| *Capra aegagrus* | Wild goat | 2,829 | 91,317,560 | 84× | Illumina HiSeq2000 |
| *Capra hircus* | Domestic goat | 2,923 | 87,277,232 | 50.0× | PacBio |
| *Bos mutus* | Wild yak | 2,645 | 1,407,960 | 130× | Illumina HiSeq; Illumina GA |
| *Ovis aries* | Sheep | 2,616 | 100,009,711 | 166× | Illumina GAII; 454; PacBio RSII |
| *Ovis canadensis* | Bighorn sheep | 2,591 | 100,190,483 | 12× | PacBio |
| *Capreolus capreolus* | Western roe deer | 2,785 | 10,458 | 50× | Illumina HiSeq |
| *Odocoileus hemionus* | Mule deer | 3,310 | 156,284 | 50× | Illumina HiSeq |
| *Okapia johnstoni* | Okapi | 2,878 | 111,538 | 30× | Illumina HiSeq |
| *Giraffa camelopardalis* | Giraffe | 2,705 | 212,164 | 37× | Illumina HiSeq |
| *Sus scrofa* | Pig | 2,809 | 576,008 | 24× | Illumina GAII |
| *Ceratotherium simum* | Southern white rhino | 2,464 | 26,277,727 | 91× | Illumina HiSeq |
| *Equus asinus* | Ass | 2,391 | 3,776,412 | 42× | Illumina MiSeq; 454; Illumina HiSeq |
| *Equus caballus* | Horse | 2,475 | 46,749,900 | 6.8× | ABI 3730 |
| *Equus przewalskii* | Przewalski's horse | 2,396 | 513,800 | 86× | Illumina HiSeq |
| *Loxodonta africana* | Elephant | 3,197 | 46,401,353 | 7× | Sanger |
| *Bos taurus* | Cow | 2,650 | 6,806,220 | 19× | Sanger, PacBio RS II |
| *Bos indicus* | Zebu | 2,674 | 106,310,653 | 52× | SOLiD |
| *Vicuna pacos* | Alpaca | 2,014 | 5,303,709 | 73× | Illumina HiSeq2000 |
| *Pantholops hodgsonii* | Chiru | 2,697 | 2,772,860 | 67× | Illumina |
| *Elaphurus davidianus* | Père David's deer (Milu) | 2,580 | 2,850,000 | 114× | Illumina HiSeq2000 |

Assembly length is reported in megabases (Mb) and N50 scaffold is equal to the length of the longest sequence where the sum of the lengths is greater than or equal to half the length of the genome being assembled

and Latch 2012; Iacolina et al. 2016), scan for adaptive loci (Sim et al. 2016; Powell et al. 2016; Roffler et al. 2016a, b), and assay functional gene variation (Slate et al. 2009; Shafer et al. 2012). Moreover, both population monitoring and management of wild ungulates have routinely relied on markers developed in the non-focal species (Corti et al. 2011; Ogden et al. 2012; Olson et al. 2012). The cross-species application of resources remains prevalent in basic and applied research of ungulates (Fig. 2).

Despite multiple de novo assemblies for ungulates, quality varies as evidenced by the N50 scaffold estimates (Table 2). This is because for mammals, de novo genome assembly and annotation are still prohibitive in terms of cost, computational resources, and required genetic material for a high-quality draft genome (Ekblom and Wolf 2014). There are multitudes of less direct benefits of being genome-enabled taxa. For example, de novo genome assemblies can use the domestic reference as a backbone to make assemblies more tractable, as was the case in bighorn sheep (Miller et al. 2015) and elk (Brauning et al. 2015). Similarly, genome annotation can take advantage of protein-based homology approaches using the wide array of domesticated ungulate data (Fitak et al. 2016). In silico digests of domestic genomes also provide important information for genomic assay design (Van Tassell et al. 2008; Wiedmann et al. 2008) and techniques like optical mapping (Dong et al. 2013): this information can easily be incorporated into experimental designs and analyses of wild populations with limited resources. From a practical standpoint, sequencing facilities require basic genome information that more often than not can simply be borrowed from the most closely related wild or domestic relative (Fig. 2; Table 2).

## 4 Alternatives to Whole-Genome Sequencing

Reduced representation strategies have emerged as valuable alternatives to whole-genome sequencing (Narum et al. 2013), though they have been used sparsely in ungulates. By targeting a subset of the genome for resequencing, reduced representation strategies facilitate HTS in non-model species even in the absence of genomic information. Reduced representation approaches largely fall into two categories, genotyping by sequencing (GBS) approaches that target for DNA variants distributed across the genome and transcriptomes that provide insight into functional variation in transcribed DNA. Transcriptome sequencing has typically focused on domestic species (Jäger et al. 2011; McLoughlin et al. 2014) or farmed species to identify differentially expressed genes associated with disease or economically valuable traits (Box 1, Marfell et al. 2013; Yao et al. 2012a). A transcriptome of white-tailed deer has also been generated (Genomic Resources Development Consortium et al. 2014) that will be useful for annotation and assay designs. GBS assays allow for a targeted subset of genome to be sequenced across multiple individuals or populations and can be used to target anonymous loci, candidate loci, or a combination of both, each coming with important considerations for implementation

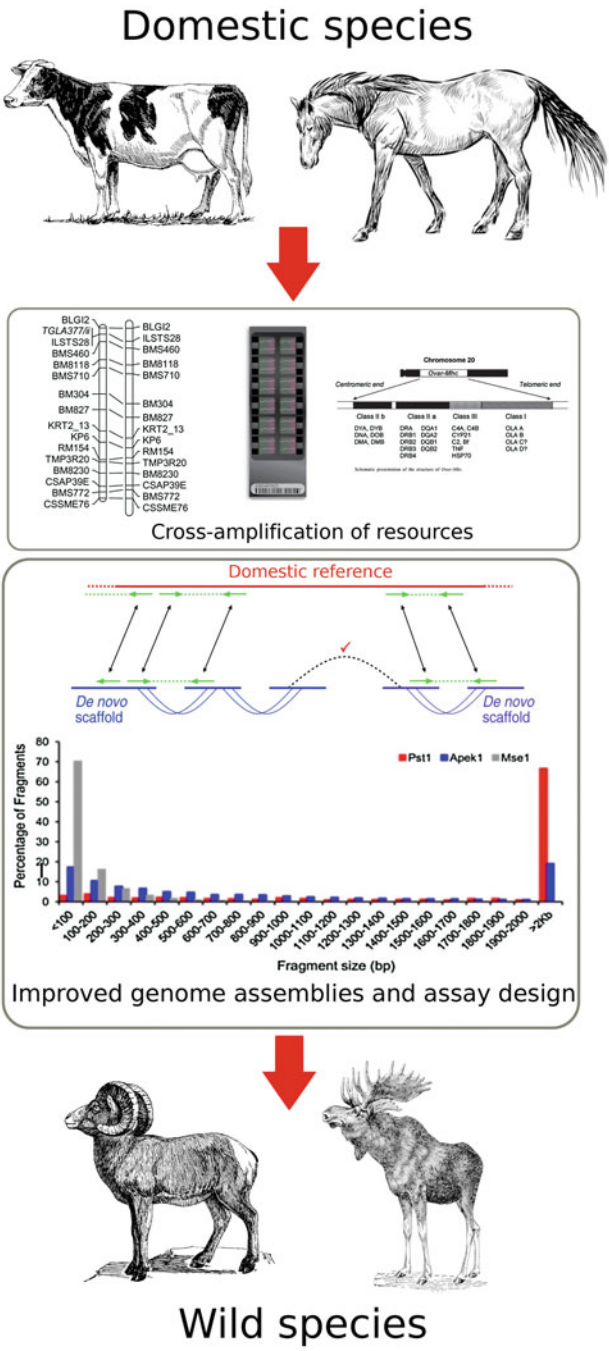


**Fig. 2** The genome-enabled nature of ungulate genomics. Genomic resources developed in domestic animals (top) have been, and still are, applied to wild ungulate species (bottom). This includes cross-amplification of markers for building linkage maps, SNP chips for genotyping, and candidate genes for sequencing. Domestic reference genomes can also be used to improve de novo assemblies of wild ungulates and help with the design assays (i.e., in silico digest of domestic genome to select restriction enzymes for a reduced representation sequencing approach)

(Andrews et al. 2016; Jones and Good 2016). Surprisingly few studies focusing on an ungulate taxon have actually employed a GBS assay. To our knowledge GBS assays have been limited to cattle (De Donato et al. 2013), pygmy hippo (Senn et al. 2014), and addax (Ivy et al. 2016). In part, readily available SNP chips from domesticated species have fettered the de novo development of GBS assays in wild ungulates.

Candidate genes of interest can be surveyed through targeted resequencing. Here, cross-species exon capture, in which coding sequences from a model species are used to create hybridization probes to sequence homologous exons in non-model species, has seen recent application in ungulates. Cosart et al. (2011) first demonstrated the utility of this approach using the cattle genome to capture exons in bison and zebu. Powell et al. (2016) used the cattle genome to capture exons in mule deer (*Odocoileus hemionus*), while Roffler et al. (2016a) used probes designed from the domestic sheep genome to capture exons in Dall's sheep (*Ovis dalli dalli*). In recent applications where no genome was available in the target species, over half of the probes designed in cattle successfully hybridized in mule deer (~30 million years divergent; Powell et al. 2016), and 35% of probes hybridized to pigs (50–60 million years divergent; Cosart 2013). Overall, both reduced representation and exon capture data provide useful insight into population processes and adaptive and functional variation in ungulates (Cosart et al. 2011; Hodges et al. 2007; Powell et al. 2016; Roffler et al. 2016a) that can be informative for basic research and conservation and management programs.

## 5 From Phylogenomics to Population Demography

Examining newly constructed ungulate genomes has allowed for determination of the most recent common ancestor (MRCA) and reconstructing evolutionary relationships. This is particularly relevant for ungulate groups like the Caprinae, where even 6,000 base pairs of DNA sequence data could not resolve deeper phylogenetic relationships (Shafer and Hall 2010). Whole-genome approaches must consider the influence of incomplete lineage sorting (Nater et al. 2015), although this appears to be only a minor cause of phylogenetic discrepancies within mammals (Scornavacca and Galtier 2017). Among ungulates, whole-genome sequencing of the okapi and giraffe showed that giraffe species shared a MRCA ~2 mya, with the okapi roughly 11.5 mya, and with cattle 28 mya (Agaba et al. 2016). Glanzmann et al. (2016) showed that the African buffalo shared MRCA with the cow (*Bos taurus*) approximately 5–10 mya based on the two species' genome sequences.

Genomic data also offer the prospect of reconstructing population histories from a single contemporary genome, a complex task that is virtually impossible with patchy or nonexistent observational and fossil data. Two particularly exciting approaches are the pairwise and multiple sequentially Markovian coalescent models (SMCs; Li and Durbin 2011; Schiffels and Durbin 2014 see also Salmona et al. 2017) that were developed for whole-genome data and have recently been applied to GBS data

(Liu and Hansen 2017). Using SMC models, Orlando et al. (2013) compared the genomes of five domestic horse breeds, a Late Pleistocene horse, Przewalski's horse, and a donkey to reconstruct the demographic history of the modern horse. Horse population fluctuations coincided with favorable and adverse climatic conditions and were confirmed by population size estimates resulting from analysis of ancient mitochondrial DNA genomes (Orlando et al. 2013). Recent reductions correspond to the last glacial maximum with no evidence of recovery (Fig. 3a; Der Sarkissian

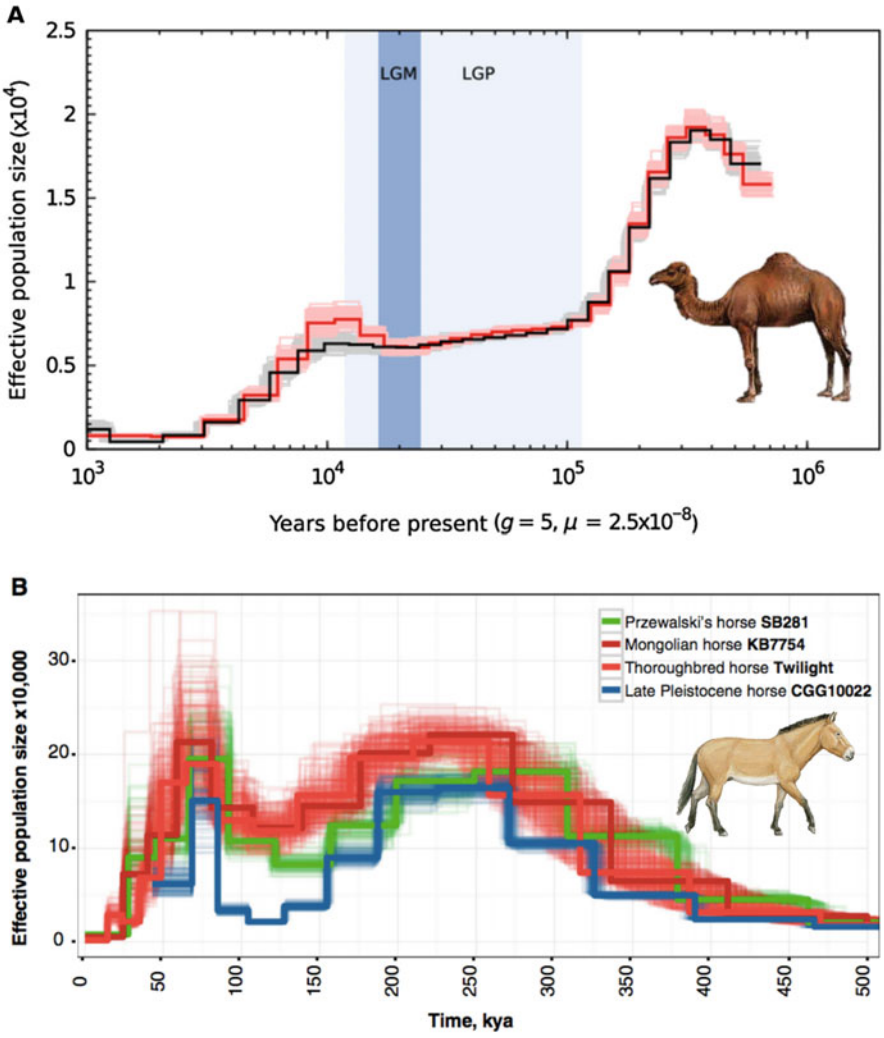


**Fig. 3** Demographic inferences using pairwise sequentially Markovian coalescent (PSMC) models in **(a)** dromedary (camel) and **(b)** horse species. PSMC plots are from Fitak et al. (2016) and Der Sarkissian et al. (2015)

et al. 2015; Librado et al. 2015). Similarly, Wu et al. (2014) characterized the demographic history of three Old World camel species over the last million years (only dromedary shown in Fig. 3b). The three camel species experienced sharp drops in the effective population size (up to 70%) from 100,000 to 20,000 ya due to climatic changes during the last glacial period (Burger 2016; Fitak et al. 2016) with the reasons for a second, recent reduction hypothesized to be from habitat loss and increased hunting pressure (Burger 2016).

## 6 Adaptive Divergence in Ungulates

Ungulates inhabit some of the most inhospitable environments on earth ranging from the Arctic to desert. These selective pressures have driven genetic changes by favoring certain phenotypes in one environment, thereby increasing corresponding allele frequencies over generations (Whitehead 2012). A plurality of genome sequencing approaches can be used to identify locally adapted genes (Hoban et al. 2016), with many being applied to ungulates. Desert species, like camels and some goat and sheep populations, have several adaptations to deal with excessive sun exposure and hot and dry conditions (Table 3). To combat low water availability in the desert, both camels and sheep have adaptations that promote vasodilation and water reabsorption (Jirimutu et al. 2012; Yang et al. 2016). Camels also show evidence for selection on genes that affect photoreception and visual protection, presumably related to high exposure to UV radiation in the desert (Wu et al. 2014). Similarly, goat and sheep populations showed selection on genes that have likely contributed to thermotolerance by regulating melanogenesis and coat color (Kim et al. 2016).

Genome sequencing of the Tibetan yak and goat has detected genetic evidence for adaptation to high-altitude environments (Qiu et al. 2012; Song et al. 2016). Both yak and goats show increased blood hemoglobin concentrations (Song et al. 2016), whereas sheep in the Himalayas appear to have evolved enhanced lung capacity (Gorkhali et al. 2016). Ungulates inhabiting the Arctic and sub-Arctic, including the Yakutian horse and even woolly mammoth (although a member of the Proboscidea order), have shown evidence for changes altering the regulation of epidermis and hair development, creating thicker, warmer hair (Librado et al. 2015; Lynch et al. 2015). These two species also have adaptations associated with their circadian clock in order to maintain normal rhythms despite the lack in daily light and dark cycles in the Arctic (Librado et al. 2015; Lynch et al. 2015).

Not all adaptations are due to natural selective pressures but rather domestication events and selective breeding. The domestication and selective breeding of goats, pigs, and horses have led toward increased productivity (e.g., milk production, body size, and fecundity; National Research Council et al. 2002). Fertility is an important factor when breeding, and genes associated with increased litter sizes have

**Table 3** Candidate genes identified in ungulates, the proposed selection pressure, and adaptive response

| Selection pressure | Adaptive response | Species | Candidate genes |
|---|---|---|---|
| Desert/low water availability | Increased water reabsorption/salt concentration | Camel, sheep | CYP2J, CYP2E, GPX3 |
| Desert/airborne dust | Increased respiratory health/lung development | Camel | FOXP3, CX3CR1, CYSLTR2, SEMA4A |
| Desert/high UV exposure | Increased visual protection | Camel | OPN1SW, CX3CR1, CNTFR |
| High altitude/low oxygen availability | Increased hypoxia response | Yak, horse, sheep | Adam17, Arg2, Mmp3 |
| High altitude/low oxygen availability | Increased hemoglobin concentration | Horse, yak | EPAS1 |
| Desert/high temperature | Increased thermotolerance (melanogenesis) | Sheep, goat | FGF2 |
| Arctic/extended periods of light and dark | Altered circadian clock response | Woolly mammoth, horse | LECT2, FBXL21, PER2 |
| Arctic/low temperature | Increased hair and epidermis development | Woolly mammoth, horse | BARX2, TRPV3 |
| Domestication | Increased fat production/composition | Pig | FASN, MOGAT2 |
| Domestication | Nervous system | Pig, horse, goat | VDAC1, GRID1, CACNA1C, CACNA1D |
| Domestication/ racing | Increased speed | Horse | MSTN |
| Agriculture/ transport | Equine physiology | Horse | NR3C2, SCPEP1, ACAD8, SGCD |
| Domestication | Higher fertility | Pigs | AHR |
| Intra-male competition and female choice | Larger horns | Bighorn sheep, Soay sheep, sheep | RXFP2 |
| Low food availability | Increased energy storage/metabolism | Camel, sheep, goat, horse, woolly mammoth | ACC2, DGKZ, GDPD4, GLUL, GCNT3,MYH, NOS3, eNOS, CPS1 |

signatures of selection (Groenen 2016). Both horses and goats have been under selection for serotonin release, as this is involved in the taming of animals (Dong et al. 2015; Schubert et al. 2014). There are also optimal or desired traits in domesticated species that have been selected for by breeders. For example, horses show selection for an allele in the MSTN gene, also known as the speed gene in

thoroughbreds (Schubert et al. 2014), and there is evidence for selection of genes like NR3C2, SCPEP1, ACAD8, and SGCD that are associated with the increased energy demand required agriculture, transport, and racing (Schubert et al. 2014). Pigs show signatures of selection due to breeding and domestication, specifically in fat composition (Molnár et al. 2014); the gene FASN, which encodes a fatty acid synthase, increases the total body weight and fat, and the gene MOGAT2 affects the absorption and digestion of fat (Molnár et al. 2014).

## 7 Future Perspectives

More than one-third of ungulate species are of conservation concern, with threats ranging from climate change to overharvest (IUCN 2017). Disease and other changing selective pressures can lead to rapid population declines in wildlife populations – often threatening population persistence, as evidenced by several emerging wildlife diseases and the increased transfer of disease between domestic and wild ungulate populations (e.g., see Forde et al. 2016; Kutz et al. 2004; Martin et al. 2011). Selective forces exerted by infectious diseases can rapidly influence the distribution of adaptive genetic variants associated with disease susceptibility over short time scales (Gallana et al. 2013). This process can result in an evolutionary rescue of a species (Carlson et al. 2014; Maslo and Fefferman 2015), where disease-resistant animals increase in number subsequent to initial population declines from strong selective sweeps from disease. Supporting this would be scans of immune genes showing signatures consistent with selection in buffalo (Lane-deGraaf et al. 2015) and more broadly ungulates (Schaschl et al. 2006). The prion gene (PRNP) is of particular relevance for screening in cervids because of its link to chronic wasting disease resistance.

Genomic data can also guide the identification of individuals most suited for translocation and introduction to genetically depauperate populations. For example, in case of a disease outbreak, genome scans can test for pathogen presence in the putative source (see also Box 2). Extending this idea, translocation candidates can also be selected based on specific genomic markers that increase fitness or just to introduce more variation to the target population (Shafer et al. 2015). Genetic and genomic profiles of endangered oryx species (Hedrick et al. 2000; Ogden et al. 2012) are being factored into breeding programs and eventual reintroduction plans. Further, monitoring the success of a translocation or reintroduction can be bolstered by genome data, where Miller et al. (2012) used genomic data to document the positive effect of introductions, including the reversal of deleterious effects of inbreeding, in a population of bighorn sheep.

**Box 1 Genetic Basis to Horns and Antlers**

Multiple ungulate species have evolved cranial appendages in the form of horns and antlers (Fig. 4). Understanding the genetic basic to these traits is important for our understanding of both sexual selection (Poissant et al. 2008) and artificial selection (Hengeveld and Festa-Bianchet 2011), with work on the impact of trophy hunting (Coltman et al. 2003) and the downstream management consequences being particularly polarizing. However, pedigree (Coltman et al. 2003) and harvest data (Pigeon et al. 2016) data show an unequivocal genetic component to horn growth. Likewise, antler characteristics are heritable, although estimates vary (Lukefahr and Jacobson 1998; Michel et al. 2016; Williams et al. 1994). Targeted sequence of immune genes and their influence on antler development have produced mixed results (Buczek et al. 2016; Ditchkoff et al. 2001), but genomic studies have identified multiple candidate genes (Johnston et al. 2011; Poissant et al. 2008), with the relaxin-like receptor 2 (RXFP2) linked to horn size in Soay sheep (Johnston et al. 2013) and bighorn sheep (Kardos et al. 2015; Roffler et al. 2016a, b). Transcriptome sequencing has also identified candidate genes associated with antler development, with genes involved in signaling (growth) pathways and extracellular matrix proteins upregulated (Yao et al. 2012b). While genome scans for genes underlying traits in natural populations will have limitations (Kardos et al. 2016), the array of long-term studies with pedigree information (Table 1) hold considerable promise for elucidating the genomic architecture of horns and antlers. Once detected, management and conservation agencies could easily screen such genes – especially in scenarios of suspected artificial selection – to inform harvest quotas and breeding designs.

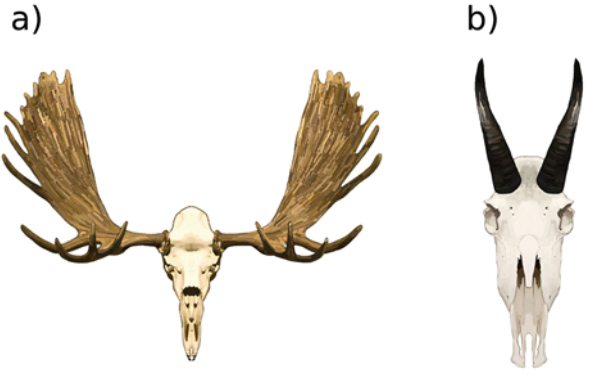


**Fig. 4** Many ungulate studies have focused on detecting the genetic basis to (**a**) antlers and (**b**) horns. Antlers consist of bone and are shed annually by members of the cervid family; horns have a permanent bony core that is covered by keratin and is found in the bovid family. Images courtesy of Peter Mills

**Box 2 Metagenomic Assessments of Ungulate Microbiomes and Viromes**

Metagenomes are defined as all the genetic material found within an environmental sample but have a general focus on microbiota (Marchesi and Ravel 2015). As such, the source of this genetic material falls into two groups: viral, including bacteriophages and RNA and DNA viruses that make up the virome, and microbiota that include bacteria, archaea, and fungi that make up the microbiome. Genomic tools include targeted amplicon sequencing of conserved genes, such as 16S in bacteria, and shotgun sequencing techniques that rely on the similarity of sequences to define operational taxonomic units (OTUs). These molecular approaches to assessing metagenomes are important in that many taxa are either very difficult or too dangerous to culture, and even when cultured, can be very difficult to identify phenotypically (Budowle et al. 2007).

The importance of elucidating the diversity of different microbiomes and viromes is related to the fact that this nonhost genetic material is integral to species health and can be commensal, symbiotic, and pathogenic in nature. The roles of these nonhost taxa are as fundamental as conveying an ability to process and obtain nutrients from food to protecting or harming the host organism through varied immune function. Specifically, ungulate microbiomes play central roles in digestion by providing the metabolic capabilities required to digest the ingested plant material (Yoon et al. 2015). Importantly, understanding the microbial communities of commercial ungulates is front and center when it comes to improving production (Alexander and Plaizier 2016). It is not surprising then, that much of the literature on ungulate metagenomics is focused on applications within animal husbandry. For example, bovine respiratory disease is a costly yet poorly understood disease for the cattle industry. Mitra et al. (2016) characterized the virome of cattle and found that a suite of both previously known and uncharacterized viruses contribute to the etiology of this disease. Magistrelli et al. (2016) investigated the influence of different diets on the microbial populations of pig fecal matter using targeted oligonucleotide probes to gain insight into the respective bacteriomes and infer health. Similar work on captive musk deer identified a shift in microbial communities with age (Hu et al. 2017).

Metagenomic approaches have been used to illustrate not only the diversity but also the evolutionary history, geographic range, and spread of various pathogens. Schirtzinger et al. (2015) gained insight into the evolutionary history and spread of viruses associated with respiratory syndrome in pigs through metagenomic sequencing, whereas Dupuy et al. (2015) documented

(continued)

**Box 2** (continued)
the spread of contagious caprine pleuropneumonia that infects both wild and domestic goats in Africa and Asia. As such, metagenomic tools have been identified as being critical to enhanced surveillance and control measures for these devastating diseases and begin to highlight the real threats of transmission of infectious disease between wildlife and domestic ungulates (Martin et al. 2011; Maclachlan and Mayo 2013). For wild populations, beyond identifying health parameters, varying metagenomic profiles have the potential to give higher spatial and temporal resolution to the movements of their host species and spread of emerging diseases, although to date, such approaches have been limited. Forde et al. (2016) used bacterial genomics to understand the epidemiology of *Erysipelothrix rhusiopathiae* in Arctic and boreal ungulates, including caribou, moose, and muskox. In yaks, further insight was gleaned from bacterial genomic analyses of a diarrhetic disease (Chen et al. 2015). Beyond applications toward enhanced understanding of animal health and enhancing commercial applications, ungulate metagenomics, specifically those from pigs, are also often used as surrogates for human research given the similarity of their internal flora and fauna relative to humans. McIntyre et al. (2016) used 16S rDNA profiling to identify the cutaneous microbiome of pigs on healthy skin and at the sites of wounds and found subsets of bacteria accelerate healing from injury. Human health is also heavily impacted by zoonoses, such as identified by Roth et al. (2016) where hepatitis E viral infections in humans were traced back to wild boar strains. The applications and implications of metagenomics work are largely nascent, but the aforementioned spectrum of studies provides insight into the importance of continued research in this field.

Genome-wide SNP arrays can be useful for estimating relationships among individual animals and for reconstructing pedigrees, which has advanced conservation and management of both captive and wild ungulates (Box 3). Genomic data also provide accurate estimates of inbreeding and can identify the specific genes contributing to inbreeding depression (Kardos et al. 2016). This can have important consequences, as, for example, in Soay sheep, the estimated rate of inbreeding detection was higher based on genomic data compared to more traditional molecular markers (Bérénos et al. 2016). In red deer only three fitness components were significantly correlated with the pedigree inbreeding estimate compared to six based on SNP data (Huisman et al. 2016). Genomic data provide some key advantages over pedigree estimates in terms of the sensitivity and accuracy of inbreeding estimates while permitting the potential identification potential genes underlying inbreeding depression (Kardos et al. 2016).

**Box 3 Genomics for Conservation and Management of Captive and Intensively Managed Ungulates**

Genomic data has improved the management of captive and intensively managed ungulates. Zoo-based captive breeding programs typically rely on accurate pedigrees to maintain genetic variation and prevent close inbreeding over the long term. Breeding strategies that minimize the average kinship (coancestry) in a population are an effective way to retain diversity and limit the accumulation of inbreeding (Lacy 1995; Fernández and Toro 1999; Sonesson and Meuwissen 2001). In the absence of complete data on parentage, molecular data can be used to resolve unknown relationships. High-throughput sequencing methods have the potential to revolutionize the genetic management of populations with incomplete or poorly known pedigrees, because we can use a large number of DNA markers to calculate very accurate estimates of kinship between animals (Jones et al. 2010; Santure et al. 2010; Skare et al. 2009). In captive ungulates, genome-wide SNPs have been used to accurately estimate relationships among individuals (Bosse et al. 2015; Ivy et al. 2016). Ivy et al. (2016) used both the BovineHD BeadChip and RAD-derived SNPs to resolve unknown relationships within the addax (*Addax nasomaculatus*) captive breeding program. Both approaches produced sufficient data to accurately estimate relationships, even in this severely bottlenecked population. Molecular coancestry estimates are also improving genetic diversity retention in intensively managed wild populations, where managers cannot dictate breeders but can remove less valuable individuals from the breeding pool (Eggert et al. 2010; Giglio et al. 2016, 2018). In bison, a kinship-based removal strategy outperformed alternative removal strategies at retaining genome-wide variation over the long term and limiting the accumulation of inbreeding (Giglio et al. 2016). Long-term population viability relies on the maintenance of genome-wide variation, and HTS has the potential to revolutionize the genetic management of captive and wild populations by supplying a very large number of markers distributed throughout the genome to calculate very accurate empirical estimates of genetic variation within and among populations and genetic relationships between individual animals.

Conservation and management of ungulates stands to benefit from identifying geographic patterns of local adaptation, the environmental drivers of divergent selection among populations, and genes and their variants involved in local adaptation, especially in the context of changing environmental conditions. Individuals with adaptive genetic variants leading to higher fitness in local environments could be used in conservation actions such as breeding programs, assisted gene flow, genetic rescue, or reintroduction programs to help ensure success of those programs. Managers could monitor the frequency of adaptive variants over time to assess the genetic health of a population following management interventions (Schwartz et al. 2007; Shafer et al. 2015; Flanagan et al. 2017).

## 8 Summary

Ungulates are a diverse taxonomic group that can be found in some of the most extreme environments. For most of human history, people have relied on ungulates for food, clothing, and transportation, with more contemporary uses including sport hunting and ecotourism. In the late twentieth century, ungulate management programs began to incorporate genetic data; with technological advancements and lowering costs of genome sequencing, it has become possible to address more detailed questions about the evolution, conservation, and management of ungulates, both in captivity and the wild. The first ungulate genomes sequenced were those of domestic animals due to economic importance. Wild ungulate genomes have started to emerge, and in many cases rely on domestic genome assemblies. Genomic data allows for reconstruction of the evolutionary and population histories from a few contemporary genomes while providing the tools to uncover the genes underlying adaptive divergence. In particular, both natural selection and artificial selection have left distinct signatures on ungulate genomes, and there is a wide array of candidate genes identified. For conservation and management, genomic data provides more accurate estimates of inbreeding, allows for the reconstruction of pedigrees in wild populations, and facilitates the design of breeding and management programs to retain genetic variation, based on markers distributed at high density across the genome. As more ungulate genomes are sequenced and the existing assemblies are improved, our understanding of population processes and the genetic basis of key traits will steadily improve, impacting regional economies, and the conservation and management of this important group.

## References

Agaba M, Ishengoma E, Miller WC, McGrath BC, Hudson CN, Bedoya-Reina OC, et al. Giraffe genome sequence reveals clues to its unique morphology and physiology. Nat Commun. 2016;7:11519.

Alexander TW, Plaizier JC. The importance of microbiota in ruminant production. Anim Front. 2016;6(2):4–7.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17(2):81–92.

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM. Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. Mol Ecol. 2016;25 (13):3152–68.

Bosse M, Megens H-J, Madsen O, Crooijmans RPMA, Ryder OA, Austerlitz F, et al. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Res. 2015;25(7):970–81.

Brauning R, Fisher PJ, McCulloch AF, Smithies RJ, Ward JF, Bixley MJ, et al. Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk. bioRxiv. 2015;23:27318.

Brown RD. The history of wildlife conservation in North America. In: Krausman PR, Cain JW, editors. Wildlife management and conservation: contemporary principles and practices. Baltimore, MD: Johns Hopkins University Press; 2013. p. 6–23.
Buczek M, Okarma H, Demiaszkiewicz AW, Radwan J. MHC, parasites and antler development in red deer: no support for the Hamilton & Zuk hypothesis. J Evol Biol. 2016;29(3):617–32.
Budowle B, Beaudry JA, Barnaby NG, Giusti AM, Bannan JD, Keim P. Role of law enforcement response and microbial forensics in investigation of bioterrorism. Croat Med J. 2007;48 (4):437–49.
Burger PA. The history of old world camelids in the light of molecular genetics. Tropl Anim Health Prod. 2016;48:905–13.
Carlson SM, Cunningham CJ, Westley PAH. Evolutionary rescue in a changing world. Trends Ecol Evol. 2014;29(9):521–30.
Chen X, Zhang B, Yue H, Wang Y, Zhou F, Zhang Q, et al. A novel astrovirus species in the gut of yaks with diarrhoea in the Qinghai–Tibetan Plateau, 2013. J Gen Virol. 2015;96(12):3672–80.
Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, et al. Revealing the history of sheep domestication using retrovirus integrations. Science. 2009;324(5926):532–6.
Coltman DW, O'Donoghue P, Jorgenson JT, Hogg JT, Strobeck C, Festa-Bianchet M. Undesirable evolutionary consequences of trophy hunting. Nature. 2003;426(6967):655–8.
Corti P, Shafer ABA, Coltman DW, Festa-Bianchet M. Past bottlenecks and current population fragmentation of endangered huemul deer (Hippocamelus bisulcus): implications for preservation of genetic diversity. Conserv Genet. 2011;12(1):119–28.
Cosart T. Evaluation of a new method for large-scale and gene-targeted next generation DNA sequencing in nonmodel species. Graduate Student Theses, Dissertations, & Professional Papers. 2013.
Cosart T, Beja-Pereira A, Chen S, Ng SB, Shendure J, Luikart G. Exome-wide DNA capture and next generation sequencing in domestic and wild species. BMC Genomics. 2011;12:347.
De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS One. 2013;8(5).
Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol. 2015;25 (19):2577–83.
Ditchkoff SS, Lochmiller RL, Masters RE, Hoofer SR, Van Den Bussche RA. Major-histocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (odocoileus virginianus): evidence for good-genes advertisement. Evolution. 2001;55(3):616–25.
Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotechnol. 2013;31 (2):135–41.
Dong Y, Zhang X, Xie M, Arefnezhad B, Wang Z, Wang W, et al. Reference genome of wild goat (capra aegagrus) and sequencing of goat breeds provide insight into genic basis of goat domestication. BMC Genomics. 2015;16:431.
Dou H, Jiang G, Stott P, Piao R. Climate change impacts population dynamics and distribution shift of moose (Alces alces) in Heilongjiang Province of China. Ecol Res. 2013;28(4):625–32.
Dupuy V, Verdier A, Thiaucourt F, Manso-Silván L. A large-scale genomic approach affords unprecedented resolution for the molecular epidemiology and evolutionary history of contagious caprine pleuropneumonia. Vet Res. 2015;46:74.
Eggert LS, Powell DM, Ballou JD, Malo AF, Turner A, Kumer J, et al. Pedigrees and the study of the wild horse population of Assateague Island National Seashore. J Wildl Manag. 2010;74 (5):963–73.
Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl. 2014;7(9):1026–42.
Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol. 2014;29(1):51–63.

Elsik CG, Tellam RL, Worley KC, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science. 2009;324(5926):522–8.
Fernández BJ, Toro MA. The use of mathematical programming to control inbreeding in selection schemes. J Anim Breed Genet. 1999;116(6):447–66.
Fitak RR, Mohandesan E, Corander J, Burger PA. The de novo genome assembly and annotation of a female domestic dromedary of North African origin. Mol Ecol Resour. 2016;16(1):314–24.
Flanagan SP, Forester BR, Latch EK, Aitken S, Hoban S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol App. 2017:1–18.
Forde TL, Orsel K, Zadoks RN, Biek R, Adams LG, Checkley SL, et al. Bacterial genomics reveal the complex epidemiology of an emerging pathogen in arctic and boreal ungulates. Front Microbiol. 2016;7:1759.
Gallana M, Ryser-Degiorgis M-P, Wahli T, Segner H. Climate change and infectious diseases of wildlife: altered interactions between pathogens, vectors and hosts. Curr Zool. 2013;59 (3):427–37.
Gedir JV, Cain JW, Harris G, Turnbull TT. Effects of climate change on long-term population growth of pronghorn in an arid environment. Ecosphere. 2015;6(10):1–20.
Genomic Resources Development Consortium, Bensch S, Coltman DW, Davis CS, Hellgren O, Johansson T, et al. Genomic resources notes accepted 1 June 2013–31 July 2013. Mol Ecol Resour. 2014;14(1):218.
Giglio RM, Ivy JA, Jones LC, Latch EK. Evaluation of alternative management strategies for maintenance of genetic variation in wildlife populations. Anim Conserv. 2016;19(4):380–90.
Giglio RM, Ivy JA, Jones LC, Latch EK. Pedigree-based genetic management improves bison conservation. J Wildl Manag. 2018;82:766–74.
Glanzmann B, Möller M, le Roex N, Tromp G, Hoal EG, van Helden PD. The complete genome sequence of the African buffalo (Syncerus caffer). BMC Genomics. 2016;17:1001.
Gordon IJ, Hester AJ, Festa-Bianchet M. The management of wild large herbivores to meet economic, conservation and environmental objectives. J Appl Ecol. 2004;41(6):1021–31.
Gorkhali NA, Dong K, Yang M, Song S, Kader A, Shrestha BS, et al. Genomic analysis identified a potential novel molecular mechanism for high-altitude adaptation in sheep at the Himalayas. Sci Rep. 2016;6:29963.
Groenen MAM. A decade of pig genome sequencing: a window on pig domestication and evolution. Genet Sel Evol. 2016;48:23.
Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491 (7424):393–8.
Groves C, Grubb P. Ungulate taxonomy. Baltimore, MD: Johns Hopkins University Press; 2011.
Haynes GD, Latch EK. Identification of Novel Single Nucleotide Polymorphisms (SNPs) in Deer (Odocoileus spp.) using the BovineSNP50 BeadChip. PLoS One. 2012;7(5):e36536.
Hedrick PW, Parker KM, Gutiérrez-Espeleta GA, Rattink A, Lievers K. Major histocompatibility complex variation in the arabian oryx. Evolution. 2000;54(6):2145–51.
Hengeveld PE, Festa-Bianchet M. Harvest regulations and artificial selection on horn size in male bighorn sheep. J Wildl Manag. 2011;75(1):189–97.
Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188(4):379–97.
Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007;39(12):1522–7.
Hogg JT, Forbes SH, Steele BM, Luikart G. Genetic rescue of an insular population of large mammals. Proc Biol Sci. 2006;273(1593):1491–9.
Hu X, Liu G, Shafer ABA, Wei Y, Zhou J, Lin S, Wu H, Zhou M, Hu D, Liu S. Comparative analysis of the gut microbial communities in forest and alpine musk deer using high-throughput sequencing. Front Microbiol. 2017;8:572.

Huang J, Zhao Y, Shiraigol W, Li B, Bai D, Weixing Y, et al. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. Sci Rep. 2014;4:4958.
Huang J, Zhao Y, Bai D, Shiraigol W, Li B, Yang L, et al. Donkey genome and insight into the imprinting of fast karyotype evolution. Sci Rep. 2015;5:14106.
Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM. Inbreeding depression across the lifespan in a wild mammal population. Proc Natl Acad Sci. 2016;113(13):3585–90.
Iacolina L, Scandura M, Goedbloed DJ, Alexandri P, Crooijmans RPMA, Larson G, Archibald A, Apollonio M, Schook LB, Groenen MAM, Megens HJ. Genomic diversity and differentiation of a managed island wild boar population. Heredity. 2016;116(1):60–7.
IUCN 2017. The IUCN red list of threatened species. Version 2017-3. http://www.iucnredlist.org. Downloaded on 05 December 2017.
Ivy JA, Putnam AS, Navarro AY, Gurr J, Ryder OA. Applying SNP-derived molecular coancestry estimates to captive breeding programs. J Hered. 2016;107(5):403–12.
Jäger M, Ott C-E, Grünhagen J, Hecht J, Schell H, Mundlos S, et al. Composite transcriptome assembly of RNA-seq data in a sheep model for delayed bone healing. BMC Genomics. 2011;12:158.
Janeway CA Jr. How the immune system protects the host from infection. Microbes Infect. 2001;3:1167–71.
Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344(6188):1168–73.
Jirimutu, Wang Z, Ding G, Chen G, Sun Y, Bactrian Camels Genome Sequencing and Analysis Consortium, et al. Genome sequences of wild and domestic bactrian camels. Nat Commun. 2012;3:1202.
Johnston SE, JC MEWAN, Pickering NK, Kijas JW, Beraldi D, Pilkington JG, et al. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol Ecol. 2011;20(12):2555–66.
Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, et al. Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature. 2013;502 (7469):93–5.
Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25 (1):185–202.
Jones AG, Small CM, Paczolt KA, Ratterman NL. A practical guide to methods of parentage analysis. Mol Ecol Resour. 2010;10(1):6–30.
Kardos M, Luikart G, Bunch R, Dewey S, Edwards W, McWilliam S, et al. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol Ecol. 2015;24(22):5616–32.
Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9(10):1205–18.
Kim E-S, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity. 2016;116(3):255–64.
Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. Genomics and conservation genetics. Trends Ecol Evol. 2006;21(11):629–37.
Kutz SJ, Hoberg EP, Nagy J, Polley L, Elkin B. "Emerging" parasitic infections in arctic ungulates. Integr Comp Biol. 2004;44(2):109–18.
Lacy R. Clarification of genetic terms and their use in the management of captive populations. Zoo Biol. 1995;14:565–78.
Lane-deGraaf KE, Amish SJ, Gardipee F, Jolles A, Luikart G, Ezenwa VO. Signatures of natural and unnatural selection: evidence from an immune system gene in African buffalo. Conserv Genet. 2015;16(2):289–300.
Leopold A. Game management. Madison: University of Wisconsin Press; 1987.
Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6.

Librado P, Sarkissian CD, Ermini L, Schubert M, Jónsson H, Albrechtsen A, et al. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. Proc Natl Acad Sci. 2015;112(50):E6889–97.
Liu S, Hansen MM. PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. Mol Ecol Resour. 2017;4:631–41.
Lukefahr SD, Jacobson HA. Variance component analysis and heritability of antler traits in white-tailed deer. J Wildl Manag. 1998;62(1):262–8.
Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, Perry GH, et al. Elephantid genomes reveal the molecular bases of Woolly Mammoth adaptations to the arctic. Cell Rep. 2015;12(2):217–28.
Maclachlan NJ, Mayo CE. Potential strategies for control of bluetongue, a globally emerging, Culicoides-transmitted viral disease of ruminant livestock and wildlife. Antiviral Res. 2013;99 (2):79–90.
Magistrelli D, Zanchi R, Malagutti L, Galassi G, Canzi E, Rosi F. Effects of cocoa husk feeding on the composition of swine intestinal microbiota. J Agric Food Chem. 2016;64(10):2046–52.
Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. 2015;3:31.
Marfell BJ, O'Brien R, Griffin JFT. Global gene expression profiling of monocyte-derived macrophages from red deer (Cervus elaphus) genotypically resistant or susceptible to Mycobacterium avium subspecies paratuberculosis infection. Dev Comp Immunol. 2013;40(2):210–7.
Martin C, Pastoret P-P, Brochier B, Humblet M-F, Saegerman C. A survey of the transmission of infectious diseases/infections between wild and domestic ungulates in Europe. Vet Res. 2011;42:70.
Maslo B, Fefferman NH. A case study of bats and white-nose syndrome demonstrating how to model population viability with evolutionary effects. Conserv Biol. 2015;29(4):1176–85.
McIntyre MK, Peacock TJ, Akers KS, Burmeister DM. Initial characterization of the pig skin bacteriome and its effect on in vitro models of wound healing. PLoS One. 2016;11(11): e0166176.
McLoughlin KE, Nalpas NC, Rue-Albrecht K, Browne JA, Magee DA, Killick KE, et al. RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with Mycobacterium bovis. Front Immunol. 2014;5:396.
Michel ES, Demarais S, Strickland BK, Smith T, Dacus CM. Antler characteristics are highly heritable but influenced by maternal factors. J Wildl Manag. 2016;80(8):1420–6.
Miller JM, Poissant J, Kijas JW, the International Sheep Genomics Consortium, Coltman DW. A genome wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Mol Ecol Resour. 2011;11:314–22.
Miller JM, Poissant J, Hogg JT, Coltman DW. Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). Mol Ecol. 2012;21(7):1583–96.
Miller JM, Moore SS, Stothard P, Liao X, Coltman DW. Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (Ovis canadensis). BMC Genomics. 2015;16:397.
Mitra N, Cernicchiaro N, Torres S, Li F, Hause BM. Metagenomic characterization of the virome associated with bovine respiratory disease in feedlot cattle identified novel viruses and suggests an etiologic role for influenza D virus. J Gen Virol. 2016;97(8):1771–84.
Molnár J, Nagy T, Stéger V, Tóth G, Marincs F, Barta E. Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. BMC Genomics. 2014;15:761.
Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. Mol Ecol. 2013;22(11):2841–7.
Nater A, Burri R, Kawakami T, Smeds L, Ellegren H. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Syst Biol. 2015;64(6):1000–17.
National Research Council, Division on Earth and Life Studies, Board on Agriculture and Natural Resources, Board on Life Sciences, Committee on Defining Science-Based Concerns Associated with Products of Animal Biotechnology, Committee on Agricultural Biotechnology,

Health, and the Environment. Animal biotechnology: science based concerns. Washington, DC: National Academies Press; 2002.
Ogden R, Baird J, Senn H, McEwing R. The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar-horned oryx. Conserv Genet Resour. 2012;4(2):471–3.
Olson ZH, Whittaker DG, Rhodes OE. Evaluation of experimental genetic management in reintroduced bighorn sheep. Ecol Evol. 2012;2(2):429–43.
Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early middle Pleistocene horse. Nature. 2013;499 (7456):74–8.
Paterson S, Wilson K, Pemberton JM. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (Ovis aries L.). Proc Natl Acad Sci. 1998;95(7):3714–9.
Pigeon G, Festa-Bianchet M, Coltman DW, Pelletier F. Intense selective hunting leads to artificial evolution in horn size. Evol Appl. 2016;9(4):521–30.
Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW. Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. Proc R Soc Lond B Biol Sci. 2008;275(1635):623–8.
Poissant J, Hogg JT, Davis CS, Miller JM, Maddox JF, Coltman DW. Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. BMC Genomics. 2010;11:524.
Post E, Pedersen C, Wilmers CC, Forchhammer MC. Warming, plant phenology and the spatial dimension of trophic mismatch for large herbivores. Proc R Soc Lond B Biol Sci. 2008;275 (1646):2005–13.
Powell JH, Amish SJ, Haynes GD, Luikart G, Latch EK. Candidate adaptive genes associated with lineage divergence: identifying SNPs via next-generation targeted resequencing in mule deer (Odocoileus hemionus). Mol Ecol Resour. 2016;16(5):1165–72.
Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, et al. The yak genome and adaptation to life at high altitude. Nat Genet. 2012;44(8):946–9.
Quéméré E, Galan M, Cosson J-F, Klein F, Aulagnier S, Gilot-Fromont E, et al. Immunogenetic heterogeneity in a widespread ungulate: the European roe deer (Capreolus capreolus). Mol Ecol. 2015;24(15):3873–87.
Rhodes OE Jr, Latch EK. Wildlife reintroductions: conceptual development and application of theory. Pgs. 296–319. In: DeWoody JA, et al., editors. Molecular approaches to natural resource conservation. New York: University of Cambridge Press; 2010. p. 374.
Ripple WJ, Newsome TM, Wolf C, Dirzo R, Everatt KT, Galetti M, et al. Collapse of the world's largest herbivores. Sci Adv. 2015;1(4):1–12.
Roffler GH, Amish SJ, Smith S, Cosart T, Kardos M, Schwartz MK, et al. SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. Mol Ecol Resour. 2016a;16(5):1147–64.
Roffler GH, Amish SJ, Smith S, Cosart T, Kardos M, Schwartz MK, et al. SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. Mol Ecol Resour. 2016b;16(5):1147–64.
Roth A, Lin J, Magnius L, Karlsson M, Belák S, Widén F, et al. Markers for ongoing or previous hepatitis E virus infection are as common in wild ungulates as in humans in Sweden. Viruses. 2016;8(9):259.
Salmona J, Heller R, Lascoux M, Shafer ABA. Inferring population and demographic history with genomic data. In: Rajora OP, editor. Population genomics: population genomics concepts, approaches and applications. New York: Springer; 2017. https://doi.org/10.1007/13836_2017_1.
Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol Ecol. 2010;19(7):1439–51.

Schaschl H, Wandeler P, Suchentrunk F, Obexer-Ruff G, Goodman SJ. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. Heredity. 2006;97 (6):427–37.
Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014;46(8):919–25.
Schirtzinger EE, Suddith AW, Hause BM, Hesse RA. First identification of porcine parvovirus 6 in North America by viral metagenomic sequencing of serum from pigs infected with porcine reproductive and respiratory syndrome virus. Virol J. 2015;12:170.
Schubert M, Jónsson H, Chang D, Sarkissian CD, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci. 2014;111 (52):E5661–9.
Schwartz MK, Luikart G, Waples RS. Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol. 2007;22(1):25–33.
Scornavacca C, Galtier N. Incomplete lineage sorting in mammalian phylogenomics. Syst Biol. 2017;66:112–20.
Scribner KT. Conservation genetics of managed ungulate populations. Acta Theriol. 1993;38(Suppl 2):89–101.
Senn H, O'Donoghue P, McEwing R, Ogden R. Hundreds of SNPs for the endangered pygmy hippopotamus (Choeropsis liberiensis). Conserv Genet Resour. 2014;6(3):535–8.
Shafer ABA, Hall JC. Placing the mountain goat: a total evidence approach to testing alternative hypotheses. Mol Phylogenet Evol. 2010;55(1):18–25.
Shafer ABA, Fan CW, Côté SD, Coltman DW. (Lack of) genetic diversity in immune genes predates glacial isolation in the north American Mountain goat (Oreamnos americanus). J Hered. 2012;103(3):371–9.
Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30(2):78–87.
Shafer ABA, Miller JM, Kardos M. Cross-species application of SNP chips is not suitable for identifying runs of Homozygosity. J Hered. 2016;107(2):193–5.
Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, et al. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature. 1997;388 (6643):666–70.
Sim Z, Hall JC, Jex B, Hegel TM, Coltman DW. Genome-wide set of SNPs reveals evidence for two glacial refugia and admixture from postglacial recolonization in an alpine ungulate. Mol Ecol. 2016;25(15):3696–705.
Skare Ø, Sheehan N, Egeland T. Identification of distant family relationships. Bioinformatics. 2009;25(18):2376–82.
Slate J, Stijn TCV, Anderson RM, McEwan KM, Maqbool NJ, Mathias HC, et al. A deer (subfamily Cervinae) genetic linkage map and the evolution of ruminant genomes. Genetics. 2002;160 (4):1587–97.
Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM. Gene mapping in the wild with SNPs: guidelines and future directions. Genetica. 2009;136(1):97–107.
Sonesson AK, Meuwissen TH. Minimization of rate of inbreeding for small populations with overlapping generations. Genet Res. 2001;77(3):285–92.
Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, et al. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). BMC Genomics. 2016;17:122.
Truett JC, Phillips M, Kunkel K, Miller R. Managing bison to restore biodiversity. Gt Plains Res. 2001;11(1):123–44.
Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods. 2008;5(3):247–52.
Vors LS, Boyce MS. Global declines of caribou and reindeer. Glob Chang Biol. 2009;15 (11):2626–33.

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science. 2009;326 (5954):865–7.

Whitehead A. Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. J Exp Biol. 2012;215(6):884–91.

Wiedmann RT, Smith TP, Nonneman DJ. SNP discovery in swine by reduced representation and high throughput pyrosequencing. BMC Genet. 2008;9:81.

Williams JD, Krueger W, Harmel DH. Heritabilities for antler characteristics and body weight in yearling white-tailed deer. Heredity. 1994;73(1):78–83.

Winternitz JC, Minchey SG, Garamszegi LZ, Huang S, Stephens PR, Altizer S. Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. Proc R Soc Lond B Biol Sci. 2013;280(1769):20131605.

Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, et al. Camelid genomes reveal evolution and adaptation to desert environments. Nat Commun. 2014;5:5188.

Yang J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33 (10):2576–92.

Yao B, Zhao Y, Wang K, Zhang M, Liu M, Liu H, et al. De novo characterization of the antler tip of Chinese Sika deer transcriptome and analysis of gene expression related to rapid growth. Mol Cell Biochem. 2012a;364(1–2):93–100.

Yao B, Zhao Y, Zhang H, Zhang M, Liu M, Liu H, et al. Sequencing and de novo analysis of the Chinese Sika deer antler-tip transcriptome during the ossification stage using Illumina RNA-Seq technology. Biotechnol Lett. 2012b;34(5):813–22.

Yoon SS, Kim E-K, Lee W-J. Functional genomic and metagenomic approaches to understanding gut microbiota-animal mutualism. Curr Opin Microbiol. 2015;24:38–46.

# Advancing Understanding of Amphibian Evolution, Ecology, Behavior, and Conservation with Massively Parallel Sequencing


**W. Chris Funk, Kelly R. Zamudio, and Andrew J. Crawford**

**Abstract** Genomics has great potential to advance understanding of amphibian evolution, ecology, and behavior, as well as to improve conservation of this highly imperiled class of vertebrates. However, application of new massively parallel sequencing technology to amphibians lags behind its application to other vertebrates, due in part to their large, repetitive genomes, making genome assembly challenging. The goal of our chapter is to outline ways in which population genomics – coupled with field biology, experiments, and modeling – can deepen our understanding of basic and applied questions in amphibian evolutionary ecology and conservation. We start by discussing potential applications of genomics to several long-standing questions in amphibian evolution, ecology, and behavior, including phylogenetic relationships, phylogeography, sex chromosome evolution, population structure and demography, local adaptation, and mating systems and sexual selection. We then highlight opportunities for improving amphibian conservation with genomics, focusing on hybridization, disease evolution and ecology, and captive breeding programs. Next, we provide strategies for moving amphibian genomics forward in the face of challenges such as few available reference genomes and large repetitive genomes, including a bold proposal for whole genome sequencing of a minimum of one species per amphibian family. We conclude by providing suggestions for maximizing the potential of genomics to advance understanding of amphibian evolutionary ecology and conservation and recommendations for getting started in genomics.

**Keywords** Amphibian · Local adaptation · Massively parallel sequencing · Population genomics

W. C. Funk (✉)
Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA
e-mail: Chris.Funk@colostate.edu

K. R. Zamudio
Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

A. J. Crawford
Department of Biological Sciences, Universidad de Los Andes, Bogotá, Colombia

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_61,

## 1 Introduction

Massively parallel sequencing (MPS) has enormous yet largely untapped potential to advance understanding of the evolution, ecology, behavior, and conservation of amphibians, including frogs (Anura), salamanders (Caudata), and caecilians (Gymnophiona). Advances in sequencing technology and computational power have enabled the genomics era, in which vast quantities of DNA or RNA can be sequenced relatively quickly and cheaply to address questions in new, creative, and more powerful ways than was possible with traditional dye-termination sequencing technology like Sanger sequencing (Rokas and Abbot 2009; Allendorf et al. 2010). MPS platforms, such as Illumina's sequencing by synthesis technology, PacBio's single molecule real-time (SMRT) sequencing, or the Oxford Nanopore PromethION system, generate hundreds of millions of sequence reads – typically in the range of 100 to 10,000+ base pairs (bp) – from DNA or RNA (Glenn 2011). Depending on which library preparation protocol is used, the researcher can sequence the entire genome or target specific regions or loci (Andrews et al. 2016; Jones and Good 2016). By ligating unique individual barcodes to DNA libraries, it is possible to multiplex (pool) dozens to hundreds of individuals in a single MPS run, increasing efficiency and reducing costs (Baird et al. 2008). This flexibility in MPS has resulted in a plethora of different genomic techniques suited for addressing a wide variety of questions (Table 1).

Analysis of MPS data using population genomics – a subfield of genomics – is a particularly powerful framework for answering many open questions in amphibian evolution, ecology, behavior, and conservation. Population genomics is defined as the study of numerous loci (hundreds of genes to millions of polymorphisms) to understand the microevolutionary processes (mutation, genetic drift, gene flow, and selection) that influence genetic variation within and among populations (Black et al. 2001; Luikart et al. 2003). The advantage of population genomics over traditional population genetics, which involves the use of fewer loci (typically 1–20), is twofold. First, population genomics allows identification and analysis of loci under natural selection, providing a new window into patterns and genetic mechanisms of adaptation (Beaumont and Nichols 1996; Beaumont and Balding 2004; Foll and Gaggiotti 2008). Second, by allowing the identification and filtering of genetic variation under selection, population genomics allows unbiased inference of evolutionary history via accurate estimation of genome-wide, neutral demographic parameters such as effective population size ($N_e$) and gene flow ($N_em$), without the confounding effects of natural selection (Luikart et al. 2003).

Amphibians have many characteristics that make them an excellent taxonomic group for genomic studies. First, unlike mammals, most amphibians have nucleated red blood cells, allowing extraction of large quantities of high-quality DNA. Second, many amphibian species are abundant (Burton and Likens 1975) and often congregate in dense breeding associations, providing sufficient sample sizes for population genomic analyses. Third, most species of amphibians have exposed eggs that allow

**Table 1** Massively parallel sequencing (MPS) methods used commonly for addressing questions in evolution, ecology, behavior, and conservation

| Method name | Brief description | Amphibian applications | Amphibian examples |
|---|---|---|---|
| Whole-genome sequencing (WGS), assembly, and annotation (Pritchard 2011) | Sequencing of nearly all base pairs in the genome. Can range from reference-standard genome to sequencing at low depth (genome skimming) | Local adaptation (to map outliers and infer function); historical demography; phylogenomics (to identify independent markers); chromosome evolution | *Xenopus tropicalis* (Hellsten et al. 2010); *Nanorana parkeri* (Sun et al. 2015); *Xenopus laevis* (Session et al. 2016); *Rana catesbeiana* (Hammond et al. 2017) |
| Whole-genome resequencing (WGR; Fuentes-Pardo and Ruzzante 2017) | Short-read MPS sequencing nearly all base pairs in the genome from multiple individuals and assembling reads to a high-quality reference genome | Historical demography, speciation, hybridization, genomic bases of adaptation, speciation, and introgression | *Nanorana parkeri* (Wang et al. 2018) |
| Restriction site-associated DNA sequencing (RADseq; Andrews et al. 2016) | A suite of reduced representation methods that sequence and genotype loci adjacent to restriction sites | Population structure; test for signatures of local adaptation; markers for pedigrees; characterization of hybridization | *Rhinella marina* (Trumbo et al. 2016); *Bufo andrewsi* (Guo et al. 2016); *Ambystoma talpoideum*, *A. opacum* (Nunziata et al. 2017) |
| Targeted capture (Jones and Good 2016) | A reduced representation method that enriches for targeted regions of the genome using labeled oligonucleotides | Population structure; test for signatures of local adaptation; markers for pedigrees; characterization of hybridization; targeting functional genes; phylogenomics | Anura (Portik et al. 2016); *Ambystoma californiense*, *A. mavortium* (McCartney-Melstad et al. 2016) |
| Ultraconserved elements (UCE; Faircloth et al. 2012) | Highly conserved regions of the genome that can be used to generate sequence data at orthologous loci from evolutionarily distant taxa | Phylogenomics; phylogeography | *Kaloula* spp. (Alexander et al. 2017); *Plethodon serratus* (Newman and Austin 2016) |
| Anchored phylogenomics (Lemmon et al. 2012) | Sequencing and genotyping of libraries enriched for conserved, anonymous, and/or functional loci | Phylogenomics; phylogeography | Microhylidae (Peloso et al. 2016); Hylidae, Bufonidae, Ranidae (Barrow et al. 2018); Terraranae (Heinicke et al. 2018) |

(continued)

**Table 1** (continued)

| Method name | Brief description | Amphibian applications | Amphibian examples |
|---|---|---|---|
| Transcriptomics (e.g., RNAseq; Wang et al. 2009) | Analysis of gene expression levels, usually conducted by sequencing cDNA from RNA (RNAseq); obtaining complete coding sequence of expressed genes | Local adaptation (e.g., gene expression in different populations); disease ecology (e.g., gene expression with and without pathogen); obtaining candidate genes (e.g., sex determination gene ID); scans for increased rates of non-synonymous substitutions | *Bufo viridis* (Gerchen et al. 2016); *Lithobates clamitans*, *Pseudacris regilla* (Robertson and Cornman 2014); *Andrias davidianus* (Che et al. 2014); *Rana chensinensis*, *R. kukunoris* (Yang et al. 2012) |
| Metabarcoding (Caporaso et al. 2011) | MPS of DNA barcode genes for species delineation | Characterization of amphibian skin or gut microbial diversity | *Lithobates pipiens*, *Pseudacris maculata*, *Ambystoma tigrinum* (McKenzie et al. 2012) |
| Environmental DNA (eDNA) metabarcoding (Taberlet et al. 2012) | MPS of DNA barcode from an environmental sample (e.g., skin sloughed off in water) | Monitoring of amphibian diversity in streams or ponds | *Bufo bufo*, *B. calamita*, *Hyla meridionalis*, *Pelobates cultripes*, *Pelodytes punctatus*, *Pelophylax* sp., *Rana dalmatina*, *Lissotriton helveticus*, *Salamandra salamandra*, *Triturus marmoratus* (Valentini et al. 2016); *Hylodes phyllodes*, *H. asper*, *Cycloramphus boraceiensis*, *Thoropa taophora*, *Vitreorana uranoscopa*, *Scinax trapicheiroi*, *Bokermannohyla* sp. aff. *circumdata*, *Bokermannohyla*, *Aplastodiscus eugenioi*, *Phasmahyla cruzi* (Lopes et al. 2017) |

For each MPS method, we provide its name (and reference for method), a brief description, most appropriate uses for amphibian research, and examples of its use in amphibians

for application of gene-editing techniques such as CRISPR/Cas for gain or loss of function studies (Fei et al. 2014; Bhattacharya et al. 2015; Elewa et al. 2017).

At the same time, amphibians have posed some significant challenges for genomic studies. The main challenge is that because of the large size of most amphibian genomes, with median sizes of 4.1 gigabases (Gb) for frogs, 5.6 Gb for caecilians, and 32 Gb for salamanders (Gregory 2011; Liedtke et al. 2018), fewer amphibian genomes have been sequenced relative to other vertebrates (Table 2). Compounding the problem of large genomes is the highly repetitive structure of many amphibian genomes, making genome assembly computationally challenging. Having a reference genome available for aligning reads or contigs improves the accuracy of genotyping and facilitates determining the potential function of loci (Manel et al. 2016; Toews et al. 2016). A third challenge of population genomic studies of amphibians is that, since amphibians often have low gene flow and high population structure (Crawford 2003; Zeisset and Beebee 2008), it can be more challenging to identify loci with a genetic signature of divergent selection using genome scans (Francois et al. 2016). Nonetheless, these challenges of studying amphibian genomics – large genomes with repetitive elements and high population structure – are also some of the reasons that they are interesting taxa for genomic studies.

Although MPS and genomic analyses are powerful new tools for understanding amphibian biology, we argue that the best research is integrative, combining new genomic technology with tried-and-true approaches such as field observations, controlled experiments, and modeling. We (WCF, KRZ, and AJC) are field biologists in addition to evolutionary geneticists/genomicists. We are not proposing that genomics will replace classic approaches for studying amphibian biology. Rather, genomics will expand what is possible to know about these fascinating organisms by allowing us to characterize genetic variation across a much larger proportion of the genome to understand evolutionary and ecological processes more deeply than previously imaginable.

The goal of our chapter is to provide an overview of potential applications of MPS and genomics to advance studies of amphibian evolution, ecology, behavior, and conservation. A handful of recent reviews on amphibian genetics and genomics have been published, but none focuses exclusively on genomics and all focus on amphibian conservation (Storfer et al. 2009; McCartney-Melstad and Shaffer 2015; Shaffer et al. 2015). Here, we restrict our discussion to the potential of MPS and genomics to advance understanding of basic questions in evolution, ecology, and behavior, as well as applied questions in amphibian conservation and management. We first highlight opportunities for applying genomics to basic questions about amphibian evolutionary ecology, particularly in the areas of phylogenomics, phylogeography, chromosome evolution, population structure and demography, local adaptation, and mating systems. We then highlight opportunities for improving amphibian conservation with genomics, focusing on characterizing hybridization between invasive and native amphibian species, understanding disease dynamics, and designing captive breeding programs. Next, we discuss challenges of genomic studies of amphibians in more depth and provide possible solutions. Finally, we conclude by discussing how genomics can best be harnessed to advance

**Table 2** List of target amphibian families for whole genome sequencing (WGS), with one suggested species for each taxonomic family

| Order | Family | Genus | Species | Motive | Progress |
|---|---|---|---|---|---|
| Anura | Allophrynidae | *Allophryne* | *ruthveni* | Sister to Centrolenidae | None |
| Anura | Alsodidae | *Eupsophus* | *calcaratus* | PD | None |
| Anura | Alytidae | *Alytes* | *obstetricans* | Males carry fertilized eggs | Pending |
| Anura | Aromobatidae | *Rheobates* | *palmatus* | Adaptation to elevation | Pending |
| Anura | Arthroleptidae | *Arthroleptis* | *poecilonotus* | Direct development; long finger | Pending |
| Anura | Ascaphidae | *Ascaphus* | *truei* | Co-sister to all other anurans | In progress |
| Anura | Batrachylidae | *Batrachyla* | *leptopus* | PD | None |
| Anura | Bombinatoridae | *Bombina* | *variegata* | Hybrid zones | In progress |
| Anura | Brachycephalidae | *Brachycephalus* | *didactylus* | Very small tetrapod | None |
| Anura | Brevicipitidae | *Breviceps* | *macrops* | Direct development. Adapted to sand | None |
| Anura | Bufonidae | *Atelopus* | *zeteki* | Conservation | In progress |
| Anura | Bufonidae | *Rhinella* | *marina* | Invasive species | Published (Edwards et al. 2018) |
| Anura | Calyptocephalellidae | *Calyptocephalella* | *gayi* | PD | Pending |
| Anura | Centrolenidae | *Hyalinobatrachium* | *fleischmanni* | Parental care | In progress |
| Anura | Ceratobatrachidae | *Platymantis* | *montanus* | Direct development; speciation | None |
| Anura | Ceratophryidae | *Lepidobatrachus* | *laevis* | Small genome | In progress |
| Anura | Conrauidae | *Conraua* | *goliath* | World's biggest frog. Tadpole diet specialization | None |
| Anura | Craugastoridae | *Craugastor* | *fitzingeri* | Resistant to disease | RNAseq data |
| Anura | Cycloramphidae | *Cycloramphus* | *boraceiensis* | PD | None |
| Anura | Dendrobatidae | *Oophaga* | *pumilio* | Rapid color evolution | Published (Rogers et al. 2018) |
| Anura | Dicroglossidae (Dicroglossinae) | *Nanarana* | *parkeri* | Adaptation to high elevation | Published (Sun et al. 2015) |
| Anura | Eleutherodactylidae | *Eleutherodactylus* | *coqui* | Model of evodevo in frogs | In progress |

| | | | | | |
|---|---|---|---|---|---|
| Anura | *Geobatrachus* sedis incertis | *Geobatrachus* | *walkeri* | PD. Direct development | None |
| Anura | Heleophrynidae | *Hadromophryne* | *natalensis* | PD | None |
| Anura | Hemiphractidae | *Gastrotheca* | *cornuta* | Frog "placenta." Conservation | Pending |
| Anura | Hemisotidae | *Hemisus* | *marmoratus* | PD | None |
| Anura | Hylidae | *Hyla* | *arborea* | Well studied | In progress |
| Anura | Hylidae | *Pseudacris* | *feriarum* | Hybrid zones | In progress |
| Anura | Hylidae | *Dendropsophus* | *ebraccatus* | Phenotypic plasticity | In progress |
| Anura | Hylodidae | *Hylodes* | *japi* | PD | None |
| Anura | Hyperoliidae | *Hyperolius* | *riggenbachi* | Color pattern polymorphism | None |
| Anura | Leiopelmatidae | *Leiopelma* | *hochstetteri* | Co-sister to all other anurans | Pending |
| Anura | Leptodactylidae (Leiuperinae) | *Physalaemus* | *pustulosus* | Mate choice, sexual selection | In progress |
| Anura | Leptodactylidae (Leptodactylinae) | *Leptodactylus* | *pentadactylus* | Sex chromosome evolution | Pending. RNAseq data |
| Anura | Limnodynastidae | *Limnodynastes* | *dumerilii* | PD | In progress |
| Anura | Mantellidae | *Mantella* | *aurantiaca* | Convergent evolution. Breeds in captivity | In progress |
| Anura | Megophryidae | *Scutiger* | *boulengeri* | Highest elevation frog | RNAseq data |
| Anura | Micrixalidae | *Micrixalus* | *kottigeharensis* | PD | None |
| Anura | Microhylidae | *Microhyla* | *fissipes* | Thyroid hormone receptors and metamorphosis | RNAseq data |
| Anura | Myobatrachidae | *Platyplectrum* | *ornatum* | Very small amphibian genome | In progress |
| Anura | Myobatrachidae | *Rheobatrachus* | *silus* | Gastric brooder | None |
| Anura | Nasikabatrachidae | *Nasikabatrachus* | *sahyadrensis* | EDGE Top-100 Amphibian #04 | None |
| Anura | Nyctibatrachidae | *Nyctibatrachus* | *humayuni* | PD | None |
| Anura | Odontobatrachidae | *Odontobatrachus* | *natator* | PD | None |
| Anura | Odontophrynidae | *Proceratophrys* | *moratoi* | PD | None |

(continued)

**Table 2** (continued)

| Order | Family | Genus | Species | Motive | Progress |
|---|---|---|---|---|---|
| Anura | Pelobatidae | *Pelobates* | *cultripes* | Developmental plasticity | In progress |
| Anura | Pelodryadidae | *Ranoidea* (*Litoria*) | *alboguttata* | Cell metabolism and dormancy | RNAseq data |
| Anura | Pelodytidae | *Pelodytes* | *punctatus* | PD | Pending |
| Anura | Petropedetidae | *Petropedetes* | *parkeri* | Striking sexual dimorphism | None |
| Anura | Phrynobatrachidae | *Phrynobatrachus* | *auritus* | Ecological gradients | None |
| Anura | Phyllomedusidae | *Agalychnis* | *callidryas* | Hatchling plasticity | RNAseq data |
| Anura | Pipidae | *Hymenochirus* | *boettgeri* | Small genome | In progress |
| Anura | Pipidae | *Xenopus* | *laevis* | Model organism. Tetraploid | Published (Session et al. 2016) |
| Anura | Pipidae | *Xenopus* | *tropicalis* | Model organism | Published (Hellsten et al. 2010) |
| Anura | Ptychadenidae | *Ptychadena* | *pumilio* | PD | None |
| Anura | Pyxicephalidae | *Pyxicephalus* | *adspersus* | ZW sex chromosomes | Published (Denton et al. 2018) |
| Anura | Ranidae | *Pelophylax* | *lessonae* | hybrid species | RNAseq data |
| Anura | Ranidae | *Rana* | *mucosa* | Conservation | In progress |
| Anura | Ranidae | *Rana* | *temporaria* | Hybrid species | In progress |
| Anura | Ranidae | *Rana* | *sylvatica* | Freeze tolerance | In progress |
| Anura | Ranidae | *Rana* | *catesbeiana* | Invasive species | Published (Hammond et al. 2017) |
| Anura | Ranixalidae | *Indirana* | *gundia* | EDGE Top-100 Amphibian #88 | None |
| Anura | Rhacophoridae | *Chiromantis* | *xerampelina* | Desiccation resistance | None |
| Anura | Rhinodermatidae | *Rhinoderma* | *darwinii* | Brooding pouch | None |
| Anura | Rhinophrynidae | *Rhinophrynus* | *dorsalis* | Deep PD | Pending |
| Anura | Scaphiopodidae | *Spea* | *bombifrons* | Phenotypic plasticity, speciation | In progress |
| Anura | Sooglossidae | *Sooglossus* | *sechellensis* | EDGE Top-100 Amphibian #70 | None |

| | | | | | |
|---|---|---|---|---|---|
| Caudata | Ambystomatidae | *Ambystoma* | *mexicanum* | Model organism | Published (Nowoshilow et al. 2018) |
| Caudata | Amphiumidae | *Amphiuma* | *means* | Elongation. PD | None |
| Caudata | Cryptobranchidae | *Andrias* | *davidianus* | Biggest amphibian | RNAseq data |
| Caudata | Hynobiidae | *Hynobius* | *chinensis* | Salamander model | RNAseq data |
| Caudata | Plethodontidae | *Desmognathus* | *fuscus* | Reduced genome size | None |
| Caudata | Proteidae | *Necturus* | *alabamensis* | EDGE Top-100 Amphibian #27 | None |
| Caudata | Rhyacotritonidae | *Rhyacotriton* | *olympicus* | PD | None |
| Caudata | Salamandridae | *Pleurodeles* | *waltl* | Limb regeneration | Published (Elewa et al. 2017) |
| Caudata | Salamandridae | *Salamandra* | *salamandra* | Viviparity | Pending |
| Caudata | Sirenidae | *Pseudobranchus* | *striatus* | PD | None |
| Gymnophiona | Siphonopidae | *Microcaecilia* | *unicolor* | PD | Pending |
| Gymnophiona | Chikilidae | *Chikila* | *fulleri* | PD | None |
| Gymnophiona | Dermophiidae | *Geotrypetes* | *seraphinii* | PD | In progress |
| Gymnophiona | Herpelidae | *Herpele* | *squalostoma* | PD | None |
| Gymnophiona | Ichthyophiidae | *Ichthyophis* | *bannanicus* | PD | Pending |
| Gymnophiona | Indotyphlidae | *Grandisonia* | *sechellensis* | PD | None |
| Gymnophiona | Rhinatrematidae | *Rhinatrema* | *bivittatum* | Co-sister to all other caecilians | Published[a] |
| Gymnophiona | Scolecomorphidae | *Crotaphatrema* | *tchabalmbaboensis* | PD | None |
| Gymnophiona | Siphonopidae | *Microcaecilia* | *dermatophaga* | PD | None |
| Gymnophiona | Typhlonectidae | *Typhlonectes* | *compressicauda* | PD | Pending |

Some families are represented more than once since these have additional species with WGS projects already in progress or published. Species are recommended based on the following criteria: (1) preliminary data, (2) scientific interest or interesting biology, (3) conservation interest, and, by default, (4) phylogenetic diversity (PD). "EDGE" refers to rank among amphibian species based on evolutionarily distinct and globally endangered criteria (Isaac et al. 2012). While additional WGS projects are likely going on that we are unaware of, we attempted to indicate Progress as follows. "Published" includes BioRxiv or publicly available data. "In progress" implies a status between initial data collection and a complete but unpublished draft. "RNAseq data" indicates species with transcriptomic data posted in the Short Read Archive at NCBI. "Pending" implies there is a laboratory or research community in place with interest in starting a WGS effort

[a]Publication not yet available, but very high-quality assembly posted online at https://vgp.github.io/genomeark/Rhinatrema_bivittatum/

understanding of amphibians and providing advice for researchers considering applying genomics for their amphibian study species.

## 2 Opportunities for Advancing Understanding of Amphibian Evolution, Ecology, and Behavior with Genomics

### *2.1 Phylogenomics*

Phylogenomics, the analysis of genomic data for phylogenetic inference, has grown rapidly in the last decade, providing increased resolution in the tree of life for many vertebrate lineages (Jarvis et al. 2014; Ruane et al. 2015; Streicher et al. 2018; Chakrabarty et al. 2017; Irisarri et al. 2017). The advances in phylogenomic data acquisition have been propelled primarily by sequence-capture methods that selectively capture previously identified genomic regions (Faircloth et al. 2012; Lemmon et al. 2012; McCormack and Faircloth 2013), by transcriptomic datasets based on expressed gene sequences (Irisarri et al. 2017), and by the availability of large-scale PCR-based nuclear protein-coding gene panels that can be sequenced with MPS (Shen et al. 2013). Compared to earlier multilocus methods that relied on a few to dozens of genes, these new methods now allow for analyses of hundreds of nuclear loci across a large sample of individuals, which has proven useful in resolving some of the most problematic nodes in the tree of life.

The application of phylogenomics to amphibians is rapidly growing. Recently, phylogenomic methods have helped resolve higher-level relationships among salamanders (Shen et al. 2013) and frogs (Feng et al. 2017). The salamander study addressed controversial relationships within Lissamphibia. The close relationship between frogs and salamanders (the clade Batrachia) is repeatedly recovered in molecular studies (Frost et al. 2006; Roelants et al. 2007; Zhang and Wake 2009; Pyron and Wiens 2011). However, one study based on 26 nuclear genes (Fong et al. 2012) supported a caecilian–salamander sister relationship. A phylogenomic analysis confirmed the monophyly of Batrachia and confirmed the monophyly of the internally fertilizing salamanders (Salamandroidea; all salamanders exclusive of Hynobiidae, Cryptobranchidae, and Sirenidae; Shen et al. 2013) in contrast to earlier studies based on a smaller number of markers (Frost et al. 2006). This newer study also provided a strongly supported phylogeny of all major frog lineages and estimated a much younger divergence time for frog lineages than inferred by earlier studies (Feng et al. 2017). In particular, divergence-time analyses indicated that three species-rich clades (Hyloidea, Microhylidae, and Natatanura), which comprise ~88% of extant anuran species, underwent simultaneous and rapid diversification at the Cretaceous–Paleogene (K–Pg) boundary. Thus, the K–Pg mass extinction may have triggered frog radiations by creating new ecological opportunities, as has been suggested for other animal groups (Feng et al. 2017).

An accurate inference of the amphibian tree of life provides the framework for important studies of macroevolution, diversification, and biogeography. These two recent studies demonstrate that phylogenomics has the potential to greatly increase resolution of our inferred topologies.

## *2.2 Genomic Data for Phylogeographic Inference*

The field of phylogeography focuses on the evolutionary and ecological processes that shape the spatial distribution of genetic variation within species. At its inception, the field bridged microevolutionary processes within populations and macroevolutionary patterns at larger scales (Avise et al. 1987), providing a framework to examine the factors influencing population divergence, persistence, and change over time. This framework spurred a large number of comparative studies that elucidated common landscape barriers impeding gene flow, identified "suture zones" in regional faunas of diverse taxa, and detailed the historical spatial and demographic processes acting on populations (Soltis et al. 2006; Bell et al. 2012; Barrow et al. 2017). The focus on population histories placed phylogeography squarely between population genetics and systematics, and many of the earliest studies on amphibians relied on a combination of rapidly evolving loci, including mtDNA markers, some nuclear genes, and microsatellites (Crawford 2003; Zamudio and Savage 2003; Funk et al. 2007, 2008; García-R et al. 2012; Lemmon and Juenger 2017). Genomic-scale data derived from target capture of orthologous loci, either as sequences or SNPs, have lifted previous limitations on the availability of adequate markers, including for amphibians, which have large genomes compared to other tetrapods. Phylogeographic studies are now expanding their reach by incorporating genome-scale data, providing an unprecedented level of genetic detail, fostering new techniques for tests of divergence, and synergies with landscape genetics and population genetics (McCormack et al. 2013; Barrow et al. 2014; Bell et al. 2015; Garrick et al. 2015; Pie et al. 2018).

A primary goal of phylogeography is the test of concordance of divergence among species, based on the hypothesis that co-distributed organisms should exhibit a concerted response to the same historical processes. The recent study by Barrow et al. (2018) exemplifies the application of genomic data to the test of concordance. Using target capture, they compared orthologous loci across 36 populations of 4 frog species distributed across known biogeographic barriers in the southeastern USA. Target capture, combined with thorough population sampling, allowed for tests of concordance at various levels of variation: among sites within a locus, among multiple loci within a species, among multiple species within a region, and between established biogeographic provinces (Barrow et al. 2018). The study found similar patterns within species, but high discordance among species, with little correspondence of genetic patterns with putative biogeographic barriers.

Discordance in phylogeographic structure is in some ways a more interesting outcome, because it points to differences among species in traits that mediate their

response to landscape barriers (Bell et al. 2017; Barrow et al. 2018; Polato et al. 2018). This is an exciting area of active research and uncovering what those traits are, and identifying the genes that underlie them, has the potential to complete the links between selection on phenotypes, mechanisms of divergence, and species differences in phylogeographic structure and speciation (Crispo 2008; Zamudio et al. 2016a; Polato et al. 2017). This focus, which has been termed "trait-based phylogeography" (Paz et al. 2015), will become possible by including genomic-scale data across populations and the measurement of genetic variability in functional traits that accelerate or deter divergences within species.

## 2.3 *Sex Chromosome Evolution*

The evolution of sex chromosomes offers a valuable opportunity to study how genomes respond to changes in gene copy number. The process of gene duplication is fundamental to the origin of new genes and novel phenotypes, yet changes in amounts of gene product may also cause gene dose problems (Bachtrog 2006). The challenge of dosage compensation is faced by all species with heteromorphic sex chromosomes, where the two sexes have zero, one, or two copies of each sex chromosome. All mammals (except monotremes) have an XX/XY genetic sex-determining mechanism, and all birds have the opposite, ZZ/ZW system. Thus, these clades solved the dose problem early in their history and are recalcitrant to further change in sex chromosomes. Amphibians and nonavian reptiles, in contrast, have evolved and re-evolved sex chromosomes many times in their history, providing researchers with replicated potential case studies of the evolution of sex chromosomes (Hillis and Green 1990; Ezaz et al. 2009; Nakamura 2009; Abbott et al. 2017).

As of 2014, systems of sex determination had been resolved in 173 species of amphibians, including one caecilian, revealing 28 species with XY male heterogamety, 16 species with ZW female heterogamety, and 1 case of OW female heterogamety, implying at least 18 independent evolutionary transitions (Ashman et al. 2014). The neotropical frog, *Leptodactylus pentadactylus*, now holds the record for most sex chromosomes in a vertebrate, with six X and six Y, accompanied by just ten autosomes (Gazoni et al. 2018). Mechanisms of sex determination evolve quickly in amphibians, with multiple systems found within taxonomic families and genera, or even within a single species. Conspecific populations of the Japanese frog, *Rana rugosa*, have one of three sex-determining systems: XX/XY, ZZ/ZW, or homomorphic sex chromosomes (Miura et al. 1998). Sex is determined genetically in all amphibians, thus even species with homomorphic sex chromosomes likely have a heterogametic sex (male or female). Evolutionarily, transitions are equally common between homomorphic versus heteromorphic sex chromosomes and between male versus female heterogamety (Pennell et al. 2018). Many species of amphibians, therefore, may be found in some initial state of evolution from homomorphic to

heteromorphic sex chromosomes, or low amounts of recombination may prevent divergence between homomorphic sex chromosomes (Guerrero et al. 2012).

Roughly 95% of amphibian species still lack information on which is the heterogametic sex. Progress has been slow because only specialized laboratories possess the knowledge and the dedication to produce karyotypes to search for heteromorphic versus homomorphic sex chromosomes in amphibians (Schmid et al. 2010). As most amphibians fall in the latter category, determining which sex is the heterogametic sex requires genetic tools. Traditionally, sex-linked markers are obtained by creating linkage maps. These maps can be developed from molecular genotyping including MPS approaches, but most organisms are not readily amenable to crossing experiments.

Recently, MPS genotyping has been used to develop sex-linked markers directly from a collection of DNA samples from multiple individuals of known sex, without the need for linkage maps or cytogenetics. Gamble and Zarkower (2014) outline a workflow based on restriction site-associated DNA sequencing (RADseq), which generates SNP genotypes from anonymous loci throughout the genome (Table 1), and applied this method to anoles and geckos (Gamble et al. 2015). By performing standard RADseq experiments on males and females, sex-specific markers can be recovered, screened for false positives, and validated on additional samples using PCR and Sanger sequencing. The required density of markers will depend on the absolute size of the sex-specific region and its size relative to the pseudoautosomal region (PAR) of the homomorphic sex chromosome.

A similar approach to recover sex-linked markers from the North American green frog (*Rana clamitans melanota*) involved DArT complexity reduction combined with MPS in a proprietary technology called DArTseq™ (Lambert et al. 2016). This genotyping by sequencing (GBS)-type method is similar to RADseq except that single-nucleotide polymorphisms (SNPs) are obtained preferentially from gene-rich regions, avoiding repetitive regions which can be especially problematic in amphibians. The authors found 15 SNPs and 8 presence-absence markers that together established that this species shows an XY/XX or male-heterogametic system. This finding confirmed a much earlier study based on allozyme data (Elinson 1983).

MPS approaches provide more than just information on the genetic basis of sex determination. Lambert et al. (2016) also found that sex-linked markers had variable levels of female homozygosity and male heterozygosity, reflecting variation in distances to the putative sex locus on the otherwise homomorphic sex chromosomes. In other words, loci located between the PAR and the sex-determining genes may experience some recombination and show intermediate levels of sex linkage. Additionally, since RADseq and DArTseq methods provide a few hundred base pairs of DNA sequence, the marker sequences themselves can be compared to reference genomes. In the case of the green frog, one marker was a putative paralog of *DMRT1*, a gene related to sex determination in many metazoans [including in *Xenopus laevis*, but not in *X.* (*Silurana*) *tropicalis*; Lambert et al. 2016].

Studies of other vertebrates have yielded a few dozen candidate genes involved in sex determination and sex differentiation. These genes can be accessed in amphibians with no previous genetic information through the application of RNAseq

(Wang et al. 2009). Furthermore, the relative positions of these genes may potentially be inferred from their positions in published genomes. *Xenopus* and *Nanorana* genomes show remarkable levels of synteny despite 266 million years of divergence (Sun et al. 2015). Gerchen et al. (2016) took advantage of this conservation of gene function and gene order to obtain DNA sequences of candidate sex genes using RNAseq of a single individual of *Bufo viridis*. From the resulting transcriptome, they further developed microsatellite loci located in coding regions of these genes in *Xenopus*. Finally, sex linkage was confirmed in *Bufo viridis* by genotyping these variable markers in parents and their offspring of known sex.

Not only do sex-determining mechanisms evolve rapidly, including sex chromosome turnover (Miura et al. 1998), sex steroids and steroid mimics can override sex-determining genes and reverse the gonadal sex of adults of some amphibian species (Hayes 1998). Combining these unusual amphibian traits with genetic recombination between sex chromosomes (see above) should be a warning to researchers using MPS methods based on reduced representation to study sex chromosome evolution. Carefully designed experiments should sample numerous adults of each sex to minimize false positives and catch possible sex reversals. Surveying multiple populations may reveal environmental correlates of sex reversal, such as contamination by endocrine disruptors, or may reveal additional cases of sex chromosome turnover within species (Lambert et al. 2016).

Whole genome resequencing (WGR) is currently nontrivial in amphibians; thus, until cheaper and more powerful sequencing methods become available, the study of sex chromosome evolution will benefit from new MPS-based genome reduction methods, such as RADseq. While a reference genome is an invaluable tool for any evolutionary genetic study, polymorphism data are even more important to link phenotypic and gonadal sex of the individual with potential sex-determining genes. Sex-linked loci identified through RADseq approaches could be further screened in more individuals and species using exon capture; although without a reference genome, complete gene sequences would be difficult to obtain. Alternatively, RNAseq could provide complete coding sequences, but the genes obtained may depend on the tissue, ontogeny, condition, and environment of the donor animal. In the study by Gerchen et al. (2016), RNA was extracted from six tissues from one adult male toad, providing 37 candidates genes but some with only partial coverage. While RADseq-based approaches for finding sex-linked markers obviate the need for linkage maps generated from experimental crosses between individuals (Gamble and Zarkower 2014), such maps may be helpful in assembling large, repetitive genomes, as demonstrated for the 32 Gb axolotl genome (*Ambystoma mexicanum*; Smith et al. 2018).

## *2.4 Population Structure and Demography*

The most common application of genetics to amphibians is the study of population structure, which involves characterizing the distribution of genetic variation within and among populations (Wright 1965; Allendorf and Phelps 1981), inferring the

evolutionary processes contributing to these patterns (primarily genetic drift and gene flow) (Slatkin 1981), and estimating important population genetic parameters such as effective population size ($N_e$) (Kimura and Crow 1963; Do et al. 2014). A related field is landscape genetics, which combines population genetics, landscape ecology, and spatial statistics to understand how complex landscapes affect patterns and rates of gene flow (Manel et al. 2003; Balkenhol et al. 2016). Understanding population structure and demography of amphibians is especially important in light of amphibian population declines, given that $N_e$ and gene flow mold the distribution of genetic variation, which in turn influences inbreeding depression, adaptive potential, and population persistence (Allendorf et al. 2013).

Population genomics will provide greater accuracy and power than ever to characterize population structure and demography of amphibians. First, population genomics allows identification of loci with a signature of divergent selection (Hohenlohe et al. 2010). It has been shown that inclusion of these non-neutral loci can severely bias estimates of population structure, gene flow, and other demographic parameters (Luikart et al. 2003). Thus, identification and removal of these loci should increase accuracy of estimates. Second, population genomics simply provides more independent loci from a larger proportion of the genome with which to characterize population structure and estimate population genetic parameters, increasing precision of estimates, as well (Luikart et al. 2003).

Characterization of population structure and estimation of $N_e$, gene flow, and related demographic parameters requires neutral genetic markers. Thus, MPS approaches that provide neutral markers are required. SNP data generated from RADseq are appropriate for any analyses that require data from independent loci spread across the genome, such as characterizing population structure, estimating $N_e$, or testing for population bottlenecks. Moreover, if longer contigs are generated from paired-end RADseq libraries or other reduced representation approaches such as anchored phylogenomics (Lemmon et al. 2012), which provide haplotype blocks with multiple SNPs, then it is possible to use coalescent-based approaches to infer population divergence, gene flow, $N_e$, and changes in $N_e$ through time simultaneously (Drummond et al. 2012).

Field approaches should ideally be combined with inference from population genomics to understand amphibian demography and dispersal. For example, field estimates of dispersal based on multistate capture-mark-recapture (CMR) analysis have been successfully combined with population genetic estimates of gene flow to more fully understand contemporary and historic patterns and rates of movements among amphibian populations (Funk et al. 2005; Lowe et al. 2006). Each approach has its strengths and limitations, but integrating both provides a more complete picture of movement across the landscape. CMR estimates of movement provide a detailed snapshot of contemporary movement for a limited number of populations, whereas genetic or genomic approaches are more useful for understanding deeper historic gene flow over a broader geographic sampling area.

A handful of studies have taken advantage of the power of MPS to address questions about demography and gene flow in amphibians. Nunziata et al. (2017) genotyped two species of salamanders (*Ambystoma talpoideum* and *A. opacum*)

with double-digest RADseq (ddRAD; Peterson et al. 2012) and tested whether coalescent-based analysis could detect changes in population sizes documented in the field. For both species, coalescent models largely agreed with CMR estimates of population declines or increases, demonstrating the utility of population genomics for detecting changes in population size over ecological time scales. Trumbo et al. (2016) also used a population genomic dataset consisting of over 20,000 SNPs to test the central marginal hypothesis (CMH) in the invasive range of *Rhinella marina* in Australia. The CMH predicts that genetic variation decreases from the core to edge of species' range, potentially limiting adaptation to new environmental conditions at the range margin (Eckert et al. 2008). They found support for the CMR in the southern portion of the species' range, but not in the northwestern or northeastern part of its range, which has important implications for management of this damaging, invasive species in Australia. These two studies provide evidence of the huge potential of MPS and population genomics to understand amphibian demography and population structure.

## 2.5 *Local Adaptation*

Amphibians occupy heterogeneous environments and have relatively low dispersal and gene flow compared to other vertebrates such as birds and mammals (Ward et al. 1992). This combination of high habitat heterogeneity and low gene flow suggests they will often be highly locally adapted. The fact that many amphibian species live in extreme environments (e.g., deserts, tree canopy, alpine ponds and lakes, high latitudes, caves) despite being ectothermic and having permeable skin speaks to their adaptive potential (Duellman and Trueb 1986). The observation that many amphibian species span dramatic environmental gradients also indicates adaptive divergence within species (Berven 1982; Palo et al. 2003; Funk et al. 2016). Understanding the ultimate environmental drivers and proximate genetic mechanisms of adaptive divergence is a fundamentally important question in evolutionary biology. Characterizing patterns of adaption across real-world landscapes is also of the utmost importance in conservation for assuring that the maximum amount of additive genetic variation is conserved (McKay and Latta 2002; Funk et al. 2012). Understanding patterns of adaptation is also critical for making sure that source populations for augmentation of declining populations are not adaptively divergent from the target population, which can lead to outbreeding depression rather than the desired outcome of genetic rescue (Edmands 2007; Frankham et al. 2011).

Multiple analytical approaches are available in the field of population genomics for studying local adaptation, and the best MPS approach depends on which of these analyses will be applied. One analytical approach is to identify loci under divergent selection using genome scans, for example, to detect locus-specific $F_{ST}$ values significantly higher than the baseline genome-wide average $F_{ST}$ value (Beaumont and Nichols 1996; Beaumont and Balding 2004). Genome scans can be performed using a variety of marker types, but since these methods are designed to identify loci

with values higher than those observed at neutral loci, they require that the majority of loci are not under directional selection (Luikart et al. 2003). Given this requirement, WGR (Table 1) or RADseq (and other related reduced representation restriction enzyme-based methods) are two appropriate choices, but each has its strengths and limitations. WGR provides complete or nearly complete coverage of the entire genome but requires a reference genome, which currently exist for few amphibian species (Table 2). However, this approached could theoretically allow detection of natural section across all types of variation (not just SNPs) including structural variants (Toews et al. 2016; Fuentes-Pardo and Ruzzante 2017). In perhaps the first-ever application of WGR to amphibians, Wang et al. (2018) took advantage of the published reference genome of the Tibetan frog (*Nanorana parkeri*; Sun et al. 2015) by resequencing 63 more frogs at a depth of 6- to 17-fold coverage and recovered almost 9 million SNPs to infer historical demography, speciation, hybridization, and potential genomic bases of adaptation to high elevation environments. For a given budget, however, complete coverage of the entire genome comes at the cost of fewer individuals that can be sequenced, potentially resulting in lower power to detect selection at any given locus. In contrast, since RADseq is a reduced representation approach, many more individuals can be genotyped for a given budget than with WGR but at the cost of no coverage of a sizable percent of the genome (Baird et al. 2008; Andrews et al. 2016). If a reference genome is available, it is possible to map RADseq loci to the genome so that linkage disequilibrium can be calculated and the protocol can be fine-tuned to make sure that marker density is high enough to detect most loci under selection (Lowry et al. 2017; Catchen et al. 2017).

Another genomic analysis available for dissecting the genetic basis of adaptation is genome-wide association studies (GWAS; Stinchcombe and Hoekstra 2008). The basic premise behind GWAS is to identify loci and alleles correlated with variation in phenotypes. If populations have different values of phenotypic traits that are hypothesized to be adaptive in their respective environments, GWAS can estimate the presence and strength of statistical correlations between phenotypic differences and the frequency of alternative alleles at a locus. A conceptually similar analytical framework is genotype-environment association (GEA) methods, which are designed to test for correlations between genotypic variants and environmental variation (e.g., Joost et al. 2007; Coop et al. 2010; Frichot et al. 2013). As with genome scans, the most important criterion for choosing an appropriate MPS approach for GWAS and GEA is that it provides high-density coverage of the genome so that most loci influencing the phenotype can be detected. Whole-genome resequencing and the family of RADseq methods both fit this bill but with the same pros and cons discussed above for genome scans.

Yet another genomic approach for studying adaptive divergence is transcriptomics based on RNAseq (including the Iso-Seq method implemented by PacBio). Experiments provide three types of information: DNA and inferred amino acid sequences, gene diversity such as duplications, and quantitative gene expression patterns such as differences among populations or different environmental conditions (Zhen et al. 2012). The strength of this approach is that it focuses exclusively on functional, expressed genes that might underlie adaptive phenotypic differences

(Ghalambor et al. 2015). Differences in gene expression patterns could be caused by environmental or genetic differences, or both; thus, inferring adaption can be challenging. On the other hand, RNAseq data can be used for comparative analyses of the adaptive basis of natural selection, such as dN/dS ratios, which may have more power than some population genetic tests of selection (Zhai et al. 2009). One example of this approach applied to frogs looked for accelerated rates of non-synonymous substitution across expressed genes in Himalayan versus lowland species of *Rana* (Yang et al. 2012). Combining transcriptomics with other tools such as GWAS can also be a powerful integrative approach for understanding adaptive divergence.

The study of adaptation is a prime example of an area of study that requires both classic field observations and experiments (Endler 1986) in addition to population genomics for robust inferences. The first criterion for local adaptation is that phenotypes differ in different environments, which cannot be determined with genomics. Secondly, evidence is required that these phenotypic differences are adaptive (increase fitness) in the local environment. The gold standard for testing this is a reciprocal transplant experiment (Claussen et al. 1948), which is feasible for some amphibian species but not others (Urban et al. 2017). If a reciprocal transplant experiment is not feasible for a given species, then the combination of fieldwork showing among-population phenotypic differences, genome scans identifying loci under divergent selection, and GWAS showing that some of these same loci are related to the observed phenotypic differences provides compelling evidence that the phenotypic differences are adaptive. The key for successful studies of adaptation in nature is to combine traditional field and cutting-edge genomic approaches in creative and well-designed ways.

The application of genomics to studies of adaptation is increasing dramatically but is still in its infancy in amphibians. For example, Richter-Boix et al. (2011) used genome scans to identify a locus under divergent selection among ponds associated with variation in tadpole life history characteristics (Ficetola and Bonin 2011). Since Richter-Boix et al.'s analysis was based on only 15 microsatellite loci, they almost certainly missed many other loci under divergent selection. In a more recent study, Guo et al. (2016) used genome-wide scans of over 15,000 SNP loci obtained using RADseq to test for loci with signatures of divergent selection between low and high elevation populations of *Bufo andrewsi*. They found many SNPs associated with differences in elevation, temperature, or both hypothesized to be involved in adaptation to high elevations. These studies pave the way for future work harnessing MPS and genomics to understand adaptation in amphibians.

## *2.6 Mating Systems and Sexual Selection*

Amphibians have some of the most diverse and complex reproductive modes of all vertebrates, including eggs versus live birth, terrestrial versus aquatic oviposition sites and larval development, and, sometimes, parental care and even feeding of

offspring (Salthe and Duellman 1973). Although all amphibian orders show diversity of reproductive modes, the patterns of evolution in these modes have best been characterized in frogs (Anura). Two evident patterns in the evolution of these traits are the higher diversity of reproductive modes in the tropics and the apparent progression from aquatic to terrestrial reproduction, often attributed to higher fitness resulting from decreased predation on terrestrial eggs and tadpoles (Gomez-Mestre et al. 2012) or to reduced loss of fitness due to polyandry in terrestrial breeders (Zamudio et al. 2016b). Thus, reproductive modes of frogs offer an excellent opportunity to genetically characterize mating systems, measure reproductive fitness, and quantify the selective advantage of different traits or behaviors during reproductive events. To date, very few studies have taken advantage of the diversity in amphibian reproduction or used genetic or genomic techniques to characterize mating outcomes in species with different modes. The few studies that have done so have typically used few microsatellite markers for paternity assignment and estimates of relatedness; thus, the power of genomics has not yet been harnessed in this field. Nonetheless, every study that has genetically assessed reproductive outcomes in amphibians has found surprising results such as high degrees of multiple paternity (Laurila and Seppa 1998; Myers and Zamudio 2004; Kupfer et al. 2008; Adams et al. 2005), evidence for "good genes" and heritability of fitness traits (Welch et al. 1998), novel reproductive strategies and mate choice (Vieites et al. 2004; Ringler et al. 2012), and a high degree of parental care relative to parentage (Summers and Amos 1997; Chen et al. 2011; Muralidhar et al. 2014).

Data on individual relationships is essential to studies of the behavioral ecology of wild organisms. Advances in molecular and analytical techniques have enhanced our ability to test hypotheses about reproductive modes and mating systems by providing information on the genetic relationships among individuals (Hughes 1998; Avise et al. 2002; Griffith et al. 2002; Myers and Zamudio 2004; Thrasher et al. 2018). The application of genome-wide SNPs to analyses of parentage and relatedness has received greatest attention (Glaubitz et al. 2003; Thrasher et al. 2018). Studies in birds (Weinman et al. 2015; Kaiser et al. 2017), fish (Hauser et al. 2011), and domesticated species (Tokarska et al. 2009; Fernandez et al. 2013) have developed large SNP panels with power comparable to or higher than polymorphic microsatellites. Likewise, targeted amplicon resequencing of large panels of microsatellites also permits rapid and accurate genotyping (Andrés and Bogdanowicz 2011; Nali et al. 2014). Once polymorphic loci have been identified, each individual can be genotyped at >150 microsatellite loci using multiplex PCR reactions. Multiplexed loci are then pooled for each individual, barcoded, and sequenced on next-generation sequencing platforms. Targeted amplicon sequencing offers a couple of advantages. First, once the loci have been identified, this method is fast, allowing for genotyping of a large number of individuals, which is often required in amphibian parentage studies. Second, resequencing microsatellite loci allows for the identification of homoplasy caused by flanking mutations or reversals, thus reducing assignment error in parentage analyses. This method has been applied in a variety of taxa (Nali et al. 2014; D'Aloia et al. 2017) and provides paternity assignments with

high probabilities. MPS methods for parentage and relatedness assays, once applied to amphibians, have the potential to reveal the selective contexts and mechanisms leading to their unusually high diversity of reproductive modes.

## 3 Opportunities for Improving Amphibian Conservation with Genomics

### *3.1 Hybridization*

Hybridization between species often results in offspring that are less fit than parental forms, which may result in selection for traits that enhance prezygotic barriers to gene flow ("reinforcement"). Alternatively, in the absence of reproductive barriers, hybridization may result in the "genetic swamping" of one of the parental forms, due to extensive introgression (Rhymer and Simberloff 1996). This duality in the nature of hybridization, potentially enhancing or decreasing biodiversity, has been a focus of much genetic work, especially in the context of conservation. Many studies have identified a role for interspecific hybridization in promoting the evolution of novel adaptive forms (Anderson and Stebbins 1954; Harrison 1993; Arnold 1997). Natural hybridization occurs relatively frequently among divergent populations of animal species (Barton and Bengtsson 1986; Grant and Grant 1992; Mallet 2005), and only a few hybridization events are needed to allow the exchange of advantageous alleles between species. The historical admixture of genomes has also contributed to speciation, especially in plants, but also in some animal taxa (Arnold 1997; Dowling and Secor 1997; Mavárez et al. 2006; Gompert et al. 2006; Grant and Grant 2008). Therefore, genomic studies of hybrid zones have the potential to inform not only the causes but also the consequences of hybridization.

In amphibians, genetic or genomic approaches have been used to identify the extent of hybridization between endangered and non-endangered species to guide conservation and management actions (Austin et al. 2011; Zamudio et al. 2010). For example, natural hybridization has been detected between endangered *Ambystoma tigrinum stebbinsi* and the widespread barred tiger salamander *A. t. mavortium* (Storfer et al. 2004), raising concern for the persistence of *A. t. stebbinsi* populations in Arizona. In some cases, anthropogenic translocation of one species outside its natural range causes population dynamics that can favor hybrids over pure parental forms, as is the case with the hybridogenetic frogs *Rana lessonae* and *R. ridibunda* in Europe. Hybridogenesis is an unusual form of reproduction in which a hybrid persists and spreads in populations with just one parent, with which it backcrosses over multiple generations (Beebee 2005). Introduced *R. ridibunda* have replaced *R. lessonae* in several areas of Western Europe in recent decades (Vorburger and Reyer 2003). Likewise, nonnative *A. t. mavortium* introduced into Central California in the 1950s have led to the formation of a hybrid swarm within the range of the federally protected California tiger salamander (*A. californiense*). Genomic analyses

show that a small fraction of superinvasive genes are introgressing more rapidly into the native species (Riley et al. 2003; Fitzpatrick et al. 2009, 2010). The hybrids have higher fitness than the native *A. californiense*, raising concerns of the possibility of genetic extinction of populations of the native species (Fitzpatrick and Shaffer 2007; Box 1).

**Box 1 Anthropogenically Mediated Hybridization in the Critically Endangered *Ambystoma californiense***

A well-characterized example of hybridization is found in the California tiger salamander (*Ambystoma californiense*), an endangered species that hybridized with the more widespread barred tiger salamander (*Ambystoma tigrinum mavortium*), following anthropogenic introductions of the barred tiger salamander within the breeding range of the formerly allopatric California tiger salamander (Riley et al. 2003; Fitzpatrick and Shaffer 2007). The introduced species has spread and the hybrid swarm currently occurs throughout 25% of the native species' original range (Shaffer et al. 2015). The conservation challenge is exacerbated by the fact that hybrids seem to have higher fitness, especially in disturbed environments (Fitzpatrick and Shaffer 2007; Ryan et al. 2013). Using a panel of 64 genome-wide SNP markers and a survey of hybrid and pure populations, researchers found that 3 introduced alleles have largely displaced native alleles within the hybrid populations, likely due to strong selection favoring allelic fixation at those loci (Fitzpatrick et al. 2009, 2010). Introgression of a few, strongly selected introduced alleles may not directly affect the persistence of California tiger salamanders, but these patterns underscore how selection can rapidly promote introgression and pose challenges for maintaining pure populations of endangered species threatened by hybridization (Fitzpatrick et al. 2009).

*Ambystoma californiense* (Photo credit: Ian J. Wang)

Not all hybridization results in negative consequences for populations, of course. Hybrid-origin or allopolyploidization is an important mechanism of hybrid speciation with strong, although often incomplete, postzygotic reproductive barriers between the polyploid hybrid and its diploid parents (Ficetola and Stock 2016). In amphibians, allopolyploids often have novel phenotypes that differ from those of the

ancestral lineages, leading to the hypothesis that niche shifts could be one of the consequences of hybrid speciation. A recent review of hybridization in toads showed that some allopolyploids occupy intermediate niches to those of the diploid parental lineages, but in other cases, allopolyploids showed transgressive niche evolution; they inhabited environments that were more arid and with colder winters than either of their parental species (Ficetola and Stock 2016). This leads to the possibility that endangered species with hybrid genomes might show higher fitness in anthropogenically modified environments, suggesting that some degree of hybridization might provide greater adaptive potential with which to respond to environmental change (Zamudio et al. 2010). A fruitful avenue of research in conservation will be characterizing the genomic architecture and evolutionary potential of fit hybrids.

## *3.2 Disease Evolution and Ecology: Lessons from Chytridiomycosis*

Genomic studies have been used effectively to disentangle host-pathogen interactions and disease dynamics in amphibians (Longo et al. 2014). The amphibian fungal disease, chytridiomycosis, which affects species across all continents, recently emerged as a major threat to biodiversity. The disease is caused by the chytrid fungus, *Batrachochytrium dendrobatidis*, [hereafter *Bd* (Longcore et al. 1999)], a generalist amphibian pathogen which has extirpated populations of diverse species from around the globe. *Bd* has now been reported from over 500 amphibian host species and has a cosmopolitan distribution (Olson et al. 2013; James et al. 2015). The emergence of *Bd* has shown that host-pathogen interactions can play a major role in species declines and even extinctions (Crawford et al. 2010). Genomic approaches have elucidated aspects of the biology and evolution of frogs (hosts) and the pathogen itself.

The first genetic assessments of pathogen diversity in this system showed a surprising absence of genetic variability, leading researchers to propose a recent emergence of *Bd* followed by a rapid spread of a Global Panzootic Lineage (*Bd*-GPL) around the world (Morgan et al. 2007; James et al. 2009). More recently, novel genotypes putatively endemic to the Cape of South Africa (*Bd*-Cape), Switzerland (*Bd*-CH), Brazil (*Bd*-Brazil), and Korea (*Bd*-Korea) were found, and whole genome sequencing showed that those lineages are basal divergences within the *Bd* phylogenetic tree (Farrer et al. 2011; Schloegel et al. 2012; Rosenblum et al. 2013; Bataille et al. 2013). By increasing the sampling of global *Bd* strains, it became apparent that the earlier perspective on low genetic diversity was due to biased sampling of only epizootic strains (James et al. 2015). The discovery of enzootic lineages of *Bd* that are more restricted in their distribution contrasts with the broad distribution and spread of the virulent genotype (*Bd*-GPL).

A recent study has probed the functional genomics of *Bd* virulence (Ellison et al. 2017). Using laser capture microdissection (LCM), which allows for analysis of

pathogen gene expression in infected tissues of living hosts, Ellison et al. (2017) performed the first in vivo functional assays of *Bd*-GPL infection in the amphibian skin. Using sequencing of *Bd* RNA from infected epidermal cells in two different hosts and in culture, that study identified more than 2,000 differentially expressed genes between *Bd* in tissues and in culture that included key *Bd* defense and host exploitation mechanisms. Significant enrichments of genes with increased expression in both host species compared to *Bd* cultures were those related to proteolysis and membrane transport activity, both important during the host invasion and infection process (Ellison et al. 2017). In contrast, variation in *Bd* transcriptomes from different amphibian hosts demonstrates shifts in pathogen resource allocation. *Bd* genes more highly expressed in *Atelopus zeteki* were those related to cilium organization and cilium morphogenesis, suggesting a greater investment in motile zoospore production. In contrast, *Bd* samples collected from *Hylomantis lemur* were predominantly enriched for biosynthetic and amino acid metabolic processes (Ellison et al. 2017). Earlier studies of *Bd* in culture show that abiotic conditions can significantly alter life history trade-offs in the pathogen (Woodhams et al. 2008). Different gene expression profiles in different hosts indicate that host species also influences the relative investment of *Bd* in growth and reproduction, likely as a response to the host's defensive capabilities (Ellison et al. 2015). Thus, the selective environment provided by different host species has a strong effect, and the pathogen may respond by shifts in resource allocation rather than evolutionary changes. This level of flexibility, revealed by functional genomic assays, is most likely part of the strategy ultimately underlying the success of *Bd* as a generalist pathogen.

In the field, hosts vary in their disease outcomes with some populations persisting after the arrival of *Bd*, while others go extinct. These variable outcomes suggest potential differences in host genotype and potential for evolution of resistance to *Bd*. Amphibians can rely on innate or adaptive immune responses to manage *Bd* infections, and at least some of these immune responses have a genetic basis (Savage and Zamudio 2011; Ellison et al. 2014), suggesting host genotypic variation may be an important factor explaining persistence or mortality. For instance, alleles of the major histocompatibility complex (MHC), a family of genes in the adaptive immune response, were significantly associated with resistance and survival in *Lithobates yavapaiensis* (Savage and Zamudio 2011) and *Litoria verreauxii* (Bataille et al. 2015). Various immunogenetic studies have reported either a strong or weak adaptive immune response post-*Bd* infection (Rosenblum et al. 2013; Ellison et al. 2014), underscoring variation among species in their potential for the evolution of resistance. A comparative study of amphibian functional immunogenomic responses to *Bd* provided some insights into key genetic mechanisms underlying variation in disease outcomes among amphibian species (Ellison et al. 2015). That study challenged four Central American frog species that vary in *Bd* susceptibility with a sympatric virulent strain of *Bd*. Comparison of host gene expression profiles showed that resistant species have reduced skin inflammatory responses and increased expression of genes involved in skin integrity. In contrast, only highly susceptible species exhibited suppression of splenic T-cell genes, likely in response to the suppressive action of *Bd* on host immune function (Fites et al. 2013). Thus,

resistance to chytridiomycosis may be related to a species' ability to escape the immunosuppressive activity of the fungus.

In summary, genomic surveys suggest natural variation in both pathogen virulence and host immunity, but the interactions between these two components have not been adequately addressed to allow predictions of which species or communities have the potential to recover after exposure to *Bd*. Nonetheless, genomic approaches have begun to unravel the mechanisms underlying the evolution of pathogen virulence and host resistance not only in *Bd*, but now in the newly emerged *B. salamandrivorans* that is specific to salamanders (Box 2).

**Box 2 Genomics of the Newly Emerged *Batrachochytrium salamandrivorans* (*Bsal*)**

A major breakthrough in amphibian disease ecology was the discovery of a new chytrid species, *Batrachochytrium salamandrivorans* (*Bsal*) that infects salamanders. The new species is morphologically, genetically, and functionally distinct from *B. dendrobatidis* and was discovered as a pathogen of fire salamanders (*Salamandra salamandra*) in Europe (Martel et al. 2013) but was likely introduced from eastern Asia (Martel et al. 2014). The recent arrival of *B. salamandrivorans* in Europe was followed by rapid expansion of its geographical distribution and host range, confirming the unprecedented threat that this chytrid fungus poses to amphibians globally (Stegen et al. 2017). *Bd* and *Bsal* diverged an estimated 67 million years ago, and they have different host species ranges, with *Bsal* mostly infecting a single order, Caudata (salamanders; Martel et al. 2014), while *Bd* infects many species across all three orders of Amphibia (Fisher et al. 2009). Scientists predicted that this evolutionary jump to amphibian hosts was facilitated by the acquisition of common ancestral traits, whereas subsequent differentiation of infection strategies has been the result of lineage-specific adaptations (Farrer et al. 2017). To test this prediction, they sequenced the genomes of *Bd* and *Bsal* and compared them to those of two related saprobic chytrids. The results show that evolutionary adaptation to infect amphibians is correlated with the acquisition of genes encoding secreted proteins that are unique to the genus *Batrachochytrium. Bd* and *Bsal* share 542 gene clusters that are not found in the two saprobic chytrids and include specific functions related to cell wall modification and candidate secreted effectors (Farrer et al. 2017). Several of these lineage-specific protein families are highly expressed during in vivo infection of salamanders; these upregulated genes likely include key virulence factors. Among these are the M36 metalloproteases implicated in chytrid pathogenicity (Farrer et al. 2013; Martel et al. 2014) as well as at least two large families of secreted proteins with no recognizable functional domains, which are very highly expressed and may represent novel virulence factors unique to *Bsal*. Characterizing the genomic architecture of pathogens with different virulence, and the expression

(continued)

**Box 2** (continued)
of pathogen and host genes during the infection process, will yield important information on how these pathogens have specialized and their potential impact on other salamander species.

*Salamandra salamandra fastuosa* from Guipúzcoa, Basque country, Spain (Photo credit: Guillermo Velo-Antón)

## *3.3 Captive Breeding*

Captive breeding programs are a stopgap approach to conserving those populations that have declined to the point where their persistence in the wild is in doubt. With the catastrophic declines of many amphibian species and populations in the last few decades associated with *Bd* and other factors, ex situ captive breeding has become an important management strategy for amphibian conservation (Harding et al. 2016). Although a suite of complex technical and ethical issues surround captive breeding (Seigel and Dodd 2002; Trenham and Marsh 2002), it will likely remain an important conservation strategy for endangered amphibian species for the foreseeable future.

MPS and population genomics can maximize the potential for success of captive breeding programs by identifying mating pairs that are not too closely related – potentially resulting in inbreeding depression – or from divergent populations, which could result in outbreeding depression. When individuals are brought into captivity, their relatedness is typically unknown, meaning that closely related individuals

might be mated, resulting in inbred offspring with low fitness (Kardos et al. 2016). Genomic datasets consisting of genotypes at hundreds or thousands of SNP loci generated from RADseq allow precise estimation of relatedness (Weir et al. 2006), facilitating selective breeding to maximize success of ex situ breeding programs.

Genomics can also be used to more precisely estimate genetic divergence among populations to identify pairs of populations that are so divergent that crosses between individuals from these populations could result in outbreeding depression (Frankham et al. 2011). Several genomic approaches are available for characterizing the degree of adaptive divergence among populations (see Sect. 2.5); thus, genomics can play a key role in making sure that only individuals from adaptively similar populations are crossed as part of captive breeding programs.

## 4 Challenges and Solutions for Genomic Studies of Amphibians

### *4.1 Few Reference Genomes*

Increasing numbers of mammal and bird species (Zhang et al. 2014) are becoming genome enabled, yet amphibians remain largely "genome disabled." Even for nonavian reptiles and fishes (Bernardi et al. 2012), the number of published reference genomes is increasing rapidly. Extant amphibians, or Lissamphibia, represent over 300 million years of evolution and include 7,934 species as of September 30, 2018 (AmphibiaWeb 2018). Additionally, amphibians arguably are the most endangered clade of vertebrates as well, with one third of species listed as endangered by the IUCN (Hoffmann et al. 2010). Despite this tremendous age, diversity, and threat level, only four reference genomes had been published by the end of 2017: the model organisms *Xenopus tropicalis* (diploid) and *X. laevis* (a tetraploid), the Tibetan frog, *Nanorana parkeri*, and finally, the American bullfrog, *Rana catesbeiana* (see Table 1 for citations). Fortunately, 2018 has seen this number double, with publication or prepublication of reference genomes announced for the African bullfrog, *Pyxicephalus adspersus* (Denton et al. 2018), cane toad, *Rhinella marina* (Edwards et al. 2018), strawberry dart-poison frog (Rogers et al. 2018), and two independent efforts at assembling the enormous axolotl genome (Nowoshilow et al. 2018; Smith et al. 2018). Reference genomes provide a deeper understanding of demographic history and are a key resource for studying functional variation that may be adaptive (Steiner et al. 2013). Knowing the genomic position of genes is essential for understanding the interaction between natural selection and recombination including gene conversion (Hoban et al. 2016). Genomic information can greatly enhance management efforts for captive and wild populations of threatened species by identifying loci involved in inbreeding depression and disease susceptibility (Johnson and Koepfli 2014).

In prioritizing amphibian species for whole genome sequencing (WGS), the Genome 10K consortium suggested three criteria (Haussler et al. 2009). First, species of special conservation concern may receive an immediate and applied benefit to becoming genome enabled. Ironically, the most endangered amphibian species may be so rare that access to genomics-quality samples becomes very difficult. Second, WGS projects should maximize phylogenetic diversity to provide genomic "outposts" across the phylogeny of Lissamphibia. Each WGS will facilitate genomic studies of a constellation of related species, and any comparative genomics questions will require data from as many distinctive lineages as possible. Thus, initial WGS projects should maximize phylogenetic diversity (Faith 1992), e.g., by sequencing at least one species per taxonomic family of amphibians (Table 2). Third, and arguably the most important, each WGS requires a community of curators and users attracted to a given species for scientific questions and dedicated to providing support for quality control and continued development of annotation and other resources. Reference genomes that are not curated and updated risk becoming inaccessible or obsolete. In fact, reference genomes are almost never "done," rather they progressively improve in terms of length, contiguity, accuracy, and annotations. So too, not all genomes need to be completed to chromosome-level, phased scaffolds. Genome 10K has recommended that each major taxonomic group (clades with a stem age of up to 50 million years, e.g., taxonomic families; Table 2) should be represented by a very high-quality reference referred to as a "platinum genome" (Koepfli et al. 2015). Under the aegis of Genome 10K, Phase I and Phase II of the Vertebrate Genomes Project (VGP) aim to produce at least one near-gapless, chromosome-level, phased genome assembly representing each vertebrate lineage, including amphibians, which predates the K–Pg mass extinction event (https://www.rockefeller.edu/research/vertebrate-genomes-project/vertebrate-genomes-project-plan/).

## *4.2 The Genome Size Problem*

Amphibian genomes are undoubtedly among the largest of all organisms (Elliott and Gregory 2015). While genome size information is lacking for the majority of amphibian species, and most of these data come from microscopy rather than from more precise flow cytometry methods, preliminary trends can be inferred from the www.genomesize.com database (Gregory 2011) augmented by a recent study (Liedtke et al. 2018). Salamander genomes start at around 14 Gb with an interquartile range of 25–44 Gb ($n = 170$), with waterdog or mudpuppy (*Necturus* spp.) genome sizes ranging from 84 to an astounding 118 Gb, the equivalent of 37 human genomes! Frog genome sizes show an interquartile range of about 3–5 Gb but range from 1 to 11 Gb ($n = 272$: Fig. 1). Caecilian genome sizes show an interquartile range of about 5–8 Gb, but range from 3 to 12 Gb ($n = 22$; Fig. 1). While amphibian genomes can tolerate a surprising amount of polyploidy (Schmid et al. 2015; Session et al. 2016), gigantism in amphibian genomes appears to be more commonly caused by proliferation of transposable elements (TE) such as long

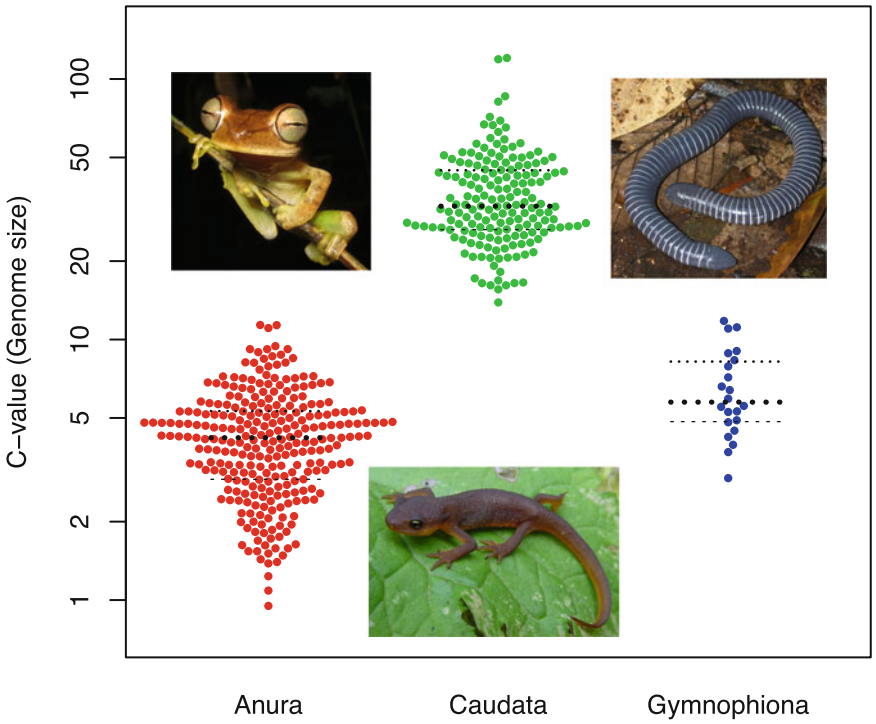


**Fig. 1** Variation in genome size within and among taxonomic orders of amphibians. Data consist of 464 species in Gregory (2011) and Liedtke et al. (2018), combining data across multiple techniques for genome size estimation. When multiple estimates existed for a species, the mean of the estimated genome sizes was assumed. Heavy dashed lines are the median values; light dashed lines are the first and third quartiles. Y-axis is on a natural log scale and reports C-values in picograms (pg), where 1 pg = 978 megabases of DNA sequence. The largest values are approximately 120 pg (118 Gb) for waterdogs (salamanders of the diploid genus, *Necturus*), while the smallest genome at 0.95 pg (929 Mb) was reported for the ornate burrowing frog, *Platyplectrum* (*Limnodynastes*) *ornatum*, of Australia. These largest and smallest values were obtained by Feulgen densitometry on red blood cells. As a point of reference, a human genome is 3.3 pg (3.23 Gb). Frog, salamander, and caecilian photos are of *Boana* aff. *fasciata* from Município Cotriguaçu, Mato Grosso, Brazil (Photo credit: Kelly R. Zamudio), *Taricha granulosa* from Benton County, Oregon, USA (Photo credit: W. Chris Funk), and *Siphonops annulatus* from Panguana, Huánuco, Peru (Photo credit: Andrew J. Crawford), respectively

terminal repeat (LTR) retrotransposons (Sun et al. 2012). Some frog species have secondarily evolved smaller genomes, potentially driven by natural selection for faster development time in xeric habitats (Liedtke et al. 2018), as species with larger genomes have slower embryonic development times (Jockusch 1997).

An estimated 31% of frog species have a genome size smaller than that of a human (data from Liedtke et al. 2018), which might suggest that size is not the only stumbling block to producing reference genomes for amphibians. Indeed, sequencing all the DNA in a genome is simple and each year less expensive. Assembling a

genome, however, is vastly more challenging. The worst-case scenario for whole genome assembly would be a genome rife with recently proliferated (i.e., very similar) repetitive elements. This is what has been found in the strawberry dart-poison frog (*Oophaga pumilio*): 4.76 of its 6.76 Gb genome is made up of recently expanded TE families, including 1 Gb of just the *Gypsy* family, making assembly a colossal challenge (Rogers et al. 2018).

In addition to the size, ploidy, and abundant TEs in their genomes, amphibians face further challenges to WGS. Many MPS and scaffolding techniques, such as optical mapping, require large amounts of ultrahigh molecular weight (uHMW) DNA as input (Wong et al. 2012), yet some amphibians rank among the smallest vertebrates, making it difficult to obtain large quantities of DNA for scaffolding (e.g., Rittmeyer et al. 2012). Finally, amphibians are in sharp decline around the world, and some of the most fascinating frogs have not been seen in decades (e.g., the gastric-brooding frog, *Rheobatrachus silus* of Australia; Table 2). Access to abundant, fresh, uHMW DNA is already impossible for many of the species that could most benefit from genomic tools.

We predict that the next few years will witness a renaissance in WGS and related approaches to amphibian genomics. We suggest three strategies that will help drive global amphibian genomics. First, we need more data on genome size variation in amphibians, i.e., a "1,000 genome sizes" project. Closely related species may vary substantially in genome size; thus, within each clade of interest, WGS efforts can be focused on species with smaller genomes with lower ploidy. We recommend flow cytometry (Vinogradov 1998), but imaging (Hardie et al. 2002) and quantitative PCR can also be used (Wilhelm et al. 2003). Karyotyping is still important to evaluate ploidy, check for heteromorphic sex chromosomes, and inform studies of synteny (Bogart 1973). Blood samples for flow cytometry should be collected along with routine tissue sampling in the field. Karyotyping requires slightly more specialized preparation with live animals.

Second, for many questions in amphibian genomics, a catalog of expressed genes may be sufficient. Very large diploid amphibian genomes do not present any additional challenges for transcriptomic studies, as these genomes likely have a similar complement of protein-coding genes as other vertebrates (Sun et al. 2015). A single RNAseq experiment (Wang et al. 2009) costs only a few hundred US dollars including sequencing but can provide 1,000s of complete, de novo assembled protein-coding genes for use directly in phylogenomics, comparative genomics, and molecular evolution (Huang et al. 2016). These gene sequences can also be used for developing target capture probes for phylogenomics and population studies (McCartney-Melstad et al. 2016; Portik et al. 2016). Perhaps the major limitation to implementing RNAseq studies is access to fresh, well-preserved RNA samples, e.g., flash-frozen, stored in nucleic acid preservation (NAP) buffer (Camacho-Sanchez et al. 2013), or both. RNA sampling should be a routine part of field collecting, whenever possible. RNAseq data also greatly improves eukaryotic reference genome annotation and can improve scaffolding (Yandell and Ence 2012). Transcriptomic data collection should be a first step in the genetic characterization of every amphibian species of special scientific interest or conservation concern (Table 2).

Third, amphibians may be among the greatest beneficiaries of recent advances in genome scaffolding. While MPS platforms, such as PacBio and Oxford Nanopore, offer long-read technology in the 10's of Kb range, additional service providers, such as 10X Genomics and Dovetail's Chicago method, are focusing on using linked DNA to assemble the resulting contigs into scaffolds in the 100's of Kb range. Hi-C methods fix full chromosomes in situ with the goal of creating Mb-scale assemblies, and optical mapping can be used to create chromosome-scale assemblies. We should know soon whether these new approaches using medium- and long-range positional information will finally achieve the assembly of very large and highly repetitive genomes such as those of frogs or perhaps even of salamanders (Koepfli et al. 2015).

### *4.3 High Population Structure*

The tendency of amphibians to have low dispersal, and therefore high population structure, can reduce power to detect loci under positive selection using genome scans. The premise of genome scans is that loci with "outlier behavior," such as $F_{ST}$ values significantly higher than the genome-wide average site-wise $F_{ST}$ value, might be under positive selection (Beaumont and Nichols 1996). If genome-wide average $F_{ST}$ values are low due to high gene flow, then the power to identify high $F_{ST}$ outliers should be high. But if all loci have high $F_{ST}$ values because of restricted gene flow, then the power to identify loci under divergent selection may be reduced. Conversely, if most loci have high $F_{ST}$ values, some may be incorrectly identified as high $F_{ST}$ outliers, resulting in type I errors, a potentially more insidious problem.

Due to the potential for type I errors, multiple lines of independent evidence should be used to test whether outlier loci are truly under divergent selection and adaptive. First, if a reference genome is available for the study species or close relative, loci can be mapped to determine if they fall in or near exons or regulatory regions (Manel et al. 2016). Second, multiple independent environmental gradients of a given type of gradient (e.g., multiple, similar elevational gradients) should be sampled to test whether the same loci are identified as outliers consistently, providing additional evidence that the gradient of interest is the cause of directional selection. Third, genotype-by-environment association (GEA) tests can be used to determine whether allele frequencies at loci identified as outliers are correlated with the gradient (s) hypothesized to cause adaptive evolution (Joost et al. 2007; Coop et al. 2010; Frichot et al. 2013). Finally, experimental approaches should ideally be used to determine potential adaptive functions of the candidate loci. For example, reciprocal transplant experiments could be conducted in which tadpoles are moved to ponds in different environments to test how allele frequencies change at outliers in mismatched environments (Soria-Carrasco et al. 2014). The prediction is that allele frequency of immigrants at loci involved in local adaptation should become more similar to the frequency of residents. Alternatively, it may soon be feasible to use CRISPR/Cas gene editing to silence genes or replace specific alleles to test their effects on fitness in different environments. Thus, there are many different follow-up analyses that can be conducted to increase confidence that loci truly are involved in local adaptation.

# 5 Recommendations

## *5.1 Maximizing the Potential of Genomics to Transform Amphibian Research*

The possibilities for harnessing genomics to advance understanding of amphibian evolution, ecology, behavior, and conservation are essentially limitless. Ultimately, the general idea of MPS is not complex: MPS provides huge quantities of sequence data from wherever a researcher wants to look in the genome. It is up to us as scientists to select the best species and approaches to address long-standing questions in novel and exciting ways.

In Sects. 2 and 3 above, we pointed out different ways in which genomics will advance understanding of amphibian ecology, evolution, and behavior and improve conservation management of this imperiled taxon. For example, genomics provides tremendous opportunities to test for and understand the genetic basis of local adaptation. Many amphibian species occur across heterogeneous environments, suggesting the capacity to adapt to diverse habitat and climatic conditions, which can now be dissected using MPS. Genomics also provides unprecedented power to infer the history of diversification and speciation, providing the opportunity to understand how the diverse traits of amphibians (reproductive modes, behavior, habitat, size, morphology, etc.) influence speciation mechanisms. Amphibians have been the unfortunate victims of one of the worst disease epidemics ever known: chytridiomycosis caused by *Batrachochytrium dendrobatidis* and *B. salamandrivorans*. Some populations, however, have recovered after epidemics caused by these diseases, suggesting the evolution of resistance. Genomics can help identify individuals and populations that are resistant, for example, using genome-wide association tests (GWAS), and that could therefore serve as sources for supplementation of populations without resistance. Our newfound ability to peer deeply and broadly into the genomes of amphibians thanks to the emergence of MPS will greatly improve our understanding of their biology and our capacity to make informed decisions to conserve them.

To maximize the potential of genomics to transform our understanding of amphibian biology, we argue that herpetologists and genomicists should strive for three things: taxonomic diversity, multiple scales, and integration. By "taxonomic diversity," we mean that for strong inferences about evolutionary and ecological processes across all Amphibia, we need to investigate these processes in as many groups of amphibians as possible. Too often, a handful of species serve as models on which conclusions for an entire class of organisms are based. Yet amphibians are extremely diverse in terms of their life histories, reproductive modes, habitats, genome sizes and structure, and interactions with other species (Wells 2007; Zamudio et al. 2016b). The best way to understand ecological and evolutionary processes for all Lissamphibia is to study these processes in multiple taxonomic families.

"Multiple scales" refer to investigating questions at different levels of biological organization, from species to populations to individuals to tissues, which we argue provides the deepest understanding. For example, the strongest studies of adaptation examine (1) how species' traits are related to environmental variation using comparative phylogenetics at the species level, (2) patterns and function of phenotypic and genetic divergence among populations in relation to environmental heterogeneity, (3) the genetic basis of variation among individuals in traits, and (4) organismal physiology.

Finally, as we have highlighted throughout our chapter, integration of genomics with other established methods in ecology and evolution will provide the strongest inferences. These include field observations, field and laboratory experiments, and modeling. As scientists, we are attracted to the power of new technologies and have a tendency to dismiss "old" scientific approaches, but we must remember that in evolution, ecology, behavior, and conservation, no technological breakthrough can replace astute and careful observation, controlled experiments, and clear quantitative thinking formalized with models.

## *5.2 Getting Started*

We conclude with recommendations for those interested in applying MPS and genomics to a question about their favorite amphibian species, but who do not yet have experience in this area. Our first recommendation is to simply recognize that developing a genomics component to a research program will take time, patience, and determination. Learning how to prep MPS libraries and especially learning bioinformatics takes significant investments of time. Many things can go wrong with library preparations and computational analyses. Expect problems and allow time for troubleshooting. Second, team up with other researchers who are experts in genomics and bioinformatics. Although anyone who is determined and persistent can learn these approaches, the learning curve will be less steep with the guidance of an experienced collaborator. Third, get formal training in genomics and bioinformatics in courses or workshops. Not only will this rapidly increase one's knowledge of the field, it will also provide the opportunity to meet experts in the field, providing you with a network of colleagues whom you can call on to answer questions. The fourth and most important recommendation is to make sure one understands the fundamental principles of population genetics (Allendorf 2017). Ultimately, population genomics is nothing more than population genetics "writ large" and is based on the same powerful population genetics theory developed by early pioneers in the field (Fisher 1930; Haldane 1930; Wright 1931) and by more recent theoreticians (Kingman 1982). Piles of DNA sequence data are useless without an understanding of theory to come up with well-grounded, interesting hypotheses to test by rigorously analyzing data and correctly interpreting the results.

**Acknowledgments** We thank Daryl Trumbo, Emily Moriarty Lemmon, Mathew Fujita, and Guillermo Velo-Antón for providing helpful comments on the manuscript and Guillermo Velo-Antón and Ian J. Wang for photos. We acknowledge funding from the National Science Foundation Rules of Life grant (DEB 1838282) to WCF.

## References

Abbott JK, Norden AK, Hansson B. Sex chromosome evolution: historical insights and future perspectives. Proc R Soc B Biol Sci. 2017;284:20162806.

Adams EM, Jones AG, Arnold SJ. Multiple paternity in a natural population of a salamander with long-term sperm storage. Mol Ecol. 2005;14:1803–10.

Alexander AM, Su YC, Oliveros CH, Olson KV, Travers SL, Brown RM. Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. Evolution. 2017;71:475–88.

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30.

Allendorf FW, Phelps SR. Use of allelic frequencies to describe population structure. Can J Fish Aquat Sci. 1981;38:1507–14.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

Allendorf FW, Luikart G, Aitken SN. Conservation and the genetics of populations. 2nd ed. Oxford: Wiley-Blackwell; 2013.

AmphibiaWeb. University of California, Berkeley, 2018. http://amphibiaweb.org/. Accessed 30 Sept 2018.

Anderson E, Stebbins GL. Hybridization as an evolutionary stimulus. Evolution. 1954;8:378–88.

Andrés JA, Bogdanowicz SM. Isolating microsatellite loci: looking back, looking ahead. Methods Mol Biol. 2011;772:211–32.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.

Arnold ML. Natural hybridization and evolution. Oxford: Oxford University Press; 1997.

Ashman TL, Bachtrog D, Blackmon H, Goldberg EE, Hahn MW, Kirkpatrick M, et al. Tree of sex: a database of sexual systems. Sci Data. 2014;1:140015.

Austin JD, Gorman TA, Bishop D, Moler P. Genetic evidence of contemporary hybridization in one of North America's rarest anurans, the Florida bog frog. Anim Conserv. 2011;14:553–61.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 1987;18:489–522.

Avise JC, Jones AG, Walker D, DeWoody JA. Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. Annu Rev Genet. 2002;36:19–45.

Bachtrog D. A dynamic view of sex chromosome evolution. Curr Opin Genet Dev. 2006;16:578–85.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3:e3376.

Balkenhol N, Cushman SA, Storfer AT, Waits LP. Landscape genetics: concepts, methods, applications. Oxford: Wiley-Blackwell; 2016.

Barrow LN, Ralicki HF, Emme SA, Lemmon EM. Species tree estimation of north American chorus frogs (Hylidae: *Pseudacris*) with parallel tagged amplicon sequencing. Mol Phylogenet Evol. 2014;75:78–90.

Barrow LN, Soto-Centeno JA, Warwick AR, Lemmon AR, Lemmon EM. Evaluating hypotheses of expansion from refugia through comparative phylogeography of south-eastern Coastal Plain amphibians. J Biogeogr. 2017;44:2692–705.

Barrow LN, Lemmon AR, Lemmon EM. Targeted sampling and target capture: assessing phylogeographic concordance with genome-wide data. Syst Biol. 2018; https://doi.org/10.1093/sysbio/syy021/4948751.

Barton N, Bengtsson BO. The barrier to genetic exchange between hybridizing populations. Heredity. 1986;57:357–76.

Bataille A, Fong JJ, Cha M, Wogan GOU, Baek HJ, Lee H, et al. Genetic evidence for a high diversity and wide distribution of endemic strains of the pathogenic chytrid fungus *Batrachochytrium dendrobatidis* in wild Asian amphibians. Mol Ecol. 2013;22:4196–209.

Bataille A, Cashins SD, Grogan L, Skerratt LF, Hunter D, McFadden M, et al. Susceptibility of amphibians to chytridiomycosis is associated with MHC class II conformation. Proc R Soc B Biol Sci. 2015;282:20143127.

Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 2004;13:969–80.

Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc B Biol Sci. 1996;263:1619–26.

Beebee TJC. Conservation genetics of amphibians. Heredity. 2005;95:423–7.

Bell RC, MacKenzie JB, Hickerson MJ, Chavarria KL, Cunningham M, Williams S, et al. Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. Proc R Soc B Biol Sci. 2012;279:991–9.

Bell RC, Drewes RC, Zamudio KR. Reed frog diversification in the Gulf of Guinea: overseas dispersal, the progression rule, and in situ speciation. Evolution. 2015;69:904–15.

Bell RC, Parra JL, Badjedjea G, Barej MF, Blackburn DC, Burger M, et al. Idiosyncratic responses to climate-driven forest fragmentation and marine incursions in reed frogs from Central Africa and the Gulf of Guinea Islands. Mol Ecol. 2017;26:5223–44.

Bernardi G, Wiley EO, Mansour H, Miller MR, Orti G, Haussler D, et al. The fishes of genome 10K. Mar Genomics. 2012;7:3–6.

Berven KA. The genetic basis of altitudinal variation in the wood frog *Rana sylvatica*. I. An experimental analysis of life history traits. Evolution. 1982;36:962–83.

Bhattacharya D, Marfo CA, Li D, Lane M, Khokha MK. CRISPR/Cas9: an inexpensive, efficient loss of function tool to screen human disease genes in *Xenopus*. Dev Biol. 2015;408:196–204.

Black WC IV, Baer CF, Antolin MF, DuTeau NM. Population genomics: genome-wide sampling of insect populations. Annu Rev Entomol. 2001;46:441–69.

Bogart JP. Evolution of anuran karyotypes. In: Vial JL, editor. Evolutionary biology of the anurans. Missouri: University of Missouri Press; 1973. p. 337–49.

Burton TM, Likens GE. Energy flow and nutrient cycling in salamander populations in the Hubbard Brook Experimental Forest, New Hampshire. Ecology. 1975;56:1068–80.

Camacho-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA. Preservation of RNA and DNA from mammal samples under field conditions. Mol Ecol Resour. 2013;13:663–73.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 2011;108:4516–22.

Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362–5.

Chakrabarty P, Faircloth BC, Alda F, Ludt WB, McMahan CD, Near TJ, et al. Phylogenomic systematics of Ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of Characiformes. Syst Biol. 2017;66:881–95.

Che RB, Sun YN, Wang RX, Xu TJ. Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers. PLoS One. 2014;9:e87940.

Chen YH, Cheng WC, Yu HT, Kam YC. Genetic relationship between offspring and guardian adults of a rhacophorid frog and its care effort in response to paternal share. Behav Ecol Sociobiol. 2011;65:2329–39.

Claussen J, Keck DD, Hiesey WM. Experimental studies on the nature of species. III. Environmental responses of climatic races of *Achillea*. Washington: Carnegie Institution of Washington Publication; 1948. p. 581.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185:1411–23.

Crawford AJ. Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. Mol Ecol. 2003;12:2525–40.

Crawford AJ, Lips KR, Bermingham E. Epidemic disease decimates amphibian abundance, species diversity, and evolutionary history in the highlands of central Panama. Proc Natl Acad Sci U S A. 2010;107:13777–82.

Crispo E. Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. J Evol Biol. 2008;21:1460–9.

D'Aloia CC, Bogdanowicz SM, Harrison RG, Buston PM. Cryptic genetic diversity and spatial patterns of admixture within Belizean marine reserves. Conserv Genet. 2017;18:211–23.

Denton RD, Kudra RS, Malcom JW, Du Preez L, Malone JH. The African Bullfrog (*Pyxicephalus adspersus*) genome unites the two ancestral ingredients for making vertebrate sex chromosomes. bioRxiv. 2018; https://doi.org/10.1101/329847.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NEESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size ($N_e$) from genetic data. Mol Ecol Resour. 2014;14:209–14.

Dowling TE, Secor CL. The role of hybridization and introgression in the diversification of animals. Annu Rev Ecol Syst. 1997;28:593–619.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73.

Duellman WE, Trueb L. Biology of amphibians. New York: McGraw-Hill Book; 1986.

Eckert CG, Samis KE, Lougheed SC. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol. 2008;17:1170–88.

Edmands S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol. 2007;16:463–75.

Edwards RJ, Tuipulotu DE, Amos TG, O'Meally D, Richardson MF, Russell TL, et al. Draft genome assembly of the invasive cane toad, *Rhinella marina*. GigaScience. 2018;7:giy095.

Elewa A, Wang H, Talavera-López C, Joven A, Brito G, Kumar A, Hameed LS, Penrad-Mobayed-M, Yao Z, Zamani N, Abbas Y, Abdullayev I, Sandberg R, Grabherr M, Andersson B, Simon A. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. Nat Commun. 2017;8:2286.

Elinson RP. Inheritance and expression of a sex-linked enzyme in the frog, *Rana clamitans*. Biochem Genet. 1983;21:435–42.

Elliott TA, Gregory TR. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos Trans R Soc B Biol Sci. 2015;370:20140331.

Ellison AR, Savage AE, DiRenzo GV, Langhammer P, Lips KR, Zamudio KR. Fighting a losing battle: vigorous immune response countered by pathogen suppression of host defenses in the chytridiomycosis-susceptible frog *Atelopus zeteki*. G3: Genes Genom Genet. 2014;4:1275–89.

Ellison AR, Tunstall T, DiRenzo GV, Hughey MC, Rebollar EA, Belden LK, et al. More than skin deep: functional genomic basis for resistance to amphibian chytridiomycosis. Genome Biol Evol. 2015;7:286–98.

Ellison AR, DiRenzo GV, McDonald CA, Lips KR, Zamudio KR. First in vivo *Batrachochytrium dendrobatidis* transcriptomes reveal mechanisms of host exploitation, host-specific gene expression, and expressed genotype shifts. G3: Genes Genom Genet. 2017;7:269–78.

Endler JA. Natural selection in the wild. Monogr Popul Biol. 1986;21:1–336.

Ezaz T, Sarre SD, O'Meally D, Graves JAM, Georges A. Sex chromosome evolution in lizards: independent origins and rapid transitions. Cytogenet Genome Res. 2009;127:249–60.
Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol. 2012;61:717–26.
Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61:1–10.
Farrer RA, Weinert LA, Bielby J, Garner TWJ, Balloux F, Clare F, et al. Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. Proc Natl Acad Sci U S A. 2011;108:18732–6.
Farrer RA, Henk DA, Garner TWJ, Balloux F, Woodhams DC, Fisher MC. Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. PLoS Genet. 2013;9:e1003703.
Farrer RA, Martel A, Verbrugghe E, Abouelleil A, Ducatelle R, Longcore JE, et al. Genomic innovations linked to infection strategies across emerging pathogenic chytrid fungi. Nat Commun. 2017;8:14742.
Fei JF, Schuez M, Tazaki A, Taniguchi Y, Roensch K, Tanaka EM. CRISPR-mediated genomic deletion of Sox2 in the axolotl shows a requirement in spinal cord neural stem cell amplification during tail regeneration. Stem Cell Rep. 2014;3:444–59.
Feng YJ, Blackburn DC, Liang D, Hillis DM, Wake DB, Cannatella DC, et al. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. Proc Natl Acad Sci U S A. 2017;114:E5864–E70.
Fernandez ME, Goszczynski DE, Liron JP, Villegas-Castagnasso EE, Carino MH, Ripoli MV, et al. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genet Mol Biol. 2013;36:185–U94.
Ficetola GF, Bonin A. Conserving adaptive genetic diversity in dynamic landscapes. Mol Ecol. 2011;20:1569–71.
Ficetola GF, Stock M. Do hybrid-origin polyploid amphibians occupy transgressive or intermediate ecological niches compared to their diploid ancestors? J Biogeogr. 2016;43:703–15.
Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930.
Fisher MC, Garner TWJ, Walker SF. Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. Annu Rev Microbiol. 2009;63:291–310.
Fites JS, Ramsey JP, Holden WM, Collier SP, Sutherland DM, Reinert LK, et al. The invasive chytrid fungus of amphibians paralyzes lymphocyte responses. Science. 2013;342:366–9.
Fitzpatrick BM, Shaffer HB. Hybrid vigor between native and introduced salamanders raises new challenges for conservation. Proc Natl Acad Sci U S A. 2007;104:15793–8.
Fitzpatrick BM, Johnson JR, Kump DK, Shaffer HB, Smith JJ, Voss SR. Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. BMC Evol Biol. 2009;9:176.
Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB. Rapid spread of invasive genes into a threatened native species. Proc Natl Acad Sci U S A. 2010;107:3606–10.
Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180:977–93.
Fong JJ, Brown JM, Fujita MK, Boussau B. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. PLoS One. 2012;7:e48990.
Francois O, Martins H, Caye K, Schoville SD. Controlling false discoveries in genome scans for selection. Mol Ecol. 2016;25:454–69.
Frankham R, Ballou JD, Eldridge MDB, Lacy RC, Ralls K, Dudash MR, et al. Predicting the probability of outbreeding depression. Conserv Biol. 2011;25:465–75.
Frichot E, Schoville SD, Bouchard G, Francois O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 2013;30:1687–99.

Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, et al. The amphibian tree of life. Bull Am Mus Nat Hist. 2006;297:1–370.
Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol Ecol. 2017;26:5369–406.
Funk WC, Greene AE, Corn PS, Allendorf FW. High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation. Biol Lett. 2005;1:13–6.
Funk WC, Caldwell JP, Peden CE, Padial JM, Riva IDL, Cannatella DC. Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. Mol Phylogenet Evol. 2007;44:825–37.
Funk WC, Pearl CA, Draheim HM, Adams MJ, Mullins TD, Susan MH. Range-wide phylogeographic analysis of the spotted frog complex (*Rana luteiventris* and *Rana pretiosa*) in northwestern North America. Mol Phylogenet Evol. 2008;49:198–210.
Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.
Funk WC, Murphy MA, Hoke KL, Muths E, Amburgey SM, Lemmon EM, et al. Elevational speciation in action? Restricted gene flow associated with adaptive divergence across an altitudinal gradient. J Evol Biol. 2016;29:241–52.
Gamble T, Zarkower D. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Mol Ecol Resour. 2014;14:902–13.
Gamble T, Coryell J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D. Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. Mol Biol Evol. 2015;32:1296–309.
García-R JC, Crawford AJ, Mendoza AM, Ospina O, Cardenas H, Castro F. Comparative phylogeography of direct-developing frogs (Anura: Craugastoridae: *Pristimantis*) in the southern Andes of Colombia. PLoS One. 2012;7:e46077.
Garrick RC, Bonatelli IAS, Hyseni C, Morales A, Pelletier TA, Perez MF, et al. The evolution of phylogeographic data sets. Mol Ecol. 2015;24:1164–71.
Gazoni T, Haddad CFB, Narimatsu H, Cabral-de-Mello DC, Lyra ML, Parise-Maltempi PP. More sex chromosomes than autosomes in the Amazonian frog *Leptodactylus pentadactylus*. Chromosoma. 2018;127:269–78.
Gerchen JF, Reichert SJ, Rohr JT, Dieterich C, Kloas W, Stock M. A single transcriptome of a green toad (*Bufo viridis*) yields candidate genes for sex determination and -differentiation and non-anonymous population genetic markers. PLoS One. 2016;11:e0156419.
Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature. 2015;525:372–5.
Glaubitz JC, Rhodes OE, Dewoody JA. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol Ecol. 2003;12:1039–47.
Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11:759–69.
Gomez-Mestre I, Pyron RA, Wiens JJ. Phylogenetic analyses reveal unexpected patterns in the evolution of reproductive modes in frogs. Evolution. 2012;66:3687–700.
Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC. Homoploid hybrid speciation in an extreme habitat. Science. 2006;314:1923–5.
Grant PR, Grant BR. Hybridization of bird species. Science. 1992;256:193–7.
Grant BR, Grant PR. Fission and fusion of Darwin's finches populations. Philos Trans R Soc B Biol Sci. 2008;363:2821–9.
Gregory TR. Animal genome size database. 2011. http://www.genomesize.com/.
Griffith SC, Owens IPF, Thuman KA. Extra pair paternity in birds: a review of interspecific variation and adaptive function. Mol Ecol. 2002;11:2195–212.
Guerrero RF, Kirkpatrick M, Perrin N. Cryptic recombination in the ever-young sex chromosomes of hylid frogs. J Evol Biol. 2012;25:1947–54.
Guo BC, Lu D, Liao WB, Merila J. Genomewide scan for adaptive differentiation along altitudinal gradient in the Andrew's toad *Bufo andrewsi*. Mol Ecol. 2016;25:3884–900.

Haldane JBS. A mathematical theory of natural and artificial selection. VI. Isolation. Proc Camb Philos Soc. 1930;26:220–30.
Hammond SA, Warren RL, Vandervalk BP, Kucuk E, Khan H, Gibb EA, et al. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nat Commun. 2017;8:1433.
Hardie DC, Gregory TR, Hebert PDN. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. J Histochem Cytochem. 2002;50:735–49.
Harding G, Griffiths RA, Pavajeau L. Developments in amphibian captive breeding and reintroduction programs. Conserv Biol. 2016;30:340–9.
Harrison RG. Hybrids and hybrid zones: historical perspective. In: Hybrid zones and the evolutionary process. Oxford: Oxford University Press; 1993. p. 3–12.
Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Mol Ecol Resour. 2011;11:150–61.
Haussler D, O'Brien SJ, Ryder OA, Barker FK, Clamp M, Crawford AJ, et al. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered. 2009;100:659–74.
Hayes TB. Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. J Exp Zool. 1998;281:373–99.
Heinicke MP, Lemmon AR, Lemmon EM, McGrath K, Hedges SB. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). Mol Phylogenet Evol. 2018;118:145–55.
Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, et al. The genome of the western clawed frog *Xenopus tropicalis*. Science. 2010;328:633–6.
Hillis DM, Green DM. Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. J Evol Biol. 1990;3:49–64.
Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188:379–97.
Hoffmann M, Hilton-Taylor C, Angulo A, Bohm M, Brooks TM, Butchart SHM, et al. The impact of conservation on the status of the world's vertebrates. Science. 2010;330:1503–9.
Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010;6:e1000862.
Huang L, Li J, Anboukaria H, Luo ZH, Zhao M, Wu H. Comparative transcriptome analyses of seven anurans reveal functions and adaptations of amphibian skin. Sci Rep. 2016;6:24069.
Hughes C. Integrating molecular techniques with field methods in studies of social behavior: a revolution results. Ecology. 1998;79:383–99.
Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire JY, Kupfer A, et al. Phylotranscriptomic consolidation of the jawed vertebrate timetree. Nat Ecol Evol. 2017;1:1370–8.
Isaac NJB, Redding DW, Meredith HM, Safi K. Phylogenetically-informed priorities for amphibian conservation. PLoS One. 2012;7:e43912.
James TY, Litvintseva AP, Vilgalys R, Morgan JAT, Taylor JW, Fisher MC, et al. Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. PLoS Pathog. 2009;5:e1000458.
James TY, Toledo LF, Rodder D, Leite DD, Belasen AM, Betancourt-Roman CM, et al. Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. Ecol Evol. 2015;5:4079–97.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014;346(6215):1320–31.
Jockusch EL. An evolutionary correlate of genome size change in plethodontid salamanders. Proc R Soc B Biol Sci. 1997;264(1381):597–604.

Johnson WE, Koepfli K. The role of genomics in conservation and reproductive sciences. Adv Exp Med Biol. 2014;753:71–96.
Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25:185–202.
Joost S, Bonin A, Bruford MW, Despres L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol. 2007;16(18):3955–69.
Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Mol Ecol Resour. 2017;17:183–93.
Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9:1205–18.
Kimura M, Crow JF. The measurement of the effective population number. Evolution. 1963;17:279–88.
Kingman JFC. The coalescent. Stoch Process Appl. 1982;13:235–48.
Koepfli KP, Paten B, O'Brien SJ, Scientists GKC. The Genome 10K Project: a way forward. Annu Rev Anim Biosci. 2015;3:57–111.
Kupfer A, Wilkinson M, Gower DJ, Muller H, Jehle R. Care and parentage in a skin-feeding caecilian amphibian. J Exp Zool A Ecol Genet Physiol. 2008;309a:460–7.
Lambert MR, Skelly DK, Ezaz T. Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DArTseq provide early insight into sex chromosome evolution. BMC Genomics. 2016;17:844.
Laurila A, Seppa P. Multiple paternity in the common frog (*Rana temporaria*): genetic evidence from tadpole kin groups. Biol J Linn Soc. 1998;63:221–32.
Lemmon EM, Juenger TE. Geographic variation in hybridization across a reinforcement contact zone of chorus frogs (*Pseudacris*). Ecol Evol. 2017;7:9485–502.
Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol. 2012;61:727–44.
Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat Ecol Evol. 2018;2:1792–9. https://doi.org/10.1038/s41559-018-0674-4.
Longcore JE, Pessier AP, Nichols DK. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. Mycologia. 1999;91:219–27.
Longo AV, Burrowes PA, Zamudio KR. Genomic studies of disease-outcome in host-pathogen dynamics. Integr Comp Biol. 2014;54:427–38.
Lopes CM, Sasso T, Valentini A, Dejean T, Martins M, Zamudio KR, et al. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Mol Ecol Resour. 2017;17:904–14.
Lowe WH, Likens GE, McPeek MA, Buso DC. Linking direct and indirect data on dispersal: isolation by slope in a headwater stream salamander. Ecology. 2006;87(2):334–9.
Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2017;17:142–52.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94.
Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20:229–37.
Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 2003;18:189–97.
Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, et al. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. Mol Ecol. 2016;25:170–84.

Martel A, Spitzen-van der Sluijs A, Blooi M, Bert W, Ducatelle R, Fisher MC, et al. Batrachochytrium salamandrivorans sp. nov causes lethal chytridiomycosis in amphibians. Proc Natl Acad Sci U S A. 2013;110:15325–9.

Martel A, Blooi M, Adriaensen C, Van Rooij P, Beukema W, Fisher MC, et al. Recent introduction of a chytrid fungus endangers Western Palearctic salamanders. Science. 2014;346:630–1.

Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M. Speciation by hybridization in *Heliconius* butterflies. Nature. 2006;441:868–71.

McCartney-Melstad E, Shaffer HB. Amphibian molecular ecology and how it has informed conservation. Mol Ecol. 2015;24:5084–109.

McCartney-Melstad E, Mount GG, Shaffer HB. Exon capture optimization in amphibians with large genomes. Mol Ecol Resour. 2016;16:1084–94.

McCormack JE, Faircloth BC. Next-generation phylogenetics takes root. Mol Ecol. 2013;22:19–21.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol. 2013;66:526–38.

McKay JK, Latta RG. Adaptive population divergence: markers, QTL and traits. Trends Ecol Evol. 2002;17:285–91.

McKenzie VJ, Bowers RM, Fierer N, Knight R, Lauber CL. Co-habiting amphibian species harbor unique skin bacterial communities in wild populations. ISME J. 2012;6:588–96.

Miura I, Ohtani H, Nakamura M, Ichikawa Y, Saitoh K. The origin and differentiation of the heteromorphic sex chromosomes Z, W, X, and Y in the frog *Rana rugosa*, inferred from the sequences of a sex-linked gene, ADP/ATP translocase. Mol Biol Evol. 1998;15:1612–9.

Morgan JAT, Vredenburg VT, Rachowicz LJ, Knapp RA, Stice MJ, Tunstall T, et al. Population genetics of the frog-killing fungus *Batrachochytrium dendrobatidis*. Proc Natl Acad Sci U S A. 2007;104:13845–50.

Muralidhar P, de Sa FP, Haddad CFB, Zamudio KR. Kin-bias, breeding site selection and female fitness in a cannibalistic Neotropical frog. Mol Ecol. 2014;23:453–63.

Myers EM, Zamudio KR. Multiple paternity in an aggregate breeding amphibian: the effects of reproductive skew on estimates of male reproductive success. Mol Ecol. 2004;13:1951–63.

Nakamura M. Sex determination in amphibians. Semin Cell Dev Biol. 2009;20(3):271–82.

Nali RC, Zamudio KR, Prado CPA. Microsatellite markers for *Bokermannohyla species* (Anura, Hylidae) from the Brazilian Cerrado and Atlantic Forest domains. Amphibia-Reptilia. 2014;35:355–60.

Newman CE, Austin CC. Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Mol Ecol. 2016;25:6162–74.

Nowoshilow S, Schloissnig S, Fei JF, Dahl A, Pang AWC, Pippel M, et al. The axolotl genome and the evolution of key tissue formation regulators. Nature. 2018;554:50–5.

Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. Mol Ecol. 2017;26:1060–74.

Olson DH, Aanensen DM, Ronnenberg KL, Powell CI, Walker SF, Bielby J, et al. Mapping the global emergence of *Batrachochytrium dendrobatidis*, the amphibian chytrid fungus. PLoS One. 2013;8:e56802.

Palo JU, O'Hara RB, Laugen AT, Laurila A, Primmers CR, Merilä J. Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. Mol Ecol. 2003;12:1963–78.

Paz A, Ibanez R, Lips KR, Crawford AJ. Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. Mol Ecol. 2015;24:3723–37.

Peloso PLV, Frost DR, Richards SJ, Rodrigues MT, Donnellan S, Matsui M, et al. The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). Cladistics. 2016;32:113–40.

Pennell MW, Mank JE, Peichel CL. Transitions in sex determination and sex chromosomes across vertebrate species. Mol Ecol. 2018;27:3950–63. https://doi.org/10.1111/mec.14540.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7:e37135.
Pie MR, Faircloth BC, Ribeiro LF, Bornschein MR, McCormack JE. Phylogenomics of montane frogs of the Brazilian Atlantic Forest is consistent with isolation in sky islands followed by climatic stability. Biol J Linn Soc. 2018;125:72–82.
Polato NR, Gray MM, Gill BB, Becker CG, Casner KL, Flecker AS, et al. Genetic diversity and gene flow decline with elevation in montane mayflies. Heredity. 2017;119:107–16.
Polato NR, Gill BA, Shah AA, Gray MM, Casner KL, Barthelet A, et al. Narrow thermal tolerance and low dispersal drive higher speciation in tropical mountains. Proc Natl Acad Sci U S A. 2018; https://doi.org/10.1073/pnas.1809326115.
Portik DM, Smith LL, Bi K. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol Ecol Resour. 2016;16:1069–83.
Pritchard JK. Whole-genome sequencing data offer insights into human demography. Nat Genet. 2011;43:923–5.
Pyron AR, Wiens JJ. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol. 2011;61:543–83.
Rhymer JM, Simberloff D. Extinction by hybridization and introgression. Annu Rev Ecol Syst. 1996;27:83–109.
Richter-Boix A, Quintela M, Segelbacher G, Laurila A. Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. Mol Ecol. 2011;20:1582–600.
Riley SPD, Shaffer HB, Voss SR, Fitzpatrick BM. Hybridization between a rare, native tiger salamander (*Ambystoma californiense*) and its introduced congener. Ecol Appl. 2003;13:1263–75.
Ringler E, Ringler M, Jehle R, Hodl W. The female perspective of mating in *A. femoralis*, a territorial frog with paternal care – a spatial and genetic analysis. PLoS One. 2012;7:e40237.
Rittmeyer EN, Allison A, Grundler MC, Thompson DK, Austin CC. Ecological guild evolution and the discovery of the world's smallest vertebrate. PLoS One. 2012;7:e29797.
Robertson LS, Cornman RS. Transcriptome resources for the frogs *Lithobates clamitans* and *Pseudacris regilla*, emphasizing antimicrobial peptides and conserved loci for phylogenetics. Mol Ecol Resour. 2014;14:178–83.
Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, et al. Global patterns of diversification in the history of modern amphibians. Proc Natl Acad Sci U S A. 2007;104:887–92.
Rogers RL, Zhou L, Chu C, Márquez R, Corl A, Linderoth T, et al. Genomic takeover by transposable elements in the strawberry poison frog. Mol Biol Evol. 2018; https://doi.org/10.1093/molbev/msy185.
Rokas A, Abbot P. Harnessing genomics for evolutionary insights. Trends Ecol Evol. 2009;24:192–200.
Rosenblum EB, James TY, Zamudio KR, Poorten TJ, Ilut D, Rodriguez D, et al. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proc Natl Acad Sci U S A. 2013;110:9385–90.
Ruane S, Raxworthy CJ, Lemmon AR, Lemmon EM, Burbrink FT. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evol Biol. 2015;15:221.
Ryan ME, Johnson JR, Fitzpatrick BM, Lowenstine LJ, Picco AM, Shaffer HB. Lethal effects of water quality on threatened California salamanders but not on co-occurring hybrid salamanders. Conserv Biol. 2013;27:95–102.

Salthe SN, Duellman WE. Quantitative constraints associated with reproductive mode in anurans. In: Vial JL, editor. Evolutionary biology of anurans: contemporary research on major problems. Columbia: University of Missouri Press; 1973. p. 229–49.
Savage AE, Zamudio KR. MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci U S A. 2011;108:16705–10.
Schloegel LM, Toledo LF, Longcore JE, Greenspan SE, Vieira CA, Lee M, et al. Novel, panzootic and hybrid genotypes of amphibian chytridiomycosis associated with the bullfrog trade. Mol Ecol. 2012;21:5162–77.
Schmid M, Steinlein C, Bogart JP, Feichtinger W, Leon P, La Marca E, et al. The chromosomes of Terraranan frogs: insights into vertebrate cytogenetics. Cytogenet Genome Res. 2010;130:15–557.
Schmid M, Evans B, Bogart JP. Polyploidy in Amphibia. Cytogenet Genome Res. 2015;145:315–30.
Seigel RA, Dodd CK Jr. Translocation of amphibians: proven management method or experimental technique? Conserv Biol. 2002;16:552–4.
Session AM, Uno Y, Kwon T, Hapman JAC, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature. 2016;538:336–43.
Shaffer HB, Gidis M, McCartney-Melstad E, Neal KM, Oyamaguchi HM, Tellez M, et al. Conservation genetics and genomics of amphibians and reptiles. Annu Rev Anim Biosci. 2015;3:113–38.
Shen XX, Liang D, Feng YJ, Chen MY, Zhang PA. Versatile and highly efficient toolkit including 102 nuclear markers for vertebrate phylogenomics, tested by resolving the higher level relationships of the Caudata. Mol Biol Evol. 2013;30:2235–48.
Slatkin M. Estimating levels of gene flow in natural populations. Genetics. 1981;99:323–35.
Smith JJ, Timoshevskaya N, Timoshevskiy VA, Keinath MC, Hardy D, Voss SR. A chromosome-scale assembly of the enormous (32 Gb) axolotl genome. bioRxiv. 2018:373548. https://doi.org/10.1101/373548.
Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 2006;15:4261–93.
Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick insect genomes reveal natural selection's role in parallel speciation. Science. 2014;344:738–42.
Stegen G, Pasmans F, Schmidt BR, Rouffaer LO, Van Praet S, Schaub M, et al. Drivers of salamander extirpation mediated by *Batrachochytrium salamandrivorans*. Nature. 2017;544:353–6.
Steiner CC, Putnam AS, Hoeck PEA, Ryder OA. Conservation genomics of threatened animal species. Annu Rev Anim Biosci. 2013;1:261–81.
Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity. 2008;100:158–70.
Storfer A, Mech SG, Reudink MW, Ziemba RE, Warren J, Collins JP. Evidence for introgression in the endangered Sonora tiger salamander, *Ambystoma tigrinum stebbinsi* (Lowe). Copeia. 2004;2004:783–396.
Storfer A, Eastman JM, Spear SF. Modern molecular methods for amphibian conservation. Bioscience. 2009;59:559–71.
Streicher JW, Miller EC, Guerrero PC, Correa C, Ortiz JC, Crawford AJ, et al. Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. Mol Phylogenet Evol. 2018;119:128–43.
Summers K, Amos W. Behavioral, ecological, and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*. Behav Ecol. 1997;8:260–7.
Sun C, Shepard DB, Chong RA, Arriaza JL, Hall K, Castoe TA, et al. LTR Retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol. 2012;4:168–83.
Sun YB, Xiong ZJ, Xiang XY, Liu SP, Zhou WW, Tu XL, et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proc Natl Acad Sci U S A. 2015;112:E1257–E62.
Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. Environmental DNA. Mol Ecol. 2012;21:1789–93.

Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. Mol Ecol Resour. 2018;18:953–65.
Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr Biol. 2016;26:2313–8.
Tokarska M, Marshall T, Kowalczyk R, Wojcik JM, Pertoldi C, Kristensen TN, et al. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity. 2009;103:326–32.
Trenham PC, Marsh DM. Amphibian translocation programs: reply to Seigel and Dodd. Conserv Biol. 2002;16:555–6.
Trumbo DR, Epstein B, Hohenlohe PA, Alford RA, Schwarzkopf L, Storfer A. Mixed population genomics support for the central marginal hypothesis across the invasive range of the cane toad (Rhinella marina) in Australia. Mol Ecol. 2016;25:4161–76.
Urban MC, Richardson JL, Freidenfelds NA, Drake DL, Fischer JF, Saunders PP. Microgeographic adaptation of wood frog tadpoles to an apex predator. Copeia. 2017;105:451–61.
Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016;25:929–42.
Vieites DR, Nieto-Román S, Barluenga M, Palanca A, Vences M, Meyer A. Post-mating clutch piracy in an amphibian. Nature. 2004;431:305–8.
Vinogradov AE. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry. 1998;31:100–9.
Vorburger C, Reyer HU. A genetic mechanism of species replacement in European waterfrogs? Conserv Genet. 2003;4:141–55.
Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.
Wang G-D, Zhang B-L, Zhou W-W, Li Y-X, Jin J-Q, Shao Y, et al. Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*. Proc Natl Acad Sci U S A. 2018;115:E5056–E65.
Ward RD, Skibinski DOF, Woodwark M. Protein heterozygosity, protein-structure, and taxonomic differentiation. Evol Biol. 1992;26:73–159.
Weinman LR, Solomon JW, Rubenstein DR. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. Mol Ecol Resour. 2015;15:502–11.
Weir BS, Anderson AD, Hepler AB. Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet. 2006;7:771–80.
Welch AM, Semlitsch RD, Gerhardt HC. Call duration as an indicator of genetic quality in male gray tree frogs. Science. 1998;280:1928–30.
Wells KD. The ecology and behavior of amphibians. Chicago: University of Chicago Press; 2007.
Wilhelm J, Pingoud A, Hahn M. Real-time PCR-based method for the estimation of genome sizes. Nucleic Acids Res. 2003;31:e56.
Wong PBY, Wiley EO, Johnson WE, Ryder OA, O'Brien SJ, Haussler D, et al. Tissue sampling methods and standards for vertebrate genomics. Gigascience. 2012;1:8.
Woodhams DC, Alford RA, Briggs CJ, Johnson M, Rollins-Smith LA. Life-history trade-offs influence disease in changing climates: strategies of an amphibian pathogens. Ecology. 2008;89:1627–39.
Wright S. Evolution in Mendelian populations. Genetics. 1931;16:97–159.
Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 1965;19:395–420.
Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. Nat Rev Genet. 2012;13:329–42.

Yang WZ, Qi Y, Bi K, Fu JZ. Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. BMC Genomics. 2012;13:588.

Zamudio KR, Savage WK. Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in spotted salamanders (*Ambystoma maculatum*). Evolution. 2003;57:1631–52.

Zamudio KR, Harrison RG, Matocq M. Hybridization in threatened and endangered animal taxa: implications for conservation and management of biodiversity. In: DeWoody A, Bickham J, Michler C, Nichols K, Rhodes G, Woeste K, editors. Molecular approaches in natural resource conservation and management. Cambridge: Cambridge University Press; 2010. p. 169–89.

Zamudio KR, Bell RC, Mason NA. Phenotypes in phylogeography: species' traits, environmental variation, and vertebrate diversification. Proc Natl Acad Sci U S A. 2016a;113:8041–8.

Zamudio KR, Bell RC, Nali RC, Haddad CFB, Prado CPA. Polyandry, predation, and the evolution of frog reproductive modes. Am Nat. 2016b;188:S41–61.

Zeisset I, Beebee TJC. Amphibian phylogeography: a model for understanding historical aspects of species distributions. Heredity. 2008;101:109–19.

Zhai WW, Nielsen R, Slatkin M. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol Biol Evol. 2009;26:273–83.

Zhang P, Wake DB. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. Mol Phylogenet Evol. 2009;53:492–508.

Zhang GJ, Li C, Li QY, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346(6215):1311–20.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. Parallel molecular evolution in an herbivore community. Science. 2012;337:1634–7.

# Population Genomics of Birds: Evolutionary History and Conservation


**David P. L. Toews, Jennifer Walsh, and Leonardo Campagna**

**Abstract** The use of genome-scale data to understand the evolutionary history of birds has provided important progress in the field of evolutionary biology and conservation. Here we review the conceptual advances of avian genomics, along with key examples from the literature. In each section, we contrast studies that utilized only a small number of genetic markers to studies that incorporated many independent loci across the genome. We discuss the important characteristics of avian genome architecture, and we explore the connections between DNA sequence variation and ecologically relevant phenotypes, such as color and morphology. We ask how environmental factors have left their mark on the genomes of birds and how genomic data can be used to reconstruct histories across multiple species. We outline how admixture and reticulate evolutionary histories have been an important source of variation and review cases in which hybridization has possibly led to the formation of new species. Finally, we discuss how genomic data have helped delineate population structure and inform conservation actions in declining avian species. Like in other taxonomic groups, the ever-expanding molecular toolbox for avian biologists is at once becoming more accessible in cost and more powerful in its applications. Therefore, the study of avian genomes will continue to provide important insights into many aspects of ecology, evolutionary history, and conservation biology.

**Keywords** Aves · Birds · Evolution · Genomics · High-throughput sequencing

D. P. L. Toews (✉) · J. Walsh · L. Campagna
Cornell Lab of Ornithology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA
e-mail: toews@cornell.edu

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_33,

## 1 Introduction

Birds have long been central to key concepts in evolutionary and conservation biology. Moreover, many central thinkers in these fields have been inspired by – and draw key examples from – avian systems. The beaks of Galápagos finches adorn the pages of most introductory biology textbooks and illustrate rapid morphological evolution under natural selection (Futuyma 2013). Bird-of-paradise plumes exemplify sexual selection in the extreme (Diamond 1986). The declining greater sage-grouse is a conservation icon of the sagebrush steppe in Western North America (Connelly et al. 2011). Given the fundamental importance of avian systems in answering scientific questions, it seems fitting then that two birds – the chicken (*Gallus gallus*) and the zebra finch (*Taeniopygia guttata*) – were among the first nonhuman vertebrates to have their genomes fully sequenced (in 2004 and 2010, respectively). Building upon the foundational resources of these two genomes has been important for the development of genomic tools for researchers and managers working in wild populations of other avian taxa. Compared to other animals, we have now generated substantial knowledge about the natural history of many bird species, which has only grown with the integration of massive electronic datasets gathered from citizen scientists across the world. The combination of these three facets – a long history within evolutionary thought, strong genomic foundations, and a deep knowledge of natural history – suggests that birds are positioned to provide key insights into evolutionary biology and wildlife genomics in years to come.

In the last decade, the application of genomic methods to the study of avian systems has burgeoned, as it has with many other taxonomic groups. Here we discuss how genomic data have been used to address important basic and applied questions regarding avian evolution and conservation. Our focus is mainly conceptual, and we frame our discussion around a selection of carefully chosen examples from the avian literature to illustrate important ideas. For a more technical summary of genomic methods, and how they have been applied to questions of avian population genetics and speciation, we refer readers to the reviews by Toews et al. (2016a) and Kraus and Wink (2015). These are complemented by the work of Oyler-McCance et al. (2016), which reviews several practical considerations for ecologists and conservationists in the context of avian genomics.

In this chapter, we begin with a discussion of the very basis for genomic analysis: the genome itself. We give an overview of the important characteristics of avian genomes, how these characteristics have helped genomic analyses in birds, and then discuss instances of within-species variation in genome architecture that has given rise to relevant phenotypic variation. We then explore the important connections between DNA sequence variation and ecologically relevant phenotypes in more detail, asking how genomic data have helped link the two. We then turn to how environmental factors, such as dispersal barriers or climatic shifts, have left their mark on the genomes of birds and how genomic data can be used to reconstruct these histories across multiple species. We then discuss the genomic consequences of hybridization between related taxa and the role that this source of genetic variation plays in avian evolution. Finally, we emphasize how these new data and analyses

have helped delineate relevant units of biodiversity and, subsequently, inform conservation actions.

In each section, we contrast studies that applied genetic tools that focused on a handful of markers, usually in predetermined regions of the genome, to studies that incorporate many independent genetic loci scattered across the genome. This will become an overarching theme of the chapter: with the increased power afforded by genome-wide data, the ever-expanding molecular toolbox for biologists is at once becoming more accessible in cost and more powerful in its applications.

## 2 The Architecture of Avian Genomes

For many biologists generating genomic data, the resulting genetic variation, usually in the form of thousands of single nucleotide polymorphisms (SNPs), can be used to address questions regarding population structure or evolutionary history. However, it is also important to consider the ultimate substrate of this variation: the genome itself and how it is arranged. This structure, known as genomic architecture, broadly refers to a genome's overall size and the number and morphology of chromosomes. At a finer scale, genomic architecture also refers to the arrangement of genes into the different chromosomes and their organization with respect to each other within each chromosome. When these characteristics are shared across distinct groups, this is known as shared or conserved synteny. In this section, we summarize several aspects of avian genome architecture, focusing on aspects where bird genomes are unique, how high-throughput sequencing data have been used to investigate genome architecture, and finally how in some cases, unique phenotypes are associated with large-scale genomic rearrangements.

### *2.1 Avian Genome Size and Arrangement*

A prominent aspect of the genomic architecture of birds is its general conservation in several characteristics, even across the deep time scale spanned by avian evolution, and a relatively small genome size (Ellegren 2013). Both factors make birds particularly suitable for genomic studies, where costs scale directly with genome size. For example, a genomic study of the Japanese flower *Paris japonica* – a species with one of the largest known genomes (Pellicer et al. 2010) – would cost over two orders of magnitude more than sequencing any bird genome. The conserved avian genomic architecture also allows researchers interested in wild birds to utilize resources, like gene annotations, that have been developed for model taxa (e.g., the zebra finch), or species of commercial importance, like the chicken. This kind of sharing of genomic resources would not be as reliable in other taxa, such as fishes or angiosperms, where between-species genome size, synteny, and even ploidy vary widely.

In the past, a variety of indirect methods that do not involve DNA sequencing have been used to estimate genome sizes. For example, genome size can be estimated using flow cytometry in combination with DNA staining (Hare and Johnston 2011), a process that involves staining the nuclei of individual cells with a fluorescent dye and comparing the levels of fluorescence with reference cells, such as those of species for which genome size is known. In the last decade, reference genome sequences have been obtained for many avian species, allowing a direct estimation of genome size. Most avian genomes published to date show a similar size to those estimated with flow cytometry, averaging around 1.2 Gb (Zhang et al. 2014a), which is relatively small compared to other vertebrate taxa (Gregory et al. 2007). One reason for this conservation across broad time scales is that bird genomes have fewer repetitive elements than other taxa in their genomes – DNA sequences that replicate and increase the overall size of the genome – and they also do not appear to have undergone whole-genome duplications, like is common in plants and ray-finned fishes (Zhang et al. 2014a). An interesting exception to this general pattern is the finding that woodpeckers and allies have a large number of transposable elements repeated across their genomes, at least as compared to other bird groups (Zhang et al. 2014a, b; Manthey et al. 2018). Birds also have nucleated red blood cells, and, given the energetic demands of powered flight, some research has suggested that a small genome size might be advantageous for efficient gas exchange, although this idea is controversial (Organ et al. 2007).

## *2.2 Avian Chromosomes*

Although there is some variation, the absolute number of chromosomes is fairly conserved among birds, with a haploid average of 40 chromosomes (Ellegren 2010). These insights into karyotypes come from chromosome squashes. Chromosome painting, which involves hybridizing chromosomes-specific probes designed in the chicken to the chromosomes of other species, has also revealed a high level of conserved synteny in bird genomes (Ellegren 2010). Presumably these chromosome-level inferences should be improved by obtaining whole-genome sequences of many taxa. However, while the number of available avian genomes has grown by over an order of magnitude in the last 5 years (e.g., Jarvis et al. 2014), the quality of these assemblies is far from chromosome level. Instead, most draft avian genomes can only be confidently assembled into linear strings that are small and where the relative order and orientation of these segments are unknown. These recent draft assemblies – based on short-read Illumina sequences – are also of lower quality than the first bird genomes that were assembled using traditional (but more labor-intensive) sequencing methods (e.g., Sanger sequencing; International Chicken Genome Sequencing Consortium 2004). However, despite the fragmented nature of these draft genomes, in many cases the order of genes appears to be conserved (Zhang et al. 2014a).

The sex chromosomes of birds are another aspect of their genomic architecture where they differ compared to most vertebrates, as birds have female heterogamety (Ellegren 2011). This means that male birds have two copies of the Z chromosome,

while females have one Z and one W (in contrast to the XY system in mammals, for instance, where females have two copies of the X chromosome). Sex chromosomes are also presumed to have a strong influence on the evolution of new species (Charlesworth et al. 1987), possibly because they contain genes involved in generating reproductive barriers. They are also the focus of much hybridization research investigating the effects of Haldane's rule, which states that in hybrid crosses, the heterogametic sex is usually infertile or inviable (Haldane 1922). Consistent with this postulate, many studies have shown that hybrid female birds are more likely to be inviable or sterile than males, and this pattern likely speaks to the importance of sex chromosomes in avian speciation (Price and Bouvier 2002; Tubaro and Lijtmaer 2002; Lijtmaer et al. 2003; Arrieta et al. 2013). Sex chromosomes also show a faster rate of evolution than autosomes because their effective population size is smaller – generally by three fourths for the Z and one fourth for the W – than that of autosomes (Charlesworth 2001). Indeed, genome comparisons across several avian taxa have found higher levels of differentiation on the Z chromosome than on the autosomes, consistent with what is expected from a reduced effective population size (e.g., Ruegg et al. 2014; Irwin 2018).

## *2.3 The Recombination Landscape*

Another important aspect of architecture that is important in shaping genomic variation are differences in genetic recombination. Recombination is the sharing of genetic information between the two homologues of the same chromosome during meiosis. Importantly, recombination is not uniform across the genome, with certain areas showing high recombination rates, while others experience very low levels (Kawakami et al. 2014; Burri et al. 2015). For example, a consistent feature of avian genomes is the presence of many microchromosomes. Microchromosomes – also present in many reptiles and amphibians (Burt 2002) – are much smaller than the other chromosomes and have relatively higher recombination rates, in addition to higher substitution rates and a higher gene density than macrochromosomes (Ellegren 2013). Because of the small size of microchromosomes, they have been difficult to study, and sequence data from them are absent from many draft genome assemblies (Ellegren 2013). More generally, recombination hotspots within chromosomes appear to be conserved across independent avian species, separated by up to tens of millions of years (Singhal et al. 2015). Although it remains to be determined how conserved recombination is among all birds, the conservation of the recombination landscape can determine the rate at which different areas of the genome evolve.

The recombination landscape appears to have structured much of the broad patterns of divergence between pied and collared *Ficedula* flycatchers, two species that hybridize across a large region of Europe. In comparisons of these species, several studies have found that the genome is heterogeneously differentiated between the two: certain genomic regions are highly divergent, while other regions show few sequence differences (Ellegren et al. 2012). Notably, in studies of multiple

species in the genus *Ficedula*, the areas of high divergence appear coincident among independent pairs of species and coincide with known areas of low recombination (Burri et al. 2015). A possible explanation for this pattern is that low recombination is amplifying the effects of purifying selection, which acts on eliminating deleterious alleles (Cruickshank and Hahn 2014; Burri et al. 2015). Whether this is a general pattern remains to be seen.

## 2.4 Structural Rearrangements Associated with Complex Traits

Mutations that occur at the level of individual nucleotides are referred to as point mutations. These mutations are the type that, when occurring in a coding sequence, may lead to the replacement of one amino acid for another in a protein or produce an abnormal termination of translation. However, mutations can also involve large tracts of DNA, and these are called macromutations or structural rearrangements. Macromutations include deletions or duplications of large portions of DNA; translocations, which involve the movement of a segment of DNA to another region of a chromosome or another chromosome; and inversions, which happen when DNA segments are flipped in their orientation. Despite the generally conserved nature of the avian genomic architecture, a large number of macromutations have been identified across several species. For example, inversions have been found to be relatively common among estrildid finches (Hooper and Price 2015). Indeed, the individual zebra finch that was selected for whole-genome sequencing and assembly is heterozygous for an inversion – on the Z chromosome – and the frequency at which this inversion exists in wild finch populations is highly variable (Itoh et al. 2011).

Inversions can also play a role in the evolution of complex traits by "capturing" genes and allowing them to coevolve, generating what is known as a supergene (Schwander et al. 2014). Supergenes can evolve because recombination is largely suppressed within an inversion, leading to tight linkage among the genes that are involved in the structural rearrangement. This suppression of recombination occurs because most crossing over within an inversion in an individual that is heterozygous for the structural rearrangement will lead to unviable gametes (reviewed by Rieseberg 2001). Here we discuss two well-studied examples of complex traits in birds that are generated by the presence of inversions: the reproductive biology of the white-throated sparrow (*Zonotrichia albicollis*) and the ruff (*Philomachus pugnax*; Fig. 1).

The white-throated sparrow has two color morphs, white and tan-striped, and these color morphs exhibit alternative reproductive strategies. White males are more behaviorally aggressive than tan males, provide little parental care, and are more promiscuous. In contrast, tan males are monogamous and invest more in rearing their nestlings. Cytological research, as well as recent genomic studies, have allowed researchers to understand the genetic origin of these two sparrow morphs. They appear to be determined by the presence of a large inversion on chromosome two: this large inversion encompasses approximately 10% of the entire sparrow genome

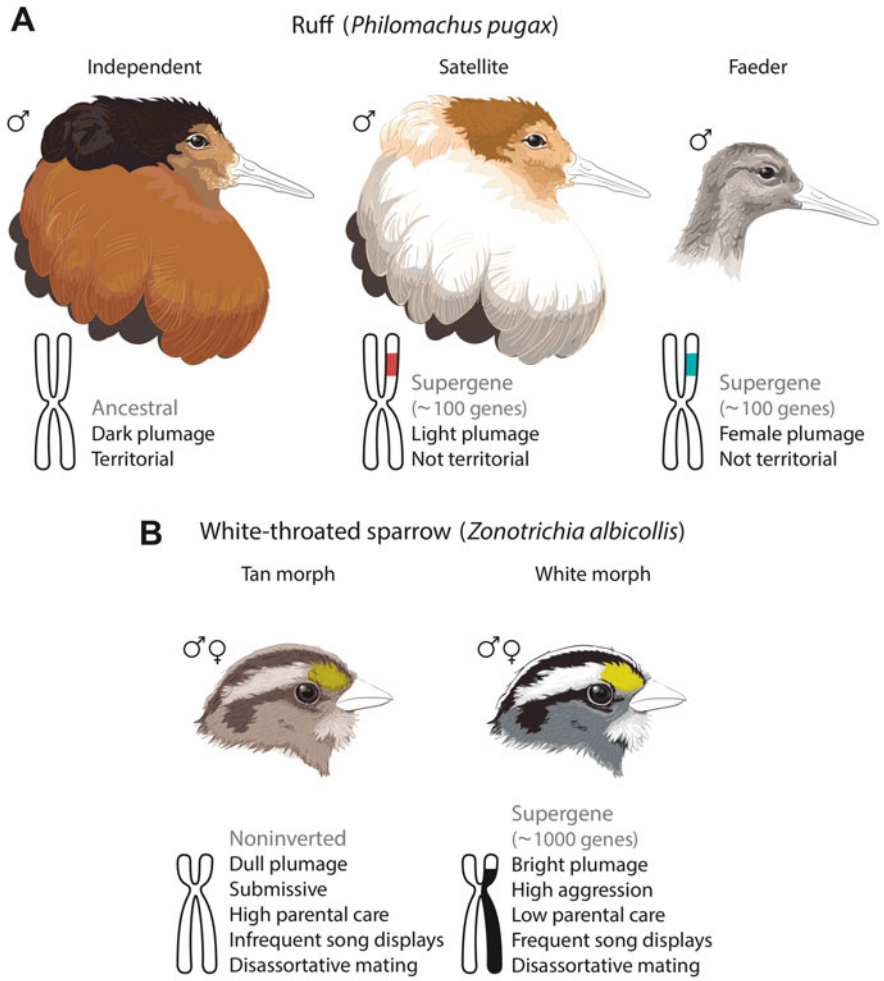


**Fig. 1** Genomic architecture and breeding phenotypes. An example of inversion-generated supergenes in ruffs (panel **a**) and white-throated sparrows (panel **b**). In ruffs, an inversion harboring approximately 100 genes results in morphological and behavioral traits in "satellite" and "faeder" males. In white-throated sparrows, a larger inversion (approximately 1,000 genes) results in the white morph, which differs morphologically and behaviorally from the tan morph. Figure from Taylor and Campagna (2016), illustration by K. Sutliff/Science. Reprinted with permission from AAAS

and includes over 1000 genes (Thomas et al. 2008; Tuttle et al. 2016). Tan birds do not carry the inversion, whereas white birds are heterozygous for the rearrangement. The inversion is maintained in the population by a puzzling aspect of the white-throated sparrow's reproductive behavior: male white sparrows mate almost exclusively with tan female birds, whereas female white sparrows mate almost exclusively with tan male birds, therefore passing on the inversion and maintaining it as a stable polymorphism in the species. Many genes within the inversion are candidates for

generating both the behavior and the coloration patterns that are specific to each morph, but the exact mechanism by which they do this is not yet fully understood.

Another fascinating combination of coloration pattern variation and behavior that is controlled by the presence of an inversion is the complex breeding system of the ruff (Fig. 1). Two independent studies have linked different mating strategies in the ruff to a large genomic inversion (Lamichhaney et al. 2016a; Küpper et al. 2016). Male ruffs, known as "independents," constitute the majority of males in a population and are ornamented individuals that control the breeding sites that females visit (i.e., leks). Satellite males have different ornaments and, instead of holding leks, steal copulations from the independent males. Finally, a small proportion of the adult males are known as "faeders." These birds mimic females in appearance and size and also steal copulations. Genomic analysis suggests that these complex behaviors – and the associated morphological traits – are determined by a ~4.5 Mb inversion encompassing around 100 genes. Independent males do not carry the inversion, while satellites and faeders are heterozygous for different inversions. In this case, it appears that there are two different alleles of the inversion: one generates the traits associated with satellite males, while the other determines the faeder morph. As in the white-throated sparrow, there are genes captured by the inversion in the ruff that are good candidates for generating the different mating behaviors. The mechanism by which this is achieved, however, is also not yet understood. It appears that, in both the sparrow and the ruff, coloration genes and genes that mediate the response to hormones have coevolved within the inversions to generate different reproductive strategies (Taylor and Campagna 2016).

## 3 Linking Genotype to Phenotype in Avian Genomes

More commonly than from large-scale macromutation, differences in phenotype arise from DNA sequence variation. However, at a basic level, why might it be useful to understand the genetic basis for relevant phenotypic traits? First, finding the genes responsible for a given phenotype can empower functional research that attempts to understand broader molecular and biochemical actions. For example, the genes involved in allowing birds to convert dietary yellow carotenoids to red ketocarotenoids in developing feathers were recently identified (Mundy et al. 2016; Lopes et al. 2016; Toomey et al. 2017). This finding sets the stage for additional characterization of the pathways and enzymes involved in the processing of carotenoid molecules. The discovery of the genes coding for traits also allows for indirect tests of natural selection, by quantifying patterns of genetic variation within and near important genes: selection leaves its mark on the genome, in part by reducing genetic variation, although there are many ways to quantify molecular signatures of selection. The signature of reduced variation can be a first step in addressing whether selection was involved in shaping phenotypic variation or whether other processes are more likely. In addition, a central question in evolutionary biology is in the directionality and repeatability of evolution. As it relates to genomics, how often are similar genes used in generating similar phenotypes in independent groups?

Knowing the relevant genes involved allows for this kind of broader, comparative analysis. Finally, by understanding the genes responsible for generating phenotypes that are involved in adaptation, conservation practitioners can better understand how phenotypes are generated, with the goal of conserving them for the future. With these concepts in mind, in this section we review cases where genomic data have been used to link genetic and phenotypic variation in avian systems.

## 3.1 Candidate Genes Versus Genome-Wide Associations

Across a broad array of genomic applications, the massive amount of sequence data produced from modern short-read technologies has revolutionized the strength and feasibility of genotype-phenotype association studies. Prior to broad adoption of genome-wide sequencing data, progress in this area relied almost exclusively on candidate gene approaches. Candidate gene approaches are characterized by studying genes with known functions in one species but which were originally studied in more detail in another, often distantly related, group. This usually takes the form of comparing short sequences to test for molecular signatures of selection or examining the patterns of gene expression profiles, commonly for a small number of loci. For example, by quantifying the expression patterns of several candidate craniofacial genes in the bills of developing Galápagos finches, Abzhanov et al. (2004) found a conspicuous pattern of upregulation of the *Bmp4* gene in finches with large versus small bills. Indeed, in this study, like in many others, a candidate gene approach was fruitful in connecting genotype to phenotype.

However, there is a clear drawback of the candidate gene approach: by only focusing on a subset of genes with known function, these studies were not designed to find genes with more poorly characterized functions. The application of whole-genome sequencing has helped address this limitation. By using data from sequences distributed more or less randomly throughout the genome or representing the entire genome, genome-wide association studies (GWAS) are poised to explore uncharted genomic territory (e.g., Brelsford et al. 2017). As we discuss in more detail by reviewing several case studies of genotype-phenotype links in birds, in several instances these new approaches have highlighted novel and previously uncharacterized genes. However, in several cases, those familiar genes and gene families at the heart of many candidate gene studies were also identified by these new genomic approaches.

## 3.2 Genomics of Feather Coloration

Coloration is one of the most conspicuous phenotypes to differ between closely related birds. The melanogenesis pathway – a pathway that gives rise to eumelanin (brown or black) and pheomelanin (tan or yellows) pigments – is a very well-characterized biochemical pathway in vertebrates (Hubbard et al. 2010). Several

candidate gene studies in birds found strong links to genes – *agouti signaling protein* (*ASIP*) and *melanocortin receptor-1* (*MC1R*) – that had previously been linked to pelage differences in mammals (Hoekstra et al. 2006). For example, amino acid differences in *MC1R* are associated with melanistic versus pale plumage across several divergent bird taxa (e.g., bananaquits, lesser snow geese, and Arctic skuas), although in each case there is a different mutation responsible for the phenotypic shift within the same gene (reviewed by Mundy 2005; Cuthill et al. 2017).

Uy et al. (2016) set out to ask whether a very recent case of melanistic evolution might also show evidence of different mutations conferring parallel phenotypic evolution. To address this, the authors combined a candidate gene approach with reduced-representation genome sequencing to study melanic variation in the chestnut-bellied monarch (*Monarcha castaneiventris*), a flycatcher found on the Solomon Islands. Restriction enzyme-based reduced-representation sequencing (e.g., Elshire et al. 2011; Peterson et al. 2012) is a genomic method designed to represent a small fraction of the genome, a tradeoff associated with sequencing common genomic regions across many individuals in an affordable way. Uy et al. (2016) compared chestnut-bellied birds from a large, central island to all-black forms, which inhabit several satellite islands. Candidate gene sequencing revealed amino acid substitutions in the coding regions of both *ASIP* and *MC1R*, which likely confer the melanin-based phenotypic changes. However, their data suggest that different mutations in the two genes were responsible for the parallel evolution of all-black phenotypes on the satellite islands – melanic birds on two islands had mutations in *ASIP*, whereas birds on a third island had a mutation in *MC1R*. The genomic data were important for Uy et al. (2016), as they were used to quantify the genomic background upon which these mutations arose. By using outlier analyses, Uy et al. (2016) found that polymorphisms in the candidate genes were significantly more differentiated in allele frequencies when compared to this overall background. Therefore, this study shows that even over apparently short time scales, different mutations were recruited to generate similar phenotypes.

This novel combination of candidate genes and a genome scan is likely to be an attractive approach for similar studies, although it will still be missing other genes that may be involved in generating relevant phenotypes. To address this, whole-genome comparisons can be used, which survey variation across all coding and noncoding regions of the genome. When levels of background divergence are high between groups, associations between genotype and phenotype can be weaker. This is because when many genomic regions differ between groups, identifying which regions include putatively causative genes can be more difficult. In these cases, artificial crossing and/or pedigrees may be required to associate genomic regions with phenotypes. However, there are logistical challenges associated with breeding experiments, often making them unfeasible with wild birds.

In the absence of artificial crosses, four studies comparing phenotypically distinct birds have taken advantage of naturally low levels of background divergence, either due to recent ancestry or extensive natural hybridization: hooded and carrion crows (Poelstra et al. 2014), golden-winged and blue-winged warblers (Toews et al. 2016b), *Sporophila* seedeaters (Campagna et al. 2017), and Munia finches (Stryjewski and Sorenson 2017). In each of these studies, only a small portion of

the genome was found to be differentiated between the phenotypically distinct groups. Moreover, across each of the comparisons, genes within the melanogenesis pathway were found either within or near the few peaks of divergence across their genomes. In both the warblers and seedeaters (groups of birds separated by millions of years of independent evolution), a small region upstream of *ASIP* – within the presumed regulatory region of this gene – was strongly divergent between birds that differ in the extent and pattern of melanistic feathers and, more specifically, was perfectly correlated with black versus pale throats in the warblers (Toews et al. 2016b; Campagna et al. 2017; Fig. 2). Genomic divergence upstream of these genes

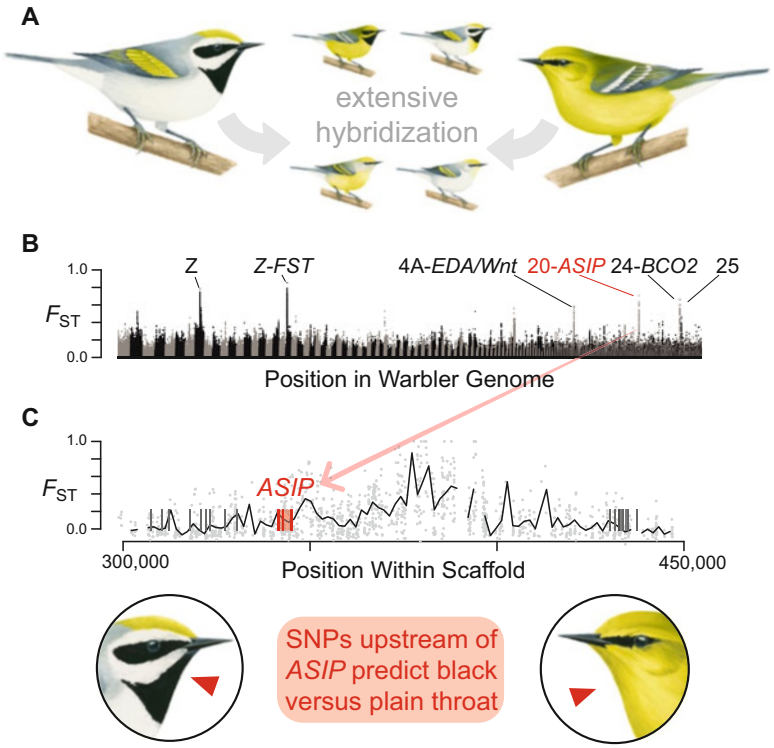


**Fig. 2** Plumage genes and wood warblers. Peaks of divergence in $F_{ST}$ corresponding to plumage variation between golden-winged (panel **a**, left) and blue-winged warblers (panel **a**, right). Panel (**a**) shows the pure individuals as well as a range of possible hybrid phenotypes. Panel (**b**) shows genome-wide patterns of differentiation between pure blue-winged and golden-winged warblers, with few peaks of divergence representing candidate genes (labeled peaks). Panel (**c**) shows a detailed view of divergence peak in the presumed promoter of the *ASIP* gene, which contains SNPs that are predictive of black versus yellow or white throat in the two species. Figure from Toews et al. (2016b), warbler illustrations by Liz Clayton Fuller. Reprinted with permission from Current Biology

is relevant for two reasons. First, it suggests that regulatory changes, as opposed to coding variation, are involved in generating these coloration differences; any candidate gene study that assayed coding sequence variation in this gene would have overlooked this connection. Second, these genome-wide surveys were not obviously biased toward finding any sets of genes with given functions. By identifying genes in the melanogenesis pathway – genes that were the focus of many candidate genes studies – these anonymous genome scans highlight the evolutionary importance of these pathways.

Bird coloration is not only due to differences in melanin: the diversity of avian coloration is also due to carotenoid-based colors (Barsh 2016). Carotenoids are pigment molecules that cannot be synthesized endogenously by vertebrates and are instead acquired from their environment. Whereas some animal species express unmodified carotenoids, many others utilize a complex suite of enzymes and biochemical pathways to alter the molecules. Until recently, however, the genetic basis for processing carotenoid-derived pigment molecules has been unclear. Lopes et al. (2016) and Toomey et al. (2017) used different color variants of domesticated canaries to make genotype-phenotype associations and try to identify the genes involved in carotenoid coloration. For example, several decades ago a "red factor canary" was bred by bird fanciers by crossing wild red siskins (*Spinus cucullata*) with domesticated yellow common canaries (*Serinus canaria*) and then subsequently backcrossing the offspring to canaries (Fig. 3). Whole-genome comparisons between the resulting red and yellow canaries highlighted two genomic regions that appear to have introgressed from red siskins, and these regions are likely responsible for red feather coloration in the backcrossed canaries. A gene encoding a cytochrome P450 enzyme, *CYP2J19*, is within one of these candidate regions, and this gene is significantly upregulated in both the skin and liver. Moreover, this same gene was found in a similar coloration study of zebra finches, which suggests this might be involved in a common mechanism for carotenoid metabolism across birds (Mundy et al. 2016).

### *3.3 Genes Underlying Morphometric Differences*

There has also been much progress in identifying the genes involved in different morphological characteristics among wild birds. This is particularly true for Galápagos finches, which are a celebrated example of evolution on islands. These birds are also notable because of their historical connection to Charles Darwin's exploration of the Galápagos and, more recently, what they have taught us about the rapid tempo of evolution (Grant and Grant 2014). They are also one of the few studies of a wild organism to connect phenotypic variation – bill size within and between different species – with reproductive fitness (Boag and Grant 1981). A logical next step was to ask whether it was also possible to identify the genes favored by natural selection during environmental shifts. As discussed earlier, Abzhanov et al. (2004) found that the gene *Bmp4* was upregulated during bill development of large-beaked finches. A subsequent study, relying on microarray

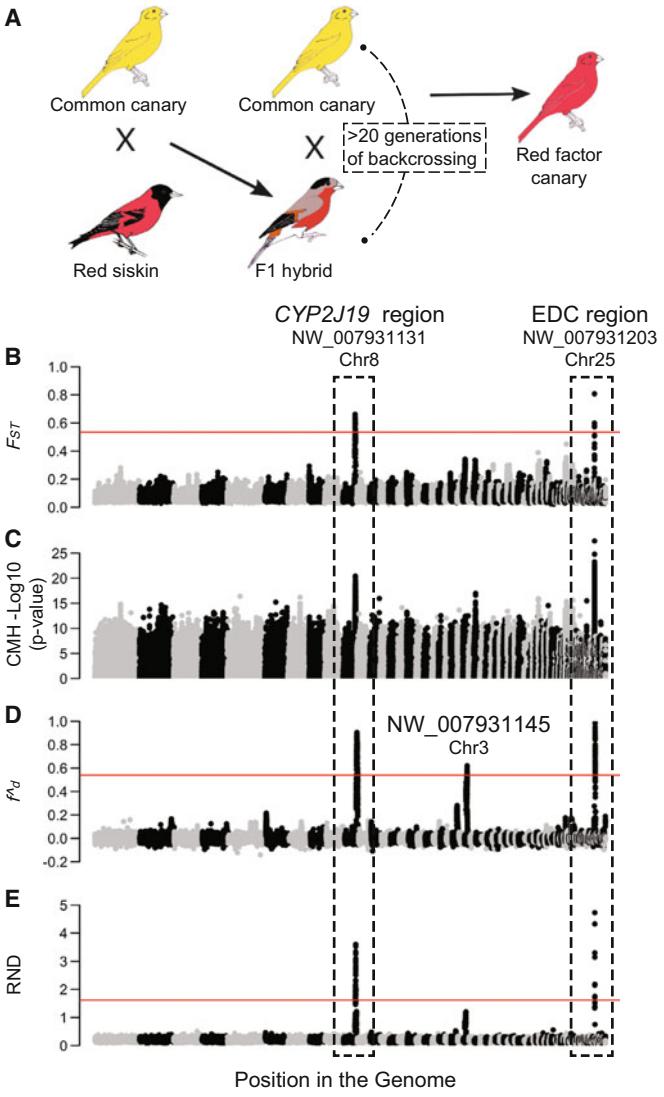


**Fig. 3** Carotenoid genes in hybrid canaries. Through artificial hybridization with red siskins, canary breeders generated a "red factor" canary (panel **a**). Unlike mammals, birds color their feathers with both melanins and carotenoids; the latter gives many birds their dramatic yellow, red, and orange colors. Unlike melanin pigments, however, the molecular machinery and genes involved in carotenoid pigmentation have been unclear. Lopes et al. (2016) used genomic data to identify the genomic regions associated with red versus yellow pigmented birds, identifying *CYP2J19* as a strong candidate causal gene. Panels (**b**–**e**) show genomic measures of: divergence between the red and non-red canaries (panels **b** and **e**); evidence of association between SNPs and phenotypes (panel **c**); and signatures of introgression from red siskin (panel **d**). Figure from Lopes et al. (2016), reprinted with permission from Current Biology

technologies to quantify differential gene expression, identified a separate gene, calmodulin (*CaM*), also involved in bill morphology (Abzhanov et al. 2006).

These original findings have now been complemented by two whole-genome analyses comparing finches with different bill morphologies. These studies were attempting to confirm earlier research, in addition to identifying novel genetic variants associated with bill morphology. Lamichhaney et al. (2015, 2016b) compared the genomes of finches that differed both in beak shape (blunt versus pointed beaks) and size (large versus small beaks). They found several associations for beak shape, including a region containing the gene *ALX homeobox 1* (*ALX1*). *ALX1* has a central role in the development of the craniofacial embryonic tissues, and, at least in humans, *ALX1* mutations disrupt craniofacial development. When comparing finches that differed in bill size, the best candidate was identified as *high-mobility AT-hook 2* (*HMGA2*). *HMGA2* has also been implicated in craniofacial features as well as tooth eruption in humans (Lamichhaney et al. 2016b). Notably, *HMGA2* showed a marked change across a severe drought on the Galápagos Islands between 2004 and 2005, where birds with homozygous "small beak" genotypes were more likely to survive compared to the homozygous "large beak" genotypes (Lamichhaney et al. 2016b). This is a striking connection between genotype, phenotype, and fitness – but did these new analyses also find associations with the previously identified genes, *Bmp4* or *CaM*? *CaM* was identified in a divergence peak associated with bill shape, but in both new genomic studies, *Bmp4* was not identified as a strong candidate. The authors stress that this result is not to imply that *Bmp4* is not involved in bill development, but more likely that its function is modified by upstream transcription factors. It also highlights the utility – and challenges – of reconciling genome-wide patterns of divergence with gene expression studies.

## *3.4 Genes for Avian Migration*

The study of migratory behavior has a strong history in attempting to connect genotype and phenotype, with many new studies incorporating whole-genome sequencing as well as expression analysis. Biannual migration is an extraordinary feat exhibited by migratory birds that links their breeding and wintering ranges, which in many cases are separated by thousands of kilometers. During these migrations, many small birds travel between these distant areas taking only a few weeks to reach their destinations (Alerstam 2006). Many young birds also migrate between their breeding and wintering areas for the first time alone, at night, without any assistance from experienced birds. There is much evidence to suggest that migratory behaviors are under genetic control, providing impetus to identify the genes involved in this complex behavior (Liedvogel et al. 2011).

Early genotype-phenotype association studies of migration behavior in birds focused on a subset of genes thought to be involved in circadian rhythms. Migration timing is presumed to be mostly hardwired – individuals respond to seasonal changes

in ambient light (Alerstam 2006) – making genes involved in circadian rhythms excellent candidates. This has been exemplified by research in European blackcaps (*Sylvia atricapilla*), which have long been a focus of studies in avian migration (e.g., Helbig 1991). Indeed, the classic demonstrations of genetically inherited migration directionality were completed on lab-reared blackcaps. To begin identifying the genetic basis of migration phenotypes, Mueller et al. (2011) identified a gene, *ADCYAP1*, that explained a small but significant proportion of variation in migratory restlessness. This behavior, also known as "zugunruhe," is a frenetic behavior observed in captivity during the periods that a bird would naturally be migrating. Most notably, this gene has an unusual microsatellite within it – these are small tandemly repeated portions of the genome, usually thought to have little functional implications. However, in this case, the number of repeats within the microsatellite, which is within a regulatory region of the gene, is positively correlated with increased restlessness within and between populations, implying a functional role.

There have not been genome-wide studies of the blackcap system to date. However, in North America, recent genomic research has focused on divergent migratory populations of the Swainson's thrush (*Catharus ustulatus*). Two separate studies by Ruegg et al. (2014) and by Delmore et al. (2015) mapped out patterns of genomic divergence between two subspecies – coastal and inland – that hybridize in a narrow region and employ divergent migratory routes. Migration is thought to play a role in generating reproductive barriers between these two groups, as hybrid birds may inherit an intermediate (and possibly less fit) migratory behavior. There is also a marked pattern of heterogeneous divergence across the genomes of the two subspecies. Heterogeneous divergence, in this case, is characterized by discrete chromosomal blocks, which are highly divergent in allele frequencies, but that occur on a background of much lower divergence. Delmore et al. (2016) used a method known as admixture mapping – the association of genetic variants with phenotypic traits across hybrid offspring – to find which of these divergent regions of the genome might also be involved with migratory orientation. Indeed, Delmore et al. (2016) found a cluster of genes within one of these divergent regions that is significantly associated with distinct orientation behaviors. This cluster of genes includes genes involved in circadian rhythm but also several involved in other aspects of the nervous system as well as cell signaling. In a gene expression analysis of this system, Johnston et al. (2016) analyzed tissues presumably important in circadian rhythms and, subsequently, migration behaviors. In this case, the goal was not necessarily to identify genetic variants that were different between the groups but those genes that might be differentially expressed. Notably, there were few if any genes that overlapped between the studies of Delmore et al. (2016) and Johnston et al. (2016) (genome sequencing versus expression, respectively). This highlights the complementary nature of these two approaches, as well as the complexities in attempting phenotype-genotype associations for a multifaceted – and polygenic – behavior such as migration.

## 4 Phylogenomics and Demographic Histories of Birds

Birds occur in an amazing array of environments and live on all of the continents. They have also evolved in response to environmental changes and dramatic differences in climate. In this section, we ask how genomic data have been used to provide inferences in the different ways that bird populations and species have changed over time and how genetic structure is partitioned over space, across time, and in different environments. Historically, this field of study – known as phylogeography – relied on mitochondrial markers to make inferences about the connectivity and affinities of populations and species across their ranges (Avise et al. 1987). However, while there are numerous aspects of mtDNA that make it tractable for addressing these questions – its maternal mode of inheritance, the lack of recombination, and its small effective population size – many researchers have also raised concerns that mitochondrial genomes might not be representative of other important processes that influence the nuclear genome (Ballard and Whitlock 2004; Edwards and Bensch 2009; Toews and Brelsford 2012). Moreover, because of the lack of recombination in mtDNA, admixed individuals cannot be easily detected. Therefore, it is not ideal for studies investigating cases of gene flow or hybridization.

Despite these potential pitfalls, many valuable insights have been derived from studies of mtDNA, and researchers have long argued about its utility (Zink and Barrowclough 2008; Edwards and Bensch 2009). In efforts to improve phylogeographic inferences, researchers began to incorporate more nuclear markers into their studies, either in the form of nuclear introns or developing microsatellite DNA loci. From an empirical standpoint, the ability to sequence nuclear markers was limited by the availability of primers that would work successfully across a wide range of taxa (Kimball et al. 2009). This was a less important issue for avian mtDNA, where the early development of nearly "universal" primers allowed the amplification of mitochondrial markers across many divergent species (Sorenson et al. 1999). For DNA microsatellites, however, cross-species amplification is not always successful. In addition, if the species used for marker design and the target species are very distantly related, this can result in lower levels of polymorphism (Primmer et al. 1996). Despite these technical hurdles, researchers aimed to obtain the largest sample of nuclear markers to test increasingly more complex phylogeographic scenarios (Hare 2001).

Applying genomics to phylogeographic studies of birds (and other taxa) has overcome the main limitations of working with species with no prior genetic resources. For example, reduced-representation genome sequencing approaches (reviewed in Andrews et al. 2016) aim to randomly subsample the genome using restriction enzymes and therefore require very little, if any, prior knowledge about the target species. These techniques have revolutionized phylogeography by allowing researchers to obtain thousands of nuclear markers at low cost. Reduced-representation approaches also have the ability to uncover loci that may be under selection. However, the fraction of the genome that is sampled is usually very low, and thus these techniques are better suited for obtaining neutral loci (Lowry et al. 2017).

### 4.1 Biogeographic Inferences in the Era of Phylogenomics

Being able to generate similar markers across many species has allowed researchers to tackle important biogeographic questions using a comparative approach. One example of this kind of approach, as applied to avian systems, is the study of the drivers of Amazonian bird diversity by Weir et al. (2015). The Amazonian basin harbors one of the largest and most diverse avifaunas of the world. The drivers of the high species richness of the region have long been debated and include several non-mutually exclusive factors like the age of species assemblages and the landscape features that may isolate populations and promote allopatric speciation (reviewed by Rull 2011). One hypothesis – the forest refugia hypothesis – suggests that forests contracted into isolated refugia during dry climatic periods (Haffer 1969) and, consequently, isolated populations of many different taxa, promoting allopatric speciation. However, it is widely recognized that closely related taxa tend to replace each other across the major tributaries of the Amazon River, a pattern that holds true for many bird species (e.g., Ribas et al. 2012). This observation has led to a second hypothesis – the river barrier hypothesis – that suggests that the major tributaries of the Amazon River constitute barriers to gene flow that can promote allopatric speciation.

However, the Amazon tributaries vary in width, ranging from many kilometers where they drain into the Amazon River, to less than 100 m in their headwaters. It is possible, therefore, that these tributaries also vary in their ability to isolate taxa and may not represent impermeable barriers to gene flow in their headwaters. Weir et al. (2015) set out to test the effectiveness of rivers as barriers to gene flow using a comparative phylogenomic approach. Their study included seven pairs of taxa (both recognized species and subspecies) that occur mostly on either side of the Tapajós river in Brazil. They sampled areas where the river was wide and expected to prevent the contact of individuals restricted to either side, but also sampled the narrow headwaters where taxon pairs were more likely to come into contact. The pairs of taxa were found to be deeply diverged in mtDNA – corresponding to millions of years of separation – and in almost all cases were reciprocally monophyletic. The authors tested whether the pairs were equally differentiated in their nuclear genome and if there was evidence of hybridization, especially in those areas where the river was narrow and presumably represented a weaker barrier to dispersal and therefore gene flow. To answer this question, the researchers needed to sample the nuclear genomes of the focal species and test for evidence of admixed individuals. Weir et al. (2015) used a reduced-representation approach called genotyping-by-sequencing (GBS; Elshire et al. 2011) to produce thousands of nuclear loci per taxon pair. Surprisingly, even given the substantial levels of isolation and divergence, they found evidence for introgression between all of the taxon pairs. As predicted, the origin of the admixed individuals coincided with the narrow headwaters of the river. The authors interpret this pattern to suggest that rivers are not absolute barriers to gene flow, and therefore cannot be the sole promoters of allopatric speciation. More generally, this example illustrates the benefits of using reduced-representation genomic approaches in phylogeography. The authors obtained thousands of markers for

many different species, without the need to test specific primers on a locus per locus basis. In addition, it shows the value of the information obtained from the nuclear genome: the detection of admixed individuals which were the product of hybridization, not apparent from mtDNA data alone.

## *4.2 Inferring Population Demographic Histories from Genomic Data*

What are the major environmental and climate related factors that influence population declines or increases? How important are population bottlenecks in eroding genetic and phenotypic variation? These questions all fall within a central theme: how much have historical processes influenced the species and the population dynamics that we observe today? Until the advent of genetic markers, these kinds of inferences relied mostly on indirect historical reconstructions or evidence left within the fossil record. With genetic tools, it is now possible to indirectly mine this kind of historical information from the genomes of extant individuals. However, most of these methods provide only a snapshot of the recent past, particularly when analyses have historically been based on a single genetic marker, namely, mitochondrial DNA. For example, sequence mismatch distributions – the number of pairwise differences between sequences in a population – were historically used to distinguish between historically expanding versus stable populations, such as those of Eurasian and Pacific wrens (Drovetski et al. 2004). Estimates of Tajima's D – a metric sensitive to population expansions or contractions – have been invoked to make inferences about population stability, as in the green-backed tit (Wang et al. 2013). However, in both examples, numerous other population processes can likely influence the estimates of these summary statistics. The incorporation of genomic data has expanded both the resolution and the breadth of demographic inferences that can be gleaned from avian systems.

One central demographic parameter that genomic data can estimate with higher resolution compared to traditional methods is effective population size ($N_e$), a measure of the number of individuals in a population that pass on genes to the next generation. There are several computational methods that have been used to estimate effective population sizes, integrating historical estimates over different time periods at varying degrees into the past. For instance, Bayesian skyline plots based on mtDNA data or small numbers of nuclear markers have been used to quantify historical trends in the effective population sizes across several avian taxa (e.g., Qu et al. 2011). However, as discussed above, there are significant limitations to statistics gleaned from a single genetic marker. For example, many studies have relied upon mitochondria DNA to estimate patterns of historical population sizes. Yet it has been recognized that these estimates from mitochondrial DNA will eventually hit a significant roadblock. For example, in humans, no mtDNA data

are available from mitochondrial DNA beyond about 200 kyr ago, when we all shared a common maternal ancestor (i.e., mitochondrial eve; Li and Durbin 2011).

Recently, avian researchers have used multi-locus genomic data to work around some of these limitations, although the analyses can be computationally intensive. Reconstructions of ancestral effective population sizes for birds, relying on high-resolution genome-wide data, have been conducted in kiwi (Weir et al. 2016), falcons (Zhan et al. 2013), flycatchers (Nadachowska-Brzyska et al. 2016), crows (Vijay et al. 2016), turtle doves (Calderón et al. 2016), and passenger pigeons (Hung et al. 2014). One such analytical method – pairwise sequentially Markovian coalescent modeling – generates inferences about historical effective population sizes by using genome-wide sequencing data from a single individual (Li and Durbin 2011; Nadachowska-Brzyska et al. 2015; Toews 2015; Fig. 4). This method relies on the premise that different genomic regions within an individual's DNA provide semi-independent pieces of information from distinct historical time periods. Several dozen full bird genomes have now been published, with varying degrees of assembly quality (Jarvis et al. 2014). Applying this whole-genome method to these published avian genomes, Nadachowska-Brzyska et al. (2015) asked how population sizes

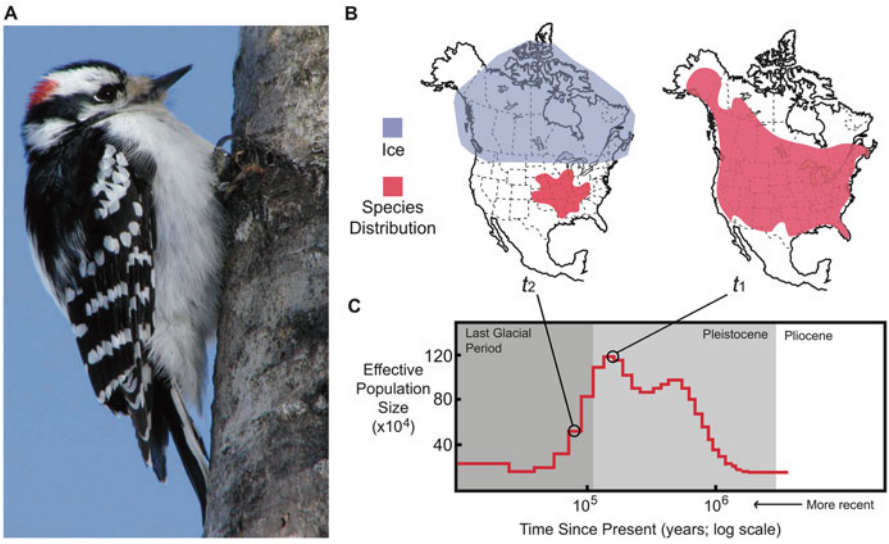


**Fig. 4** Demographic history of birds revealed in their genomes. An example depicting how genomic data can be used to reconstruct the demographic histories of avian taxa. The downy woodpecker (panel **a**) is one of many bird genomes analyzed by Nadachowska-Brzyska et al. (2015) using an algorithm (PSMC) that infers historical population sizes from whole-genome data. Panel (**b**) shows a possible scenario for dynamic range shifts at high latitudes corresponding to glaciation by expanding ($T_1$) and later contracting ($T_2$) during more favorable times. Panel (**c**) shows example output of the PSMC analysis from Nadachowska-Brzyska et al. (2015) showing variation in historical population size through time for the downy woodpecker. Figure from Toews (2015). Photograph by Peter de Wit (Wikimedia Commons), figure reprinted with permission from Current Biology

changed in different bird species over millions of years. In many species, there was a clear increase in population sizes during the Pleistocene, followed by steep declines, likely correlated with the most recent glacial advances. It is thought that many species, particularly those at high latitudes, were confined to small refugia during glacial maxima. These periods of reduced population size appear to correspond to periods where suitable habitat was significantly restricted to small, ice-free regions. While the importance of the glacial cycles in promoting divergence in avian systems is well established (e.g., Haffer 1969; Weir and Schluter 2004), having accurate inferences of population sizes through time is an important confirmation of many of these biogeographic ideas.

## *4.3 Timing of Divergence and Gene Flow Using Genomic Data*

In addition to being able to estimate historical effective population sizes, modeling other processes that generate genomic variation can provide insight into a wide variety of population demographic parameters, including the timing of population splits and the tempo and extent of gene flow. This can be useful for delineating cryptic species, dating important biogeographic events, and understanding the population histories for declining species to ascertain the causes that might have put these taxa at risk. We illustrate how demographic modeling of genomic data can address evolutionary and conservation questions by highlighting two illustrative case studies: one in kiwi (Weir et al. 2016) and one in warblers (Toews et al. 2016b).

Kiwis are a charismatic, flightless group of birds that are endemic to New Zealand. Unlike many other diversification events in birds, kiwi lineages are rather cryptic in their plumage and morphology. Weir et al. (2016) used mitochondrial DNA and hundreds of nuclear markers evaluated in kiwis sampled across New Zealand to (1) try and understand levels of species diversity, including possible cryptic lineages, and (2) to reconstruct population histories in an attempt to understand the biogeographic processes that may have contributed to kiwi diversity. To their surprise, the mitochondrial and nuclear genomic data revealed a clear pattern of cryptic variation. Previously there were only five described species of kiwi, whereas this new genomic analysis strongly suggested the existence of (at least) 11 extant species. Moreover, by including ancient DNA samples, their analyses suggest that between five and six lineages may have gone extinct following the arrival of humans. Using a suite of demographic models, Weir et al. (2016) found that the timing of lineage diversification aligned strongly to the encroachment of large glaciers across the south and north islands of New Zealand during the Pleistocene. Moreover, the very small ranges of extant kiwi lineages are distributed in areas that were likely ice-free during the glaciations. This example highlights how genomic data can provide multifaceted insights into the drivers of diversification and can help uncover cryptic lineages, which is critically important for biodiversity science and conservation.

An example of how complex demographic modeling of genomic data were used to address more contemporary conservation concerns is illustrated by the study of hybridization dynamics between declining golden-winged and blue-winged warblers (Toews et al. 2016b). These warblers illustrate a notable contrast to the kiwi system: golden-winged and blue-winged warblers are quite different in how they look and sound but are very similar genetically (Vallender et al. 2007). This similarity is likely due to hybridization between them – and studies have found little or no detectable fitness reduction in hybrid individuals, which appear fully fertile. Both species are declining – golden-winged warblers precipitously so – and are the focus of conservation efforts (Buehler et al. 2007). This decline is due in part to forest regeneration, which has reduced the availability of the early-successional habitats that both taxa rely on during their breeding season. Golden-winged warblers are also thought to be threatened by displacement and hybridization due to expanding ranges of blue-winged warblers: in many locations, golden-winged warblers have been replaced by hybrids and subsequently by phenotypic blue-winged warblers. It has been thought that humans have influenced both aspects contributing to this decline: anthropogenic habitat change may have facilitated the expansion of blue-winged warblers, accelerating hybridization and subsequently contributing to the decline of golden-winged warblers. Toews et al. (2016b) used demographic modeling and whole-genome data to estimate the timing of when the species began hybridizing. In this case, if hybridization appeared older than anthropogenic habitat alteration, then this factor may not be as important in the species' declines as previously thought. Alternately, if the genomic homogeneity in this system was a consequence of recent hybridization, then management plans for this group might focus more intently on reducing the opportunities for admixture and removing hybrids from the populations. Indeed, the analysis found that hybridization between the two groups better fit a model where hybridization was old and ongoing. Therefore, Toews et al. (2016b) suggest that while humans may have facilitated some recent admixture, this was likely an ongoing feature of these warblers' evolutionary histories, as is thought for many other avian species pairs in North America (Swenson and Howard 2005).

## *4.4 Understanding Deep Patterns of Divergence in the Avian Tree*

Inferences about how environmental factors shaped avian evolutionary histories are generally focused on events that took place over the last several million years. This focus is, in part, due to the technical limitation in quantifying deep phylogenetic relationships. To obtain genetic markers that are informative at a deep scale, they must be sufficiently conserved to provide confidence in their homology – and therefore align correctly across species – while at the same time provide enough variation to be informative. While reduced-representation genomic techniques have allowed researchers to quantify variation in recently diverged taxa with limited

genomic resources, they are less useful when divergence times are deep. This is because these reduced-representation methods require that enzyme cut sites are conserved across the taxa of interest, which may not be the case between highly divergent groups. Therefore, for questions at a deep phylogenetic scale, a different strategy has been used to subsample the genome. Researchers have focused on areas of the genome that are highly conserved, known as ultra-conserved elements (UCEs; Faircloth et al. 2012). The general workflow consists of developing DNA probes that target these conserved areas and hybridize them to the DNA from the different samples that will be sequenced. The probes are then used to isolate the DNA fragments, a strategy known as sequence capture. One benefit of this technique is that it performs well with fragmented DNA, allowing researchers to work with degraded samples such as those obtained from museum study skins (McCormack et al. 2016). While less common, UCEs have also been used to address questions in groups that show shallower scale divergence, at the level that is normally encountered in phylogeographic studies (Smith et al. 2013). This is because sequence variation increases with the distance from the core of the UCE, and therefore the probes can be modified to mine variation appropriate for different scales of divergence.

One long-standing problem in avian phylogenetics where these probes have been applied has been in resolving the relationships among the major bird families and their timing of divergence. Neoaves encompass all birds to the exclusion of tinamous, flightless ratites, game birds, and waterfowl. The relationships within Neoaves have been difficult to establish because these taxa are thought to have diversified very quickly but very deep in the past (Hackett et al. 2008). Two studies have used genomic data in different ways to try and resolve the avian tree of life. Prum et al. (2015) used a sequence capture technique (like the one described above) to sample and compare conserved elements of the genome. Jarvis et al. (2014) adopted a different strategy, sequencing the genomes of four dozen species but pruning the data and retaining only the most conserved regions that could be aligned across taxa (<5%). While Prum et al. (2015) interrogated a smaller proportion of sites than Jarvis et al. (2014) – by approximately two orders of magnitude – it included a much broader representation of avian taxa. The trees from the two studies differ somewhat in their topologies and divergence times. However, what is clear from both is that several deep relationships are *still* unresolved, even with substantial amounts of novel data, and debate continues regarding the best methods and data types (e.g., Reddy et al. 2017). Taken together, these analyses suggest that birds radiated very rapidly after the Cretaceous–Paleogene mass extinction – but that the relationships among them continues to be a mystery even with genomic data.

## 5 Detecting Signals of Divergence and Hybridization

As we have discussed, many studies have used genomic data to identify the environmental factors that shaped the evolutionary histories of wild birds. How these evolutionary histories mix and intertwine among diverging species is another

important area of study. In this way, genomic data have provided many insights into the role of hybridization between closely related taxa. Hybridization – reproduction between members of genetically distinct populations resulting in the production of offspring of mixed ancestry – is influential in shaping the dynamics of many interacting species (Barton and Hewitt 1985). While many commonly think of hybridization as a rare occurrence and the outcomes as an evolutionary dead end (i.e., the textbook example of hybridization is a sterile mule, the offspring from a donkey and a horse), hybridization is quite common in nature. In avian systems, hybridization is especially widespread, occurring in approximately one in ten species (Grant and Grant 1992) and has gained much attention for its role in adaptation and introgression (Harrison 1993; Rheindt and Edwards 2011) as well as the formation of new species (i.e., hybrid speciation; Hermansen et al. 2011; Lamichhaney et al. 2018; Barrera-Guzmán et al. 2017).

Scientists have long been drawn to hybrid zones, as they provide "windows into the evolutionary process" (Harrison 1993). Hybridizing taxa that maintain genetic differences in the face of ongoing gene flow also provide insight into the processes responsible for maintaining species boundaries (Abbott et al. 2013; Carneiro et al. 2014; Vijay et al. 2016) and offer a direct measure of reproductive isolation. Furthermore, hybrid zones are semipermeable boundaries between divergent genomes, where the exchange of genes important in maintaining reproductive isolation is selected against, while movement of others is permitted (Barton and Hewitt 1981; Harrison 1986; Payseur 2010). Because of these differential patterns of introgression, hybrid zones provide the opportunity to identify the genetic and phenotypic traits influencing species divergence. In this section, we review cases where genomic data have been used to understand hybrid zone dynamics and patterns of divergence in avian systems. We contrast genomic approaches to more traditional molecular techniques. We do this to highlight how increasing access to genomic tools has advanced our understanding of divergence between hybridizing species and, in turn, the mechanisms responsible for maintaining hybrid zones through time and space.

To better understand the mechanisms responsible for maintaining reproductive barriers between diverging populations, evolutionary biologists have been turning to molecular markers for several decades. The utility of genetic markers in hybrid zone research is considerable, and our understanding of hybridization has been largely paced by the generation of genetic data (Abbott et al. 2016). Until recently, most hybrid zone research relied on a few genetic markers, typically mitochondrial DNA, microsatellites, or AFLP markers. While these approaches have advanced our understanding of hybrid zone dynamics considerably, the availability and quality of genetic resources can pose major limitations when assessing patterns of introgression (Twyford and Ennos 2012). For instance, the choice of genetic markers can be particularly important for the identification and classification of admixed individuals within a hybrid zone, which has major implications for taxonomy and conservation management (Dupuis and Sperling 2016).

The most significant contributions to avian hybrid zone research have come from the increased resolution gained from genome-wide data. Compared to traditional approaches, this increased resolution allows for more robust estimates of hybrid

indices (Pacheco-Sierra et al. 2016; Dupuis and Sperling 2016), more power to identify genetic regions under selection (Poelstra et al. 2014; Toews et al. 2016b; Campagna et al. 2017), greater sensitivity for detecting instances of cryptic introgression (Weir et al. 2015), and more general insights into the mechanisms driving genetic differentiation across the genome (Burri et al. 2015; Vijay et al. 2016). Genomic tools have been particularly useful for identifying regions of genetic differentiation between species exhibiting shallow levels of divergence, likely a result of hybridization (e.g., Poelstra et al. 2014; Toews et al. 2016b), as traditional approaches using only a few, randomly selected markers can miss differences when they are small or few. Here we discuss the benefits of genome-wide data in characterizing hybridization and divergence in birds. In many instances, these new approaches discovered complex patterns of differentiation, which were largely missed by previous genetic analyses. We illustrate this by providing a detailed discussion of the carrion and hooded crow hybrid zone.

## *5.1 Hybridization in Recently Diverged Taxa*

The hybrid zone between carrion and hooded crows illustrates several important insights that genomic data can provide into the dynamics of hybridization and speciation in birds (Poelstra et al. 2014). In this case, the increased resolution gained from genome-wide data allowed researchers to uncover divergent regions of the genome that likely play a critical role in the maintenance of reproductive isolation. The hybrid zone between the carrion (*Corvus corone corone*) and the hooded (*C. c. cornix*) crows spans a narrow region across Europe and has remained relatively stable over time (Wolf et al. 2010). The two species differ in plumage – carrion crows are all black, while hooded crows have a gray breast and coat – and based on morphological, ecological, and behavioral evidence, they appear to exhibit strong reproductive isolation where they co-occur (Moore 1977; Saino et al. 1998; Wolf et al. 2010). Taxonomic authorities have considered them both as subspecies or full species, with the assumption that strongly diverged genomes should result in hybrid inviability and complete reproductive isolation. Despite these assumptions, however, field studies concluded that heterospecific pairs exhibit little (Saino and Villa 1992) or no (Picozzi 1976) reductions in reproductive success. Furthermore, variation in plumage traits suggests that hybrids successfully backcross with the parental species (Wolf et al. 2010). Molecular studies similarly failed to support the assumption of strong divergence and genetic incompatibilities between them. Studies using mitochondrial markers (Kryukov and Suzuki 2000; Haring et al. 2007), restriction fragment length polymorphism analysis (Kryukov et al. 1992), and allozyme markers (Saino et al. 1992) found no evidence of genetic differentiation. While genetic studies relying on microsatellite markers (Haas et al. 2009) found significant differentiation between populations, the genetic differentiation within species was larger than the genetic differences between the two species. Taken together, these findings suggested that the divergence time between the two taxa is very low, in the

range of thousands of years as opposed to millions, as is likely for other avian hybrid zones (Price 2008).

Poelstra et al. (2014) used a comparative whole-genome approach to identify regions of divergence between carrion and hooded crows. Using a high-quality reference genome, coupled with resequenced individuals from pure populations, the authors identified millions of SNPs across the genome. Their findings support substantial genome-wide gene flow across the hybrid zone, with extremely low levels of differentiation. Genome scans found only a small number of fixed sites between the two taxa, and nearly all of these differences were found within a single region associated with genes involved in pigmentation and visual perception (Poelstra et al. 2014). These findings point to a genomic signal between carrion and hooded crows that may reflect color-mediated prezygotic isolation (Poelstra et al. 2014). Vijay et al. (2016) took the analysis of the crow system one step further by investigating additional hybrid zones within the species complex. The goal here was to evaluate whether there were parallel patterns of divergence across semi-independent groups. The authors documented heterogeneous genomic landscapes across the crow genomes. Further, while they similarly identified divergent regions related to pigmentation, they found that each were specific to a given contact zone (i.e., no candidate genes were shared among all contact zones). Using this comparative, whole-genome approach, Poelstra et al. (2014) and Vijay et al. (2016) highlight the utility of using replicated hybrid zones in disentangling population-specific versus shared drivers of differentiation.

## 5.2 *Hybrid Speciation*

The prevalence of hybridization and its importance for avian evolution is an active area of research. Hybridization is widely recognized as a source of genetic variation for selection to act on. In addition, there are a few well-documented examples in which hybridization has actually led to the formation of new bird species. The process of homoploid hybrid speciation – the formation of a new hybrid species without a change in the number of chromosomes – involves the generation of a hybrid lineage that is reproductively isolated from both parental species as a consequence of the hybridization event. The best studied cases of avian hybrid species are those of the Italian sparrow (*Passer italiae*), the golden-crowned manakin (*Lepidothrix vilasboasi*), and a lineage of Darwin's finches that has recently established on the small island of Daphne Major in the Galápagos Islands, known as the "Big bird lineage" for its large size. These examples differ in the age of the hybridization events and the geographic extent of the range of the hybrid species. They also illustrate the importance of pre-mating reproductive isolation barriers in avian speciation, yet postmating barriers may also operate in some of these cases.

The most recent and geographically restricted case of a hybridization event that illustrates how a hybrid species could be formed occurred on the island of Daphne Major in 1981 (Grant and Grant 2009). An unusually large male finch with a distinct

beak morphology arrived on the island and was identified as an immigrant, itself of possible hybrid origin, using microsatellite markers. Nearly a decade later, genomic data offered the increased resolution to accurately identify the immigrant as a large cactus finch (*Geospiza conirostris*), a species that until then did not occur on Daphne Major, from the distant island of Española (over 100 km away; Lamichhaney et al. 2018). This immigrant finch mated with a local medium ground finch female (*Geospiza fortis*) and established a lineage that bred exclusively endogamously for the 31 years (six generations) in which the population was closely monitored. The hybrid birds held territories close together on a small sector of the island, showed a distinct beak morphology, and sang a unique song. The songs of *Geospiza* finches act as a pre-mating barrier that, in this case, isolated the hybrid lineage from the local medium ground finches (Grant and Grant 2009). It remains unknown, however, if the hybrid lineage is reproductively isolated from the paternal species, the large cactus finch from the island of Española. The persistence of this lineage and its establishment as a reproductively isolated species may depend on stochastic events (e.g., surviving demographic fluctuations), yet at the very least the example illustrates what the early stages of hybrid speciation could look like.

A more ancient example of avian hybrid speciation involves two Amazonian manakins, initially hybridizing during the Pleistocene. The snow-capped manakin (*Lepidothrix nattereri*) and the opal-crowned manakin (*Lepidothrix iris*) putatively hybridized to form the golden-crowned manakin (*Lepidothrix vilasboasi*). The ranges of these three species are mostly defined by barriers such as mountains and rivers, yet the parental species also currently hybridize in narrow areas where their ranges come into contact. This contemporary hybridization allows for comparison between modern hybrids and the putative hybrid species, *L. vilasboasi* individuals. Barrera-Guzmán et al. (2017) found that both types of hybrids were intermediate between the parental species, in their genomes and in structural aspects of the ornamental feathers on their crowns. Modern hybrids showed a range of admixture proportions, consistent with many types of hybrids (e.g., F1s, backcrosses, etc.). *L. vilasboasi* individuals, however, had constant admixture proportions (~80% versus ~ 20% split between *L. iris* and *L. nattereri* origin, respectively) and could be differentiated from modern hybrids. The genome of *L. vilasboasi* formed as a result of the shuffling of the genomes from the parental species, and sufficient time has gone by for there to have been sorting of ancestral alleles leading to the genetic diagnosability of *L. vilasboasi*. Both types of hybrids are also intermediate in the structure of their crown feathers. However, while the crown of modern hybrids resembles that of the parental species, *L. vilasboasi* has a yellow crown formed by the deposition of carotenoids. This unique phenotype is hypothesized to have led to the pre-mating isolation of the hybrid species.

The final example, the Italian sparrow (*Passer italiae*), has the largest range of known avian hybrid species, occupying all of Italy and a few islands in the Mediterranean Ocean (Fig. 5). This species is the product of hybridization between the Spanish sparrow (*P. hispaniolensis*) and the house sparrow (*P. domesticus*) and is thought to have formed within the last 10,000 years, when the latter species colonized the Mediterranean. The hybrid species does not mate with the Spanish

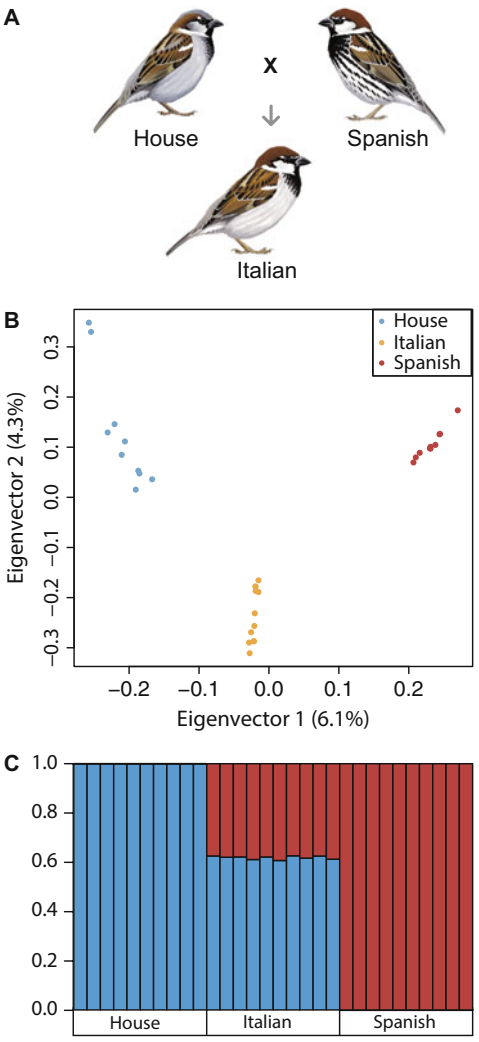


**Fig. 5** The hybrid origin of the Italian sparrow. The formation of the Italian sparrow is a result of hybridization between two parental species, the house sparrow and the Spanish sparrow (panel **a**). Panels (**b**) and (**c**) show a PCA and a structure admixture plot derived from genomic data, showing it is genetically intermediate between its parental species. Panel (**a**) is from del Hoyo et al. (2018); Panels (**b**) and (**c**) are reproduced from Elgvin et al. (2017) © The Authors, some rights reserved; exclusive licensee AAAS, distributed under CC BY-NC (http://creativecommons.org/licenses/by-nc/4.0/)

sparrow where it occurs in sympatry and forms a stable hybrid zone with the house sparrow in the Alps. The Italian sparrow is intermediate genetically and phenotypically between its parental species (Hermansen et al. 2011), and its genome is a mosaic of that of the Spanish and house sparrows (Elgvin et al. 2017). However, across the range of the Italian sparrow, there are at least four different combinations of the parental genomes (i.e., combined in different proportions), which suggests there could have been different hybridization events and shuffling of the genomes (Runemark et al. 2018). Despite this range of different proportions of the parental genomes, there are some areas that are consistently inherited from a single species.

These areas are disproportionately located on the Z chromosome and contain candidate loci for generating incompatibilities that may lead to the reproductive isolation of the Italian sparrow.

## 6 Informing Conservation Using Genomic Data

As global biodiversity loss continues at an alarming rate, conservation biologists are tasked with preserving and managing natural populations and maximizing genetic variation in vulnerable species. While the toolbox of a conservation biologist is large and diverse, genetics has long been an integral component to informing management efforts. Conservation genetics is devoted to the study of genetic and evolutionary patterns and processes within the context of biodiversity conservation (Avise 2010; Frankham 2010; Angeloni et al. 2012). Over the years, the field has provided valuable insight into several conservation issues facing birds, including the impacts of habitat fragmentation, the preservation of genetic diversity, the appropriate management of captive populations and invasive species, a better understanding of population structure, and offering new insights into taxonomic boundaries (Frankham 2010). Despite its achievements, advances in conservation genetics remain slower than in other areas of avian genomics, if only because many studies continue to rely on a few genetic markers, particularly microsatellites or mtDNA. As discussed previously, by limiting studies to only a few genetic markers, it is challenging to determine just how representative these are of the rest of the genome. Furthermore, because most traditional genetic markers are assumed to be selectively neutral, their utility in studies designed to evaluate processes such as local adaptation, evolutionary potential, or the impacts of inbreeding remains a topic of debate (Allendorf et al. 2001; Ouborg et al. 2010; Angeloni et al. 2012). Thus, the application of genomic approaches to conservation allows for both increased resolution and a more representative view into genome-wide patterns, as well as an opportunity to study adaptive variation across populations and habitats.

While genomic approaches are considered a promising tool for advancing conservation practices (Primmer 2009; Avise 2010; Frankham 2010; Angeloni et al. 2012; Shafer et al. 2015), examples of genomic data in conservation biology, particularly in the avian literature, remain rare. In many cases, this is due to several logistical challenges that accompany the shift to next generation technologies. However, the increasing accessibility to genomic resources has the potential to transform the discipline. To illustrate the diversity of questions that can be addressed using genomic data, we present three avian case studies below. These examples represent novel applications of genomics in conservation and provide examples of the research-to-application frameworks necessary for advancing the field (Shafer et al. 2015).

## 6.1 Population Structure and Delineation

A central goal in biodiversity conservation is the ability to recognize relevant biological units that act more or less independently in a management or an evolutionary context (i.e., management units or evolutionary significant units). Across all the applications in wildlife genomics, it seems like no other field of study is better suited for the inclusion of genomic data than the quantification of genetic population structure in natural systems. This is best illustrated by a study of a charismatic bird of significant conservation concern: the sage-grouse (*Centrocercus* spp.). These birds represent icons of the Western United States and are often recognized for their lek-based mating system and impressive male displays. These sagebrush obligates have also been the focus of many conservation efforts and have recently been the center of controversies regarding their listing status under the Endangered Species Act (Knick and Connelly 2011; Jahner et al. 2016). Over the past century there has been substantial reductions of sage-grouse habitat (Knick and Connelly 2011). Population declines range from 45 to 90% across their distribution, with sage-grouse currently occupying approximately one-half of their historic range (Connelly and Braun 1997; Schroeder et al. 2004). From a conservation genetics standpoint, the mating system and behavior of sage-grouse leave them particularly vulnerable to local reductions in genetic diversity. Reduced dispersal among leks due to high site fidelity (Gibson et al. 2014), coupled with reproductive skew and limited habitat availability, may also result in reduced gene flow, enhanced genetic drift, and increased sub-structuring among populations (Oyler-McCance et al. 2015; Jahner et al. 2016). Uncovering patterns of population structure within sage-grouse has several relevant management implications, particularly regarding the identification of conservation units.

Two distinct species of sage-grouse have been identified: the greater (*C. urophasianus*) and Gunnison (*C. minimus*) sage-grouse. These distinctions were primarily based on morphology, behavior, and neutral genetic variation at microsatellite markers (Oyler-McCance et al. 2015). A parapatric group of greater sage-grouse – the Bi-State population – is also genetically differentiated at neutral microsatellites despite little differentiation in behavior and morphology. To better understand patterns of genetic variation across the sage-grouse range, Oyler-McCance et al. (2015) used a reduced-representation approach to re-evaluate genetic variation among Gunnison sage-grouse, the Bi-State population of greater sage-grouse, and other greater sage-grouse populations. Using thousands of SNPs, Oyler-McCance et al. (2015) documented similar population clusters as identified using microsatellites and mtDNA (Oyler-McCance et al. 2005), but that differed in terms of the extent of differentiation among the groups. Previous work by Oyler-McCance et al. (2005) concluded that the Gunnison sage-grouse and the Bi-State population were both different from the Greater sage-grouse and that these differences were similar in magnitude. However, increased resolution gained from genome-wide data indicates that the differences between Gunnison and greater sage-grouse are substantial and greater than that observed between greater sage-grouse and

the Bi-State population. While these results provide support for current taxonomic designations and management units, this information may influence how management efforts are prioritized among these groups.

To further investigate patterns of population structuring in sage-grouse, Jahner et al. (2016) employed over tens of thousands of SNPs from multiple leks to assess patterns of fine-scale structuring across a small region in central Nevada. Jahner et al. (2016) found that sage-grouse from different lek complexes formed statistically identifiable clusters. Although some studies have found patterns of isolation by distance (Davis et al. 2015), few studies have found patterns of genetic structure on this fine of a scale. The increased resolution gained from genome-wide SNPs highlights the promise of genomics in informing the identification of conservation units. Jahner et al. (2016) concluded that the incorporation of genome-wide markers allowed for the discovery of cryptic genetic structure in sage-grouse populations. Understanding how genetic diversity is distributed across the landscape can inform conservation efforts and aid in efforts to maximize the preservation of genetic diversity for this declining species.

## 6.2 *Genomic Insights into Inbreeding and Population Decline*

Conservation genetics tends to focus on extinction risks to small and isolated populations, due to their vulnerability to genetic drift and inbreeding. Loss of genetic diversity can lead to inbreeding depression – the reduction in fitness in inbred individuals compared to their outbred relatives (Charlesworth and Willis 2009) – and can reduce the capacity of a population to adapt to changing environments (Angeloni et al. 2012). While the importance of genetic diversity is well accepted, very few studies have documented direct genetic and fitness consequences of decreased immigration into natural populations. Using genomic data, Chen et al. (2016) assessed temporal variation in gene flow, inbreeding, and fitness in a population of threatened Florida scrub jays (*Aphelocoma coerulescens*) over two decades (1995–2013; Fig. 6). Habitat loss and degradation has resulted in range-wide declines in Florida scrub jay populations (Woolfenden and Fitzpatrick 1996). Due to management efforts, the study population has remained stable, while the surrounding regions have declined. As a result, Chen et al. (2016) quantified the genetic consequences of decreased immigration into the study site from surrounding populations over a two-decade period. Immigrants into the population had significantly lower levels of heterozygosity compared to residents, indicative of immigrant birds originating from smaller and more isolated surrounding populations. Despite this reduction in heterozygosity, immigrant birds contributed genetic variation overall, which was important, as it appeared to alleviate inbreeding pressures within the focal population.

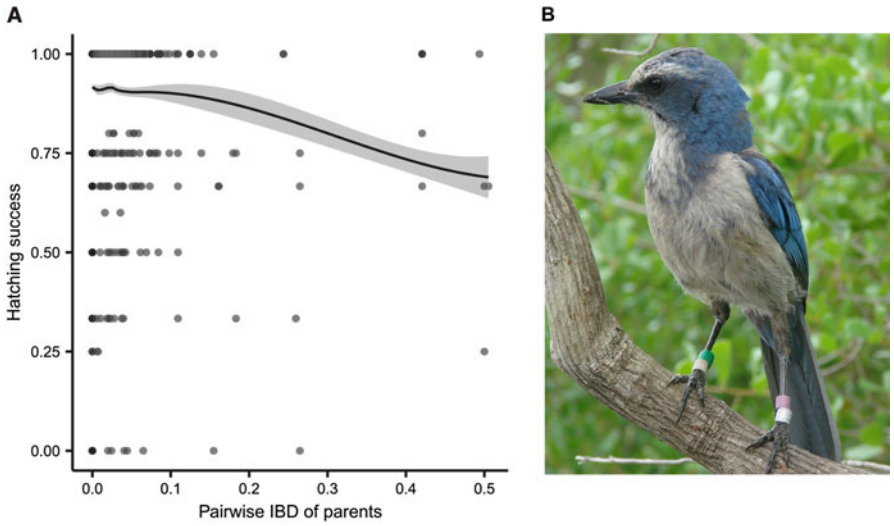


**Fig. 6** Inbreeding in Florida scrub jays. Panel (**a**) shows the proportion of the genome shared IBD (identical by descent) between parents and the correlation with hatching success. Hatching success is negatively correlated with more evidence of inbreeding. The line shows predicted values from the model, and the shaded area represents the 95% confidence interval around these values. Panel (**a**) is reproduced from Chen et al. (2016). Florida Scrub Jay image obtained from Creative Commons (reproduced under Attribution-ShareAlike 4.0 International)

Chen et al. (2016) additionally found reduced hatching success, lower nestling weights, and reduced survival in the first year after fledging in response to increased inbreeding. Using a combination of high-resolution genomic data and long-term demographic and pedigree information from a single study population, this study marks one of the first to directly document the negative fitness consequences of reduced gene flow within a relatively stable natural population. It also highlights the importance of connectivity and gene flow when managing fragmented populations. Although the focal population assayed by Chen et al. (2016) was extensively managed, reduced connectivity from neighboring populations has likely resulted in increased inbreeding at a local scale. This study provides an illustrative example of how genomic approaches can be applied within a conservation framework and can provide valuable insight into managing small, isolated populations.

## *6.3 Assessing Population Response to Climate Change*

The documented impacts of climate change are vast, with observed effects spanning from the level of individual organisms to entire communities (Scheffers et al. 2016). In fact, climate change impacts have been documented across every ecosystem on Earth (Scheffers et al. 2016), with anthropogenic changes in our environment

presenting major challenges to the long-term preservation of biodiversity. In face of these changes, species persistence will be dependent on the ability to adapt to novel and rapidly changing environmental conditions (Bay et al. 2017). Furthermore, empirical studies suggest that environmental resilience can be highly variable among individuals and populations, as well as across species (Somero 2010; Hoffmann and Sgrò 2011). There is wide acceptance among biologists that the preservation of ecological viability and evolutionary processes should be a key goal in conservation (Moritz 1994). To this end, the maintenance of genetic diversity should be prioritized in hopes of providing the building blocks for future evolutionary change (Waples 1995; Fraser and Bernatchez 2001; Walsh et al. 2017). With growing empirical evidence suggesting that species are adapting to human-induced climate change (Bradshaw and Holzapfel 2001; Kovach et al. 2012), documenting the existence and distribution of adaptive alleles associated with climate can have important implications for conservation efforts (Bay et al. 2018; Ruegg et al. 2018). High-throughput sequencing technologies have the potential to aid conservation biologists in identifying adaptive alleles that may aid in bolstering population persistence in the face of rapid environmental change.

Using a reduced-representation sequencing approach combined with environmental data, Bay et al. (2018) examined the genomic basis of climate adaptation in the yellow warbler (*Setophaga petechia*). Because these migratory warblers span a diversity of environmental conditions across their broad distribution in the United States and Canada, they are good candidates for identifying patterns of local climate adaptation. While Bay et al. (2018) found little evidence of substructure across the range of the yellow warbler (isolation by distance showed the strongest correlation with genome-wide variation), they did document differences in environmentally associated genetic variation across longitude and latitude, suggestive of local adaptation associated with the environment. The authors used this information to calculate "genomic vulnerability," which they define as the *mismatch between current and predicted future genomic variation based on genotype-habitat relationships*. Bay et al. (2018) correlated this metric of genomic vulnerability with population trend estimates and showed that regions that had high genomic vulnerability scores had experienced the largest population declines over the past century. This suggests that a mismatch between genetic variation and local climate may already be driving population declines, as these vulnerable populations are predicted to be least likely to adapt quickly enough to track climatic shifts. Lastly, Bay et al. (2018) tested for associations with genotypes and the top explanatory environmental variables and identified several SNPs that correlated strongly with local environment. One of the strongest associations that they identified was upstream of genes with known function in avian behavior and migration – two genes of interest have been linked to exploratory behavior, which may allow species to occupy new environments. This work presents a novel approach to combining environmental, demographic, and genomic data to understand patterns of local genetic adaptation in the face of climatic change. The authors illustrate how information on genetic variation and adaptation can inform species distribution models and guide conservation efforts in the face of rapid environmental change.

## 7 Concluding Remarks

A general theme across this chapter has been the increased resolution that genomic data have provided to answer evolutionary- and conservation-related questions in avian systems. This resolution has improved the ability to diagnose distinct avian populations, to generate more complex models of demographic histories, and to identify regions of the genome that confer important ecological functions. In many cases, these advances have been paced by technological improvements, and, undoubtedly, we are witness to the "era of short reads" or "second"-generation sequencing technologies. It is likely this will change in the near future, and "third"-generation sequencing technologies are already promising to provide even more, long-read high-quality data. What is notable about the methodological shift associated with high-throughput sequencing has been the empowerment of researchers to generate genetic data at a fraction of the cost that was associated with traditional sequencing. Here we have illustrated some of the most important conceptual advances in the field of avian genomics. However, the fact that these technologies can be readily adapted to more practical applications has a clear, albeit less acknowledged, benefit. Taken together, we envision that the study of avian genomes will continue to provide important insights into many aspects of ecology, evolutionary history, and conservation biology.

**Acknowledgments** DPLT was supported by an NSERC Banting Postdoctoral Fellowship. JW was supported by NSF Postdoctoral Research Fellowship in Biology. The authors would like to thank Irby J. Lovette and the Fuller Evolutionary Biology Program for intellectual support and thank Eliot Miller, Gavin M. Leighton, and Paul Hohenlohe for comments on a previous version of this chapter.

## References

Abbott RJ, Albach D, Ansell S, Arntzen JW, SJE B, Bierne N, et al. Hybridization and speciation. J Evol Biol. 2013;26:229–46.

Abbott RJ, Barton NH, Good JM. Genomics of hybridization and its evolutionary consequences. Mol Ecol. 2016;25:2325–32.

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. *Bmp4* and morphological variation of beaks in Darwin's finches. Science. 2004;305:1462–5.

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature. 2006;442:563–7.

Alerstam T. Conflicting evidence about long-distance animal navigation. Science. 2006;313:791–4.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. Trends Ecol Evol. 2001;16:613–22.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.

Angeloni F, Wagemaker N, Vergeer P, Ouborg J. Genomic toolboxes for conservation biologists. Evo Appl. 2012;5:130–43.

Arrieta RS, Lijtmaer DA, Tubaro PL. Evolution of postzygotic reproductive isolation in galliform birds: analysis of first and second hybrid generations and backcrosses. Biol J Linn Soc. 2013;110:528–42.
Avise JC. Perspective: conservation genetics enters the genomics era. Cons Gen. 2010;11:665–9.
Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 1987;18:489–522.
Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol. 2004;13:729–44.
Barrera-Guzmán AO, Aleixo A, Shawkey MD, Weir JT. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. Proc Natl Acad Sci. 2017;115:E218–25.
Barsh G. Evolution: sex, diet and red ketocarotenoids. Curr Biol. 2016;26:R1145–7.
Barton NH, Hewitt GM. The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. Heredity. 1981;47:367–83.
Barton NH, Hewitt GM. Analysis of hybrid zones. Annu Rev Ecol Syst. 1985;16:113–48.
Bay RA, Rose N, Barrett R, Bernatchez L, Ghalambor CK, Lasky JR, et al. Predicting responses to contemporary environmental change using evolutionary response architectures. Am Nat. 2017;189:463–73.
Bay RA, Harrigan RJ, Le Underwood V, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science. 2018;359:83–6.
Boag PT, Grant PR. Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. Science. 1981;214:82–5.
Bradshaw WE, Holzapfel CM. Genetic shift in photoperiodic response correlated with global warming. Proc Natl Acad Sci. 2001;98:14509–11.
Brelsford A, Toews DPL, Irwin DE. Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. Proc R Soc B. 2017. https://doi.org/10.1098/rspb.2017.1106.
Buehler DA, Roth AM, Vallender R, Will TC, Confer JL, Canterbury RA, et al. Status and conservation priorities of golden-winged warbler (*Vermivora chrysoptera*) in North America. Auk. 2007;124:1439–45.
Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res. 2015;25:1656–65.
Burt DW. Origin and evolution of avian microchromosomes. Cytogenet Genome Res. 2002;96:97–112.
Calderón L, Campagna L, Wilke T, Lormee H, Eraud C, Dunn JC, et al. Genomic evidence of demographic fluctuations and lack of genetic structure across flyways in a long-distance migrant, the European turtle dove. BMC Evol Biol. 2016;16:237.
Campagna L, Repenning M, Silveira LF, Fontana CS, Tubaro PL, Lovette IJ. Repeated divergent selection on pigmentation genes in a rapid finch radiation. Sci Adv. 2017. https://doi.org/10.1126/sciadv.1602404.
Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, et al. The genomic architecture of population divergence between subspecies of the European rabbit. PLoS Gen. 2014. https://doi.org/10.1371/journal.pgen.1003519.
Charlesworth B. The effect of life-history and mode of inheritance on neutral genetic variability. Genet Res. 2001;77:153–66.
Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10:783–96.
Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 1987;130:113–46.
Chen N, Cosgrove EJ, Bowman R, Fitzpatrick JW, Clark AG. Genomic consequences of population decline in the endangered Florida scrub-jay. Curr Biol. 2016;26:2974–9.

Connelly JW, Braun CE. Long-term changes in sage grouse *Centrocercus urophasianus* populations in western North America. Wildlife Biol. 1997;3:229–34.
Connelly JW, Knick ST, Braun CE, Baker WL, Beever EA, Christiansen T, et al. Conservation of greater sage-grouse. Studies Avian Biol. 2011;38:549–63.
Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014;23:3133–57.
Cuthill IC, Allen WL, Arbuckle K, Caspers B, Chaplin G, Hauber ME, et al. The biology of color. Science. 2017. https://doi.org/10.1126/science.aan0221.
Davis DM, Reese KP, Gardner SC, Bird KL. Genetic structure of greater sage-grouse (*Centrocercus urophasianus*) in a declining, peripheral population. Condor. 2015;117:530–44.
del Hoyo J, Elliott A, Sargatal J, Christie DA, Kirwan G, editors. Handbook of the birds of the world alive. Barcelona: Lynx Edicions; 2018. http://www.hbw.com/. Accessed 31 May 2018.
Delmore KE, Hübner S, Kane NC, Schuster R, Andrew RL, Câmara F, et al. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol Ecol. 2015;24:1873–88.
Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE. The genetics of seasonal migration and plumage color. Curr Biol. 2016;26:2167–73.
Diamond J. Biology of birds of paradise and bowerbirds. Annu Rev Ecol Syst. 1986;17:17–37.
Drovetski SV, Zink RM, Rohwer S, Fadeev IV, Nesterov EV, Karagodin I. Complex biogeographic history of a Holarctic passerine. Proc R Soc B. 2004;271:545–51.
Dupuis JR, Sperling FA. Hybrid dynamics in a species group of swallowtail butterflies. J Evol Biol. 2016;29:1932–51.
Edwards S, Bensch S. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol. 2009;18:2930–3.
Elgvin TO, Trier CN, Tørresen OK, Hagen IJ, Lien S, Nederbragt AJ, et al. The genomic mosaicism of hybrid speciation. Sci Adv. 2017. https://doi.org/10.1126/sciadv.1602996.
Ellegren H. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol. 2010;25:283–91.
Ellegren H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet. 2011;12:157–66.
Ellegren H. The evolutionary genomics of birds. Annu Rev Ecol Evol Syst. 2013;44:239–59.
Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature. 2012;491:756–60.
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011. https://doi.org/10.1371/journal.pone.0019379.
Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol. 2012;61:717–26.
Frankham R. Where are we in conservation genetics and where do we need to go? Cons Gen. 2010;11:661–3.
Fraser DJ, Bernatchez L. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol Ecol. 2001;10:2741–52.
Futuyma D. Evolution. Sunderland, MA: Sinauer Associates, Inc; 2013.
Gibson D, Blomberg EJ, Atamian MT, Sedinger JS. Lek fidelity and movement among leks by male greater sage-grouse *Centrocercus urophasianus*: a capture-mark-recapture approach. Ibis. 2014;156:729–40.
Grant PR, Grant BR. Hybridization of bird species. Science. 1992;256:193–7.
Grant PR, Grant BR. The secondary contact phase of allopatric speciation in Darwin's finches. Proc Natl Acad Sci. 2009;106:2xs0141–8.
Grant PR, Grant BR. 40 years of evolution: Darwin's finches on Daphne Major Island. Princeton, NJ: Princeton University Press; 2014.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. Nucleic Acids Res. 2007. https://doi.org/10.1093/nar/gkl828.
Haas F, Pointer MA, Saino N, et al. An analysis of population genetic differentiation and genotype–phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and MC1R. Mol Ecol. 2009;18:294–305.
Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ, et al. A phylogenomic study of birds reveals their evolutionary history. Science. 2008;320:1763–8.
Haffer J. Speciation in Amazonian forest birds. Science. 1969;165:131–7.
Haldane JB. Sex ratio and unisexual sterility in hybrid animals. J Genet. 1922;12:101–9.
Hare MP. Prospects for nuclear gene phylogeography. Trends Ecol Evol. 2001;16:700–6.
Hare EE, Johnston JS. Genome size determination using flow cytometry of propidium iodide-stained nuclei. Mol Meth Evol Gen. 2011;772:3–12.
Haring E, Gamauf A, Kryukov A. Phylogeographic patterns in widespread corvid birds. Mol Phylogenet Evol. 2007;45:840–62.
Harrison RG. Pattern and process in a narrow hybrid zone. Heredity. 1986;56:337–49.
Harrison RG. Hybrid zones and the evolutionary process. Oxford: Oxford University Press; 1993.
Helbig AJ. SE-and SW-migrating blackcap (*Sylvia atricapilla*) populations in Central Europe: orientation of birds in the contact zone. J Evol Biol. 1991;4:657–70.
Hermansen JS, Saether SA, Elgvin TO, Borge T, Hjelle E, Saetre GP. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Mol Ecol. 2011;20:3812–22.
Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006;313:101–4.
Hoffmann AA, Sgrò CM. Climate change and evolutionary adaptation. Nature. 2011;470:479–85.
Hooper DM, Price TD. Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. Evolution. 2015;69:890–903.
Hubbard JK, Uy JA, Hauber ME, Hoekstra HE, Safran RJ. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 2010;26:231–9.
Hung CM, Shaner PJ, Zink RM, Liu WC, Chu TC, Huang WS, Li SH. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc Natl Acad Sci. 2014;111:10636–41.
International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432:695–716.
Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. Mol Ecol. 2018. https://doi.org/10.1111/mec.14537.
Itoh Y, Kampf K, Balakrishnan CN, Arnold AP. Karyotypic polymorphism of the zebra finch Z chromosome. Chromosoma. 2011;120:255–64.
Jahner JP, Gibson D, Weitzman CL, Blomberg EJ, Sedinger JS, Parchman TL. Fine-scale genetic structure among greater sage-grouse leks in central Nevada. BMC Evol Biol. 2016. https://doi.org/10.1186/s12862-016-0702-4.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014;346:1320–31.
Johnston RA, Paxton KL, Moore FR, Wayne RK, Smith TB. Seasonal gene expression in a migratory songbird. Mol Ecol. 2016;25:5680–91.
Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, et al. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol. 2014;23:4035–58.
Kimball RT, Braun EL, Barker FK, Bowie RC, Braun MJ, Chojnowski JL, et al. A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol. 2009;50:654–60.

Knick S, Connelly JW. Greater sage-grouse: ecology and conservation of a landscape species and its habitats, vol. 38. Berkeley, CA: University of California Press; 2011.
Kovach RP, Gharrett AJ, Tallmon DA. Genetic change for earlier migration timing in a pink salmon population. Proc R Soc B. 2012;279:3870–8.
Kraus RH, Wink M. Avian genomics: fledging into the wild! J Ornith. 2015;156:851–65.
Kryukov AP, Suzuki H. Phylogeography of carrion, hooded, and jungle crows (Aves, Corvidae) inferred from partial sequencing of the mitochondrial DNA cytochrome b gene. Russ J Gen. 2000;36:922–9.
Kryukov AP, Uphyrkina OV, Chelomina GN. Analysis of crow genomes (Corvidae, Passeriformes) from the zone of overlapping areas of habitation and hybridization. Genetika. 1992;28:136–40.
Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, et al. A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet. 2016;48:79–83.
Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518:371–5.
Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nat Genet. 2016a;48:84–8.
Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, et al. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science. 2016b;352:470–4.
Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. Rapid hybrid speciation in Darwin's finches. Science. 2018;359:224–8.
Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475:493–6.
Liedvogel M, Åkesson S, Bensch S. The genetics of migration on the move. Trends Ecol Evol. 2011;26:561–9.
Lijtmaer DA, Mahler B, Tubaro PL. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution. 2003;57:1411–8.
Lopes RJ, Johnson JD, Toomey MB, Ferreira MS, Araujo PM, Melo-Ferreira J, et al. Genetic basis for red coloration in birds. Curr Biol. 2016;26:1427–34.
Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2017;17:142–52.
Manthey JD, Moyle RG, Boissinot S. Multiple and independent phases of transposable element amplification in the genomes of Piciformes (woodpeckers and allies). Genome Biol Evol. 2018. https://doi.org/10.1093/gbe/evy105.
McCormack JE, Tsai WL, Faircloth BC. Sequence capture of ultraconserved elements from bird museum specimens. Mol Ecol Resour. 2016;16:1189–203.
Moore WS. An evaluation of narrow hybrid zones in vertebrates. Q Rev Biol. 1977;52:263–77.
Moritz C. Defining 'evolutionarily significant units' for conservation. Trends Ecol Evol. 1994;9:373–5.
Mueller JC, Pulido F, Kempenaers B. Identification of a gene associated with avian migratory behaviour. Proc R Soc B. 2011;278:2848–56.
Mundy NI. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc R Soc B. 2005;272:1633–40.
Mundy NI, Stapley J, Bennison C, Tucker R, Twyman H, Kim KW, et al. Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster. Curr Biol. 2016;26:1435–40.
Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr Biol. 2015;25:1375–80.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol Ecol. 2016;25:1058–72.
Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. Origin of avian genome size and structure in non-avian dinosaurs. Nature. 2007;446:180–4.
Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. Conservation genetics in transition to conservation genomics. Trends Genet. 2010;26:177–87.
Oyler-McCance SJ, Taylor SE, Quinn TW. A multilocus population genetic survey of the greater sage-grouse across their range. Mol Ecol. 2005;14:1293–310.
Oyler-McCance SJ, Cornman RS, Jones KL, Fike JA. Genomic single-nucleotide polymorphisms confirm that Gunnison and greater sage-grouse are genetically well differentiated and that the bi-state population is distinct. Condor. 2015;117:217–27.
Oyler-McCance SJ, Oh KP, Langin KM, Aldridge CL. A field ornithologist's guide to genomics: practical considerations for ecology and conservation. Auk. 2016;133:626–48.
Pacheco-Sierra G, Gompert Z, Domínguez-Laso J, Vázquez-Domínguez E. Genetic and morphological evidence of a geographically widespread hybrid zone between two crocodile species, *Crocodylus acutus* and *Crocodylus moreletii*. Mol Ecol. 2016;25:3484–98.
Payseur BA. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Mol Ecol Resour. 2010;10:806–20.
Pellicer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? Botan J Linn Soc. 2010;164:10–5.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012. https://doi.org/10.1371/journal.pone.0037135.
Picozzi N. Hybridization of carrion and hooded crows *Corvus c. corone* and *Corvus c. cornix*, in northeastern Scotland. Ibis. 1976;118:254–7.
Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science. 2014;344:1410–4.
Price TF. Speciation in birds. Greenwood Village, CO: Roberts and Company; 2008.
Price TD, Bouvier MM. The evolution of F1 postzygotic incompatibilities in birds. Evolution. 2002;56:2083–9.
Primmer CR. From conservation genetics to conservation genomics. Annals NY Acad Sci. 2009;1162:357–68.
Primmer CR, Møller AP, Ellegren H. A wide-range survey of cross-species microsatellite amplification in birds. Mol Ecol. 1996;5:365–78.
Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 2015;526:569–73.
Qu Y, Luo X, Zhang R, Song G, Zou F, Lei F. Lineage diversification and historical demography of a montane bird *Garrulax elliotii*-implications for the Pleistocene evolutionary history of the eastern Himalayas. BMC Evol Biol. 2011. https://doi.org/10.1186/1471-2148-11-174.
Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, et al. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. Syst Biol. 2017;66:857–79.
Rheindt FE, Edwards SV. Genetic introgression: an integral but neglected component of speciation in birds. Auk. 2011;128:620–32.
Ribas CC, Aleixo A, Nogueira AC, Miyaki CY, Cracraft J. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Proc R Soc B. 2012;279:681–9.
Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001;16:351–8.
Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol. 2014;23:4757–69.

Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, et al. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecol Lett. 2018. https://doi.org/10.1111/ele.12977.

Rull V. Neotropical biodiversity: timing and potential drivers. Trends Ecol Evol. 2011;26:508–13.

Runemark A, Trier CN, Eroukhmanoff F, Hermansen JS, Matschiner M, Ravinet M, Elgvin TO, Sætre GP. Variation and constraints in hybrid genome formation. Nat Ecol Evo. 2018;2:549–56.

Saino N, Villa S. Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. Auk. 1992;109:543–55.

Saino N, Lorenzini R, Fusco G, Randi E. Genetic-variability in a hybrid zone between carrion and hooded crows (*Corvus corone corone* and *C. c. cornix*, Passeriformes, Aves) in north-western Italy. Biochem Syst Ecol. 1992;20:605–13.

Saino N, Wuster W, Thorpe RS. Congruence between morphological variation and altitudinal gradient across a hybrid zone between carrion and hooded crows. Ital J Zool. 1998;65:407–12.

Scheffers BR, De Meester L, Bridge TC, Hoffmann AA, Pandolfi JM, Corlett RT, et al. The broad footprint of climate change from genes to biomes to people. Science. 2016. https://doi.org/10.1126/science.aaf7671.

Schroeder MA, Aldridge CL, Apa AD, Bohne JR, Braun CE, Bunnell SD, et al. Distribution of sage-grouse in North America. Condor. 2004;106:363–76.

Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Curr Biol. 2014;24:R288–94.

Shafer AB, Wolf JB, Alves PC, Bergström L, Bruford MW, Brännström I, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30:78–87.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable recombination hotspots in birds. Science. 2015;350:928–32.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. Target capture and massively parallel sequencing of ultraconserved elements (UCEs) for comparative studies at shallow evolutionary time scales. Syst Biol. 2013;63:83–95.

Somero GN. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. J Exp Biol. 2010;213:912–20.

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol Phylo. 1999;12:105–14.

Stryjewski KF, Sorenson MD. Mosaic genome evolution in a recent and rapid avian radiation. Nat Eco Evo. 2017. https://doi.org/10.1038/s41559-017-0364-7.

Swenson NG, Howard DJ. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am Nat. 2005;166:581–91.

Taylor S, Campagna L. Avian supergenes. Science. 2016;351:446–7.

Thomas JW, Cáceres M, Lowman JJ, Morehouse CB, Short ME, Baldwin EL, et al. The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. Genetics. 2008;179:1455–68.

Toews DPL. Evolution: a genomic guide to bird population history. Curr Biol. 2015;25:R465–7.

Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 2012;21:3907–30.

Toews DPL, Campagna L, Taylor SA, Balakrishnan CN, Baldassarre DT, Deane-Coe PE, et al. Genomic approaches to understanding population divergence and speciation in birds. Auk. 2016a;133:13–30.

Toews DP, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr Biol. 2016b;26:2313–8.

Toomey MB, Lopes RJ, Araújo PM, Johnson JD, Gazda MA, Afonso S, et al. High-density lipoprotein receptor *SCARB1* is required for carotenoid coloration in birds. Proc Natl Acad Sci. 2017;114:5219–24.

Tubaro PL, Lijtmaer DA. Hybridization patterns and the evolution of reproductive isolation in ducks. Biol J Linn Soc. 2002;77:193–200.

Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, et al. Divergence and functional degradation of a sex chromosome-like supergene. Curr Biol. 2016;8:344–50.

Twyford AD, Ennos RA. Next-generation hybridization and introgression. Heredity. 2012;108:179–89.

Uy JA, Cooper EA, Cutie S, Concannon MR, Poelstra JW, Moyle RG, Filardi CE. Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. Proc R Soc B. 2016. https://doi.org/10.1098/rspb.2016.0731.

Vallender R, Robertson RJ, Friesen VL, Lovette IJ. Complex hybridization dynamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. Mol Ecol. 2007;16:2017–29.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JB. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun. 2016. https://doi.org/10.1038/ncomms13195.

Walsh J, Lovette IJ, Winder V, Elphick CS, Olsen BJ, Shriver G, Kovach AI. Subspecies delineation amid phenotypic, geographic and genetic discordance in a songbird. Mol Ecol. 2017;26:1242–55.

Wang W, McKay BD, Dai C, Zhao N, Zhang R, Qu Y, et al. Glacial expansion and diversification of an east Asian montane bird, the green-backed tit (*Parus monticolus*). J Biogeogr. 2013;40:1156–69.

Waples RS. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. Am Fish Soc Symp. 1995;17:8–27.

Weir JT, Schluter D. Ice sheets promote speciation in boreal birds. Proc R Soc B. 2004;271:1881–7.

Weir JT, Faccio MS, Pulido-Santacruz P, Barrera-Guzmán AO, Aleixo A. Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. Evolution. 2015;69:1823–34.

Weir JT, Haddrath O, Robertson HA, Colbourne RM, Baker AJ. Explosive ice age diversification of kiwi. Proc Natl Acad Sci. 2016;113:E5580–7.

Wolf JB, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. Mol Ecol. 2010;19:162–75.

Woolfenden GE, Fitzpatrick JW. Florida Scrub-Jay (*Aphelocoma coerulescens*), version 2.0. In: Poole AF, Gill FB, editors. The birds of North America. Ithaca, NY: Cornell Lab of Ornithology; 1996. https://doi.org/10.2173/bna.228.

Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG, et al. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. Nat Genet. 2013;45:563–6.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014a;346:1311–20.

Zhang G, Li B, Li C, Gilbert MT, Jarvis ED, Wang J. Comparative genomic data of the Avian Phylogenomics Project. Gigascience. 2014b. https://doi.org/10.1186/2047-217X-3-26.

Zink RM, Barrowclough GF. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol. 2008;17:2107–21.

# Population Genomics Provides Key Insights into Admixture, Speciation, and Evolution of Closely Related Ducks of the Mallard Complex


**Philip Lavretsky**

**Abstract** The ability to identify population structure, estimate rates of hybridization, and genetic sources of gene flow is critical when attempting to conserve wild populations. Recently diverged species with few pre- or post-zygotic isolating mechanisms are prone to exchange genetic material during secondary contact events, potentially causing the breakup of important coadapted genes and resulting in maladapted populations. Such events are especially exacerbated when domestic versions come into contact with their wild congeners and exchange genetic variation that had been under artificial selection. Being able to genetically identify individuals to populations or species, and thus potential hybrids, is essential when attempting to assess impacts from hybridization. Until recently, molecular methods often resulted in insufficient marker coverage to confidently assign individuals to populations for organisms comprised of closely related taxa. Advances in partial genome sequencing methods (e.g., ddRAD-seq, sequence capture) and decreasing sequencing costs have made it possible to readily access thousands of genetic markers across hundreds of samples, providing a population genomics across the landscapes of wild systems. Here, I review what landscape-level sampling coupled with thousands of nuclear markers has uncovered for a group of recently radiated ducks of the Mallard Complex (genus *Anas*). Deploying the latest population genomics approaches, researchers have been able to reconstruct complex evolutionary histories, assign individuals to species with confidence, as well as identify genetic hybrids. These population genomics studies have produced findings that are in contrast to what was thought to be known for many of these species. Among results, studies consistently found that the problem of hybridization for many of these species was due to feral mallard populations. In fact, the result of these anthropogenic hybridization events is the formation of hybrid swarms on Hawaii, North America, Eurasia, and New Zealand. Wildlife biologists are now incorporating these population genomics-based results into their management planning, demonstrating the need and importance of population genomics in wildlife conservation.

P. Lavretsky (✉)
Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX, USA
e-mail: plavretsky@utep.edu

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_76, 

**Keywords** *Anas* · Conservation genetics · ddRAD-seq · Hybridization · Partial genome sequencing · Population genomics · Wildlife management

# 1 Introduction

## *1.1 Background*

For conservation to work, it is essential to understand the biological unit for which efforts are being undertaken; however, this is often easier said than done. Specifically, management of wild populations requires an understanding of how many potentially evolutionarily independent groups exist as to devise specific conservation plans for each group (Oyler-McCance et al. 2016; Peters et al. 2016; Allendorf 2017; Ralls et al. 2018). Among issues faced for many organisms as habitats and environments continue to change is the increasing events of secondary contact with other closely related sister species or even their domestic conspecifics (Randi 2008; Kidd et al. 2009; Tufto 2017; Wang et al. 2017; Heikkinen et al. 2019). Thus, determining the source of hybridization and potential outcomes of admixture are important aspects in the field of conservation genetics (Crispo et al. 2011; Nadeau and Kawakami 2019).

The exchange of genetic material between two populations, or gene flow, is an important and sometimes necessary process in the speciation process (Dobzhansky 1940; Hoskin et al. 2005; Rundle and Nosil 2005; Schluter 2009) and can even result in positive outcomes by increasing the adaptive potential of a population or species (i.e., adaptive introgression; Hedrick 2013; Hamilton and Miller 2016; Nadeau and Kawakami 2019; vonHoldt et al. 2018; Nadachowska-Brzyska et al. 2019; Qiao et al. 2019; Owens and Samuk 2020) or even increasing overall biodiversity (i.e., hybrid speciation; Mallet 2007; Jacobsen and Omland 2011a; Schumer et al. 2014; Lavretsky et al. 2015b; Lamichhaney et al. 2018). In addition, conservation efforts for highly fragmented populations often require the direct movement of individuals to artificially reinvigorate gene flow (Ralls et al. 2018). Typically, however, gene flow is considered as a negative player in conservation as the most foreseen consequence(s) are often the creation of perpetual hybrid zone(s) (Barton and Hewitt 1989) and inhibition of the speciation process (Mallet 2005, 2007), including a reversal of speciation (Seehausen 2006; Webb et al. 2011; Kearns et al. 2018) and outright extinction through introgressive hybridization (Rhymer 2006) (also see Fig. 1 in Crispo et al. 2011). Though gene flow is becoming more relevant and clearly an important player in the evolution of many organisms (Mallet 2007; Nadeau and Kawakami 2019; vonHoldt et al. 2018), it is artificially induced secondary contact events (a.k.a. anthropogenic gene flow; McFarlane and Pemberton 2019) that are now of conservation concern for many organisms (Lande 1998; Puigcerver et al. 2014; Lavretsky et al. 2015b; Skoglund et al. 2015; Wayne and Shaffer 2016; Söderquist et al. 2017; Leitwein et al. 2018; McFarlane and Pemberton 2019).

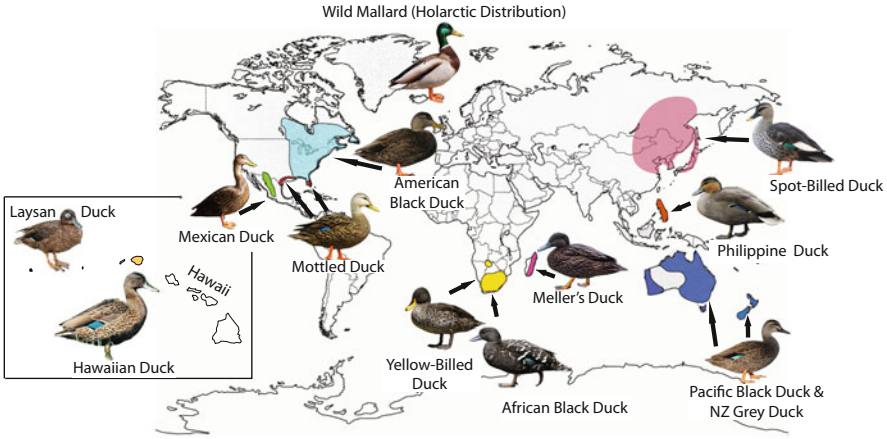


**Fig. 1** Picture representations of all mallard-like ducks and their respective distributions; note wild mallards have a Holarctic distribution. Hawaiian Islands are within the inset

Identifying genetically admixed individuals and their true parental populations with confidence is essential when attempting to conserve populations facing high frequencies of hybridization. Moreover, understanding the true biological outcome of hybridization (see Crispo et al. 2011), including whether such events have negative (i.e., extinction by introgressive hybridization; Rhymer 2006) or potentially positive (i.e., adaptive introgression; vonHoldt et al. 2018; Qiao et al. 2019; Owens and Samuk 2020) outcome(s), is paramount. However, determining shared genetic variability due to admixture is often complicated for recent radiations, including between wild and domestic congeners as much of the genome may be shared simply due to ancestry (Orr et al. 2004; Seehausen 2004; Wu and Ting 2004; Nosil et al. 2009; Via 2009; Nosil and Schluter 2011). Being able to assign samples to their true genetic population cluster(s) can help distinguish what genetic variability is due to shared ancestry versus gene flow (Lavretsky et al. 2019b). In general, population genomics has the potential to advance our understanding of even the most complex interactions (DaCosta and Sorenson 2014; Andrews et al. 2016; McKinney et al. 2017; Nadeau and Kawakami 2019; McFarlane and Pemberton 2019).

In this chapter, I recount how the integration of genetic information from population genomics studies has transformed our understanding of the evolutionary histories and contemporary population structure of worldwide populations of closely related ducks of the Mallard Complex. The Mallard Complex successfully radiated around the world and is comprised of 14 species of mallard-like ducks (Fig. 1). Across the continents and Islands that they successfully adapted to, these mallard-like ducks are ecologically, culturally, and economically important. Their importance has made it a priority to understand the genetic health and adaptive potential of respective populations and species by local conservation private and public organizations.

Among the species, the mallard (*Anas platyrhynchos*) is one of the most ubiquitous and well-known ducks in the world, and wildlife, in general. Its success is best

explained by its adaptive nature and the fact that it was domesticated 4,000 years ago, providing a long, intertwined history with humans. Being domesticated, many forms of the mallard have been introduced around the world. Population genomics has revealed that the expansion of these domestic mallards resulted in secondary contact with their wild congeners and ultimately resulted in high levels of anthropogenic gene flow and the formation of feral x wild hybrid swarms in many instances. Finally, I discuss the potential that historical museum samples and ancient DNA techniques provide in understanding how populations have changed, including how hybridization has shaped the genetic diversity of contemporary populations. In general, population genomics studies coupling landscape-level sampling have been able to provide important insight into the evolution and population structure and identify previously unknown levels of hybridization for a sweep of wild populations of ducks. The Mallard Complex provides a unique system to study how radiations occur and the importance of gene flow on these processes.

## *1.2 History of the Mallard Complex*

Of the many avian orders, waterfowl (order Anseriformes) experience the highest rates of hybridization (Johnsgard 1960; Livezey 1986; Lijtmaer et al. 2003), with 30–40% of species being capable of hybridizing (Grant and Grant 1992) and about 20% producing viable hybrids (Scherer and Hilsberg 1982). For example, the Mallard Complex is comprised of 13–14 closely related species of mallard-like ducks (Fig. 1) believed to have radiated out of Africa in the last million years (Palmer 1976; Johnson and Sorenson 1999). With some hybridization events producing 100% viable offspring (Avise et al. 1990), concerns over the possibility of introgressive extinction for many of the endemic mallard-like ducks have been raised over the years. Therefore, there has been a growing need to assess and delimitate individuals to species when assessing conservation risks for many of these species.

First proposed by Palmer (1976), the "out of Africa hypothesis" suggests an African origin for the mallard clade, followed by a northward and eastward radiation through Eurasia, with a stepwise progression through the South Pacific, and perhaps a single colonization event into North America. Several phylogenies have been reconstructed from either plumage characteristics (Livezey 1991), allozymes (Browne et al. 1993) or mitochondrial molecular markers (Johnson and Sorenson 1999; McCracken et al. 2001; Kulikova et al. 2004, 2005). However, these have been largely inconclusive, and it was not until Lavretsky et al. (2014b) who coupled multiple nuclear introns, multiple samples per species, and the coalescence method as implemented in the program ∗BEAST (Drummond and Rambaut 2007; Drummond et al. 2012; Bouckaert et al. 2014) that provided more robust phylogenetic relationships that largely confirmed the out of Africa evolutionary history for the Mallard Complex.

The onset for speciation of the Mallard Complex is estimated to be approximately 1 million years ago, with the most recent divergences estimated at 150,000 years ago (Lavretsky et al. 2014a, b). Within this single radiation, you can find species divergence proceeding in allopatry, parapatry, undergoing secondary contact, and potentially evolving via hybrid speciation (Lavretsky et al. 2014a, b, 2015a, b; Peters et al. 2016). With the exception of the dichromatic mallard, the remaining species are all monochromatic, where males and females show a similar phenotype (Fig. 1). With mallards being the only species with obvious sex-based plumage differences, phylogenetic comparisons have led to several speculations into the evolution of dimorphism in the mallard clade including that it was gained once (Johnson and Sorenson 1999; McCracken et al. 2001; Kulikova et al. 2004, 2005) or lost several times (Omland 1997). As a result of their relatively recent divergence, widespread incomplete lineage sorting has resulted in the retention of much of the genome among taxa, and likely contributing to the fact that viable hybrid offspring are produced in sympatry (Avise et al. 1990; Rhymer 2006; Lavretsky et al. 2014b).

Currently, with the exception of the Holarctic mallard, the remaining species are endemic to a single continent or island group (Haddon 1984; Rhymer et al. 1994; Kulikova et al. 2004) (Fig. 1). Unfortunately, environmental degradation and both human-facilitated (i.e., release programs) and natural expansion of the mallard's range have caused formerly allopatric species to come into secondary contact leading to hybridization. For example, mallard introductions in the Hawaiian Islands and New Zealand have resulted in nearly complete introgression of mallard alleles into the endemic species (i.e., Hawaiian duck or koloa (*A. wyvilliana*; Fowler et al. 2009; Wells et al. 2019) and New Zealand grey duck (*A. superciliosa superciliosa*; Rhymer et al. 1994)), respectively. Furthermore, the mallard has expanded its range westward in both Eurasia and North America, and accumulating evidence demonstrates that mallards are likely to outcompete and to hybridize with native species, including spot-billed ducks (*A. zonorhyncha*; Kulikova et al. 2004) and American black ducks (*A. rubripes*; Rhymer and Simberloff 1996), respectively. With a variety of evolutionary histories and ability to successfully interbreed, the Mallard Complex is an ideal example of an adaptive or rapid radiation in which secondary contact events can have real implications into the adaptive potential of the invaded species.

Advancing our understanding of the evolutionary history and consequences of interspecific gene flow was not possible with confidence until advancement in partial genome sequencing attained sufficient marker coverage to be able to taxonomically assign samples with confidence. Specifically, coupling landscape-level sampling with double-digest restriction site-associated DNA sequencing (ddRAD-seq) methods (Peterson et al. 2012; Catchen et al. 2017; McKinney et al. 2017) has been truly transformative for gaining insight into the genetic histories of species within the Mallard Complex (Fig. 2). Finally, with ducks being well represented in museum collections, using ancient DNA methods on historical specimens have recently permitted researchers to understand genetic change through time, helping to understand timing, cause, and extent in changes of population structure and contemporary gene flow (Lavretsky et al. 2020).

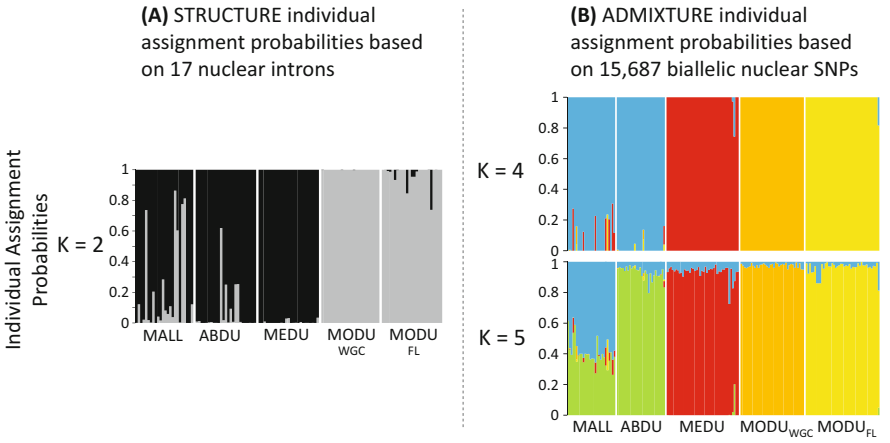


**Fig. 2** Population structure of the New World mallard clade that includes mallards (MALL), American black duck (ABDU), Mexican duck (MEDU), West Gulf Coast mottled duck ($MODU^{WGC}$), and Florida mottled duck ($MODU_{FL}$) and as determined using (**a**) 17 nuclear introns (adapted from Lavretsky et al. 2014a) or (**b**) 15,687 biallelic ddRAD-seq nuclear SNPs (adapted from Lavretsky et al. 2019a). Note the clear increased resolution of population structure using thousands of ddRAD-seq loci. For American black ducks and mallards, genetic diagnosability was achieved once a sufficient sample size of pure parental was sampled (see Fig. 4; also see Lavretsky et al. (2019b)). These results demonstrate the importance of attaining sufficient statistical diagnosability across genetic markers when dealing with systems as closely related as these mallard-liked ducks

## *1.3 Mitochondrial DNA and the Onset of Fear of Hybrid Swarms*

Avian researchers have generally focused on mtDNA. Maternally inherited and having no recombination (Giles et al. 1980; Watanabe et al. 1985), mtDNA has a more rapid sorting rate and shorter coalescent intervals relative to biparentally inherited, recombining nuclear DNA. This makes it particularly useful for recently diverged populations (Moore 1995; Zink and Barrowclough 2008). However, being maternally inherited and potentially under strong selection, its appropriateness for phylogenetics and phylogeography has been questioned (Hurst and Jiggins 2005; Bazin et al. 2006; Edwards and Bensch 2009; Jacobsen and Omland 2011b).

Mitochondrial DNA has played a significant role in shaping our understanding of the evolution of the Mallard Complex. In particular, mtDNA was important in gaining insight into the relationship between New World (NW) and Old World (OW) mallard populations (Johnson and Sorenson 1999; Kulikova et al. 2005). Whereas Eurasian mallard populations are all characterized as possessing OW A haplotypes, those in North America are paraphyletic with a substantial proportion carrying OW A and NW B mtDNA haplotypes. Hypotheses to explain the co-occurrence of OW A and NW B mtDNA haplotypes in North America

were (a) secondary contact and widespread bi-directional gene flow between invading Eurasian mallards (Johnson and Sorenson 1999; McCracken et al. 2001; Kulikova et al. 2004, 2005) and/or (b) incomplete lineage sorting from a dichromatic ancestor that invaded the New World and diverged into several monochromatic species (Omland 1997). Regardless, consensus from early work was that bi-directional gene flow with endemic NW mallard-like ducks (Fig. 1) was a big reason for why OW A and NW B mtDNA haplotypes were present in North America. In fact, there was a growing concern for the possibility of genetic extinction via widespread introgressive hybridization across monochromatic species from these results. Concern that endemic North American mallard-like ducks were likely hybrid swarms continued until recent advances in various genomics methods permitted for the access of the genome (i.e., thousands of nuclear loci), and making it possible to assign individuals to populations with confidence. Coupling population assignments with nuclear markers and mtDNA confirmed OW A mtDNA haplotypes in wild populations of mallard-like ducks were in fact the result of gene flow with domestic mallards. Note that all domestic mallards are known to carry OW A mtDNA haplotypes as the origins of domestication for mallards are in Eurasia (Kiple 2001; Huang et al. 2013). For example, comparing mitochondrial sequences to local domestic mallard populations, OW A mtDNA haplotypes recovered in Hawaii's Hawaiian duck populations (Fowler et al. 2009; Lavretsky et al. 2019a, b) and Florida's mottled ducks (*A. fulvigula*) (Bielefeld et al. 2016; Peters et al. 2016) was the result of gene flow with local populations of park mallards. Similarly, OW A mtDNA haplotypes found in North American wild mallards, American black ducks, and West Gulf Coast mottled ducks were the result of gene flow with domestic game-farm mallards, which have been released for sport hunting in North America since the 1920s (Lavretsky et al. 2019a, b, 2020). Thus, the true story was only revealed by applying landscape- and genomic-level sampling. While gene flow is the culprit for why OW A mtDNA haplotypes are now widespread in North American mallard-like ducks, it was not due to gene flow with wild mallards coming from Eurasia, but rather domestic ones that were released in the respective areas.

## *1.4 Hybridization Versus Gene Flow*

Avian lineages are especially prone to hybridization, even between species with relatively deep divergences (Grant and Grant 1997; Rheindt and Edwards 2011; Ottenburghs et al. 2015; Ottenburghs 2019). The high rates of hybridization in birds are attributable to their dispersal ability (Greenwood 1980), chromosomal stasis (Ellegren 2010), and relatively low levels of reinforcement (Grant and Grant 1997). The rapid evolutionary history, as well as the extent of gene flow among species within the Mallard Complex, makes this system ideal to study the interplay among various evolutionary mechanisms at the earliest stages of species divergence. Moreover, whereas many taxa within the Mallard Complex evolved in allopatry, the

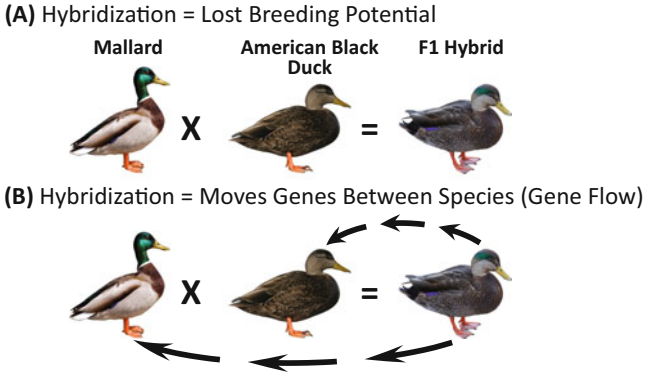


**Fig. 3** Schematic and biologically relevant potential outcomes of hybridization: (**a**) hybridization results in an F1 hybrid that does not breed into parental gene pools and, thus, is simply lost breeding potential. (**b**) Hybridization results in an F1 hybrid that does breed into parental gene pools (denoted by arrows), thus effectively moving genes between species

mallard has responded to anthropogenic influences (e.g., releases from game-farms and altered landscapes) and can now be found in sympatry with most of the other species. This secondary contact has resulted in widespread hybridization with the American black duck (Avise et al. 1990), Mexican duck (*A. diazi*; Hubbard 1977; Lavretsky et al. 2015a), mottled duck (McCracken et al. 2001; Williams et al. 2005a), Chinese spot-billed duck (*A. zonorhyncha*; Kulikova et al. 2004; Wang et al. 2019), New Zealand (NZ) grey duck (*A. superciliosa superciliosa*; Rhymer et al. 1994), Hawaiian duck (Griffin and Browne 1990; Lavretsky et al. 2015b), and yellow-billed duck (*A. undulata*; Stephens et al. 2019). Because all hybridization events involve mallards, this group provides natural replicates to understand how true gene flow impacts the genomes of endemic species and overall consequences on their adaptive and evolutionary trajectories.

From a conservation standpoint, determining the extent that hybridization events translate into true gene flow is critical (Fig. 3). I define hybridization as an event in which pure parental taxa interbreed and make a potentially viable $F_1$ hybrid. If that hybrid does not breed into either of the parental populations, then the biological outcome of said mating event is simply lost breeding potential for both parental taxa. Conversely, gene flow requires the $F_1$ hybrid to breed back into one or both parental populations, as to effectively move genes between the parental taxa (Fig. 3). While both events may be of concern for conservation biologists, the inability for the hybrid to breed and effectively move genes between parental taxa would be evidence of hybrid breakdown, assortative mating, or other potential post-zygotic isolating processes. Thus, determining the number and hybrid types ($F_1$, $F_2$-backcross, $F_3$-backcross, etc.) found on the landscape is critical when attempting to determine the extent and potential genetic polluting from intraspecific hybridization events. For example, if hybrids are relegated to the $F_1$ generation with none or few backcrosses, then one can conclude the ultimate consequence of hybridization is lost breeding

potential without gene flow. Conversely, if a variety of backcrosses are identified in a population, then the ultimate outcome of hybridization is gene flow. If this is the case, then determining the number of generational backcrossed hybrids, the geographical locations of these hybrids, as well as whether these backcrosses tend to breed with one of the parental populations are important next steps to best inform proper conservation action(s).

Often, breeding experiments are necessary to establish expected genetic assignments of various generational classes of hybrids; however, such experiments are often not possible with wild populations (Lavretsky et al. 2016). I note that breeding experiments are ideal to carefully understand the genetic admixture effect(s) on morphology and ecology of a species (Grabenstein and Taylor 2018). There are a variety of methods that allow to assign individuals to hybrid status (Nielsen et al. 2006; VÄHÄ and Primmer 2006; Corbett-Detig and Nielsen 2017; Wringe et al. 2017; Janzen et al. 2018). Recently, a method that permits the use of empirical molecular data collected from wild individuals was designed to simulate breeding experiments and based on the available dataset was created (Lavretsky et al. 2016). In short, a parental gene pool is established with samples that are genetically vetted as pure parental. A single random allele or SNP is chosen from each parental gene pool and across all available markers to simulate $F_1$ genotypes. Next, subsets of $F_1$ hybrids are then backcrossed into each parental gene pool for a determined number of times, and thus establishing gene pools for $F_2$ through $F_X$ backcrossed generations. Assignment probabilities are estimated in programs like STRUCTURE (Pritchard et al. 2000) or ADMIXTURE (Alexander et al. 2009; Alexander and Lange 2011) for a combined dataset of the simulated and empirical genotypes (Fig. 4). This can be done multiple times to establish expected average and range of assignment probabilities for each hybrid generation. For example, simulation outcomes for American black ducks and mallards resulted in six identifiable indices that included individuals with (a) $\geq$95% black duck assignment as pure black duck, (b) $\geq$98% mallard assignment as pure mallard, (c) 27–72% interspecific assignment as F1 hybrids, (d) 10–27% as F2-black duck backcrosses, (e) 2–27% black duck assignment as F2-mallard backcrosses, and (f) 5–10% mallard assignment as F3-black duck backcrosses (Fig. 4a). In the end, such analyses allow researchers to gain a more robust estimate of assignment probabilities of what a pure parental or hybrid is expected given the available molecular data.

Simulated "breeding" experiments help determine the number of hybrid classes and subsequently allow researchers to assign individuals to those classes. The proportion of individuals falling into each hybrid class provides an estimate of the relative rate of hybridization across the sampled landscape of the specie(s). For example, using simulations to genetically vet North American mallards and American black ducks demonstrated the limitations of US Fish and Wildlife Service's phenotypic key being used to assign pure and hybrid status to samples (Lavretsky et al. 2019b). Specifically, of those individuals phenotypically assigned as pure American black duck or pure mallard, 20% should have been identified as hybrids in each set. Similarly, only ~60% of all samples assigned as phenotypic American black duck x mallard hybrids were correct, with the remaining samples actually

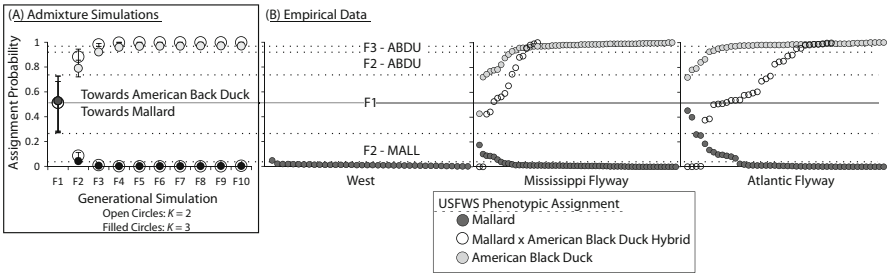


**Fig. 4** (**a**) The average and range of assignment probabilities from ADMIXTURE results at K = 2 and 3 across 25 simulated replications of hybridization (F1) and 9 generations of backcrossing (F2-F10) using genetically vetted American black ducks (ABDU) and mallards (MALL) – each K is based on 250 independent ADMIXTURE analyses. (**b**) Empirical data for western (WEST), Mississippi flyway (MISS), and Atlantic flyway (ATL) samples originally identified by USFWS as Mallards, American black ducks, or putative hybrids. Within geographical region, samples in all three phenotypic classes are aligned by interspecific assignment probability from high to low. Based on expected assignment probabilities as determined from simulations, I recategorized samples by assignment probabilities and found that ~80% of all phenotypically identified mallards and black ducks, as well as only ~60% of phenotypically identified hybrids, are correct. Figure adapted from Lavretsky et al. (2019b)

being either pure American black duck (~28% of samples) or pure mallard (12% of samples) (Fig. 4). Thus, these results clearly demonstrated the ineffectiveness in correctly identifying individuals and particularly hybrids with the current sweep of phenotypic traits for American black ducks and mallards. Finally, overlaying empirical and simulated data to determine the true number of each generational class provided the means to determine the rate of hybridization across North America. As expected, American black duck and mallard hybridization was highest in eastern North America and with evidence for a variety of hybrid classes present on the landscape. However, despite a century and a half of secondary contact between American black ducks and mallards resulting in some of the highest rates of hybridization (i.e., ~25%), Lavretsky et al. (2019b) concluded that American black ducks are not the hybrid swarm once hypothesized and that gene flow into American black duck was somehow limited (e.g., assortative mating). Once again, the ability to genetically identify samples, and even between taxa that are very closely related like American black ducks and mallards, can illuminate inconsistencies in current methods and datasets, and even previous notions (e.g., American black ducks are a hybrid swarm) that are used to guide and make important management decisions. Moreover, these results demonstrate how advances in genomic methods provide the capacity to genetically establish phenotypic traits that are truly informative (e.g., see the genetically vetted key created for Florida mottled ducks; Bielefeld et al. 2016). These and similar methods offer a powerful approach for examining concerns of hybridization in conservation efforts and without the requirement of captive breeding.

The depletion of native populations makes it susceptible to genetic swamping from even small numbers of introduced species (Childs et al. 1996; Rhymer 2006; Russo et al. 2018). Although hybridization is prevalent in birds, and ducks especially (Cade 1983; Rhymer 2006), species extinction due to complete genetic swamping, while concerning (Rhymer and Simberloff 1996; Buerkle et al. 2003), has been identified in few systems (Rhymer and Simberloff 1996; Salzburger et al. 2002; Wells et al. 2019; Lavretsky et al. 2020). Examples of true hybrid swarms may be rare because in general, the backcrossing of hybrids back into large parental populations may prevent the persistence of large numbers of admixed individuals (e.g., Lavretsky et al. 2016). For example, between American black ducks and mallards where Lavretsky et al. (2019b) found that clear outlier regions between the parental species decreased in genetic differentiation when comparing the genomic landscape of several generations of backcrossed individuals (Fig. 5). Thus, having a parental gene pool to which hybrids can continuously backcross into is not only important for conservation but may be an important mechanism that

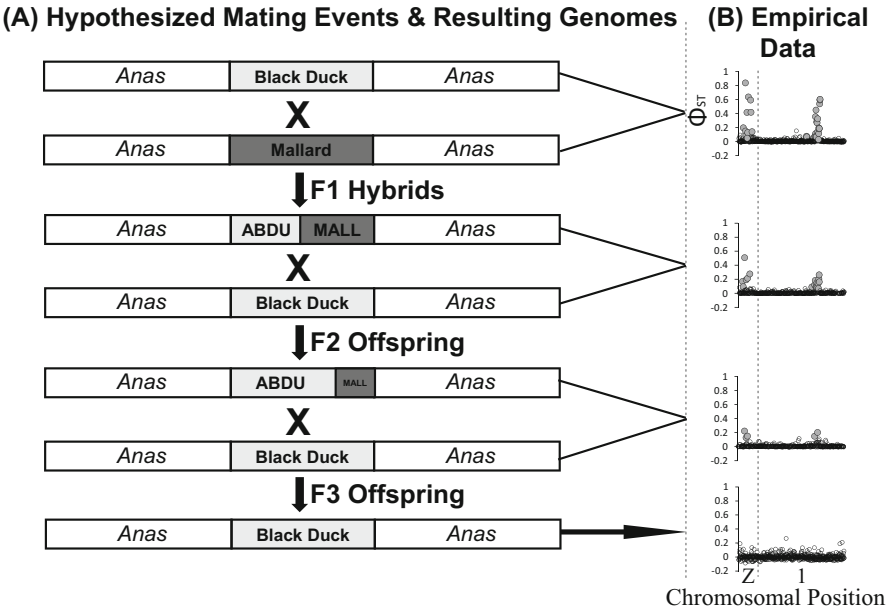


**Fig. 5** (**a**) Hypothesized mating between mallards (MALL) and American black ducks (ABDU) and subsequent backcrosses into American black ducks with expected genomic contributions of offspring. (**b**) Empirical ΦST estimates across markers along the Z-chromosome and chromosome 1 between genetically vetted samples that are consistent with expected offspring from each respective hypothesized mating event. Genomes depicted to be largely ancestrally shared (denoted as genus *Anas*), with species specificity identified at a few highly selected regions. Note the clear decrease in outlier regions with each subsequent backcross that correspond with proportional replacement of mallard alleles with American Black ducks until F3 (or second generation backcrosses), which are genetically identical to a pure American black ducks. Figure adapted from Lavretsky et al. (2019b)

decreases the potential negative effects of gene flow, in general. As habitat continues to be depleted, domestic forms increasing on the landscape, as well as overall global change is bring many closely related taxa together, and thus understanding the geographical extent of hybridization and whether these events result in true gene flow is essential. Genomics methods that readily provide sufficient marker coverage have made even the most difficult or complex relationships possible to tease apart (Wayne and Shaffer 2016; Catchen et al. 2017; McKinney et al. 2017).

## 2 How Population Genomics Has Increased Our Understanding of the Mallard Complex and Its Implications for Conservation

### *2.1 The History of the New World Mallard Complex: How the Phenotype Lies and What Genetics Has Revealed*

Seven mallard-like ducks make North America home. Of these, the mallard, American black duck, Mexican duck, and two subspecies of mottled ducks are found on mainland North America, while the Laysan and Hawaiian ducks are found on islands making up the Hawaiian Archipelago (Fig. 1). All but the mallard are endemic to either the Hawaiian Islands or mainland North America (Baldassarre 2014). Concern over genetic swamping, the resulting hybrid swarm, and eventual genetic extinction for all the endemic mallard-like ducks has spurred over four decades of research into understanding rates of hybridization between mallards and each of the other North American taxa (Rhymer and Simberloff 1996; Rhymer 2006). Cause for concern was due to the high prevalence of what appeared to be individuals carrying phenotypic traits of the mallard (e.g., green iridescence in head, top and/or bottom secondary white wing bars, curl feathers in the rump) and of the respective monochromatic taxa in populations of Mexican ducks (Hubbard 1977), American black ducks (Brodsky and Weatherhead 1984; Ankney et al. 1987; Avise et al. 1990; Kirby et al. 2000), mottled ducks (Bielefeld et al. 2010, 2016), and Hawaiian ducks (Griffin and Browne 1990; Livezey 1993; Engilis et al. 2002a). The presentation of these mallard traits in significant proportions heightened concern over the possible genetic extinction of these endemic monochromatic ducks. Any advances made in molecular methods since the 1980s have been applied towards attempting to determine rates of hybridization and gene flow. However, early attempts with allozymes (Browne et al. 1993), microsatellites (Williams et al. 2002, 2005a, b; Mank et al. 2004; Fowler et al. 2009), and Sanger sequencing of single and multiple loci (Avise et al. 1990; Johnson and Sorenson 1999; McCracken et al. 2001; Kulikova et al. 2004, 2005; Lavretsky et al. 2014a) resulted in largely inconclusive findings (see Fig. 3a as example). In each case, the authors determined that too much of the genetic variation was shared among the taxa to be able to confidently identify hybrids, let alone assign samples to their respective taxon. Paraphyly and intermixed mtDNA haplotypes further

suggested to researchers that these species likely were represented by highly admixed individuals (Avise et al. 1990; McCracken et al. 2001; Lavretsky et al. 2014a, b; Peters et al. 2014). In general, additive effects from ancestry (i.e., incomplete lineage sorting) and extensive hybridization were used to explain the extent of shared molecular diversity in this group of birds.

The lack of more definitive molecular work had important implication onto taxonomic revisions and conservation efforts for many of the taxa within the Mallard Complex. Until recently, many taxonomic and conservation decisions remained informed through largely phenotypic work and the basic premise regarding the general lack of non-paraphyletic genetic markers as putatively the result of widespread admixture. For example, the taxonomy of Mexican duck has been defined by phenotypic work done in the 1970s, which presumed that the clinal-like presence of mallard-like traits in Mexican ducks was the result of extensive introgressive hybridization (Hubbard 1977). As a result of these phenotypic-based conclusions and a general lack of definitive molecular results, the Mexican duck has been relegated to subspecies status. Similarly, conservation efforts for the endangered Hawaiian duck largely surrounded the need to remove mallard-like traits from then captive Hawaiian duck population prior to release. However, no matter the effort to breed Hawaiian duck looking individuals together, a proportion of juvenile males always displayed mallard-like phenotypic characters (Engilis and Pratt 1993; Engilis et al. 2002a). Similarly, breeding experiments attempting to "breed out" mallard characters by specifically mating individuals that were especially American black duck-looking continuously resulted in broods with a proportion of males still displaying mallard-like characters (Kirby et al. 2004). Once again, understanding whether mallard-like characters displayed in many of these monochromatic species was due to the fact that these represented hybrid swarms or simply a case of ancestry remained unknown until advances in genomic methods permitted researchers to genetically identify between pure and hybrid individuals.

The population genomics approaches required to properly answer questions about hybridization in the Mallard Complex were optimized in the mid-2010s, and specifically, it was advancements made in the reduced genomic representation methods (e.g., RADs, ddRAD-seq, SeqCap) that opened the door to accessing sufficient sized genomic datasets. First in 2015, Lavretsky et al. (2015a) applied a ddRAD-seq method to sample 3,695 polymorphic loci – 3,523 loci (316,175 base pairs (bp) assigned to autosomes and 172 loci (15,869 bp) assigned to the Z-sex chromosome) – across 105 Mexican ducks from six Mexican states (N = 92 individuals) and two US states (N = 13 individuals), as well as 17 mallards sampled across North America. The authors demonstrated that Mexican ducks and mallards were genetically distinguishable and identified no samples in Mexico and only a handful of samples from the USA as genetic hybrids. The latter finding was in stark contrast to the notion that Mexican ducks were largely a Mexican duck x mallard hybrid swarm as suggested with phenotypic data (Hubbard 1977). In fact, the landscape perspective that Lavretsky et al. (2015a) achieved showed that Mexican ducks biogeography naturally followed an isolation-by-distance pattern and was the

result of sequential founder events from north to south (Fig. 6b-3). Moreover, Lavretsky et al. (2015a) identified specific genetic regions on several autosomal and Z-sex chromosome that were under divergent selection in mallards or Mexican ducks. More recent work further demonstrated that Mexican ducks retained mallard diversity due to ancestry (i.e., ILS) and not extensive gene flow, as well as provided more definitive demarcation of several genetic markers on autosomal chromosomes 2 and 14 that were best explained by divergent selection in Mexican ducks specifically (Figs. 7 and 8a; Lavretsky et al. 2019a). Finally, ongoing work linking phenotype with genetics has surprisingly revealed that Mexican ducks display mallard-like traits and shared mallard genetic diversity that follow a cline from north to south but that this is explained by retained ancestry and not gene flow from mallards (Brown et al., unpublished data). Specifically, having genetically vetted samples, the researchers determine that juvenile or hatch-year genetically pure male Mexican ducks displayed mallard-like characters and that the proportion of hatch-year males that displayed such characters decreased southward. Thus, while Hubbard (1977) correctly characterized the clinal variation in regard to mallard-like plumage displayed by Mexican ducks across their range, these patterns were not indicative of a hybrid swarm but rather due to retained ancestry. Specifically, the evolution of the Mexican duck was likely the result of a mallard population that isolated and adapted to the Chihuahuan Desert 200,000–500,000 years before present and expanded southward through sequential founder events losing dichromatism in the process (Fig. 6b-3). In the end, coupling genomics methods with landscape-level sampling proved to resolve the evolutionary history of the Mexican duck, including establishing that Mexican ducks showed some of the lowest rates of hybridization within the Mallard Complex. Moreover, Mexican ducks harbor genomic regions under divergent selection and which are at species-level differences that would suggest that taxonomic revisions for this duck may be warranted.

Applying similar methods to study the evolution and population structure of mottled ducks (Peters et al. 2016; Ford et al. 2017) and American black ducks (Lavretsky et al. 2019b, 2020) proved to be once again fruitful, with researchers being able to establish that none of these monochromatic taxa are already or on their way to becoming a hybrid swarm. Instead, data determined that these ducks are all very closely related, with much of the genome shared due to ancestry from and not gene flow with the mallard (Fig. 6b; Lavretsky et al. 2019a). More specific analyses also revealed that while mottled ducks from Florida and the West Gulf Coast diverged in allopatry from each other and the mallard (Peters et al. 2016), the American black duck and mallard likely diverged under punctuated events of secondary contact (Lavretsky et al. 2020) (Fig. 6b). Importantly, attaining sufficient marker coverage and across hundreds of samples was required to finally determine rates of hybridization and potential effects of gene flow. In short, the rates of hybridization with mallards were highest for black ducks (i.e., ~25%; Lavretsky et al. 2019b), followed by mottled ducks (i.e., 5–8%; Peters et al. 2016; Ford et al. 2017), and Mexican ducks (i.e., 2–5%; Lavretsky et al. 2015a); however, gene flow into each of the monochromatic species remained low (Lavretsky et al. 2019a), suggesting that post-zygotic isolating mechanisms likely evolved within this recent

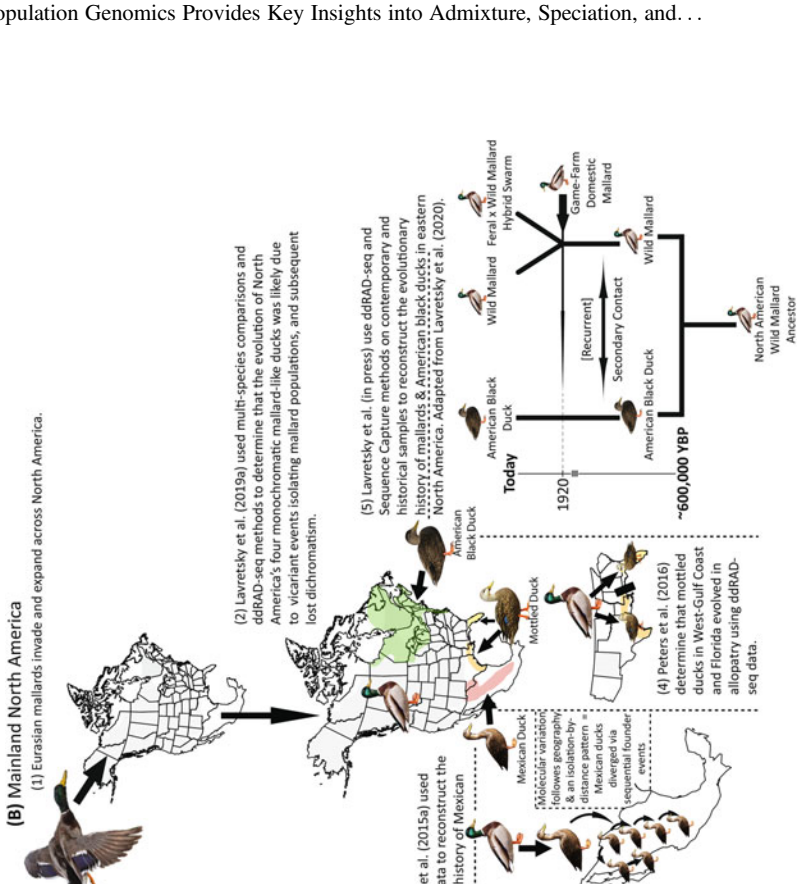


**Fig. 6** The evolutionary and contemporary histories of mallard-like ducks from North America as determined from population genomics

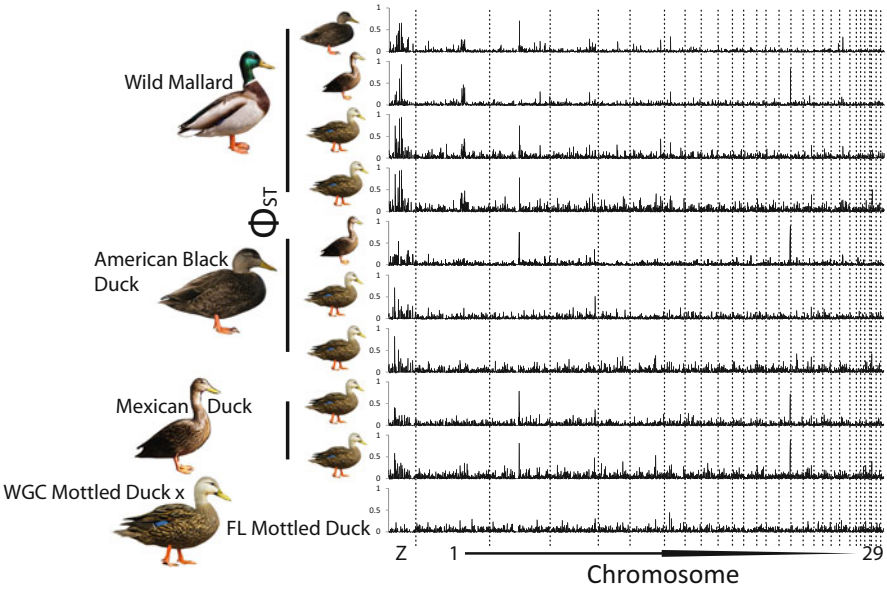


**Fig. 7** Pairwise estimates of relative differentiation ($\Phi_{ST}$) across New World mallard-like ducks for 3,017 and 177 ddRAD-seq loci that assigned to autosomes or the Z-sex chromosome, respectively. Note no ddRAD-seq loci were recovered on chromosome 17. Coupling species pair analyses and partial genome sequencing provides a mean to explore regions under divergent selection in specific species. Here, a ~21 Mbp region (positions 1.7E7–3.8E7 bp) on the Z-sex chromosome and a ~11 Mbp region (1.0E8–1.2E8 bp) on chromosome 1 were identified to be under divergent selection in mallards. An outlier locus on chromosome 14 (position ~1.6E7) was detected in all four comparisons involving Mexican ducks, suggesting directional selection at this or a linked locus in Mexican ducks only. Additional outliers were detected in pairwise comparisons involving other species. Finally, genomic patterns as observed between the two mottled ducks are consistent with instance of strict allopatric speciation. Figure adapted from Lavretsky et al. (2019a)

radiation. In fact, pairwise species comparisons across ~3,194 ddRAD-seq loci demarcated several outlier regions on the Z-sex and other autosomal chromosomes harboring genes under divergent selection in one or more of the taxa (Fig. 7; Lavretsky et al. 2019a). Among these sites, a ~21 Mbp region on the Z-sex chromosome was recovered to harbor genes under divergent selection in mallards that may be playing an important role in the evolution of dichromatism in this group (Figs. 7 and 8a). These results are consistent with a growing body of evidence suggesting that sex chromosomes are often involved in early stages of species divergence, including harboring genes linked to phenotypic variation in other taxa (Minvielle et al. 2000; Sæther et al. 2007; Phadnis and Orr 2009; Pryke 2010; Ellegren et al. 2012; Martin et al. 2013; Sutter et al. 2013; Ruegg et al. 2014; Lavretsky et al. 2015a).

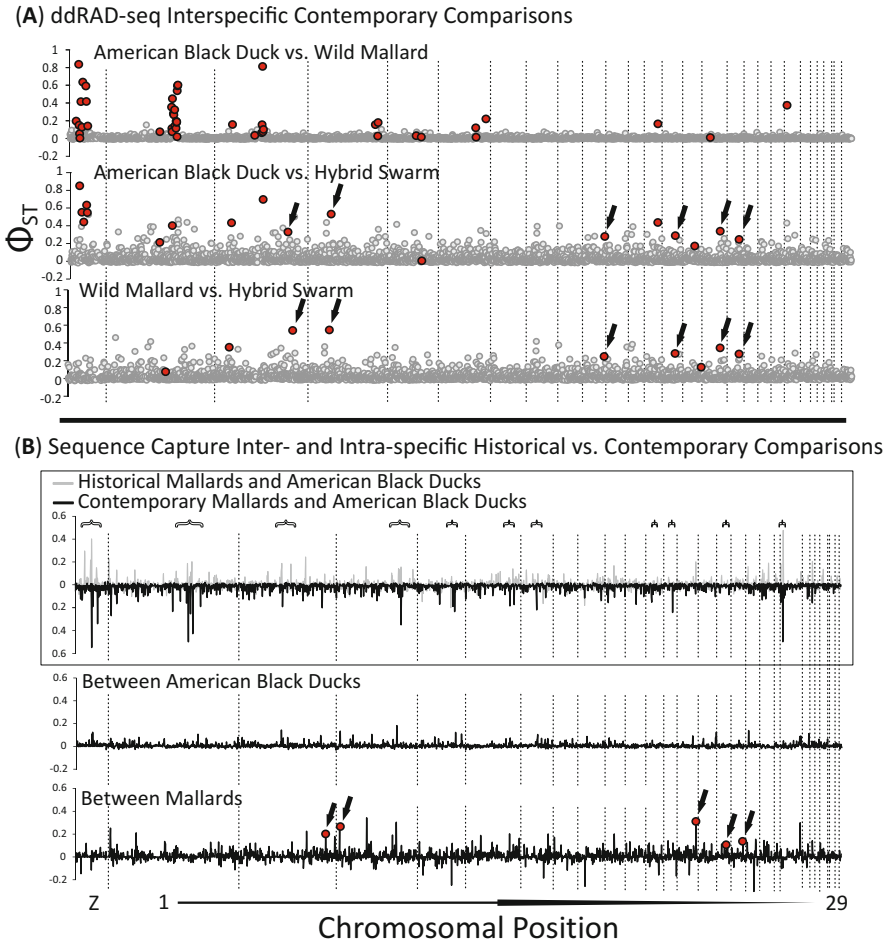


**Fig. 8** Chromosomally aligned $\Phi_{ST}$ estimates of (**a**) 3,037 autosomal and 163 Z-linked ddRAD-seq loci for pairwise comparisons between genetically vetted wild mallards, wild American Black Ducks, and feral x wild mallard hybrids (i.e., hybrid swarm). Markers identified in BayeScan analyses as putatively under divergent selection are denoted in red, arrows denoting loci potentially under artificial selection. Adapted from Lavretsky et al. (2019b). (**b**) Sequence captured 2,103 autosomal and 99 Z-linked loci for pairwise comparisons between historical and contemporary mallards and American black ducks. Brackets denote the same loci demarcated as outliers between American black ducks and wild mallards, and red circles + arrows denote loci found as under artificial selection in (**a**), which was adapted from Lavretsky et al. (2020). Note no markers were recovered for chromosome 17

## 2.2 *The Curious Case of the Hawaiian Duck: Conservation Implications When a Hybrid Species Meets Its Feral Parent*

With Laysan ducks now relegated to the Laysan and Midway Atoll Islands, the endangered Hawaiian duck is the only remaining endemic duck on the main Hawaiian Islands (Engilis et al. 2004; Pyle and Pyle 2017). Recent molecular work provided strong evidence that Hawaiian ducks represent a homoploid hybrid species (Fig. 6a). Specifically, Lavretsky et al. (2015b) determined that the Hawaiian duck's evolutionary history was the result of an ancestral hybridization event between vagrant mallards and once prevalent Laysan ducks and dated the admixture event to the Pleistocene-Holocene boundary; the authors analyzed molecular variation across 19 nuclear introns and used coalescent analyses to estimate the ancestral gene flow event to ~3,000 years before present (95% HPD = 0–207,000 years before present). More recent work by Wells et al. (2019) using thousands of ddRAD-seq loci also found evidence that pure Hawaiian ducks shared near 50:50 coancestry with Laysan ducks and mallards and the only one to show such patterns of all mallard-like ducks (Lavretsky et al. 2014b, 2015b). Together, Hawaiian ducks are genetically (Lavretsky et al. 2014b, 2015b), phenotypically, and ecologically (Engilis et al. 2002b; Uyehara et al. 2008) distinct from all other mallard-like ducks and thus satisfy all primary criteria used in avian taxonomy for species designations (Gill 2014; Sangster 2014) and as a result likely represent a young hybrid species. In fact, it was likely the combination of mallard and Laysan duck molecular variation, including predator aversion that would only be innate in the mallard that permitted the Hawaiian duck to endure past Polynesian settlement where other Island life (e.g., Laysan ducks) could not.

The Hawaiian duck remained largely in allopatry from both of its wild parental taxa, until domestic mallards were first imported to the Hawaiian Islands for food and hunting beginning in the 1800s (Engilis et al. 2004; Pyle and Pyle 2017). Later, mallards were commercially farmed on O'ahu during the 1930s and 1940s, and multiple feral populations became established on Kaua'i, O'ahu, Maui, and Hawai'i (Engilis and Pratt 1993). As a result of these actions, genetic extinction through ongoing hybridization with feral mallards has been primary concern for their conservation (USFWS 2012). Historically, Hawaiian ducks occurred on the main Hawaiian Islands of Kaua'i, Ni'ihau, O'ahu, Maui, Moloka'i, and Hawai'i but were extirpated from all islands except Kaua'i and Ni'ihau by the 1960s (Engilis et al. 2002a). Hawaiian ducks were captive-reared and reintroduced onto O'ahu and Hawai'i until the late 1980s and onto Maui in 1989. However, feral populations of domestic mallards were not dealt with prior to attempted reintroductions and may have been the reason that biannual waterbird surveys suggested an increasing number of Hawaiian duck-mallard hybrids through time (USFWS 2012). In fact, earlier molecular work confirmed that hybridization between Hawaiian ducks and local mallards was occurring on O'ahu (Browne et al. 1993; Fowler et al. 2009). However, despite early molecular efforts to understand hybridization, much of the

hybrid identification done by USFWS and state agency personnel was based on phenotype. As with the case for other New World mallard-like ducks (see above), we now know that Hawaiian ducks, and primarily first year males, naturally display mallard-like characters as by-product of their recent mallard ancestry (Lavretsky et al. 2015b). Thus, whether the increasing number of hybrids across Hawaiian Islands was real or simply due to overestimation resulting from nondiagnostic phenotypic traits remained unknown until recently.

Wells et al. (2019) set to determine the extent of true hybridization rates between Hawaiian ducks and mallards across Hawaiian Islands, potential mallard source (i.e., domestic vs. wild), whether hybridization rates have increased through time, and whether the presence/absence of mallard-like traits in the phenotype of an individual can be confidently applied to identify hybrids. Sampling included assaying 3,114 autosomal and 194 Z-linked ddRAD-seq loci across 425 Hawaiian ducks obtained across 5 Hawaiian Islands, as well as 30 samples of each Laysan ducks and wild North American mallards. Sampling effort of Hawaiian ducks was nearly a decade apart, permitting the researchers to test for changing hybridization rates across time. First, the authors confirmed that pure Hawaiian ducks persist on Kaua'i in large numbers and with relatively little evidence of mallard-Hawaiian duck hybridization during the past decade (Fig. 9). This finding was incredibly informative for conservation biologists attempting to determine whether pure Hawaiian ducks even existed (USFWS 2012). Unfortunately, Wells et al. (2019) reported that all reintroduced populations on Hawai'i, Maui, and O'ahu constituted hybrid swarms – i.e., not a single sample among reintroduced populations was identified as a pure Hawaiian duck (i.e., $\geq$95% Kaua'i Hawaiian duck ancestry; Fig. 9). The authors were able to determine that the extensive hybridization was primarily with non-wild, local feral mallards. Furthermore, the authors were able to genotype ducks collected from 1998 to 2015. First, a decreasing trend in the number of hybrids on Kaua'i where removal efforts of mallards and potential hybrids were underway was found. Conversely, all sampled reintroduced population showed no change in overall admixture proportions across samples or in the overall proportion of admixed individuals. Thus, the authors could conclude that all sampled sites across Hawaiian Islands in which reintroductions were attempted eventually failed due to extensive hybridization with these feral mallards and untimely became true hybrid swarms as early as 1998. Finally, further molecular assessment of Hawaiian ducks that were culled due to the presence of mallard-like traits on Kaua'i revealed only a handful of these to be true hybrids and the remaining as hatch-year males. Once again, mallard-like traits displayed by Hawaiian ducks were found to be due to shared mallard ancestry and not contemporary hybridization as with the other monochromatic mallard-like taxa on mainland North America (see above).

Island populations and those that have recently declined are more susceptible to genetic swamping by an introduced species (Childs et al. 1996; Rhymer 2006; Grabenstein and Taylor 2018). The absence of large native populations of Hawaiian duck on O'ahu, Maui, Moloka'i, and Hawai'i likely precipitated the formation of hybrid swarms on these islands. Hawaiian duck reintroductions involved relatively few individuals, and captive-reared Hawaiian ducks were introduced on islands with

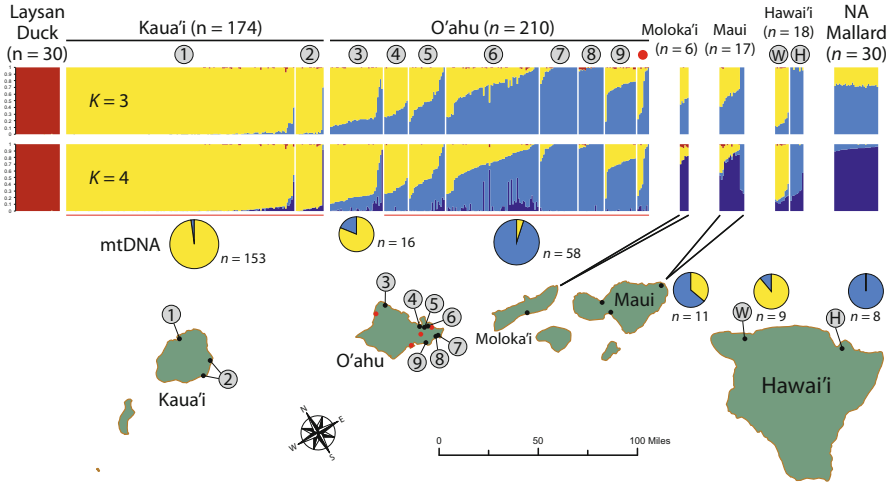


**Fig. 9** Location for 425 Hawaiian ducks and putative hybrids sampled across Hawaiian Islands. Individual ancestry proportions as estimated in the program STRUCTURE using 3,114 ddRAD-seq autosomal loci assuming both K = 3 and K = 4 populations. Included in analyses are Laysan ducks and North American mallards. Pie charts represent the proportion of samples with the putative New World "B" mtDNA lineage (yellow) and the putative Old World "A" lineage (blue) for a subset of ducks on each island. Note that all reintroduced populations are now genetically identified as hybrids (i.e., hybrid swarm; Fig. 6a). Additionally, most reintroduced populations are largely identified carrying Old World A mtDNA haplotypes that are associated with mallards from Eurasia. Figure adapted from Wells et al. (2019)

established populations of feral mallards (Engilis et al. 2002a). Together, the evidence presented by Wells et al. (2019) clearly demonstrated that all reintroduction efforts were set up to fail by not dealing with the feral mallard problem prior to attempting Hawaiian duck reintroductions. Given the lack of large, pure populations of Hawaiian ducks into which hybrids can backcross, hybrid swarms will likely persist on these islands. Thus, in the absence of additional genetic contributions from Kaua'i Hawaiian ducks, time alone is unlikely to decrease hybrid individuals comprising these hybrid swarms. Thus, these molecular results establish that future conservation efforts to reestablish pure Hawaiian ducks outside Kaua'i will need to remove the identified hybrid swarms, as well as feral mallards. Hawaiian duck reintroduction efforts represent an example of what happens to a small, isolated founder population confronted with a large population of a nonreproductively isolated congener. These results demonstrate the potential amalgamating effects of gene flow during secondary contact, spatial variation in the extent and consequences of hybridization, and the importance of considering such effects during conservation planning.

## 2.3 The Genetic and Conservation Consequences of Feral Mallards

Along with changing habitat, the direct release of domesticated individuals into the wild is a practice used worldwide to augment various wildlife populations (Laikre and Ryman 1996; Lichatowich and Lichatowich 2001; Waples and Drake 2004; Champagnon et al. 2012). Where these releases are common and intensive, concern for wild populations has increased because breeding with released or feral conspecifics can cause a loss of genetic variation leading to a loss of adaptation and overall fitness within wild population (Frankham 2005; Araki et al. 2009; Crispo et al. 2011; Grabenstein and Taylor 2018). In general, movement of these artificially selected, maladaptive traits into wild populations can reduce the adaptability and capacity of that population to survive in the wild (Araki et al. 2007; Evans and Evans 2007; Haccou et al. 2013; Corbi et al. 2018). In fact, modeling effects of gene flow between domestic and wild congeners, Tufto (2017) showed a slow, additive effect of increasing maladaptation in wild populations through continued interaction with their respective domestic counterpart and with negative outcomes taking time to be observable. As anthropogenic gene flow impacts more species, it is increasingly critical to monitor geographic regions where wild and domestic [feral] populations interact to assess for any possible genetic effects. Doing so ensures the adaptive qualities, and subsequent continued conservation of wild populations occurs.

Humans and mallards have been closely linked since their domestication in central China shortly after 500 BC (Kiple 2001; Huang et al. 2013). Around the world, domestic mallard stocking has been extensively practiced throughout history and most intense at the turn of the twentieth century (Heusmann 1991; Champagnon et al. 2012, 2013). While naturally found across the Holarctic, the intentional or accidental release of mallards has increased their range to include the entire world outside the Poles (Baldassarre 2014). Feral mallards now pose a genetic threat to global populations of wild mallard and mallard-like taxa. In all cases of mallard introductions, >25,000 domestic mallards were intentionally introduced, with the most extreme cases as within North America and Eurasia; these releases are now in the tens-of-millions (Heusmann 1974; Braithwaite and Miller 1975; Brooke and Siegfried 1991; Heusmann 1991; Tamisier 1992; Engilis and Pratt 1993; Dean 2000; Engilis et al. 2004; Guay and Tracey 2009; Bielefeld et al. 2010; Dyer and Williams 2010; Čížková et al. 2012; Pyle and Pyle 2017). For the Mallard Complex, secondary contact has always been considered to be occurring between wild populations and specifically with wild mallards. However, applying thousands of molecular markers and landscape-level sampling efforts has been transformative in correcting this early dogma, and more definitively determining that introgressive hybridization has not been with wild but rather domestic strains of the mallard. First, we learned that Kaua'i is home to the last remaining pure population of the endangered Hawaiian duck (Wells et al. 2019) and that extensive introgressive hybridization with local, feral mallards has resulted in the formation of hybrid swarms across the Islands of Hawai'i, Maui, Moloka'i, and O'ahu (Figs. 6a and 9). Similarly alarming

are the rates of widespread introgression between domestic and wild mallards in Eurasia and North America, where stocking practices still annually augment wild populations with nearly six million (Rueness et al. 2017) and two-hundred thousand (USFWS 2013) game-farm mallards, respectively, for the last 100 years. The substantial annual influx of these domesticated forms has significantly changed the genetic composition of Eurasian (Champagnon et al. 2010; Söderquist et al. 2014; Söderquist et al. 2017) and North American (Lavretsky et al. 2019b, 2020) mallards. Similarly, there is strong evidence of widespread introgression between domestic mallards and Pacific black ducks (*Anas superciliosa*) in Australia (Guay and Tracey 2009) and New Zealand (Hitchmough et al. 1990; Rhymer et al. 1994; Williams 2017). In fact, stocking practices of game-farm mallards in New Zealand have resulted in a current population of five million feral mallard birds (Williams 1981; Guay et al. 2015). A couple of recent studies assaying several molecular markers across range-wide sampled yellow-billed ducks determined that while wild mallards do not pose a genetic threat (Brown et al. 2019), yellow-billed ducks are now genetically threatened by domestic mallards that have been released and with feral populations recently establishing across Africa (Stephens et al. 2019). In general, domestic mallards differ in fertility, overall morphology, and biology from their wild counterparts (Desforges and Wood-Gush 1975a, b; Miller 1977; Paulke and Haase 1978; Cheng et al. 1979, 1982; Söderquist et al. 2013), with traits optimized for domestic settings. Understanding how the movement of their genetics and associated maladaptive traits may be decreasing the adaptability of wild populations will continue to grow in importance and particularly when devising conservation plans.

## 2.4 *Attaining a Historical Perspective to Reconstruct Evolutionary Histories*

Significant advances in ancient DNA (aDNA) extraction and sequence capture techniques have made it possible to isolate thousands of genetic markers to full genomes from historical and ancient samples, opening the possibility to attain a genetic perspective over large time scales (Grover et al. 2012; Mitchell et al. 2014; Orlando et al. 2015; Leonardi et al. 2017). These methods have been instrumental in understanding evolutionary histories of variety of organisms, including humans (Callaway 2016; Kuhlwilm et al. 2016), but are also now being used to understand the genetic turnover of populations and species (Cooper et al. 1996; Loreille et al. 2001; Leonard et al. 2002; Willerslev and Cooper 2005; Grealy et al. 2017; Leonardi et al. 2017; Rawlence et al. 2017; Lindqvist and Rajora 2019; Pont et al. 2019; Fenderson et al. 2020). In fact, the ability to not only determine the genetic diversity that was lost in a species but attaining full genomes for extinct species may make the idea of de-extinction possible (Shapiro 2017). For conservation efforts, rather than attempting to breed diversity back into an endangered species, we may be able to one day simply add the lost genetic diversity back into the species. Regardless,

increasingly efficient aDNA methods have made museum specimens even more important not only to understand evolutionary histories but also understand how populations have genetically changed through time. For example, by understanding how populations responded to changing habitat in the past can undoubtedly help refine conservation plans for those populations in the future (Fenderson et al. 2020).

The first use of aDNA methods in the Mallard Complex was by Cooper et al. (1996), who were able to isolate a small piece of mitochondrial DNA from some *Anas subfossils* on Hawai'i. Doing so, Cooper and colleagues determined that these subfossils were indeed Laysan ducks, providing the first support that Laysan ducks were once widespread across the Hawaiian Islands. Next, Mank et al. (2004) assayed three microsatellite markers in historical mallards and American blacks and reported an 18-fold reduction in differentiation (Gst) between the two sampled in 1998 (0.008) versus 1940 (0.146). The authors concluded that a century of hybridization must have led to a loss of genetic distinctiveness. The numbers of usable samples were low with these early methods as they required the creation of primers as to amplify targeted DNA using PCR. Often, however, ancient and historical samples are highly degraded, posing limitations for PCR-based methods (Keyser-Tracqui and Ludes 2005). More recently, coupling sequence capture methods with high-throughput sequencing has made the isolation and sequencing of aDNA more accessible and reliable across any specimens with endogenous DNA (Briggs and Heyn 2012; Knapp et al. 2012; Schubert et al. 2012; Lindqvist and Rajora 2019). For example, Lavretsky et al. (2020) used these recently developed aDNA methods on American black ducks and mallards from 1860 to 1915 to revisit the hypothesis that American black ducks are closely related to mallards due to widespread hybridization as suggested by Mank et al. (2004). To do so, a bait capture array was first designed from 3,446 nuclear loci initially isolated from contemporary samples using ddRAD-seq methods (Lavretsky et al. 2019b). Across the 69 historical samples and another 39 contemporary samples, a total of 2,202 markers (140,477 base pairs (bp) across the Z-sex (99 markers; 6,122 bp) and 28 autosomal (2,103 markers; 134,355 bp) chromosomes) were isolated (Fig. 8b), resulting in a recovery rate of 64%; similar recovery rates using RAD-based bait arrays were reported in early work (Souza et al. 2017). Additionally, the authors were able to off target sequence 641 base pairs of the mtDNA control region across samples. Mitochondrial DNA is often obtained as bycatch in sequence capture datasets (Griffin et al. 2014; Gasc et al. 2016) due to its stability and abundance in samples (Picardi and Pesole 2012). In contrast to the results presented by Mank et al. (2004), Lavretsky et al. (2020) reported an overall increase in divergence, including the maintenance of all known outlier positions across the genomes of these two ducks (Fig. 8b). Moreover, there was no significant change in the sampled genome of historical and contemporary American black ducks (Fig. 8b), providing additional evidence that genetically pure American black ducks today are the same as those from 150 years ago (Fig. 6b-5). These results are clearly contradicting the notion that today's American black ducks are simply a hybrid swarm. The authors suggest that earlier work by Mank et al. (2004) suffered from the total number of markers analyzed (i.e., three microsatellites). Given the evident genomic heterogeneity across their captured

markers (Fig. 8b), using a few markers as done by Mank et al. (2004) is unlikely to provide a complete picture and demonstrates the importance in maximizing genetic data. Finally, in addition to important information for the conservation of the American black ducks, the sequencing of historical mallards and comparing to contemporary populations provided further evidence that today's eastern mallards are indeed genetically different from both historical and contemporary western mallards. In the context of conservation, these results demonstrated opportunities that aDNA methods coupled with museum specimens towards understanding particular evolutionary histories, as well as determining lost genetic variation of specific populations or species.

Finally, ddRAD and related methods (e.g., RAD, GBS, etc.; Andrews et al. 2016) are inherently biased by the possibility of allelic dropout due to mutations in enzymatic cut-sites (Graham et al. 2015; Lowry et al. 2017; Catchen et al. 2017). These biases are not present in sequence capture datasets as they do not require enzymatic cut-sites to be present or intact to work. Lavretsky et al. (2020) were able to provide highly similar results between ddRAD and sequence capture datasets across a variety of estimates and analyses (e.g., Fig. 8), demonstrating that known biases with restriction enzyme-based techniques (e.g., allelic dropout; Graham et al. 2015; Lowry et al. 2017; Catchen et al. 2017) may have little or no effect on population-level statistics for species with very shallow divergence, such as between mallards and black ducks.

## 3 Integrating Population Genomics Results into Wildlife Management

Advances in molecular methods and high-throughput sequencing technology will continue to advance the field of population genomics, making the use of these methods possible for any organism. Though there is no doubt that lowering costs associated with full genomes will one day make it possible to be applied towards population genomics, partial genome sequencing methods like ddRAD-seq and related methods (e.g., RAD, GBS, etc.; Andrews et al. 2016) provide a perfect balance between data and cost. Today, a ddRAD-seq library can be attained for as little as $25–40 per sample and an Illumina HiSeq X capable of sequencing up to ~200 samples on a single lane. Importantly, methods like ddRAD-seq require no previous genetic information on the organism and are more forgiving when dealing with degraded DNA as compared to other genomic methods (Graham et al. 2015). Thus, the lowering cost and universal applicability of these partial genome sequencing methods make them ideal to study the population genetics of any wild population (Oyler-McCance et al. 2016). Moreover, although still proportionally a small amount of the genome (i.e., ddRAD datasets often represent $<0.03\%$ of the genome; Lavretsky et al. 2015a, 2019a), the thousands of markers remain a powerful means to

screen for loci putatively under selection (e.g., Figs. 7 and 8; Andrews et al. 2016; Catchen et al. 2017).

Proper wildlife conservation requires an understanding of the population in question (Oyler-McCance et al. 2016; Peters et al. 2016; Allendorf 2017; Ralls et al. 2018). The fields of population and conservation genetics have illuminated the potential pitfalls when management decisions are made without truly knowing the genetic constitution of the population or species being managed. Understanding fine-scale population structure and hybridization rates requires a population genomics approach in which datasets are represented by samples spanning the taxon's geographic range, and a maximum number of loci. Informative and decisive research into the evolution and population structure of the Mallard Complex, which harbors some of the most complex scenarios as often the case for recent and/or rapid radiations, was not possible until a landscape- and genomic-level sampling scheme was achieved (Lavretsky et al. 2019a). For example, it was not until a genomic perspective that provided sufficient marker coverage to genetically identify individuals to species with confidence that the issue of feral mallards, and not wild mallards, to the conservation of many of these mallard-like ducks was realized. Wildlife biologists now incorporated this problematic feral population into decision-making regarding future management efforts. Similarly, it was not until pure parental and genetic hybrids could be genetically determined that understanding whether the expression of particular mallard-like traits were due to ancestry or contemporary gene flow. A recent study that genetically vetted phenotypic traits between mallards, Florida mottled ducks, and their hybrids reported that a key character used to identify hybrids was in fact found in 10% of genetically "pure" mottled ducks (i.e., white-wing bar over and under secondaries; Bielefeld et al. 2016). By determining which samples were genetic hybrids, Bielefeld et al. (2016) were able to identify those phenotypic traits that were indeed diagnostic of hybrids. Doing so, the authors were able to construct a genetically vetted phenotypic field key that increased the ability of wildlife biologists to correctly identify hybrids from 60% to >90%. In general, applying a landscape- and genomic-scale approach, research into Hawaiian ducks (Wells et al. 2019), Mexican ducks (Lavretsky et al. 2015a), American black ducks (Lavretsky et al. 2019b), and yellow-billed ducks (de Souza et al. 2019) provided the same discrepancies in hybrid identification using nongenetically vetted phenotypic traits versus the traits expressed by true genetic hybrids. These studies demonstrate that many phenotypic traits once considered to be indicative of hybrids are simply due to stochastic processes independently acting on ancestral mallard variation in each of these species. Estimating rates of hybridization plays important roles in taxonomic evaluation and conservation efforts. Thus, the capacity to determine true genetic hybrids is critical and can either validate current practices or identify which species cohort requires reevaluation in regard to hybrid identification.

With lowering costs and increasing efficiency in wet lab and sequencing methods, as well as developing user-friendly bioinformatics pipelines, the future for the field of conservation genetics is bright. Integrating knowledge gained from molecular work continues to be a powerful tool for conservation (Andrews et al. 2016; Oyler-McCance et al. 2016; McKinney et al. 2017). Given that population genomics

of non-model, wild systems is achievable for any organism, attaining a molecular understanding regarding the evolutionary and population genetics of even the most complex systems can now be realized.

## 4 Future Perspectives

Landscape-level and genome-wide population genomic datasets continue to expand our understanding regarding the dynamics of wild populations, including potential issues of hybridization. Among the opportunities afforded by the growing field of population genomics is being able to assign genetic purity to samples with confidence, which has been instrumental in advancing our understanding of evolutionary and contemporary population dynamics of specific taxa. Here, I also demonstrate how identifying hybrids and establishing hybridization rates can be key in establishing whether the expression of shared non-molecular traits (phenotypic and/or biological) across species is truly due to introgressive hybridization or simply ancestry (i.e., ILS).

Although much of this chapter focuses on how advances in population genomics has opened the possibilities for conservation, these same datasets can be applied to advance our understanding of the speciation process in wild systems (Lavretsky et al. 2015a, 2019a; Nadeau and Kawakami 2019), how the domestication process impacts genomic variation (Cornejo et al. 2018; Wu et al. 2018), as well as how species may adapt to today's ever-changing climates (Fenderson et al. 2020). Among these efforts, aDNA methods are especially promising as they open historical and ancient samples for analysis, providing a means to understand how species have responded to changing climates in the past. Similarly, genomic data from landscape-level sampling of contemporary samples coupled with Gradient Forest analyses (Ellis et al. 2012) can now be used to assign the genetic niche space of a species given its standing genetic variability. Importantly, such models now make it possible to model forward expected responses to specific climatic shifts given available genetic variation of a population (Fitzpatrick and Keller 2015; Bay et al. 2018). In short, populations with lacking variation will show contracting ranges as survival in a changing landscape will require substantial increases in standing molecular variation to adapt. These models will allow researchers to build specific genetic niche maps for their favorite organism and be able to determine where habitat may be most critical under different climatic models. Such analyses are sure to be promising when attempting to predict species range responses to climate change and using this information to better inform where habitat work may be most warranted.

As full genome [re-]sequencing becomes increasingly approachable, partial genome sequencing can still offer important information to guide sampling efforts for genomic analyses. In general, if starting out with little knowledge regarding some species of interest, researchers can use partial genome sequencing on many samples to understand true population structure, hybridization rates, and hybrid identification. Having genetically vetted samples will then help inform sampling efforts for full genome sequencing by ensuring that representative pure parental and various hybrid classes are in fact used, thus decreasing the chance of mistakenly sequencing samples with incorrectly presumed origins (Lavretsky et al. 2019b; Leitwein et al. 2019). Among the opportunities afforded with full genomes is the possibility to better understand potential interactions between selection and gene flow and, in particular, the consequence of domestic variant introgression into wild populations. Having full genome sequences for pure parental and various classes of hybrids, researchers can identify haplotype block organization, number of recombination events, and types of parental variation in hybrids (Tang et al. 2006; Corbett-Detig and Nielsen 2017; Schaefer et al. 2017; Janzen et al. 2018; Leitwein et al. 2019). First, one would expect any genetic variation linked to putatively maladaptive traits to decrease in size due to recombination and lost over some number of generational backcrosses (Leitwein et al. 2019). Conversely, introgressed neutral variation is expected to simply show increasing fragmentation due to recombination events that arise with each generation of backcrossing (Janzen et al. 2018). Thus, these methods show promise to identify maladaptive versus neutral molecular variants that are moved between species during gene flow events, as well as establish the true number of generations since the initial hybridization event. Such information is invaluable when attempting to understand how hybridization may actually be impacting the adaptive potential, including survival and fecundity of a species.

## 5 Conclusions

Population genomics has opened possibilities to better refine the conservation of many organisms that was once impossible. Continued advances in wet lab and statistical analyses will undoubtedly further unlock the potential of genomics for conservation. In addition to groundbreaking and important research, translating these findings to not only the biologists themselves (Garner et al. 2016; Funk et al. 2019) but also the general public is almost as critically important as the data itself (Holderegger et al. 2019). Thus, efforts to transfer findings from genetic data to those directly implementing on the ground conservation work, as well as making the general public understand the benefits of wild lands and wildlife, are just as important endeavors.

## References

Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:246.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.

Ankney DC, Dennis DG, Bailey RC. Increasing mallards, decreasing American black ducks: coincidence or cause and effect? J Wildl Manag. 1987;51:523–9.

Araki H, Ardren WR, Olsen E, Cooper B, Blouin MS. Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the Hood River. Conserv Biol. 2007;21:181–90.

Araki H, Cooper B, Blouin MS. Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. Biol Lett. 2009;5:621–4.

Avise JC, Ankney DC, Nelson WS. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution. 1990;44:1109–19.

Baldassarre G. Ducks, geese, and swans of North America. Baltimore: Johns Hopkins University Press; 2014.

Barton NH, Hewitt GM. Adaption, speciation and hybrid zones. Nature. 1989;341:497–503.

Bay RA, Harrigan RJ, Le Underwood V, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science. 2018;359:83–6.

Bazin E, Glémin S, Galtier N. Population size does not influence mitochondrial genetic diversity in animals. Science. 2006;312:570–2.

Bielefeld RR, Brasher MG, Moorman TE, Gray PN. Mottled duck (*Anas fulvigula*). In: Poole A, editor. The birds of North America online. Ithaca: Cornell Lab of Ornithology; 2010.

Bielefeld RR, Engilis A, Feddersen JC, Eadie JM, Tringali MD, Benedict RJ. Is it a mottled duck? The key is in the feathers. Wildl Soc Bull. 2016;40:446–55.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10:e1003537.

Braithwaite L, Miller B. The mallard, *Anas platyrhynchos*, and mallard-black duck, *Anas superciliosa rogersi*, hybridization. Wildl Res. 1975;2:47–61.

Briggs AW, Heyn P. Preparation of next-generation sequencing libraries from damaged DNA. In: Ancient DNA. Berlin: Springer; 2012. p. 143–54.

Brodsky LM, Weatherhead PJ. Behavioral and ecological factors contributing to American black duck-mallard hybridization. J Wildl Manag. 1984;48:846–52.

Brooke RK, Siegfried WR. Birds introduced to the fynbos biome of South Africa. In: Groves RH, di Castri F, editors. Biogeography of mediterranean invasions. Cambridge: Cambridge University Press; 1991. p. 359–64.

Brown JI, Lavretsky P, Cumming GS, Peters JL. Strong population structure and limited gene flow between yellow-billed ducks and mallards in southern Africa. The Condor. 2019.

Browne RA, Griffin CR, Chang PR, Hubley M, Martin AE. Genetic divergence among populations of the Hawaiian duck, Laysan duck, and mallard. Auk. 1993;110:49–56.

Buerkle C, Wolf D, Rieseberg L. The origin and extinction of species through hybridization. In: Population viability in plants. Berlin: Springer; 2003. p. 117–41.

Cade TJ. Hybridization and gene exchange among birds in relation to conservation. In: Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL, editors. Genetics and conservation: a reference for managing wild animals and plant populations. Menlo Park, CA: Benjamin/Cummings; 1983. p. 288–309.

Callaway E. Ancient DNA pinpoints dawn of Neanderthals: sequencing of 430,000-year-old DNA pushes back species' divergence from humans. Nature. 2016;531:286–7.
Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362–5.
Champagnon J, Guillemain M, Elmberg J, Folkesson K, Gauthier-Clerc M. Changes in Mallard *Anas platyrhynchos* bill morphology after 30 years of supplemental stocking. Bird Study. 2010;57:344–51.
Champagnon J, Elmberg J, Guillemain M, Gauthier-Clerc M, Lebreton J-D. Conspecifics can be aliens too: a review of effects of restocking practices in vertebrates. J Nat Conserv. 2012;20:231–41.
Champagnon J, Crochet PA, Kreisinger J, Čížková D, Gauthier-Clerc M, Massez G, et al. Assessing the genetic impact of massive restocking on wild mallard. Anim Conserv. 2013;16:295–305.
Cheng KM, Shoffner RN, Phillips RE, Lee FB. Mate preference in wild and domesticated (game-farm) mallards: II. Pairing success. Anim Behav. 1979;27:417–25.
Cheng KM, Burns JT, McKinney F. Forced copulation in captive mallards (*Anas platyrhynchos*): II. Temporal factors. Anim Behav. 1982;30:695–9.
Childs MR, Echelle AA, Dowling TE. Development of the hybrid swarm between Pecos pupfish (Cyprinodontidae: *Cyprinodon pecosensis*) and sheepshead minnow (*Cyprinodon variegatus*): a perspective from allozymes and mtDNA. Evolution. 1996;50:2014–22.
Čížková D, Javůrková V, Champagnon J, Kreisinger J. Duck's not dead: does restocking with captive bred individuals affect the genetic integrity of wild mallard (*Anas platyrhynchos*) population? Biol Conserv. 2012;152:231–40.
Cooper A, Rhymer J, James HF, Olson SL, Mclntosh CE, Sorenson MD, et al. Ancient DNA and island endemics. Nature. 1996;381:484.
Corbett-Detig R, Nielsen R. A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. PLoS Genet. 2017;13:e1006529.
Corbi J, Baack EJ, Dechaine JM, Seiler G, Burke JM. Genome-wide analysis of allele frequency change in sunflower crop–wild hybrid populations evolving under natural conditions. Mol Ecol. 2018;27:233–47.
Cornejo OE, Yee M-C, Dominguez V, Andrews M, Sockell A, Strandberg E, et al. Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. Commun Biol. 2018;1:167.
Crispo E, Moore JS, Lee-Yaw JA, Gray SM, Haller BC. Broken barriers: human-induced changes to gene flow and introgression in animals. Bioessays. 2011;33:508–18.
DaCosta JM, Sorenson MD. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One. 2014;9:e106713.
de Souza SG, Symes CT, Smit-Robinson H, Mollett JM. Minimal evidence of interspecific hybridisation between the yellow-billed duck and introduced mallard in central and northwestern South Africa. Ostrich. 2019;90:285–301.
Dean W. Alien birds in southern Africa: what factors determine success? S Afr J Sci. 2000;96:9–14.
Desforges M, Wood-Gush D. A behavioural comparison of domestic and mallard ducks. Habituation and flight reactions. Anim Behav. 1975a;23:692–7.
Desforges M, Wood-Gush D. A behavioural comparison of domestic and mallard ducks. Spatial relationships in small flocks. Anim Behav. 1975b;23:698–705.
Dobzhansky T. Speciation as a stage in evolutionary divergence. Am Nat. 1940;74:312–21.
Drummond A, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.
Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73.
Dyer J, Williams M. An introduction most determined: mallard (*Anas platyrhynchos*) to New Zealand. Notornis. 2010;57:178–95.
Edwards S, Bensch S. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Mol Ecol. 2009;18:2930–3.

Ellegren H. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol. 2010;25:283–91.
Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature. 2012;491:756–60.
Ellis N, Smith SJ, Pitcher CR. Gradient forests: calculating importance gradients on physical predictors. Ecology. 2012;93:156–68.
Engilis A Jr, Pratt TK. Status and population trends of Hawaii's native waterbirds, 1977-1987. Wilson Bull. 1993;105:142–58.
Engilis A Jr, Pyle R, David R. Status and occurrence of migratory birds in the Hawaiian Islands: Part 1 – Anseriformes: Anatidae (Waterfowl). Bishop Museum Occasional Paper 81. 2004.
Engilis A, Uyehara KJ, Giffin JG. Hawaiian duck*: Anas wyvilliana* birds of North America, Incorporated. 2002a.
Engilis AJ, Uyehara K, Giffin J. Hawaiian duck (*Anas wyvilliana*). In: Poole A, Gill F, editors. The Birds of North America. Philadelphia: The Birds of North America, Inc.; 2002b.
Evans CS, Evans L. Representational signalling in birds. Biol Lett. 2007;3:8–11.
Fenderson LE, Kovach AI, Llamas B. Spatiotemporal landscape genetics: investigating ecology and evolution through space and time. Mol Ecol. 2020;29:218–46.
Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol Lett. 2015;18:1–16.
Ford RJ, Selman W, Taylor SS. Hybridization between Mottled Ducks (*Anas fulvigula maculosa*) and Mallards (*A. platyrhynchos*) in the western Gulf Coast region. Condor. 2017;119:683–96.
Fowler A, Eadie J, Engilis A. Identification of endangered Hawaiian ducks (*Anas wyvilliana*), introduced North American mallards (*A. platyrhynchos*) and their hybrids using multilocus genotypes. Conserv Genet. 2009;10:1747–58.
Frankham R. Genetics and extinction. Biol Conserv. 2005;126:131–40.
Funk W, Forester BR, Converse SJ, Darst C, Morey S. Improving conservation policy with genomics: a guide to integrating adaptive potential into US Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet. 2019;20:115–34.
Garner BA, Hand BK, Amish SJ, Bernatchez L, Foster JT, Miller KM, et al. Genomics in conservation: case studies and bridging the gap between data and application. Trends Ecol Evol. 2016;31:81–3.
Gasc C, Peyretaillade E, Peyret P. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. Nucleic Acids Res. 2016;44:4504–18.
Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A. 1980;77:6715–9.
Gill FB. Species taxonomy of birds: which null hypothesis? Auk. 2014;131:150–61.
Grabenstein KC, Taylor SA. Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. Trends Ecol Evol. 2018;33(3):198–212.
Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, et al. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Resour. 2015;15:1304–15.
Grant PR, Grant BR. Hybridization of bird species. Science. 1992;256:193–7.
Grant PR, Grant BR. Hybridization, sexual imprinting, and mate choice. Am Nat. 1997;149:1–28.
Grealy A, Rawlence NJ, Bunce M. Time to spread your wings: a review of the avian ancient DNA field. Gene. 2017;8:184.
Greenwood PJ. Mating systems, philopatry and dispersal in birds and mammals. Anim Behav. 1980;28:1140–62.
Griffin CR, Browne R. Genetic variation and hybridization in Hawaiian Duck and Mallards in Hawaii. U.S. Fish and Wildlife Service, Hawaiian and Pacific Islands National Wildlife Refuge Complex. 1990.
Griffin HR, Pyle A, Blakely EL, Alston CL, Duff J, Hudson G, et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. Genet Med. 2014;16:962–71.

Grover CE, Salmon A, Wendel JF. Targeted sequence capture as a powerful tool for evolutionary analysis. Am J Bot. 2012;99:312–9.
Guay PJ, Tracey JP. Feral Mallards: a risk for hybridisation with wild Pacific black ducks in Australia? Victorian Nat. 2009;126:87–91.
Guay P-J, Williams M, Robinson RW. Lingering genetic evidence of North American mallards (*Anas platyrhynchos*) introduced to New Zealand. N Z J Ecol. 2015;39:103–9.
Haccou P, Lu B, Snow A, Stewart C, Strasburg J, van Tienderen P, et al. Introgression of crop alleles into wild and weedy populations. Annu Rev Ecol Evol Syst. 2013;44:325–45.
Haddon M. A re-analysis of hybridization between mallards and grey ducks in New Zealand. Auk. 1984;101:190–1.
Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016;30:33–41.
Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013;22:4606–18.
Heikkinen ME, Ruokonen M, White TA, Alexander MM, Gündüz İ, Dobney KM, et al. Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history. bioRxiv. 2019:706226.
Heusmann HW. Mallard-black duck relationships in the Northeast. Wildl Soc Bull. 1974;2:171–7.
Heusmann HW. The history and status of the mallard in the Atlantic Flyway. Wildl Soc Bull. 1991;19:14–22.
Hitchmough RA, Williams M, Daugherty CH. A genetic analysis of mallards, grey ducks, and their hybrids in New Zealand. New Zeal J Zool. 1990:467–72.
Holderegger R, Balkenhol N, Bolliger J, Engler JO, Gugerli F, Hochkirch A, et al. Conservation genetics: linking science with practice. Mol Ecol. 2019;28:3848–56.
Hoskin CJ, Higgie M, McDonald KR, Moritz C. Reinforcement drives rapid allopatric speciation. Nature. 2005;437:1353–6.
Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet. 2013;45:776–83.
Hubbard JP. The biological and taxonomic status of the Mexican Duck. Santa Fe, NM: New Mexico Department of Game and Fish Bulletin; 1977. p. 1–56.
Hurst GDD, Jiggins FM. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc R Soc B Biol Sci. 2005;272:1525–34.
Jacobsen F, Omland KE. Increasing evidence of the role of gene flow in animal evolution: hybrid speciation in the yellow-rumped warbler complex. Mol Ecol. 2011a;20:2236–9.
Jacobsen F, Omland KE. Species tree inference in a recent radiation of orioles (Genus *Icterus*): multiple markers and methods reveal cytonuclear discordance in the northern oriole group. Mol Phylogenet Evol. 2011b;61:460–9.
Janzen T, Nolte AW, Traulsen A. The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. Evolution. 2018;72:735–50.
Johnsgard PA. Hybridization in the *Anatidae* and its taxonomic implications. Condor. 1960;62:25–33.
Johnson KP, Sorenson MD. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence. Auk. 1999;116:792–805.
Kearns AM, Restani M, Szabo I, Schrøder-Nielsen A, Kim JA, Richardson HM, et al. Genomic evidence of speciation reversal in ravens. Nat Commun. 2018;9:1–13.
Keyser-Tracqui C, Ludes B. Methods for the study of ancient DNA. In: Goodwin W, editor. Forensic DNA typing protocols. Berlin: Springer; 2005. p. 253–64.
Kidd A, Bowman J, Lesbarreres D, Schulte-Hostedde A. Hybridization between escaped domestic and wild American mink (*Neovison vison*). Mol Ecol. 2009;18:1175–86.
Kiple KF. The Cambridge world history of food, vol. 2. Cambridge: Cambridge University Press; 2001.
Kirby RE, Reed A, Dupuis P Obrecht HH III, Quist WJ. Description and identification of American black duck, mallard, and hybrid wing plumage. In: Biological science report. U.S. Geological Survey, Biological Resources Division. 2000. p. 26.

Kirby RE, Sargeant GA, Shutler D. Haldane's rule and American black duck × mallard hybridization. Can J Zool. 2004;82:1827–31.

Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. Setting the stage–building and working in an ancient DNA laboratory. Ann Anat. 2012;194:3–6.

Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. Nature. 2016;530:429–33.

Kulikova IV, Zhuravlev YN, McCracken KG. Asymmetric hybridization and sex-biased gene flow between eastern spot-billed ducks (*Anas zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian Far East. Auk. 2004;121:930–49.

Kulikova IV, Drovetski SV, Gibson DD, Harrigan RJ, Rohwer S, Sorenson MD, et al. Phylogeography of the mallard (*Anas platyrhynchos*): hybridization, dispersal, and lineage sorting contribute to complex geographic structure. The Auk. 2005;122:949–65.

Laikre L, Ryman N. Effects on intraspecific biodiversity from harvesting and enhancing natural populations. Ambio. 1996;25:504–9.

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. Rapid hybrid speciation in Darwin's finches. Science. 2018;359:224–8.

Lande R. Anthropogenic, ecological and genetic factors in extinction and conservation. Res Popul Ecol. 1998;40:259–69.

Lavretsky P, Hernández Baños BE, Peters JL. Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the New World mallard complex (*Anas* spp.). Auk. 2014a;131:524–38.

Lavretsky P, McCracken KG, Peters JL. Phylogenetics of a recent radiation in the mallards and allies (Aves: *Anas*): inferences from a genomic transect and the multispecies coalescent. Mol Phylogenet Evol. 2014b;70:402–11.

Lavretsky P, Dacosta JM, Hernández-Baños BE, Andrew Engilis J, Sorenson MD, Peters JL. Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. Mol Ecol. 2015a;24:5364–78.

Lavretsky P, Engilis A, Eadie JM, Peters JL. Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. J Evol Biol. 2015b;28:1005–15.

Lavretsky P, Peters JL, Winker K, Bahn V, Kulikova I, Zhuravlev YN, et al. Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. Mol Ecol. 2016;25:661–74.

Lavretsky P, DaCosta JM, Sorenson MD, McCracken KG, Peters JL. ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. Mol Ecol. 2019a;28:2594–609.

Lavretsky P, Janzen T, McCracken KG. Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of eastern North America. Ecol Evol. 2019b;9:3470–90.

Lavretsky P, McInerney NR, Mohl J, Brown JI, James H, McCracken KG, et al. Assessing changes in genomic divergence following a century of human mediated secondary contact among wild and captive-bred ducks. Mol Ecol. 2020;29:578–95.

Leitwein M, Gagnaire PA, Desmarais E, Berrebi P, Guinand B. Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. Mol Ecol. 2018;27:3466–83.

Leitwein M, Cayuela H, Ferchaud AL, Normandeau É, Gagnaire PA, Bernatchez L. The role of recombination on genome-wide patterns of local ancestry exemplified by supplemented Brook Charr populations. Mol Ecol. 2019;28:4755–69.

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vila C. Ancient DNA evidence for Old World origin of New World dogs. Science. 2002;298:1613–6.

Leonardi M, Librado P, Der Sarkissian C, Schubert M, Alfarhan AH, Alquraishi SA, et al. Evolutionary patterns and processes: lessons from ancient DNA. Syst Biol. 2017;66:e1–e29.

Lichatowich J, Lichatowich JA. Salmon without rivers: a history of the Pacific salmon crisis. Washington, DC: Island Press; 2001.

Lijtmaer DA, Mahler B, Tubaro PL, Dunn P. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution. 2003;57:1411–8.

Lindqvist C, Rajora OP. Paleogenomics: genome-scale analysis of ancient DNA. Cham: Springer; 2019.
Livezey BC. A phylogenetic analysis of recent anseriform genera using morphological characters. Auk. 1986;103:737–54.
Livezey BC. A phylogenetic analysis and classification of recent dabbling ducks (Tribe *Anatini*) based on comparative morphology. Auk. 1991;108:471–507.
Livezey BC. Comparative morphometrics of *Anas* ducks, with particular reference to the Hawaiian duck *Anas wyvilliana*, Laysan duck *A. laysanensis*, and Eaton's Pintail *A. eatoni*. Wild. 1993;44:75–100.
Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, Hänni C. Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. Curr Biol. 2001;11:200–3.
Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2017;17:142–52.
Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20:229–37.
Mallet J. Hybrid speciation. Nature. 2007;446:279–83.
Mank JE, Carlson JE, Brittingham MC. A century of hybridization: decreasing genetic distance between American black ducks and mallards. Conserv Genet. 2004;5:395–403.
Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Res. 2013;23:1817–28.
McCracken KG, Johnson WP, Sheldon FH. Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). Conserv Genet. 2001;2:87–102.
McFarlane SE, Pemberton JM. Detecting the true extent of introgression during anthropogenic hybridization. Trends Ecol Evol. 2019;34(4):315–26.
McKinney GJ, Larson WA, Seeb LW, Seeb JE. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Mol Ecol Resour. 2017;17:356–61.
Miller DB. Social displays of mallard ducks (*Anas platyrhynchos*): effects of domestication. J Comp Physiol Psychol. 1977;91:221.
Minvielle F, Ito S, Inoue-Murayama M, Mizutani M, Wakasugi N. Brief communication. Genetic analyses of plumage color mutations on the Z chromosome of Japanese quail. J Hered. 2000;91:499–501.
Mitchell KJ, Wood JR, Scofield RP, Llamas B, Cooper A. Ancient mitochondrial genome reveals unsuspected taxonomic affinity of the extinct Chatham duck (*Pachyanas chathamica*) and resolves divergence times for New Zealand and sub-Antarctic brown teals. Mol Phylogenet Evol. 2014;70:420–8.
Moore WS. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution. 1995;49:718–26.
Nadachowska-Brzyska K, Burri R, Ellegren H. Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. Mol Ecol. 2019;28(9):2290–304.
Nadeau NJ, Kawakami T. Population genomics of speciation and admixture. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 613–53.
Nielsen EE, Bach LA, Kotlicki P. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. Mol Ecol Notes. 2006;6:971–3.
Nosil P, Schluter D. The genes underlying the process of speciation. Trends Ecol Evol. 2011;26:160–7.
Nosil P, Harmon LJ, Seehausen O. Ecological explanations for (incomplete) speciation. Trends Ecol Evol. 2009;24:145–56.
Omland KE. Examining two standard assumptions of ancestral reconstructions: repeated loss of dichromatism in dabbling ducks (Anatini). Evolution. 1997;51:1636–46.
Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. Nat Rev Genet. 2015;16:395.

Orr HA, Masly JP, Presgraves DC. Speciation genes. Curr Opin Genet Dev. 2004;14:675–9.
Ottenburghs J. Multispecies hybridization in birds. Avian Res. 2019;10:20.
Ottenburghs J, Ydenberg RC, Van Hooft P, Van Wieren SE, Prins HH. The Avian Hybrids Project: gathering the scientific literature on avian hybridization. Ibis. 2015;157:892–4.
Owens GL, Samuk K. Adaptive introgression during environmental change can weaken reproductive isolation. Nat Clim Chang. 2020;10:58–62.
Oyler-McCance SJ, Oh KP, Langin KM, Aldridge CL. A field ornithologist's guide to genomics: practical considerations for ecology and conservation. Auk. 2016;133:626–48.
Palmer RS. Handbook of North American birds. New Haven, CT: Yale University Press; 1976.
Paulke E, Haase E. A comparison of seasonal changes in the concentrations of androgens in the peripheral blood of wild and domestic ducks. Gen Comp Endocrinol. 1978;34:381–90.
Peters JL, Winker K, Millam KC, Lavretsky P, Kulikova I, Wilson RE, et al. Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*). Mol Ecol. 2014;23:2961–74.
Peters JL, Lavretsky P, DaCosta JM, Bielefeld RR, Feddersen JC, Sorenson MD. Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). Biol Conserv. 2016;203:272–81.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7:e37135.
Phadnis N, Orr HA. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science. 2009;323:376–9.
Picardi E, Pesole G. Mitochondrial genomes gleaned from human whole-exome sequencing. Nat Methods. 2012;9:523.
Pont C, Wagner S, Kremer A, Orlando L, Plomion C, Salse J. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biol. 2019;20:29.
Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
Pryke SR. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. Evolution. 2010;64:1301–10.
Puigcerver M, Sanchez-Donoso I, Vilá C, Sardá-Palomera F, García-Galea E, Rodríguez-Teijeiro JD. Decreased fitness of restocked hybrid quails prevents fast admixture with wild European quails. Biol Conserv. 2014;171:74–81.
Pyle R, Pyle P. The birds of the Hawaiian Islands: occurrence, history, distribution, and status. BP Bishop Museum, Honolulu, HI, USA. Version 2 (January 1, 2017). 2017.
Qiao H, Liu W, Zhang Y, Zhang YY, Li QQ. Genetic admixture accelerates invasion via provisioning rapid adaptive evolution. Mol Ecol. 2019.
Ralls K, Ballou JD, Dudash MR, Eldridge MD, Fenster CB, Lacy RC, et al. Call for a paradigm shift in the genetic management of fragmented populations. Conserv Lett. 2018;11:e12412.
Randi E. Detecting hybridization between wild species and their domesticated relatives. Mol Ecol. 2008;17:285–93.
Rawlence NJ, Kardamaki A, Easton LJ, Tennyson AJ, Scofield RP, Waters JM. Ancient DNA and morphometric analysis reveal extinction and replacement of New Zealand's unique black swans. Proc R Soc B Biol Sci. 2017;284:20170876.
Rheindt FE, Edwards SV. Genetic introgression: an integral but neglected component of speciation in birds. Auk. 2011;128:620–32.
Rhymer JM. Extinction by hybridization and introgression in anatine ducks. Acta Zoolog Sin. 2006;52(Suppl):583–6.
Rhymer JM, Simberloff D. Extinction by hybridization and introgression. Annu Rev Ecol Syst. 1996;27:83–109.
Rhymer JM, Williams MJ, Braun MJ. Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and grey ducks (*A. superciliosa*). Auk. 1994;111:970–8.
Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol. 2014;23:4757–69.
Rueness EK, Cadahia-Lorenzo L, Hoel K, Søderquist P, Velle G, Asmyhr MG, et al. Assessment of the risks associated with the import and release of hand-reared mallards for hunting purposes.

Opinion of the Panel on Alien Organisms and Trade in Endangered Species (CITES) of the Norwegian Scientific Committee for Food Safety, VKM Report 2017:23.
Rundle HD, Nosil P. Ecological speciation. Ecol Lett. 2005;8:336–52.
Russo I-RM, Hoban S, Bloomer P, Kotzé A, Segelbacher G, Rushworth I, et al. 'Intentional Genetic Manipulation' as a conservation threat. Conserv Genet Resour. 2018;11:237–47.
Sæther SA, Sætre G-P, Borge T, Wiley C, Svedin N, Andersson G, et al. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science. 2007;318:95–7.
Salzburger W, Baric S, Sturmbauer C. Speciation via introgressive hybridization in East African cichlids? Mol Ecol. 2002;11:619–25.
Sangster G. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biol Rev. 2014;89:199–214.
Schaefer NK, Shapiro B, Green RE. AD-LIBS: inferring ancestry across hybrid genomes using low-coverage sequence data. BMC Bioinformatics. 2017;18:203.
Scherer VS, Hilsberg T. Hybridisierung und verwandtschaftsgrade innerhalb der Anatidae – eine systematische un evolutionstheoretische betrachtung. J Ornitol. 1982;123:357–80.
Schluter D. Evidence for ecological speciation and its alternative. Science. 2009;323:737–41.
Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, et al. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 2012;13:178.
Schumer M, Rosenthal GG, Andolfatto P. How common is homoploid hybrid speciation? Evolution. 2014;68:1553–60.
Seehausen O. Hybridization and adaptive radiation. Trends Ecol Evol. 2004;19:198–207.
Seehausen O. Conservation: losing biodiversity by reverse speciation. Curr Biol. 2006;16:R334–7.
Shapiro B. Pathways to de-extinction: how close can we get to resurrection of an extinct species? Funct Ecol. 2017;31:996–1002.
Skoglund P, Ersmark E, Palkopoulou E, Dalén L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr Biol. 2015;25:1515–9.
Söderquist P, Gunnarsson G, Elmberg J. Longevity and migration distance differ between wild and hand-reared mallards *Anas platyrhynchos* in Northern Europe. Eur J Wildl Res. 2013;59:159–66.
Söderquist P, Norrström J, Elmberg J, Guillemain M, Gunnarsson G. Wild mallards have more "Goose-Like" bills than their ancestors: a case of anthropogenic influence? PLoS One. 2014;9:e115143.
Söderquist P, Elmberg J, Gunnarsson G, Thulin C-G, Champagnon J, Guillemain M, et al. Admixture between released and wild game birds: a changing genetic landscape in European mallards (*Anas platyrhynchos*). Eur J Wildl Res. 2017;63:98.
Souza CA, Murphy N, Villacorta-Rath C, Woodings LN, Ilyushkina I, Hernandez CE, et al. Efficiency of ddRAD target enriched sequencing across spiny rock lobster species (Palinuridae: Jasus). Sci Rep. 2017;7:6781.
Stephens K, Measey J, Reynolds C, Le Roux JJ. Occurrence and extent of hybridisation between the invasive mallard duck and native yellow-billed duck in South Africa. Biol Invasions. 2019:1–15.
Sutter A, Beysard M, Heckel G. Sex-specific clines support incipient speciation in a common European mammal. Heredity. 2013;110:398–404.
Tamisier A. The Camargue: a model of environmental decline. In: Finlayson CM, Hollis GE, Davis TJ, editors. Managing Mediterranean wetlands and their birds. Proceedings of the Symposium of Grado, Italy. Slimbridge: IWRB; 1992. p. 106–8.
Tang H, Coram M, Wang P, Zhu X, Risch N. Reconstructing genetic ancestry blocks in admixed individuals. Am J Hum Genet. 2006;79:1–12.
Tufto J. Domestication and fitness in the wild: a multivariate view. Evolution 2017;71(9):2262–70.
U.S. Fish & Wildlife Service. Recovery plan for Hawaiian waterbirds. Portland, OR: U.S. Fish & Wildlife Service; 2012.
USFWS. Review of captive-reared mallard regulations on shooting preserves – final report. Washington: Division of Migratory Bird Management, US Fish and Wildlife Service; 2013.

Uyehara KJ, Engilis A, Dugger BD. Wetland features that influence occupancy by the endangered Hawaiian duck. Wilson J Ornithol. 2008;120:311–9.
Vähä J-P, Primmer CR. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Mol Ecol. 2006;15:63–72.
Via S. Natural selection in action during speciation. Proc Natl Acad Sci U S A. 2009;106:9939–46.
vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY. Redefining the role of admixture and genomics in species conservation. Conserv Lett. 2018;11:e12371.
Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. Genome Res. 2017;27:1029–38.
Wang W, Wang Y, Lei F, Liu Y, Wang H, Chen J. Incomplete lineage sorting and introgression in the diversification of Chinese spot-billed ducks and mallards. Curr Zool. 2019;65:589–97.
Waples RS, Drake J. Risk/benefit considerations for marine stock enhancement: a pacific salmon perspective. Stock enhancement and sea ranching: developments, pitfalls and opportunities, 2nd ed. 2004. p. 260–306.
Watanabe T, Mizutani M, Wakana S, Tomita T. Demonstration of the maternal inheritance of avian mitochondrial DNA in chicken-quail hybrids. J Exp Zool. 1985;236:245–7.
Wayne RK, Shaffer HB. Hybridization and endangered species protection in the molecular era. Mol Ecol. 2016;25:2680–9.
Webb WC, Marzluff JM, Omland KE. Random interbreeding between cryptic lineages of the Common Raven: evidence for speciation in reverse. Mol Ecol. 2011;20:2390–402.
Wells CP, Lavretsky P, Sorenson MD, Peters JL, DaCosta JM, Turnbull S, et al. Persistence of an endangered island endemic, an introduced congener, and multiple hybrid swarms across the main Hawaiian Islands. Mol Ecol. 2019;28:5203–16.
Willerslev E, Cooper A. Review paper. Ancient DNA. Proc R Soc B Biol Sci. 2005;272:3–16.
Williams M. The duckshooter's bag: an understanding of New Zealand's wetland gamebirds. Wellington: Wetland Press; 1981.
Williams M. The changing relative abundance of grey duck (*Anas superciliosa*) and mallard (*A. platyrhynchos*) in New Zealand. Notornis. 2017;64:211–28.
Williams CL, Brust RC, Rhodes OE Jr. Microsatellite polymorphism and genetic structure of Florida mottled duck populations. Condor. 2002;104:424–31.
Williams CL, Brust RC, Fendley TT, Tiller GR, Rhodes OE. A comparison of hybridization between mottled ducks (*Anas fulvigula*) and mallards (*A. platyrhynchos*) in Florida and South Carolina using microsatellite DNA analysis. Conserv Genet. 2005a;6:445–53.
Williams CL, Fedynich AM, Pence DB, Olin E, Rhodes J. Evaluation of allozyme and microsatellite variation in Texas and Florida mottled ducks. Condor. 2005b;107:155–61.
Wringe BF, Stanley RR, Jeffery NW, Anderson EC, Bradbury IR. parallel new hybrid: an R package for the parallelization of hybrid detection using new hybrids. Mol Ecol Resour. 2017;17:91–5.
Wu C-I, Ting C-T. Genes and speciation. Nat Rev Genet. 2004;5:114–22.
Wu D, Ding X, Wang S, Wójcik J, Zhang Y, Tokarska M, et al. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. Nat Ecol Evol. 2018;2:1139–45.
Zink RM, Barrowclough GF. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol. 2008;17:2107–21.

# Part IV
# Challenges Facing Wildlife Populations

# Population Genomics and Wildlife Adaptation in the Face of Climate Change


**Jacob Höglund, Anssi Laurila, and Patrik Rödin-Mörch**

**Abstract** One of the worst threats facing wildlife populations worldwide is climate change. Average temperatures have risen globally and are expected to rise even further in the near future. Thus the climate is changing at an alarming rate and hence so are the living conditions of wildlife populations. The issue then becomes how will natural populations cope and deal with these changes? Here we review how wildlife populations may respond to such changes and how genomic tools can be used to study genetic consequences of climatic changes. Such studies may either look at genetic footprints which may be linked to past climatic fluctuations, or researchers may utilize natural or artificial environmental variation such as latitudinal gradients. While still in their infancy, genomic studies of consequences of climate change provide evidence of adaptation and may provide cues to how to preserve and restore resilience to ongoing shifts in climate.

**Keywords** Adaptation · Phenotypic plasticity · Range shift

## 1 Introduction

Wildlife populations worldwide are threatened by climate change (Parmesan 2006). The climate is changing at an alarming rate and hence so are the living conditions of wildlife populations (Collins et al. 2013). How does climate change affect population viability, fitness, and ecology of wildlife populations? Impact on these fundamental factors will leave a footprint in the genomes of organisms which can be studied and interpreted. It is our belief that a fuller understanding of the genomic impact of climate change may help in mitigating adverse effects on wildlife and aid in the preservation of biodiversity.

There are three possible responses of wildlife populations to climate change: They may track the change by moving their range. Alternatively, they stay and adapt

J. Höglund (✉) · A. Laurila · P. Rödin-Mörch
Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
e-mail: jacob.hoglund@ebc.uu.se

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)], https://doi.org/10.1007/13836_2019_69, 

to the changing circumstances, via phenotypic plasticity and/or genetic changes. The third outcome is that neither of the above are possible, and populations of such species ultimately go extinct (Davis and Shaw 2001; Davis et al. 2005; de Lafontaine et al. 2018). For understanding and facilitating the first two possibilities, biologists need knowledge. For movement to be possible, there needs to be sufficient connectivity among subpopulations and corridors to allow movement into new habitats, and we need to know how much and how to aid such connectivity. For adaptation to be possible enough, standing genetic variation is needed within natural populations, but we must know how much and what kind of genetic variation. Also in this aspect, connectivity plays a crucial role. Various subpopulations of a species can harbor locally adapted genotypes which may be "preadapted" to conditions that will be more common in the future. It has been argued that gene flow may mitigate the effects of global warming but also that local adaptation may be counteracted by outbreeding depression (Aitken and Whitlock 2013).

The next concern is whether science is in a state where we can address these issues. In this chapter, we will argue that genomics and genomic techniques offer a way to study genetic and evolutionary effects of climate change at a fundamental level (Franks and Hoffman 2012; Stillman and Armstrong 2015). Such an endeavor is not without challenges, however, and in order to understand the basis of adaptation in the face of climate change, many genetic markers spread throughout the genomes of the studied species are needed – many more than what is considered relevant for an average conservation genetic study. This is especially the case if we want to understand the genetic basis behind climate adaptations (or any adaptation for that matter; Savolainen et al. 2013). Ultimately we need a well-annotated de novo assembled reference genome of our study species, but this is unlikely to be realistic in the nearest future for many wildlife species under conservation concern. However, as reviewed elsewhere and in this book (Andrews et al. 2016; Forester et al. 2018a; Hendricks et al. 2018; Weisrock et al. 2018), a plethora of new techniques falling within the realm of genomics are available at the time of writing of this text. Such techniques in combination with whole-genome studies allow for starting to address questions about the molecular genetic basis of adaptation in natural populations.

## 2 Genomics

Genomic data are presently generated by what has been called next-generation sequencing (NGS) techniques, and sometimes "genomic data" and "NGS data" are synonymous. In a strict sense, genomics is the study of the entire genetic code of organisms, i.e., a whole-genome study of an organism. Thus under such a definition genomics is restricted to cases where at least a draft genome for the study species is available. However, NGS techniques, which are constantly under development, can be used to generate data that do not cover the entire genome, and as such "genomics" in a less strict sense has come to mean massive amounts of genetic data but not necessarily studies of entire genomes.

More and more studies of whole genomes of different species are published. For example, there are efforts to complete draft genomes of all extant species of birds within the next few years (https://b10k.genomics.cn/; accessed 27/6/2019). However, a single draft genome, even though useful, may tell us little when it comes to issues of pinpointing specific adaptations within wildlife populations. For this to be possible, we need population genomic data, and to date, such data are in most cases limited to just a handful of species including humans and model organisms such as *Arabidopsis thaliana* and *Drosophila melanogaster* (Ellegren 2014). In order to obtain population genomic data in the strict sense, resequencing at shallow depth (roughly $\times$10–20 per base) of many individuals of the target species would be necessary, and depending on the number of populations and individuals required to be sequenced, such costs may still be prohibitive for budgets of many research groups.

Here reduced representation sequencing come in handy. Such techniques, various RAD-seq (restriction site-associated DNA sequencing) and GBS (genotyping by sequencing) protocols, allow the researcher to collect data on a large number of single-nucleotide polymorphisms (SNPs) within populations at moderate costs (Andrews et al. 2016; Hohenlohe et al. 2010; Luikart et al. 2019). Even if offering far from complete genome coverage (Lowry et al. 2016), such data sets consist of $10^3$–$10^5$ SNPs which is a vast improvement to a few microsatellite markers which used to be the standard for conservation genetic studies not long ago (Allendorf 2017). This is not saying microsatellite studies are not useful. For many applications such as parentage or mark-recapture studies, they are still very good and adequate, but when studying adaptation, many markers preferably spread evenly among the chromosomes of the study species are needed. Another way of retrieving vast amounts of genomic data without sequencing the entire genome is to limit the sequencing to the transcribed genes in an organism, by using RNA-seq (Wang et al. 2009).

## 3 How to Study Adaptations to Climate Change?

There are a few ways to study climate adaptations in natural populations. These approaches (reviewed below, see also Franks and Hoffmann 2012) have identified a number of phenotypes, and in some cases even genotypes, associated with variation in climate. Such studies have, for example, detected phenotypic variation associated with seasonality changes (e.g., Bradshaw and Holzapfel 2001). Various quantitative genetic techniques have also been used to uncover the quantitative genetic contribution to such phenotypes (Bradshaw and Holzapfel 2008; Bradshaw et al. 2011). The quantitative trait loci and candidate genes for climate adaptation have been more seldom identified, and the underlying genetic mechanisms remain most often unknown (see below). In fact, very few studies have gone all the way from pinpointing phenotypes, show there is a genetic basis for the phenotypes and to

identify the genes involved (see Sect. 4 below). It is in this latter respect where genomic studies show great promise.

A very useful basic dichotomy, following Merilä and Hendry (2014), when studying adaptations to climate change, is the division between synchronic (i.e., comparisons of spatial populations) and longitudinal (temporal) studies. The latter are long-term temporal population comparisons, resurrection (see below), and experimental evolution studies.

At this stage, it may be appropriate to address what is meant by "climate." In reality, most studies reduce "climate" to the mean and variance of temperature facing any study species (Stillman and Armstrong 2015). This is for good reason. Temperature is easily measured and quantified and sometimes also possible to manipulate and control. However, "climate" is a composite of many variables, and even if the ultimate driver may be temperature shifts, other climate variables than temperature (humidity, precipitation, etc.) may be more relevant for the organism studied. Temperature also affects different kinds of organism in different ways. An obvious case is that ectothermal organisms are more directly and severely affected by shifting temperatures than endotherms (Angiletta 2009). In any case, the majority of studies addressing "climate change" do so by primarily studying effects of ambient temperature.

## *3.1 Phenotypes*

Several models of adaptation to temperature have been put forward in the literature. The so-called optimality models (Gilchrist 1995; Lynch and Gabriel 1987) assume "performance" (i.e., fitness) is a function of temperature with zero performance defined by the minimum ($T_{min}$) and maximum ($T_{max}$) temperatures at which the organism can perform (Fig. 1). Optimality models provide a framework for understanding and studying "thermal phenotypes" and reaction norms. Such a framework could easily incorporate quantitative genetic aspects where both genetic and environmentally induced variations as well as genotype-environment interaction affect the phenotypes (Conover and Schultz 1995; Conover et al. 2009). Here, warm origin individuals may have an overall higher performance than those from cold origin (cogradient variation, Fig. 2b). Alternatively, in the reverse case, genetic and environmental influences are negatively associated across a thermal gradient leading to higher overall performance of cold origin individuals. In such a case, we are facing a countergradient variation (CnGV) situation (Fig. 2c). While ecological and classical quantitative genetic studies can reveal the phenotypes and the genetic contribution to individual reaction norms, the actual genotypic constitution underlying this variation is not revealed. This is where a genomic framework can contribute.

Optimality models are most applicable to ectotherms. The same rules apply to endotherms, but as they are thermal specialists, finding the limits and variation is not as easy, as the amount of food resources available to the organism is very important

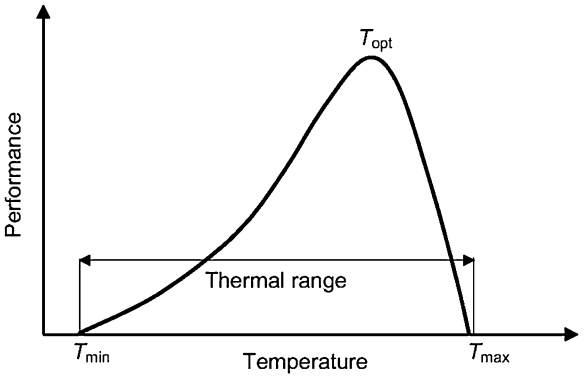


**Fig. 1** Hypothetical temperature reaction norm defining thermal breadth (range) limited by $T_{min}$, $T_{max,}$ and the temperature at which the performance is optimized ($T_{opt}$). From Denny and Dowd (2012). Thermal reaction norms are typically left skewed, and thermal breadth is limited by $T_{min}$, $T_{max,}$ and the temperature at which the performance is optimized ($T_{opt}$)

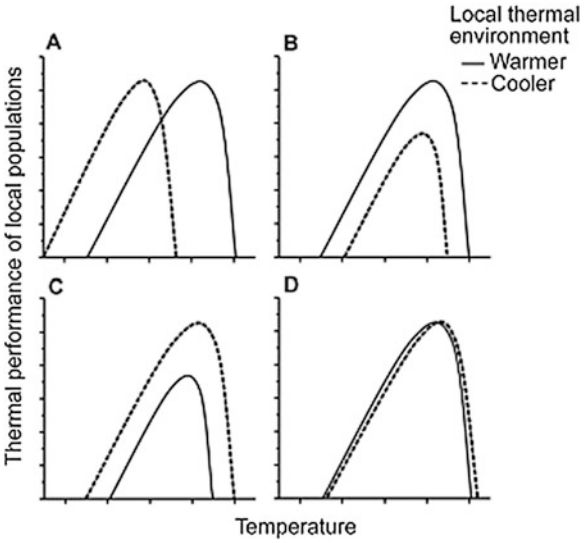


**Fig. 2** Theoretical thermal performances in two populations from a warm versus cold environment, respectively. (**a**) Local adaptation in which thermal tolerance and optima have diverged among the populations according to their local thermal experience. (**b**) Cogradient variation in which the population from the cold environment performs worse than the one from the warm at the same temperature. (**c**) Countergradient variation whereby genetic and environmental influences are negatively associated across a thermal gradient. The cold population is genetically "compensated" (e.g., has evolved a faster growth rate to compensate for environmentally induced slower growth under cold conditions and thus grows faster at the same temperature). (**d**) No divergence in thermal performance among warm and cold locations (from Gardiner et al. 2010, originally modified from Angiletta 2009)

in order to maintain body temperature. In endotherms, comparisons are mainly done between species and the amount of among-population, within-species adaptation remains largely unknown. The problem is that doing such studies in endotherms is ethically questionable, and this could be a rationale for using genomic (rather than phenotypic) methods.

Ultimately we want to understand the genetic basis for adaptation to local climates. This in essence is a special case of ecological genomics of local adaptation (Savolainen et al. 2013; Stinchcombe and Hoekstra 2008). As in any study attempting to reveal the genetic basis for phenotypes, it is essential to get the phenotyping correct. Plasticity, or norm of reaction, can potentially be studied genomically as a trait in itself, but with plasticity, this is challenging as accurate phenotyping is difficult. Without proper phenotyping, the search for the underlying genotypes is in vain.

## *3.2 Environmental Association Studies*

When it comes to studies of consequences of climate change, we may also use population samples which may give a hint on genetic differences among populations generated by differences in local conditions. Such studies attempt to disentangle the effects of genetic isolation by distance (IBD), which in theory is entirely generated by genetic drift, from isolation by environment (IBE) generated by local adaptation (Vasemägi 2006). If IBE is stronger than IBD, there is strong case for local adaptation. Unfortunately, geographic distance and environments often covary (e.g., temperatures along latitudinal and altitudinal gradients). In addition, relevant climate variables may also covary (e.g., temperature and oxygen pressure along altitudinal gradients).

A possible solution to the IBD/IBE problem is to conduct an outlier analysis. Outlier analyses compare differentiation among individual markers in order to identify selectively neutral markers from those which have been influenced by divergent or uniform selection (e.g., Beaumont and Nichols 1996). Thus differentiation calculated on strictly neutral markers would provide the baseline differentiation expectation generated by genetic drift, and functional analyses could be restricted to the loci. However, again there may be a problem with covariation between geographic distance and environment. Spatial autocorrelation can confound outlier analysis where neutral/demographic processes can result in similar signatures as selection, so what deviates from the neutral baseline might in fact be due to neutral processes and not divergent selection. IBD/IBE are essentially impossible to disentangle when everything covaries in the same dimension (Miermans 2012; Lotterhos and Whitlock 2015).

Strong evidence for local adaptation would be when the genotypes of properly phenotyped individuals can be identified. In order to control for phenotypic plasticity, it is necessary to conduct phenotyping under controlled conditions in a common garden (Merilä and Hendry 2014). Once the phenotyping has been done (e.g., using

a thermal reaction norm), the association between phenotype and genotype is no different than in any genomic association study (see Ekblom and Galindo 2011).

Another approach is to study the relationship between a set of environmental variables differing between habitats, such as temperature, and your genomic data set. These techniques, known as genotype-environment association (GEA), scan the genome in order to identify loci where the allele frequency differences between habitats show high correlations with habitat-specific environmental characteristics (Coop et al. 2010; Forester et al. 2018b). The habitat differences are assumed to promote divergent selection pressures across the heterogeneous environment facilitating local adaptation. In fewer cases, it is possible to use long-term temporal data to compare samples before and after a climate change. Such long-term studies of the same population may use a combination of quantitative (animal model) and genomic techniques to reveal allele frequency shifts due to climate change (Charmantier et al. 2016).

### *3.3 Phenology and Range Shifts*

Documented phenology and range shifts are abundant and in the direction that is often, although not always, predicted by global warming (de Lafontaine et al. 2018; Thomas 2010). Global meta-analyses have documented significant range shifts averaging 6.1 km per decade toward the poles and significant mean advancement of spring events by 2.3 days per decade (Parmesan and Yohe 2003). When combined with the so-called resurrection data (i.e., samples collected before and after an environmental change; Franks et al. 2008), such data can be used to elucidate genomic changes due to climate shifts.

For example, resurrection studies of the annual plant *Brassica rapa* have revealed rapid evolutionary response in flowering time in just a few generations demonstrating a genetic basis for the trait (Franks et al. 2007). Further studies using genomic techniques would help detecting the underlying genes and networks involved.

### *3.4 Cline Studies*

Likewise, using samples collected along latitudinal and altitudinal gradients combined with genomic data may be used to reveal responses to shifting climates (Fig. 3). Using widespread species, such as the fruit fly *Drosophila melanogaster*, researchers have revealed the molecular genetic basis for adaptive differentiation and local adaptation among populations (Turner et al. 2008). While not all adaptations revealed by this study are due to shifting climates, at least some of them are, and the study provides a set of candidate genes whose function may be tested in future studies.

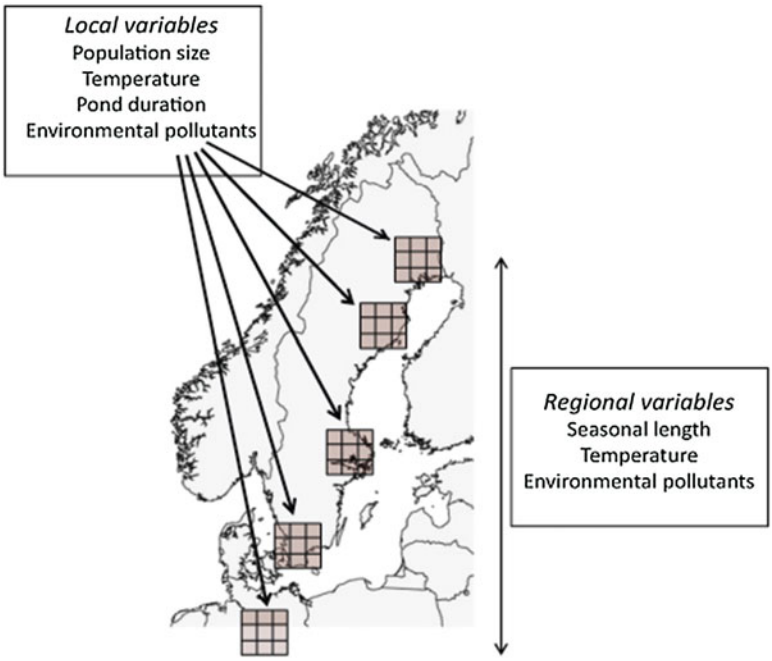


**Fig. 3** Map of Scandinavia and nearby countries showing a hypothetical study design illustrating how a cline study could be used to study climate adaptations and yet attempting to control for confounding variables. The regions (shaded boxes) differ in the environmental variables. The box to the left indicates local variables that differ within regions; the box to the right indicates the cline from south to north with shorter season, lower temperatures, and fewer environmental pollutants

Both natural shifts in phenology/range and cline studies suffer from the drawback that when relying on naturally caused variation, the ultimate factor driving the evolutionary response may be elusive (Merilä and Hendry 2014). The response could be due to climate effects but also other confounding factors. Even if a strong case for climate can be made, the precise causative agent might still remain obscure because "climate" is such a composite variable. On the other hand, recognizing that climates are complex has the consequence that the natural variation allows for responses that may be difficult to understand and replicate in a laboratory setting where multiple factors must be varied in a controlled fashion.

## *3.5 Natural and Anthropogenic Induced "Experiments"*

The use of natural or accidental anthropogenic alterations provides possibilities for studying adaptations relevant to climate change. Numerous organisms have adapted to live in local environments where the thermal regime is elevated, such as hot springs and thermal vents due to volcanic activity.

Using the freshwater gastropod *Radix balthica* in Lake Myvatn, a geothermal lake in Iceland with locally varying water temperatures, adaptations to ambient temperature were studied (Johansson et al. 2016a; Quintela et al. 2014). AFLP markers and mitochondrial haplotypes revealed strong genetic differentiation among localities. Five AFLP loci showed evidence of being under divergent selection. Variation in four of these loci was correlated with temperature differences (Johansson et al. 2016a).

Ermold (2016) took advantage of an artificially created water basin receiving cooling water discharge from a nuclear power plant on the Swedish east coast when studying the dwarf pond snail *Galba truncatula*. Using ca. 1,500 SNP loci detected by GBS, local adaptation evolved from standing genetic variation to the artificially raised temperatures in the basin within 30 years after the construction (Ermold 2016).

### *3.6 Common Garden Experiments*

Climate change adaptations are in most cases likely to consist of a combination of both genetic response and phenotypic plasticity. Thus the model shown in Fig. 1 could be interpreted both as the response of a population but also at the level of an individual. When assessing individual phenotypes, it is therefore important to screen the phenotypes under controlled common garden conditions. As a case in point, larval growth periods of amphibians in Sweden are phenotypically similar in the north compared to the south. However, when reared under common garden conditions in the laboratory, the northern populations grow and develop faster thus genetically compensating for the harsher growth conditions in the north (e.g., Laugen et al. 2003). Such countergradient variation may phenotypically mask underlying genetic differences among populations (Conover et al. 2009).

Thermal adaptation theory also predicts that thermal specialists evolve in environments with low thermal variation and generalists in environments with high. This was tested in the snail *R. balthica* in geothermally active Lake Myvatn on Iceland. Common garden experiments showed that snails originating either from cold or warm water stable temperatures differed from snails originating in variable temperatures. Predictions were only partly met since warm origin snails had poor survival when raised at cold temperatures but higher than both variable and cold origin snails at warm temperatures. However, irrespective of temperature, growth rate was highest in snails from warm origin, indicating cogradient variation. The optimal temperatures for growth and reproduction were similar irrespective of origin, but snails from variable temperature regime performed at an intermediate level compared to snails from either stable environment (Johansson et al. 2016b).

Finding the underlying genes using a genome-wide association study for phenotypes determined under common garden conditions would be major step in understanding climate adaptations. However, the fact that climate adaptations are at least partly determined by phenotypic plasticity calls for studies on gene expression and its regulation.

## 3.7 Candidate Genes

If an association study finds loci associated with climate-related phenotypes, and it is possible to map such chromosomal positions to known annotated genes, the researchers may make a case for a candidate gene. Ultimately the function of different gene variants would need to be tested in a functional genetic study using knockouts or other tools of manipulating the genetic code. However, even without a formal functional study, it is still possible to make a strong case for adaptation by using comparative genomics relying on homology.

One of the most well-studied candidate gene families are the heat shock proteins (Hsp) most well-known for detailed studies in *Drosophila*. Hsp70 is one of the heat shock proteins, later found in many organisms, which is upregulated during heat stress and other stressors (reviewed by Franks and Hoffmann 2012; see also Bentley et al. 2017; Jesus et al. 2016; Narum and Campbell 2015 for recent vertebrate examples). Heat shock proteins are a diverse group of proteins produced in response to stress and function as chaperones to stabilize new proteins and refold old damaged ones. However, expression levels are variable depending on life history stage, and, for example, variation in Hsp70 expression levels among lines correlated with heat resistance at larval stage but not in adults (Jensen et al. 2010). Thus interpreting the function of candidate genes is complex, and care is needed when interpreting data.

The thyroid hormone pathway provides another set of candidate loci relevant for climate adaptations. This pathway is, among other things, involved in phenological timing such as onset of seasonal growth (Kulkarni and Buchholz 2014). In the moor frog *Rana arvalis*, genetic variation at a transcription factor locus, C/EBP-1, probably located upstream of the thyroid hormone receptor β (THRB) gene, is associated with larval development time (Richter-Boix et al. 2010, 2013). This trait is under strong local selection in natural populations with faster development being favored in cool forest ponds (Richter-Boix et al. 2013). Variation in the thyroid hormone axis provides a promising system for disentangling the genetic mechanisms behind adaptive geographical variation in development rate masked by countergradient variation. We have made further studies of genetic variation at C/EBP-1 in *R. arvalis* which shows strong signals of selection along a latitudinal gradient across northern Europe and detected genetic variants associated with larval growth rate phenotypes (Meyer-Lucht et al. 2019).

## 3.8 Expression Studies

Most immediate responses to climate change have a large phenotypically plastic component (Merilä and Hendry 2014). Hence gene expression studies in relation to climate variables are highly valuable, also in identifying genes involved in climate response. Gene regulation and expression can be studied with NGS techniques. Previously, microarrays were the tool for studying expression variation among

different life history stages and organs. This has lately been replaced by direct sequencing of expressed genes, the so-called transcriptome sequencing or RNA-seq (Wang et al. 2009).

The molecular pathways underlying rapid adaptive phenotypic responses to global change remain poorly understood and are likely to differ between ectotherms and endotherms. Ectothermal organisms are likely to be more directly affected. However, two environmentally sensitive molecular mechanisms which affect gene expression, transposable elements (TEs) and epigenetic components (ECs), are present in both kinds of organisms. Both these mechanisms are sensitive to global change stressors and may interact with each other. It has been proposed that coupling TEs and ECs allows organisms to fine-tune phenotypes in real time, to adjust the production of phenotypic and genetic variation, and to produce heritable phenotypes with different levels of transmission fidelity (Rey et al. 2016). Clearly, screening for TEs and ECs such as DNA methylations, histone modifications, and noncoding small RNAs will be facilitated through NGS techniques.

Using genomic techniques, gene expression studies in relation to heat stress have been conducted on several fish species and turtle embryos. In redband trout (*Oncorhynchus mykiss gairdneri*), a difference in transcriptomic response among populations from montane and desert origins and their F1 hybrids was detected (Narum and Campbell 2015). Among the differentially expressed transcripts were heat shock proteins (Hsp) and also genes involved in metabolic and cellular responses. Likewise, rainbowfish (*Melanotaenia duboulayi*) displayed differential expression among different temperature treatment groups. Annotated BLAST matches revealed that differentially expressed genes corresponded to critical metabolic pathways previously shown to be important for temperature tolerance in other fish species (Smith et al. 2013). Contrasting gene expression patterns in two Iberian freshwater species (*Squalius carolitertii* and *S. torgalensis*), one from warm and the other from cold origin, showed differences among the two species but also significant upregulation of Hsp genes in both species in response to heat stress. In loggerhead turtle (*Caretta caretta*) embryos, genes enriched in a heat-shock treatment were primarily associated with the Hsp families, or other genes involved in similar protein editing and chaperone functions (Bentley et al. 2017). These studies illustrate the plastic nature of how organisms cope with heat and other stressors but also the underlying heritable genetic component of plasticity (McCairns et al. 2016).

## *3.9 Historical Reconstructions*

A slightly different use of genomic data in relation to climate is to focus on and reconstruct the demographic history of species. This is relevant because inferring demographic history can be used to examine how past climate changes have affected population dynamics (Li and Durbin 2014). Realizing that the effect of genetic drift is inversely related to population size (Kimura 1983), it follows that the effects of demography on genetic variability are stronger when effective population size ($N_e$) is

small. With genomic data, it is possible to infer the demographic history of species from a few samples by looking at the patterns of genetic variation throughout the genome of an organism. One such method is the pairwise sequentially Markovian coalescent model (PSMC; Li and Durbin 2014). It allows for the tracking of changes in species' $N_e$ from 10,000 years ago to Early Pleistocene/Late Pliocene (~3 million years ago) from the genome of just one individual (Zhao et al. 2013). Another approach is to test more complex demographic scenarios by conducting approximate Bayesian computation (ABC). This very flexible framework allows for testing demographic models involving, for example, historical gene flow and to infer demographic parameters such as $N_e$ (e.g., Weir et al. 2016). In essence, this is done by simulating genealogies under various demographic scenarios using a coalescence simulator. Summary statistics are then calculated from these simulations and compared to those obtained from empirical data in order to conduct model comparison and parameter inference (Beaumont et al. 2002; Csilléry et al. 2010). This type of computational technique provides ideal tools to study past changes in population size and migration and to relate such changes to known fluctuations of the climate.

Zhao et al. (2013) used PSMC to track the demographic fluctuations of the Giant panda (*Ailuropoda melanoleuca*) in relation to records of past climate fluctuations by analyzing 34 panda genomes. They concluded that although changes in climate were the primary drivers of population fluctuation for millions of years, recent human activities such as hunting and habitat reduction likely underlie recent population divergence and the serious decline. The same technique has been used to reconstruct the demographic history of primate species including humans (Prado-Martinez et al. 2013; Xue et al. 2015) and speciation and rapid evolutionary adaptation in polar bears (*Ursus maritimus*; Liu et al. 2014).

In birds, PSMC has been used to track changes in population size in the extinct passenger pigeon (*Ectopistes migratorius*; Hung et al. 2014), the extant turtle dove (*Streptopelia turtur*; Calderón et al. 2016), *Ficedula* flycatchers (Nadachowska-Brzyska et al. 2016), and two species of penguins (Li et al. 2014). Nadachowska-Brzyska et al. (2015) analyzed published genome data and performed PSMC analyses in several species of birds. In most studies, population size change throughout the demographic history of species is compared to changes in climate (most often reconstructed changes in temperature). We have used PSMC to track $N_e$ of three species of grouse since roughly 3 million years ago. Combined with climate modelling of species distributions (Kozma et al. 2016; Lagerholm et al. 2017), we have documented responses in population size and available habitat of all the studied species. These studies suggest general demographic responses to climate change but also differences among the species as the investigated species differ in their specialization to cool climates.

# 4 Relevant Case Studies

## *4.1 Reviews*

In general, very few studies on wildlife species (in the strict sense) have to our knowledge yet been published on using genomics to study climate adaptations. Nevertheless a few studies on what might not be considered wildlife but yet free living wild organisms have been published. Franks and Hoffman (2012) provide an excellent review of relevant work on adaptations to climate change up until 2012. Their review is focused on genetics of climate adaptation and is thus not limited to genomic studies. Nevertheless, their review also covers such studies. Moreover, Reusch (2014) in a review discusses the potential for adaptive evolution due to climate change in marine animals and plants. In another review, the impact and probably increased frequency of hybridization due to climate change and how such events could be detected using genomic techniques citing relevant case studies were assessed (Chown et al. 2015). Stillman and Armstrong (2015) reviewed the methods available and a few case studies (mainly on marine organisms) where genomic tools have been or may become useful in light of understanding responses to climate change. All these reviews state that genomic tools may be useful but have not yet been used much to study effects of climate change.

Below we highlight a few recent studies since 2012 where genomics is used to study climate adaptation. These studies are hits from Google Scholar (11 June 2017 and updated 26 June 2019) and an ISI Web of Knowledge (7 August 2017) using the keywords "taxon classifier," "genomics," and "climate."

## *4.2 Fruit Flies*

Being one of the most used genetic models, it is no surprise that the fruit fly, *Drosophila melanogaster*, was a pioneer study organism when it comes to genomic studies of climate adaptation. Recent research using genomics have involved detection of chromosomal inversion polymorphisms, known to facilitate local adaptation by preventing recombination breaking up locally adapted gene complexes (Rane et al. 2015). Furthermore, metabolic adaptation to climate in the central metabolic pathway has been studied using 127 single-nucleotide polymorphisms (SNPs) in 46 known enzyme-coding genes within the pathway enabled by the extensive genomic resources available to fruit fly researchers (Lavinton et al. 2014). Global transcriptomic profiling of diapause and climatic adaptation has been studied using pooled sequencing of cDNA (Zhao et al. 2016). Finally, clinal variation across the USA, Europe, Africa, and Australia has been studied using genomic techniques (Adrion et al. 2015; Bergland et al. 2016). Studies of this well-known model system are impressive and sometimes perhaps also discouraging for researchers working on less well-studied systems. However, fruit fly studies also show the potential and possibilities of genomic studies in other organisms.

### 4.3 Fish

In fish, climate adaptation has in the past most commonly been studied in salmonids and three-spined stickleback (*Gasterosteus aculeatus*; see references in Franks and Hoffmann 2012). Recent studies involve a genome wide phylogeographic study of golden perch (*Macquaria ambigua*) in Australia. The study showed that this widespread fish is divided in three distinct evolutionary lineages which presumably have evolved in response to climate differences (Beheregaray et al. 2017). In another Australian perch, the pygmy perch (*Nannoperca australis*), "riverscape genomics" were studied using more than 500 SNPs derived from ddRAD (Brauer et al. 2016). Again, range limits of distinct evolutionary lineages matched the climate-determined boundaries of main river basins, suggesting that also this species have adapted and diverged in response to climate factors.

### 4.4 Amphibians

In amphibians, genomic data have been used to study the interaction between climate and the amphibian-killing chytrid fungus *Batrachochytrium dendrobatidis* (Bd). Bd grows best between 15 and 25°C, and as such the disease is climate dependent (Piotrowski et al. 2004). Ellison et al. (2015) challenged four Central American frog species that vary in Bd susceptibility, with a virulent strain of the pathogen. They then compared skin and spleen orthologous gene expression using differential expression tests and coexpression gene network analyses and found that resistant species have reduced skin inflammatory responses and increased expression of genes involved in skin integrity. In contrast, only highly susceptible species exhibited suppression of splenic T-cell genes. They concluded that resistance to chytridiomycosis may be related to a species' ability to escape the immunosuppressive activity of the fungus. This study highlights that disease dynamics may be temperature/climate dependent, and since pathogens often have shorter life cycles than their hosts, they can evolve and adapt more rapidly to changing ambient temperatures, thus constituting a climate-related threat to wildlife populations by invading previously uninfected populations.

### 4.5 Li zards

Rodriguez et al. (2017) used a comparative approach to verify predicted gene functions for vertebrate thermal adaptation with observed functions underlying repeated genomic adaptations in response to elevation in the lizard *Anolis cybotes*. Using RAD-sequencing of populations at different elevations, they established a link between recurrently evolving phenotypes and functional genomics of altitude-related climate adaptation in three highland and lowland populations in Hispaniola.

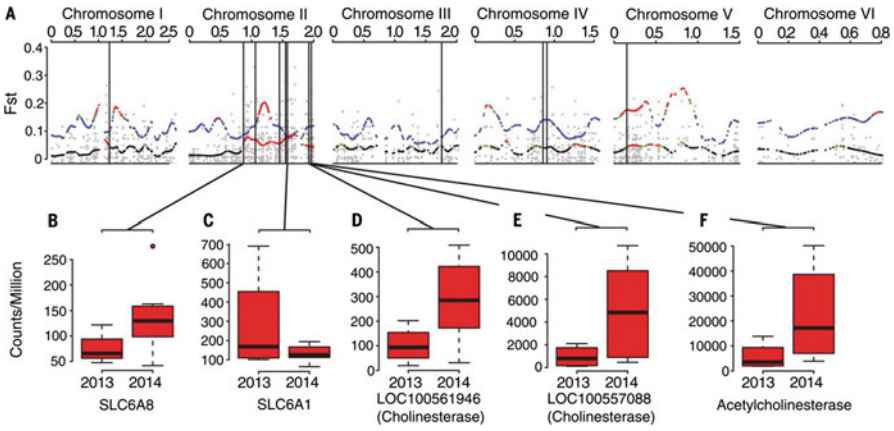


**Fig. 4** (**a**) Genomic scan for targets of storm-mediated selection in *Anolis carolinensis* after the winter of 2013–2014 in Texas, USA. Gray points represent individual values of $F_{ST}$ for each single-nucleotide polymorphism. Black dots indicate nonsignificant $F_{ST}$ values within 5-Mb windows in pre- versus post-storm comparisons. Blue dots indicate nonsignificant $F_{ST}$ values within 5-Mb windows in northern versus southern population comparisons. Red and green dots indicate regions of significantly elevated $F_{ST}$ between samples (bootstrap resampling, $P < 0.01$ and $P < 0.05$, respectively). Black lines indicate differentially expressed genes within $F_{ST}$ outlier peaks. (**b–f**) Expression differences between pre- and post-storm northern samples at gene expression outliers. Genes shown are associated with cholinesterase activity and sodium symporter activity (from Campbell-Staton et al. 2017)

They constructed a functional genomic network to infer whether newly identified protein-coding outlier genes fell into previously published functional categories for thermal adaptation in vertebrates. Their results suggest that genes for vertebrate thermal adaptation can be functionally classified and seem to be conserved through evolution. Campbell-Staton et al. (2017) used thermal performance, transcriptomics, and genome scans to measure responses of *Anolis carolinensis* populations to storm-induced selection before and after the unusually cold winter 2013–2014 in the Southeastern USA. They found 14 genomic regions which were differentiated in a surviving southern population. Four of these regions also exhibited signatures of local adaptation across a latitudinal gradient and implicated genes involved in nervous system function (Fig. 4).

## *4.6 Birds*

In an arctic seabird, the thick-billed murre (*Uria lomvia*), outlier analyses were used to test for local adaptation to breeding grounds along a latitudinal gradient (Tigano et al. 2017). These authors did not find any support for local adaptation along the gradient in this migratory seabird. However, outlier loci grouped birds according to

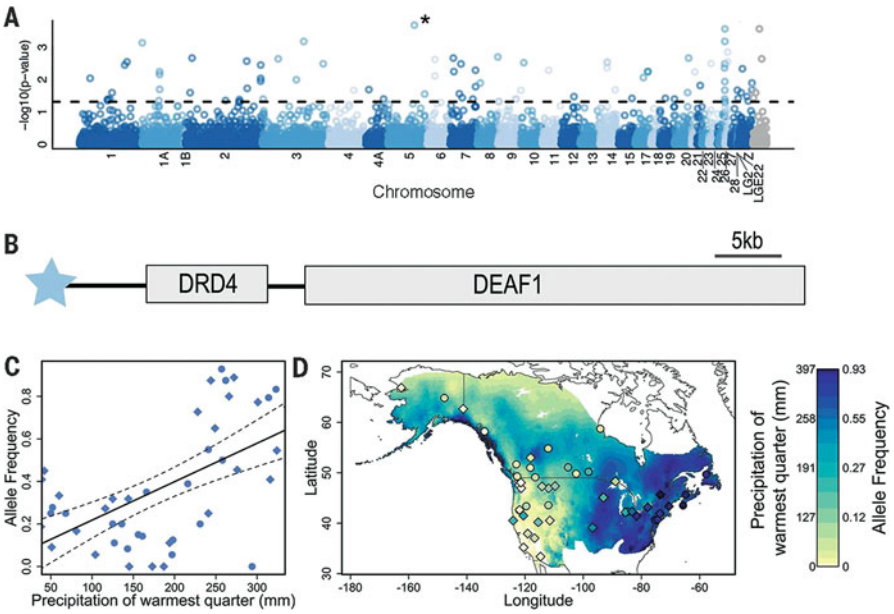


**Fig. 5** (**a**) Manhattan plot showing the significance level for SNP associations with precipitation of the warmest month in yellow warblers *Scatophaga pethecia*. The dashed line represents $P = 0.05$. Colors distinguish different chromosomes, and gray points are SNPs on scaffolds not anchored to a chromosome. (**b**) The most significant SNP association, marked with an asterisk (∗) in (**a**), is upstream from the *DRD4* and *DEAF1*genes. (**c**, **d**) Correlations between allele frequency and a climate variable for this SNP. Samples genotyped by RAD-seq are represented as circles, and samples genotyped with Fluidigm assays are shown as diamonds (from Bay et al. 2018)

their non-breeding distributions suggesting adaptations to different flyways and non-breeding grounds. The link to climate is thus obscure in this case. Likewise in blue tits (*Cyanistes caeruleus*), SNP data suggest differentiation among different habitat types in Southern France and Corsica, which may suggest local adaptation, but again the link to climate variables may be obscured by other relevant ecological variables (Charmantier et al. 2016). By integrating population genomics and climate modelling, Bay et al. (2018) identified genomic variation associated with climate across the breeding range of a migratory songbird, the yellow warbler (*Setophaga petechia*). Populations requiring the greatest shifts in allele frequencies to keep pace with climate change experienced the largest population declines, suggesting that failure to adapt may have negatively affected populations (Fig. 5).

## *4.7 Mammals*

Among mammals, the American pika (*Ochotona princeps*) has been advanced as a study species for examining biotic responses to climate change. Pooled

transcriptome sequencing, in addition to sequencing of haplotypes of the mitochondrial gene NADH, revealed differences among high and low elevation sites providing candidate loci for further studies (Lemay et al. 2013). Using amplified fragment length polymorphisms (AFLP) in the same species, Henry and Russello (2013) found 68 outlier loci out of 1,509 screened associated with altitude (and thus implicating local adaptation to climate variables). However, studies of the closely related collared pika (*O. collaris*) used RAD-seq to evaluate the effects of historical range shifts, contemporary climates, and landscape factors on gene flow. The results suggest that contemporary climate factors had little effect, and instead genetic diversity was strongly shaped by historical demographic factors (Lanier et al. 2015). This suggests that concluding that adaptive selection is behind any observed correlation with a geographic cline is premature without additional studies. Also historic geographic subdivision may establish clinal patterns (e.g., Strand et al. 2012). Nonetheless, targeted resequencing suggested environmentally driven functional genetic variation among six distinct gray wolf (*Canis lupus*) ecotypes (Schweizer et al. 2016). Also, analyses of 89 complete genomes suggest the diversification among brown (*Ursus arctos*) and polar bears (*U. maritimus*) has been shaped by climatic factors during the last 400,000 years (Liu et al. 2014).

## 5 Concluding Remarks

Although there is concern that populations are unable to track the fast environmental changes occurring in the modern world, there is some hope of evolutionary rescue. Killeen et al. (2017) investigated how different rates of temperature increase affect population persistence and evolutionary change in experimental microcosms of the protozoan *Paramecium caudatum*. Perhaps not surprisingly, they found that populations experiencing the slowest rate of temperature increase were the least likely to become extinct and tended to be best adapted to the new temperature environment. However, all high-temperature populations were more tolerant to severe heat stress, indicating a common mechanism of heat protection. These and other results reviewed in this chapter are consistent with the idea that local adaptation is possible and may safeguard against extinction, at least in short-lived organisms with abundant populations, but that such responses depend on the magnitude of the change. Adaptation to more dramatic changes is much harder and less probable. Whether adaptation is likely also in long-lived organism which are less abundant is an issue which urgently needs to be addressed in future research.

Genomic tools are likely to aid future research on climate adaptation by providing relevant data for phenotype and genotype associations and environmental associations using GWAS and by providing a powerful tool in order to track past demographic changes (such as those provided by PSMC and ABC models). While still in its infancy and mostly used on model organisms, genomic tools applied on wildlife species will most certainly advance our understanding of the consequences of climate change.

However, and more urgently, how can genomic results inform conservation or management in the face of climate change? Population genomics can inform conservation practitioners on which species and populations have lost genetic variation, the fuel of any further evolutionary change and adaptation. This can of course also be assessed with classical genetic markers. However, genomics will provide a more complete picture on which genomics regions are liable to change and which are relevant in terms of functional characteristics. Such can be identified with information on genes associated with climate or local adaptation across climate gradients. Key issues and questions to address in this area in the next 5–10 years will be to locate the genomic regions relevant for climate adaptions and if there are commonalties among taxa. It is likely that closely related species and species with similar life history characteristics will share genomic architectures of relevance to climate adaptation (such as larval growth rates, seasonal timing of reproduction, etc.). Given the vast diversity in lifestyles across the tree of life, it is likely that many organisms may reveal unique adaptations and therefore require special attention and needs. This may even be the case within species, and identifying the genes involved in local adaptation will be crucial for avoiding outbreeding depression and introgression of maladapted genotypes in genetic rescue projects when translocations may be considered.

## References

Adrion JR, Hahn MW, Cooper BS. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. Trends Genet. 2015;31:434–44. https://doi.org/10.1016/j.tig.2015.05.006.

Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. Annu Rev Ecol Evol Syst. 2013;44:367–88. https://doi.org/10.1146/annurev-ecolsys-110512-135747.

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30. https://doi.org/10.1111/mec.13948.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92. https://doi.org/10.1038/nrg.2015.28.

Angiletta MJ. Thermal adaptation. A theoretical and empirical synthesis. Oxford: Oxford University Press; 2009.

Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science. 2018;359:83–6. https://doi.org/10.1126/science.aan4380.

Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lon B Biol Sci. 1996;263:1619–26. https://doi.org/10.1098/rspb.1996.0237.

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002;162:2025–35.

Beheregaray LB, Pfeiffer LV, Attard CRM, Sandoval-Castillo J, Domingos FMCB, Faulks LK, et al. Genome wide data delimits multiple climate-determined species ranges in a widespread Australian fish, the golden perch (*Macquaria ambigua*). Mol Phylogenet Evol. 2017;111:65–75. https://doi.org/10.1016/j.ympev.2017.03.021.

Bentley BP, Haas BJ, Tedeshi JN, Berry O. Loggerhead sea turtle embryos (*Caretta caretta*) regulate expression of stress response and developmental genes when exposed to a biologically realistic heat stress. Mol Ecol. 2017;26:2978–2. https://doi.org/10.1111/mec.14087.

Bergland AO, Tobler R, González J, Schmidt P, Petrov D. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. Mol Ecol. 2016;25:1157–74. https://doi.org/10.1111/mec.13455.

Bradshaw WE, Holzapfel CM. Genetic shift in photoperiodic response correlated with global warming. Proc Natl Acad Sci U S A. 2001;98:14509–11. https://doi.org/10.1073/pnas.241391498.

Bradshaw WE, Holzapfel CM. Genetic response to rapid climate change: it's seasonal timing that matters. Mol Ecol. 2008;17:157–66. https://doi.org/10.1111/j.1365-294X.2007.03509.x.

Bradshaw WE, Emerson KJ, Holzapfel CM. Genetic correlations and the evolution of photoperiodic time measurement within a local population of the pitcher-plant mosquito, *Wyeomyia smithii*. Heredity. 2011;108:473–9. https://doi.org/10.1038/hdy.2011.108.

Brauer CJ, Hammer MP, Beheregaray LB. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Mol Ecol. 2016;25:5093–113. https://doi.org/10.1111/mec.13830.

Calderón L, Campagna L, Wilke T, Lormee H, Eraud C, Dunn JC, et al. Genomic evidence of demographic fluctuations and lack of genetic structure across flyways in a long distance migrant, the European turtle dove. BMC Evol Biol. 2016;16:237. https://doi.org/10.1186/s12862-016-0817-7.

Campbell-Staton SC, Cheviron ZA, Rochette N, Catchen J, Losos JB, Edwards SV. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. Science. 2017;357:495–8. https://doi.org/10.1126/science.aam5512.

Charmantier A, Doutrelant C, Duboc-Messier G, Fargevieille A, Szulkin M. Mediterranean blue-tits as a case study of local adaptation. Evol Appl. 2016;9:135–52. https://doi.org/10.1111/eva.12282.

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA. Biological invasions, climate change and genomics. Evol Appl. 2015;8:23–46. https://doi.org/10.1111/eva.12234.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, et al. Long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, editors. Climate change 2013: the physical science basis. Working Group I contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2013.

Conover D, Schultz ET. Phenotypic similarity and the evolutionary significance of countergradient variation. Trends Ecol Evol. 1995;10:248–52. https://doi.org/10.1016/S0169-5347(00)89081-3.

Conover D, Duffy TA, Hice LA. The covariance between genetic and environmental influences across ecological gradients reassessing the evolutionary significance of countergradient and cogradient variation. Ann N Y Acad Sci. 2009;1168:100–29. https://doi.org/10.1111/j.1749-6632.2009.04575.x.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185:1411–23. https://doi.org/10.1534/genetics.110.114819.

Csilléry K, Blum MGB, Gaggiotti OE, François O. Approximate Bayesian Computation (ABC) in practice. Trends Ecol Evol. 2010;25:410–8. https://doi.org/10.1016/j.tree.2010.04.001.

Davis MB, Shaw RG. Range shifts and adaptive responses to Quaternary climate change. Science. 2001;292:673–9.

Davis MB, Shaw RG, Etterson JR. Evolutionary responses to changing climate. Ecology. 2005;86:1704–14.

de Lafontaine G, Napier JD, Petit RJ, Hu FS. Invoking adaptation to decipher the genetic legacy of past climate change. Ecology. 2018;99:1530–46.

Denny MW, Dowd WW. Biophysics, environmental stochasticity, and the evolution of thermal safety margins in intertidal limpets. J Exp Biol. 2012;215:934–47. https://doi.org/10.1242/jeb.058958.

Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of nonmodel organisms. Heredity. 2011;107:1–15. https://doi.org/10.1038/hdy.2010.152.

Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol. 2014;29:51–63. https://doi.org/10.1016/j.tree.2013.09.008.

Ellison AR, Tunstall T, DiRenzo GV, Hughey MC, Rebollar EA, Belden LK, et al. More than skin deep: functional genomic basis for resistance to amphibian chytridiomycosis. Genome Biol Evol. 2015;7:286–98. https://doi.org/10.1093/gbe/evu285.

Ermold F. Climate change time machine. Adaptation to 30 years of warming in the Baltic Sea. PhD thesis, Uppsala University; 2016.

Forester BR, Lasky JR, Wagner HH, Urban DL. Using genotype-environment associations to identify multilocus local adaptation. Mol Ecol. 2018a;27:2215–33. https://doi.org/10.1101/129460.

Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Mol Ecol. 2018b;27:2215–33. https://doi.org/10.1111/mec.14584.

Franks SJ, Sim S, Weis AE. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc Natl Acad Sci U S A. 2007;104:1278–82. https://doi.org/10.1073/pnas.0608379104.

Franks SJ, Avise JC, Bradshaw WE, Conner JK, Etterson JR, et al. The resurrection initiative: storing ancestral genotypes to capture evolution in action. Bioscience. 2008;58:870–3. https://doi.org/10.1641/B580913.

Franks SJ, Hoffmann AA. Genetics of climate change adaptation. Annu Rev Genet. 2012;46:185–208. https://doi.org/10.1146/annurev-genet-110711-155511.

Gardiner NM, Munday PL, Nilsson GE. Counter-gradient variation in respiratory performance of coral reef fishes at elevated temperatures. PLoS One. 2010;5:e13299. https://doi.org/10.1371/journal.pone.0013299.

Gilchrist GW. Specialists and generalists in changing environments 1. Fitness landscapes of thermal sensitivity. Am Nat. 1995;146:252–70. https://doi.org/10.1086/285797.

Hendricks S, Anderson EC, Antao T, Bernatchez L, Forester BR, Garner B, et al. Recent advances in conservation and population genomics data analysis. Evol Appl. 2018;11:1197–211. https://doi.org/10.1111/eva.12659.

Henry P, Russello MA. Adaptive divergence along environmental gradients in a climate-change sensitive mammal. Ecol Evol. 2013;3:3906–17. https://doi.org/10.1111/eva.12282.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010;6:e1000862. https://doi.org/10.1371/journal.pgen.1000862.

Hung CM, Sheiner PJL, Zink RM, Liu WC, Chu TC, Huang WS, et al. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc Natl Acad Sci. 2014;111:10636–41. https://doi.org/10.1073/pnas.1401526111.

Jensen LT, Cockerell FE, Kristensen TN, Rako L, Loeschcke V, et al. Adult heat tolerance variation in *Drosophila melanogaster* is not related to Hsp70 expression. J Exp Zool Part A Ecol Genet Physiol. 2010;313A:35–44. https://doi.org/10.1002/jez.573.

Jesus TF, Grosso AR, Fonseca Almeida-Val VM, Coelho MM. Transcriptome profiling of two Iberian fresh water fish exposed to thermal stress. J Therm Biol. 2016;55:54–61. https://doi.org/10.1016/j.jtherbio.2015.11.009.

Johansson MP, Ermold F, Kristjansson BK, Laurila A. Divergence of gastropod life history in contrasting thermal environments in a geothermal lake. J Evol Biol. 2016a;29:2043–53.

Johansson MP, Quintela M, Laurila A. Genetic divergence and isolation by thermal environment in geothermal populations of an aquatic invertebrate. J Evol Biol. 2016b;29:1701–12.

Killeen J, Gougat-Barbera C, Krenek S, Kaltz O. Evolutionary rescue and local adaptation under different rates of temperature increase: a combined analysis of changes in phenotype expression and genotype frequency in *Paramecium* microcosms. Mol Ecol. 2017;26:1734–46. https://doi.org/10.1111/mec.14068.

Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.

Kozma R, Melsted P, Magnússon KP, Höglund J. Looking into the past – the reaction of three grouse species to climate change over the last million years using whole genome sequences. Mol Ecol. 2016;25:570–80. https://doi.org/10.1111/mec.13496.

Kulkarni SS, Buchholz DR. Corticosteroid signalling in frog metamorphosis. Gen Comp Endocrinol. 2014;203:225–31. https://doi.org/10.1016/j.ygcen.2014.03.036.

Lagerholm VK, Sandoval-Castellanos E, Vaniscotte A, Potapova OR, Tomek T, Bochenski ZM, et al. Range shifts or extinction? Ancient DNA and distribution modelling reveal past and future responses to climate warming in cold-adapted birds. Glob Chang Biol. 2017;23:1425–35. https://doi.org/10.1111/gcb.13522.

Lanier HC, Massatti R, Qixin H, Olson LE, Knowles LL. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). Mol Ecol. 2015;24:3688–705. https://doi.org/10.1111/mec.13270.

Laugen AT, Laurila A, Räsänen K, Merilä J. Latitudinal countergradient variation in the common frog (*Rana temporaria*) development rates – evidence for local adaptation. J Evol Biol. 2003;16:996–1005. https://doi.org/10.1046/j.1420-9101.2003.00560.x.

Lavinton E, Cogni R, Kuczinsky C, Koury S, Behrman EL, O'Brien KR, et al. A small system – high-resolution study of metabolic adaptation in the central pathway to temperature climates in *Drosophila melanogaster*. Mol Biol Evol. 2014;31:2032–41. https://doi.org/10.1093/molbev/msu146.

Lemay MA, Henry P, Lamb CT, Robson KM, Russello MA. Novel genomic resources for a climate change sensitive mammal: characterization of the American pika transcriptome. BMC Genomics. 2013;14:311. https://doi.org/10.1186/1471-2164-14-311.

Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2014;475:493–6. https://doi.org/10.1038/nature10231.

Li C, Zhang Y, Li J, Kong L, Hu H, Pan H, et al. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. GigaScience. 2014;3:27. https://doi.org/10.1186/2047-217X-3-27.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell. 2014;157:785–94. https://doi.org/10.1016/j.cell.2014.03.054.

Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015;24:1031–46. https://doi.org/10.1111/mec.13100.

Lowry DB, Hoban S, Kelley JL, Lottheros KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2016;17:142–52. https://doi.org/10.1111/1755-0998.12635.

Luikart G, Kardos M, Hand BK, Rajora OP, Aitken SN, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 3–79.

Lynch M, Gabriel W. Environmental tolerance. Am Nat. 1987;129:283–303. https://doi.org/10.1086/284635.

McCairns RJ, Smith S, Sasaki M, Bernatchez L, Behergaray LB. The adaptive potential of subtropical rainbowfish in the face of climate change: heritability and heritable plasticity for the expression of candidate genes. Evol Appl. 2016;9:531–45. https://doi.org/10.1111/eva.12363.

Merilä J, Hendry AP. Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evol Appl. 2014;7:1–14. https://doi.org/10.1111/eva.12137.

Meyer-Lucht Y, Luquet E, Johannesdottir F, Rödin-Mörch P, Quintela M, Richter-Boix A, et al. Genetic basis of amphibian larval development along a latitudinal gradient: gene diversity, selection and links with phenotypic variation in transcription factor C/EBP-1. Mol Ecol. 2019;28(11):2786–801. https://doi.org/10.1111/mec.15123.

Miermans PG. The trouble with isolation by distance. Mol Ecol. 2012;21:2839–46. https://doi.org/10.1111/j.1365-294X.2012.05578.x.

Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr Biol. 2015;25:1375–80. https://doi.org/10.1016/j.cub.2015.03.047.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 2016;25:1058–72. https://doi.org/10.1111/mec.13540.

Narum SR, Campbell NR. Transcriptomic response to heat stress among ecologically divergent populations of rainbow trout. BMC Genomics. 2015;16:103. https://doi.org/10.1186/s12864-015-1246-5.

Parmesan C. Ecological and evolutionary responses to recent climate change. Annu Rev Ecol Evol Syst. 2006;37:637–69. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100.

Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 2003;421:37–42. https://doi.org/10.1038/nature01286.

Piotrowski JS, Annis SL, Longcore JE. Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians. Mycologia. 2004;96:9–15. https://doi.org/10.2307/3761981.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013;499:471–5. https://doi.org/10.1038/nature12228.

Quintela M, Johansson MP, Kristjánsson BK, Barreiro R, Laurila A. AFLPs and mitochondrial haplotypes reveal local adaptation to extreme thermal environments in a freshwater gastropod. PLoS One. 2014;9:e101821. https://doi.org/10.1371/journal.pone.0101821.

Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA. Genomic evidence for role of inversion *3RP* of *Drosophila melanogaster* in facilitating climate change adaptation. Mol Ecol. 2015;24:2423–32. https://doi.org/10.1111/mec.13161.

Reusch TBH. Climate change in the oceans: evolutionary versus phenotypically plastic responses of marine animals and plants. Evol Appl. 2014;7:104–22. https://doi.org/10.1111/eva.12109.

Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. Adaptation to global change: a transposable element-epigenetics perspective. Trends Ecol Evol. 2016;31:514–26. https://doi.org/10.1016/j.tree.2016.03.013.

Richter-Boix A, Teplitsky C, Rogell B, Laurila A. Local selection modifies phenotypic divergence among *Rana temporaria* populations in presence of gene flow. Mol Ecol. 2010;19:716–31. https://doi.org/10.1111/j.1365-294X.2009.04502.x.

Richter-Boix A, Quintela M, Kierczak M, Franch M, Laurila A. Fine-grained adaptive divergence in an amphibian: genetic basis of phenotypic divergence and the role of nonrandom gene flow in restricting effective migration among wetlands. Mol Ecol. 2013;22:1322–40. https://doi.org/10.1111/mec.12181.

Rodriguez A, Rusciano T, Hamolton R, Holmes L, Jordran D, Wollenberg Valero KC. Genomic and phenotypic signatures of climate adaptation in an *Anolis* lizard. Ecol Evol. 2017;7:6390–403. https://doi.org/10.1002/ece3.2985.

Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:801–20. https://doi.org/10.1038/nrg3522.

Schweizer RM, Robinson J, Harrigan R, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wyane RK. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Mol Ecol. 2016;25:357–79. https://doi.org/10.1111/mec.13467.

Smith S, Bernatchez L, Beheregaray LB. RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. BMC Genomics. 2013;14:375. https://doi.org/10.1186/1471-2164-14-375.

Stillman JH, Armstrong E. Genomics are transforming our understanding of responses to climate change. Bioscience. 2015;65:237–46. https://doi.org/10.1093/biosci/biu219.

Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity. 2008;100:158–70. https://doi.org/10.1038/sj.hdy.6800937.

Strand AE, Williams LM, Oleksiak MF, Sotka EE. Can diversifying selection be distinguished from history in geographic clines: a population genomic study of killifish (*Fundulus heteroclitus*). PLoS One. 2012;7:e45138. https://doi.org/10.1371/journal.pone.0045138.

Thomas CD. Climate, climate change and range boundaries. Divers Distrib. 2010;16:488–95. https://doi.org/10.1111/j.1472-4642.2010.00642.x.

Tigano A, Schultz AJ, Edwards SJ, Robertson GJ, Friesen VL. Outlier analyses to test for local adaptation to breeding grounds in in a migratory arctic seabird. Ecol Evol. 2017;7:2370–81. https://doi.org/10.1002/ece3.2819.

Turner TL, Levine MT, Eckert ML, Begun DJ. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. Genetics. 2008;179:455–73. https://doi.org/10.1534/genetics.107.083659.

Vasemägi A. The adaptive hypothesis if clinal variation revisited: single locus clines as a result of spatially restricted gene flow. Genetics. 2006;173:2411–4. https://doi.org/10.1534/genetics.106.059881.

Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63. https://doi.org/10.1038/nrg2484.

Weir JT, Haddrath O, Robertson HA, Colbourne RM, Baker AJ. Explosive ice age diversification of kiwi. Proc Natl Acad Sci U S A. 2016;113:E5580–7. https://doi.org/10.1073/pnas.1603795113.

Weisrock DW, Hime PM, Nunziata SO, Jones KS, Murphy MO, Hotaling S, et al. Surmounting the large-genome "problem" for genomic data generation in salamanders. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG; 2018. https://doi.org/10.1007/13836_2018_36.

Xue Y, Prado-Martinez J, Sudmant P, Narashiman V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 2015;348:242–5. https://doi.org/10.1126/science.aaa3952.

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat Genet. 2013;45:67–71. https://doi.org/10.1038/ng.2994.

Zhao X, Bergland AO, Behrman EL, Gregory BD, Petrov DA, Schmidt PS. Global transcriptional profiling of diapause and climatic adaptation in *Drosophila melanogaster*. Mol Biol Evol. 2016;33(3):707–20. https://doi.org/10.1093/molbev/msv263.

# Applications of Population Genomics for Understanding and Mitigating Wildlife Disease


**Andrew Storfer, Christopher P. Kozakiewicz, Marc A. Beer, and Anna E. Savage**

**Abstract** Infectious diseases are increasingly threatening wildlife populations, domestic animal populations, and human health. Ongoing advances in genomic technologies are providing a growing set of tools for mitigating the impacts of wildlife diseases, enabling researchers to rapidly identify new pathogen threats, understand the mechanisms by which they affect populations, and predict future impacts and patterns of pathogen spread. In this chapter, we introduce a series of case studies that exemplify the broad utility of current genomic approaches for addressing the diverse challenges associated with managing wildlife diseases. The host-pathogen systems we present range from endemic to epidemic, and from those that are well-characterized to examples for which the causative agent remains unknown, necessitating a variety of genomic approaches specific to each disease system. We show how genomic approaches have been applied in each example, identifying lessons learned from each case study that can be used to guide future genomic investigations of novel disease systems. Finally, drawing upon these examples, we identify common challenges and emerging approaches through which genomics can enhance management of wildlife diseases now and into the future.

**Keywords** Chytridiomycosis · Colony collapse disorder · Emerging infectious diseases · Feline immunodeficiency virus · Host-pathogen genomics · Tasmanian devil facial tumor disease · White-nose syndrome · Wildlife disease

Andrew Storfer and Anna E. Savage contributed equally to this work.

A. Storfer (✉) · C. P. Kozakiewicz · M. A. Beer
School of Biological Sciences, Washington State University, Pullman, WA, USA
e-mail: astorfer@wsu.edu

A. E. Savage
Department of Biology, University of Central Florida, Orlando, FL, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_73, 

## 1 Introduction

Infectious diseases are increasingly appreciated as drivers of wildlife population dynamics (Daszak et al. 2000; Jones et al. 2008). The effects of infectious diseases in natural populations range from endemism, as is the case with feline immunodeficiency virus (FIV) in North American felids such as cougars and bobcats (VandeWoude and Apetrei 2006), to catastrophic declines, such as the chytrid fungus (*Batrachochytrium dendrobatidis*), in the frog species of the genus *Atelopus* (Lips et al. 2006, 2008). Due to recent increases in catastrophic disease-driven declines (Barnosky et al. 2011), emerging infectious diseases (EIDs) are now listed as the sixth most important cause of wildlife population extinctions (De Castro and Bolker 2005; Wake and Vredenburg 2008) and are thus of increasing concern for conservation and management (Smith et al. 2006; Jones et al. 2008).

EIDs are increasing globally, largely due to human-mediated changes in land use. Land use change brings together previously ecologically isolated hosts and pathogens in novel ways. Pathogens such as Nipah virus first emerged in Malaysia in 1998 as a result of a complex cascade of events that started with deforestation in SE Asia to make room for pig farms (Looi and Chua 2007). Fruit bats (Pteropodidae spp.), asymptomatic carriers of high Nipah viral loads, subsequently needed forage further, whereby they dropped partially eaten pieces of fruit into pig pens. Pigs ate the fruit, and the virus mutated in the pig and became infectious to humans with over 100 fatalities (Eaton et al. 2006). Climate change is expected to increase the geographic range of many infectious diseases, especially those vectored by mosquitoes (Lafferty 2009; Altizer et al. 2013). Wildlife trade is also a major contributor to infectious disease spread (Fèvre et al. 2006) because it facilitates global pathogen transport and disease emergence in novel, susceptible host species.

The rapid spread and emergence of new pathogens, cross-species transmission, and re-emergence of existing endemic pathogens documented in recent decades highlight the massive and increasing threat that diseases pose to wildlife populations. Taken together, these data demonstrate that rapid action is necessary to identify new pathogen threats, forecast their spread, and mitigate their effects. With the recent advances in genomics technologies, we have an increased ability to quickly identify novel pathogenic organisms, assess heritable host features that may facilitate or hinder extinctions, and predict the locations, environmental drivers, and extent of new epizootics. Efficient application of genomic analyses can address a variety of questions common to wildlife diseases and lead to effective mitigation and management of disease impacts on wildlife populations and species (Table 1). For example, whole-genome sequencing can help researchers quickly identify a new pathogen and reconstruct its phylogenetic relationships to other pathogens to help classify its origin and the extent of the threat. Comparative genomic studies of different pathogen strains can permit tracking of evolutionary dynamics and possible coevolution with particular host populations. Although primarily applied to human viruses, phylodynamic studies can help to elucidate specific mutations associated with enhanced pathogenicity and increased geographic range, such as amino acid

**Table 1** Examples of common questions in wildlife disease that have been addressed using genomic approaches

| Question | Molecular approach | Example | Reference |
|---|---|---|---|
| What is the disease origin? | Whole-genome sequencing | Ebola in West Africa originated from a single transmission event from an animal reservoir in Central Africa | Gire et al. (2014) |
| What is the causative agent? | Metagenomics | Densovirus is associated with sea star wasting disease | Hewson et al. (2014) |
| What factors influence host susceptibility/ resistance? | Genome-wide association study (GWAS) | Loci involved in cell adhesion and cell cycle regulation associated with DFTD susceptibility and survival in female Tasmanian devils | Margres et al. (2018a) |
| How do hosts respond physiologically to infection? | Comparative transcriptomics | Immune response does not determine tolerance of bats to *Pd* infection | Davy et al. (2017) |
| How has the pathogen spread spatially? | Phylogenetics/ phylogenomics | Large-scale patterns of avian influenza spread are driven by long-distance migration of wild waterfowl | Trovão et al. (2015) |
| Is the pathogen influencing host population structure? | Population and landscape genetics | Varying risk of infection by malaria species produces population structure in great tits | Garroway et al. (2013) |
| Is the pathogen exerting selective pressure on host populations? | Population and landscape genomics | DFTD is driving strong selection in devils that swamps out adaptation to local environmental factors | Fraik et al. (2019) |
| What factors led to disease emergence/ virulence? | Comparative genomics/ phylogenomics | Recent duplication of a nitrogen transporter gene in *Pd* may have led to more efficient nitrogen uptake and enhanced growth in cave environments associated with bats | Reynolds et al. (2016) |
| Is the pathogen evolving higher or lower virulence over time? | Time-series GWAS/ transcriptomics | CMB18 virulence genes upregulated in epizootic vs endemic *Bd* lineages based on global gene expression profiles | McDonald et al. (2020) |
| Is the host evolving disease resistance or tolerance? | Comparative population genomics/ transcriptomics | Skin integrity genes upregulated in *Bd*-tolerant species and T-cell genes downregulated in *Bd*-susceptible species based on comparative transcriptomics of the skin and spleen | Ellison et al. (2014) |

substitutions found to make the mosquito gut more competent for retaining Zika virus (Tham et al. 2018). Landscape genetic studies can help identify conduits or barriers to pathogen dispersal, such as rivers serving as barriers to chronic wasting disease in deer (Blanchong et al. 2008).

Here, we present a diverse range of case studies to illustrate the utility of genomics for understanding both endemic and emerging host-pathogen systems. The best approaches will depend on the nature of the study system and the extent to which the pathogen has been identified and characterized. Nonetheless, we show that, regardless of the amount of background information known, genomic approaches to studying wildlife disease have great utility for elucidating threats, improving management strategies, and enhancing wildlife conservation.

## 2 Case Studies

Wildlife disease systems are diverse and idiosyncratic. Consequently, the optimal genomic tools for understanding factors leading to disease emergence and predicting population-level consequences vary with the species, populations, and ecosystems involved. One particularly important consideration is the degree to which the causative agents behind a disease outbreak are known. For well-established host-pathogen systems with defined habitat characteristics linked to epidemics, narrower and deeper genomic investigations may be most appropriate and informative. In contrast, when widespread morbidity and mortality are detected in wildlife populations but the causative agents are mysterious, genomic investigations must necessarily begin from a broader and shallower scale of inquiry focused on identifying the pathogen. Due to the system-specific nature of how genomic tools are best applied to understanding disease, we present a series of case studies, ranging from systems with completely unknown etiological agents to systems with well-studied disease drivers to systems with established epidemiological and genomic frameworks. We highlight genomic approaches used for each scenario, lessons learned from genomic studies, and how genomic approaches have been applied for understanding and mitigating disease impacts.

### *2.1 Colony Collapse Disorder*

The honeybee (*Apis mellifera*) has experienced mass disappearance of workers from hives dating back to at least 1869 (Underwood and Vanengelsdorp 2007). This phenomenon was termed colony collapse disorder (CCD) in 2006 after a sharp, widespread increase in worker bee disappearances throughout North America that has expanded into Europe and continues to the present (Ellis et al. 2010). Diagnosis of CCD depends on specific criteria, including rapid loss of workers but presence of the queen and larvae, no pest invasion until after workers disappear, and a lack of obvious, known bacterial and mite bee pathogens (Evans et al. 2009). The *A. mellifera* genome was published in 2006 (Honeybee Genome Sequencing Consortium 2006), enabling comparative genome-scale investigations of honeybee responses to be linked back to functional and structural elements of the genome.

However, the causative biotic and/or abiotic drivers of CCD were largely unknown at the time of initial widespread occurrence, limiting the ability for host-centric investigations to be targeted at specific pathogens or environmental exposures. The first revelatory omics-focused analysis to identify correlates of CCD used RNAseq to characterize functional bacterial, fungal, and viral elements of the microbiome sampled from CCD bees, non-CCD bees, and the honeybee nutritive secretion royal jelly (Cox-Foster et al. 2007). The only pathogen showing a significant correlation with CCD was Israeli acute paralysis virus, making this organism the first candidate causative agent identified. In contrast, the three most common honeybee pathogens known prior to CCD emergence, the mites *Acarapis woodi* and *Varroa destructor* and fungus *Nosema* spp., were equally prevalent in CCD and non-CCD samples. This study highlights the utility of broad metagenomic surveys to objectively identify candidate pathogens, particularly those targeting expressed microbial transcripts to focus on functional elements of the microbiome, when disease agents are entirely unknown. Similar approaches were utilized to identify a Densovirus as the likely cause of sea star wasting disease (Hewson et al. 2014) and to identify several pathogens associated with coral bleaching disease (e.g., Vega Thurber et al. 2008; Webster et al. 2016), although in the latter case consistent pathogen links across bleaching events remain elusive.

Other lines of inquiry into the causes of CCD focused on previously known parasites and pathogens, particularly the ectoparasite *V. destructor* which is the most detrimental honeybee pest worldwide (Cornman et al. 2010). While a number of field-based and experimental studies identified *V. destructor*-vectored viruses such as deformed wing virus as significantly associated with CCD (e.g., Yang and Cox-Foster 2007; Rosenkranz et al. 2010; Martin et al. 2012; Ryabov et al. 2014), these analyses used only targeted PCR and qPCR approaches to characterize previously known mite-vectored pathogens. The lack of genomic surveillance approaches therefore eliminated the ability to present an unbiased picture of the microbial patterns associated with CCD and/or *V. destructor* parasitism.

Lack of a definitive or single pathogen linked with CCD continues to present a challenge for understanding the role of infectious agents in driving honeybee losses. However, *A. mellifera* is the only host species involved in CCD, simplifying studies of host responses to CCD in comparison to multi-host disease systems where species differences often present confounding variables (e.g., chytridiomycosis; see below). Early functional studies sequencing a handful of immune genes in *A. mellifera* sampled from *V. destructor* colonies compared to healthy colonies identified immunosuppression as a major cofactor of mite infection (e.g., Yang and Cox-Foster 2005). More recently, experimental infection (Di Prisco et al. 2016) and RNAseq analysis of mite-infested compared to mite-controlled colonies (Nazzi et al. 2012) validated viral immunosuppression of honeybees exposed to *V. destructor*-vectored deformed wing virus. In contrast, whole-genome microarrays comparing global gut gene expression from CCD compared to healthy colonies (regardless of *V. destructor* pressure) across a broad population genetic and geographic sampling range identified 65 differentially expressed transcripts as markers for CCD (Johnson et al. 2009). Remarkably, genes involved in immunity and pesticide response were

not differentially expressed in CCD afflicted compared to healthy individuals. Instead, ribosomal RNA fragments were significantly elevated in the gut of CCD bees, which is indicative of viral infections. These studies highlight the utility of honeybee functional genomics for comprehensively evaluating host responses, but they also illustrate the challenge of constructing an appropriate host genomic experimental design when the disease drivers remain elusive. We now have a clear understanding that *V. destructor*-vectored viruses immunosuppress honeybees, but we lack a definitive connection between *V. destructor*, viral pathogens, and CCD, and the only host genomic study to look directly at CCD versus control colonies found no evidence of immunosuppression.

**Lessons Learned** CCD is a disease system without any major genomic limitations; the single host species has a well-annotated reference genome, the genome size is tractable, and large sample sizes can be readily obtained. However, CCD is in some ways a cautionary tale against embarking upon large-scale, expensive genomic studies that assume certain causal factors are important and thus identify responses to the putative causal factors rather than the disease itself. Honeybee genomic resources continue to advance, with techniques available to investigate methylation differences for epigenetics (Welsh et al. 2017) and validation that CRISPR/Cas9 is an effective genome editing technique (Kohno et al. 2016). Twelve bee species genomes are also underway or complete (Trapp et al. 2017), enabling the use of comparative genomics techniques to understand potential honeybee predispositions for colony collapse, which could be a consequence of artificial selection altering the honeybee genome (Parejo et al. 2017). For CCD and other similarly complex and multifactorial disease systems, we recommend more extensive and exploratory genome-wide and community-wide approaches using these emerging techniques and established metagenomics and metatranscriptomics pipelines until consistent, unbiased patterns are uncovered.

## *2.2 Chytridiomycosis*

Among the most destructive and widespread emergent infectious wildlife diseases is amphibian chytridiomycosis (Skerratt et al. 2007). Caused by the fungus *Batrachochytrium dendrobatidis* (*Bd*; Longcore et al. 1999) and the salamander-specific congener, *B. salamandrivorans* (*Bsal*; Martel et al. 2013), chytridiomycosis is a skin-specific fungal disease that decimated amphibian populations for decades before *Batrachochytrium* was identified as a genus within the Chytridiomycota and as the definitive cause in 1998 (Berger et al. 1998). Chytridiomycosis thus represents a disease system where the hosts, pathogens, and numerous disease correlates were well-resolved by the time population genomic approaches first became feasible in non-model wildlife systems. Major findings from the first decade of chytridiomycosis research included identifying strong correlations between disease risk and environmental and ecological variables, rather than a pattern of elevated

susceptibility in particular amphibian families (Lips et al. 2006; Pounds et al. 2006; Wake and Vredenburg 2008). However, individual- and population-level studies found significant associations between *Bd* susceptibility and immune system genotypes, including class II MHC alleles (Savage and Zamudio 2011) and the diversity of skin antimicrobial peptides (Woodhams et al. 2007). Early attempts to resolve the phylogenetic diversity of *Bd* and detect a global region of origin were hindered by the small number of loci available and apparently low genetic variation among worldwide strains, with no evidence of one specific region of origin showing higher genetic diversity (Morehouse et al. 2003; Morgan et al. 2007).

The Broad Institute and the Joint Genome Institute sequenced the genome of two pathogenic *Bd* lineages in the mid-2000s (Fisher 2008), providing the first genome-wide insights into why, among thousands of saprobic and insect-pathogenic chytrid species, this single genus was causing virulent disease in amphibians. *Bd* has a dynamic genome, with extensive chromosomal copy number variation among strains and widespread loss of heterozygosity, potentially due to mitotic recombination (Farrer et al. 2011; Rosenblum et al. 2013). Whole-genome comparisons between *Bd*, *Bsal*, and two close saprophytic relatives further revealed that amphibian pathogenicity of *Batrachochytrium* is characterized by large expansions of cell wall and protease gene families (Farrer et al. 2017). Genome-wide expression profiling between zoospore and mature sporangia life stages using a custom *Bd* microarray revealed upregulation of fungalysin metallopeptidase genes (Rosenblum et al. 2008), which are linked to pathogenicity in distinct human fungal pathogens and thereby suggest convergence of fungal pathogenicity factors. More recently, RNAseq experiments using single-cell laser microdissection to compare gene expression in culture compared to live amphibian infections (Ellison et al. 2017) found major expression differences in vivo compared to in vitro. Most notably, in vivo *Bd* cells significantly upregulated expression of genes involved in proteolysis, including serine-type peptidases and metallocarboxypeptidases, which are both implicated in chytridiomycosis virulence (Rosenblum et al. 2012, 2013; Farrer et al. 2017). Significant expression changes also occurred in *Bd* cells when infecting different host species, including enriched expression of flagella-related genes in the supershedder host *Atelopus zeteki* (DiRenzo et al. 2014). Furthermore, in vitro RNAseq-based expression profiles among different *Bd* genetic lineages also found significant expression differences, with enzootic strains significantly upregulating peptidase genes compared to epizootic lineages (McDonald et al. 2020). Collectively, these recent transcriptome-based analyses highlight the potential for specific *Bd*-host combinations to produce dramatically different replication and transmission outcomes.

Despite the major insights into *Bd* genome structure and function gained from whole-genome studies, these approaches have not been able to resolve the question of when, where, and why *Bd* emerged as a virulent amphibian pathogen. Two competing ideas for the emergence of chytridiomycosis are the novel pathogen hypothesis (NPH) and the endemic pathogen hypothesis (EPH). The NPH posits that *Bd* is virulent because it has recently been spread around the world and introduced to naive amphibian populations, whereas the EPH suggests that *Bd* is

endemic worldwide and has emerged as a virulent pathogen due to environmental alterations such as climate change (Rachowicz et al. 2005). *Bd* was found to harbor extremely low genetic variability based on initial population genetic studies utilizing 10–15 loci and sampling only epizootic lineages (Morehouse et al. 2003; Morgan et al. 2007). In contrast, population genomic and phylogenomic approaches coupled with additional *Bd* strains being cultured from more species and regions have revealed deeply divergent genetic lineages from multiple global regions. Based on whole-genome analyses, *Bd* does include a global pandemic lineage (GPL) with low genetic diversity that is associated with major epizootics in Australia, North America, and Central America (Farrer et al. 2011). Additionally, some genomic lineages appear to have a hybrid origin (Schloegel et al. 2012), suggesting potential links between strain recombination and virulence.

Three other relatively distinct genomic lineages were recovered from South Africa, Switzerland, and Brazil, with the latter lineage representing the basal *Bd* clade (Rosenblum et al. 2013). With the advent of new culturing techniques, additional *Bd* genomes from South Korea were sequenced and identified as belonging to a more deeply diverged clade than anything found outside of Asia (O'Hanlon et al. 2018). Most recently, a method to genotype hundreds of loci from amphibian skin swab samples has identified an additional, basal *Bd* clade from Asia (Byrne et al. 2019; Fig. 1), reinforcing this continent as the likely place of origin for *Bd*. While these genomic studies offer considerable advances in our understanding of the global diversification of *Bd*, the totality of evidence to date includes numerous examples of endemic strains and of novel introductions and thus does not resolve whether the NPH or EPH better explains the emergence of chytridiomycosis. Ultimately, the power of genomic studies for understanding amphibian chytridiomycosis remains limited by the scale of sampling necessary, across global habitats and within thousands of host species, to create a truly global picture of *Bd* genomic variation.

Studies of host genomic susceptibility to chytridiomycosis have largely been restricted to transcriptome analyses due to the ongoing lack of robust genomic resources for amphibians. Amphibian disease genomics is particularly challenging compared to other vertebrate groups for two reasons: (1) amphibian genomes are among the largest of all organisms and larger than any other tetrapods, presenting a genome assembly challenge due to extensive repetitive and low-complexity regions, and (2) a taxonomically wide number of amphibians are susceptible to *Batrachochytrium* infection, meaning the number of host genomes to investigate is a logistical nightmare. Despite these limitations, transcriptome profiling of host responses to experimental *Bd* exposure identified some notable patterns. The first tissue-wide expression studies used microarray technologies available for the model frog *Silurana (Xenopus) tropicalis* and found lack of acquired immune responses and minimal, temperature-dependent innate immunity (Ribas et al. 2009; Rosenblum et al. 2009). Subsequent RNAseq experiments in non-model frog hosts instead identify species-specific patterns, with some experimental exposures resulting in minimal immune responses (Rosenblum et al. 2012) and others producing significant upregulation of acquired immune genes, particularly in less susceptible species

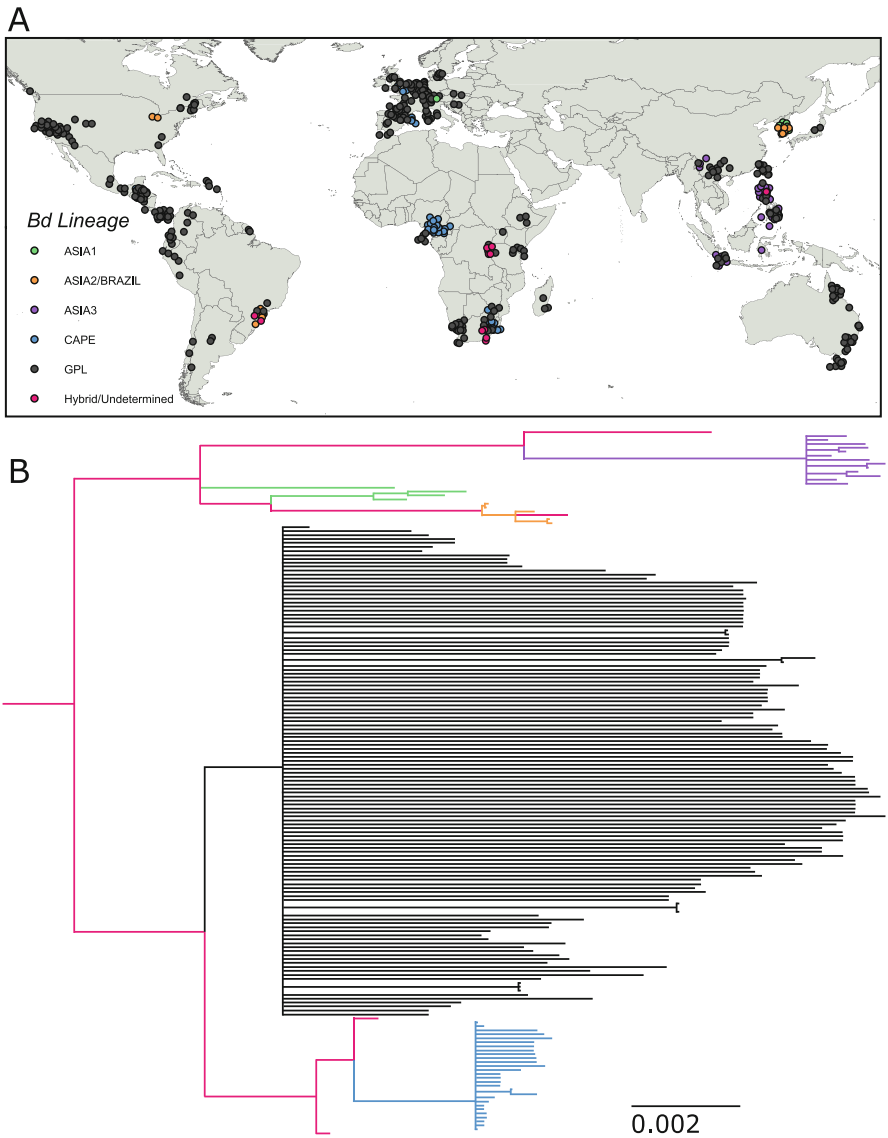


**Fig. 1** Genome-wide genotyping of *Bd* lineages has revealed multiple endemic strains and a likely Asian origin. (**a**) Global distribution of *Bd* genotypes colored by major phylogenetic lineage identifies three Asian lineages (one of which is also found in Brazil), an African lineage (that also occurs in Europe), and the global pandemic lineage (GPL). (**b**) Maximum-likelihood phylogeny based on 172 nuclear loci indicates that Asian *Bd* lineages (green, orange, purple) diverged earlier than African (blue) and GPL (black) lineages. Figure modified from Byrne et al. (2019)

(Ellison et al. 2015) and in individuals with prior pathogen exposure (Ellison et al. 2014). One of the few consistent findings from each of these transcriptome profiling studies is an increase in skin integrity gene expression when frogs become infected with *Bd*. Nevertheless, functional genomic studies to date largely serve to highlight extensive variation in host responses among individuals and across species, and pinpointing the source of this variation will require gene expression profiling across more time points, more individuals, and more amphibian families and genera.

**Lessons Learned** Genomic approaches have been powerful for elucidating the importance of pathogen structural genomic variation and gene expression variation, rather than nucleotide polymorphism, in driving *Batrachochytrium* pathogenicity. Functional genomic analyses of host responses have also been particularly valuable for revealing unbiased patterns of tissue-specific gene expression changes linked to individual- and species-level susceptibility. However, a better understanding of chytridiomycosis genomics remains hindered by our limited ability to isolate and sequence diverse pathogen strains, by the large genome size of most amphibians, and by the number of distinct host species that are susceptible to *Bd* and *Bsal* infection. Chytridiomycosis research has a problem of scale; not only is each host genome large and therefore expensive and complex to sequence robustly, but host genomic material cannot be easily isolated from pathogen genomic material, and there are thousands of host species' genomes to sequence before we can understand the relative roles of idiosyncratic and species-specific genomic variants versus conserved immune or skin sloughing pathways in driving susceptibility. For similar disease systems where isolation of the pathogen is nontrivial and/or the pathogen is an ecological generalist, we recommend an emphasis on sequence capture methods, single-cell pathogen isolation techniques such as microdissection, and on generating several robust reference genomes (e.g., by combining long-read and short-read technologies to produce large, low error scaffolds) covering the major taxonomic host lineages. Additionally, for wildlife disease systems where host genomes can be prohibitively large (i.e., amphibians, lungfishes, and some crustaceans), we recommend using reduced representation library approaches such as RADseq (Andrews et al. 2016) to best enable host population genomic analyses. We predict that advances in bioinformatics algorithms for metagenomic and metatranscriptomic datasets will improve genomic assessments of complex disease systems by better enabling shotgun sequencing of mixed host and pathogen tissue samples to be reconstructed into constituent individual genomes.

## *2.3 White-Nose Syndrome*

North American hibernating bats were first detected with a white substance on their muzzle while experiencing a mass mortality event in early 2007 (Blehert et al. 2009). The white substance was culturable on bat skin at cool temperatures and was rapidly identified as the ascomycete fungus *Geomyces destructans* (Chaturvedi et al. 2010).

Experimental exposures and fulfillment of Koch's postulates (Lorch et al. 2011) subsequently confirmed *G. destructans* was the sole causative agent of this virulent disease we now call white-nose syndrome (WNS). A striking number of parallels between WNS and chytridiomycosis are evident. Both diseases are caused by cutaneous fungal pathogens, have emerged in recent decades, and can lead to high rates of mortality in hosts during cool months when body temperature is determined externally rather than intrinsically (bats do not thermoregulate while hibernating), and both diseases affect a wide range of host species and show a clear pattern of epidemiological spread in some regions of the world. However, it is the differences between WNS and chytridiomycosis rather than the similarities that best explain variation in the success of applying genomic approaches to understanding these diseases. Here, we present major genomic advances in WNS and chytridiomycosis research in direct comparison in order to highlight the intrinsic differences that can lead to alternate success in applying the same methodologies.

Phylogenomic and population genomic analyses can be extremely powerful tools for identifying the source and spread of a pathogen, but only when clear genetic, spatial, and temporal signals of origin and subsequent spread can be detected. After identifying *G. destructans* as the cause of WNS, this newly described species was the subject of considerable molecular phylogenetic investigation. Consequences of these investigations include a taxonomy revision (*G. destructans* is now *Pseudogymnoascus destructans* [*Pd*]) and a clear point of origin and spread to North America coinciding with the emergence of WNS. *Pd* was discovered to be common, nonlethal, and naturally occurring in European bats (Puechmaille et al. 2011), and phylogenomics of ~200 Kb of the *Pd* genome isolated from bats in eight WNS die offs found that *Pd* was clonal, showing no genetic variation (Rajkumar et al. 2011). Several additional genetic and genomic analyses of *Pd* support this pattern. *Pd* has no close genetic relatives (Minnis and Lindner 2013) but is widespread throughout Europe with a single point of introduction to North America (Leopardi et al. 2015). Limited genetic variation has now been detected via genome-wide polymorphisms coinciding with the westward spread throughout North America, but no evidence exists for other introduced strains or for recombination among *Pd* lineages (Trivedi et al. 2017). In contrast, much more extensive genetic and genomic analyses of *Bd* lineages have not identified any point of origin, with clonal low diversity lineages throughout the world as well as divergent genotypes detected in Asia and Brazil and a distinct *Batrachochytrium* species identified in Europe (see above for further details). Thus, *Pd* genomic studies have been more informative for defining the epidemiology of WNS not because analyses have been more sophisticated or extensive but because they have revealed a simpler pattern of introduction and spread. Furthermore, *Pd* is easier to culture than *Bd*, providing better confidence that all lineages can be recovered and included in genomic analyses.

*Pd* functional genomic mechanisms of virulence have been elucidated via comparative phylogenomics and transcriptome-based gene expression studies. Similar to *Bd*, *Pd* has undergone expansion of some gene families that may explain higher virulence than other closely related fungi. Specifically, genome scans of 21 *Pseudogymnoascus* species revealed a recent duplication of a high-affinity nitrate

transporter that was only present in *Pd* (Reynolds et al. 2016). Also paralleling *Bd* dynamics, *Pd* in culture shows significantly different global gene expression patterns compared to *Pd* causing WNS, particularly in cell wall and micronutrient acquisition genes (Reeder et al. 2017). *Pd* and *Bd* functional genomic studies share benefits and challenges, both linked to their inherent properties as skin-infecting fungal organisms. Simultaneous RNAseq analysis of pathogen and host genes is simplified for fungi and other eukaryotic pathogens because they polyadenylate messenger RNA just like their hosts, enabling the use of standard poly-A capture protocols for capturing host and pathogen transcripts from a mixed sample. However, an ongoing limitation for studying *Pd* and *Bd* lineages causing active infections is the low abundance of pathogen transcripts compared to host transcripts, which is typically less than 5% (Enguita et al. 2016). Advances in *Bd* functional genomics have arisen from microdissection of pathogens from host tissue followed by single-cell RNAseq. This and other approaches to isolate or enrich pathogen genomes are likely to provide the next major advances in our understanding of *Pd* virulence mechanisms.

Host-centric analyses of the functional genomic response to *Pd* and *Bd* have both emphasized controlled experimental exposures followed by RNAseq-based transcriptomics. Gene expression changes in *Pd*-exposed North American bats demonstrate strikingly similar responses compared to frogs exposed to *Bd*, in that innate and inflammatory genes are upregulated but acquired and cell-mediated immune genes show no expression changes at the site of infection. Also similar is the lack of immune response in non-susceptible hosts; just as *Bd*-tolerant amphibians are nonresponsive to the presence of *Bd*, European bats exposed to *Pd* showed no change in immune gene expression compared to unexposed bats (Davy et al. 2017), highlighting that tolerance does not arise from any initiation of immune function. Amphibians and bats are both diverse taxa, providing many potential host species and therefore numerous possible host responses. This "problem of scale" presents a challenge for comprehensively understanding how idiosyncratic each host-pathogen combination might be if we were able to systematically expose every potential host under controlled conditions. Considering the individual- and population-specific responses to *Bd* within single amphibian host species (e.g., Savage and Zamudio 2016), this problem increases by orders of magnitude. However, this is one area in which WNS is a more tractable system to comprehend. *Bd* infects thousands of amphibian hosts on six continents, whereas WNS is currently restricted to North America which houses fewer than 50 bat species, meaning that comprehensive genomics of each host species is within the realm of feasibility. Furthermore, bat genomes are substantially smaller than amphibian genomes, ranging from 1.6 to 3.5 Gb (mean 2.4) compared to 1.0–117.9 Gb (mean 18.5) in amphibians. This sets a much lower threshold for robust sequencing and assembly of entire WNS host genomes using current sequencing technologies, which can still be excessively costly and present problems for assembling low-complexity regions of large genomes.

**Lessons Learned** Despite widespread similarities between WNS and chytridiomycosis, genomics has better resolved WNS biology because the spatial,

host, and pathogen components of the study system have been narrower, more straightforward, and therefore easier to elucidate. *Pd* can be identified, isolated, and studied more tractably compared to *Bd*; pathogen genetic diversity is lower, and global distribution is smaller for WNS compared to chytridiomycosis; host breadth and host genome size are both substantially smaller, and a more rapid response time from first observation of mass mortality and a better genomics toolkit at the outset enabled more rapid collection and analysis of samples critical for understanding WNS.

## 2.4 *Feline Immunodeficiency Virus*

Lentiviruses have been identified in a growing number of mammalian taxa, with the most well-known example being human immunodeficiency virus (HIV). Feline immunodeficiency virus (FIV) is endemic in many natural wild felid populations, and although it causes autoimmune disorder in domestic cats, it has little overt pathogenicity in its natural hosts (Olmsted et al. 1992; Troyer et al. 2005; VandeWoude and Apetrei 2006). Thus, FIV is not considered to be of conservation concern, unlike the other case studies in this chapter. However, it is one of the most widely studied wildlife infectious agents due to its utility as a model for understanding the dynamics of directly transmitted wildlife diseases using genetic and genomic approaches.

FIV was initially detected in domestic cats (Pedersen et al. 1987), with discovery in both captive and free-ranging non-domestic felids occurring soon after (Barr et al. 1989). Nine distinct, species-specific lineages of FIV have since been identified in Africa, Eurasia, and the Americas and are thought to have diverged from a common ancestor during felid speciation (Troyer et al. 2005; Pecon-Slattery et al. 2008). Both inter- and intraspecies divergence in even relatively conserved genomic regions are high (up to 30% and 20%, respectively; Troyer et al. 2005).

The small FIV genome (~9 Kb; Langley et al. 1994) and high mutation rate ($\sim 1.3 \times 10^{-3}$ substitutions per site, per year; Krakoff et al. 2019) enable detection of fine-scale phylogenetic relationships among FIV isolates with minimal sequencing effort. Early phylogenetic studies demonstrated that within species, not only does FIV phylogenetic structure occur at broad geographic scales (Olmsted et al. 1992; Carpenter et al. 1996; Antunes et al. 2008), but phylogeographic clustering of FIV also occurs among geographically proximal hosts within populations (Biek et al. 2003).

FIV forms chronic, lifelong infections, enabling repeated sampling post-infection (VandeWoude and Apetrei 2006). Accordingly, FIV was quickly identified as a model for studying recent transmission dynamics within and among felid populations, as well as providing an alternative means for understanding host population dynamics. In North America, phylogenetic studies of FIV have been used to elucidate recent demographic history and population structure in mountain lions (Biek et al. 2006), measure the extent to which geographic barriers limit bobcat

movements (Lee et al. 2012), and understand how urbanization shapes patterns of transmission and pathogen evolution (Fountain-Jones et al. 2017a). In African lions, FIV phylogenetic studies have helped elucidate complex inter-pride interactions (Kerr et al. 2018) and identified differential effects of host social and spatial organization on disease transmission among different FIV subtypes (Fountain-Jones et al. 2017b).

The long coevolutionary history of FIV and felids leading to the evolution of species-specific strains, the relative apathogenicity of FIV strains on their respective host species, and evidence from studies of small portions of the genome suggested that FIV strains and their native hosts have reached an evolutionary equilibrium (Biek et al. 2003; Franklin et al. 2007; Pecon-Slattery et al. 2008). However, whole-genome study of FIV strains from bobcats and mountain lions has since found high levels of recombination and positive selection, suggesting ongoing processes of host-virus adaptation (Lee et al. 2014). Cross-species transmission of bobcat-specific FIV from bobcat to mountain lion has been observed (Franklin et al. 2007; VandeWoude et al. 2010; Lee et al. 2017; Fig. 2) and is thought to be enabled by homology among bobcats and mountain lions in the APOBEC antiviral immune resistance factor (Mariani et al. 2003; Zielonka et al. 2010; Lee et al. 2014). In successful infections, APOBEC is overcome by the lentiviral Vif protein (Mariani et al. 2003). Recent work with sympatric bobcat and mountain lion populations has investigated how selection in the *vif* gene as well as host population genetic and ecological factors may influence the likelihood of cross-species transmission and establishment of intraspecific transmission within the new host (Lee et al. 2017).

The role of genetic/genomic and ecological factors in determining viral fitness, particularly in the context of pathogen spillover, will likely be a continuing focus of FIV research. Almost all our genomic knowledge of FIV in non-domestic cats has been gleaned without the use of high-throughput sequencing due to the minimal sequencing effort required to characterize FIV. However, traditional methods of molecular cloning and Sanger sequencing have limited capacity to characterize within-host FIV genetic variation and are biased toward detection of the most abundant variant within a given individual host. The ability to relatively quantify and sequence divergent viruses within single hosts will enhance our capacity to identify factors influencing the fitness of specific FIV genetic variants. Further, this will provide greater resolution to phylogenetic reconstructions of transmission within host populations, enabling multiple transmission events involving different variants to be detected within single hosts. Increasing genomic resources of host species will also benefit further study of host-virus coevolution, with the imminent publication of at least a mountain lion reference genome (Saremi et al. 2018; Ochoa et al. 2019). No detailed assessment of host genomic factors influencing FIV evolution in wild felids has thus far been undertaken.

**Lessons Learned** The FIV system demonstrates that genetic and genomic studies of apathogenic infections can generate a multitude of insights into host and parasite ecological and evolutionary processes. The HIV/AIDS pandemic motivated a substantive effort to understand lentiviral function and evolution that has provided a

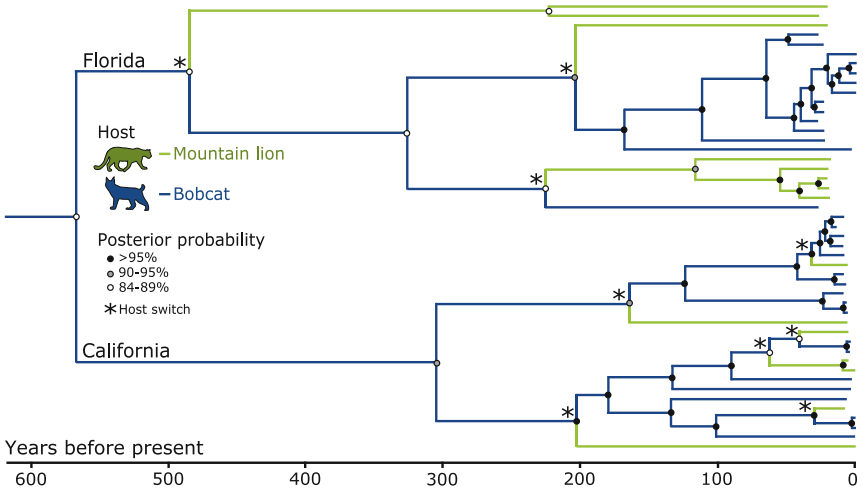


**Fig. 2** Phylogenetic approaches have revealed patterns of historic and contemporary cross-species transmission in feline immunodeficiency virus (FIV). This maximum clade credibility tree was generated using FIV *pol* gene sequences collected from bobcat and mountain lion populations in California and Florida. Ancestral host states have been reconstructed across the phylogeny, with asterisks indicating predicted cross-species transmission events. Figure modified from Lee et al. (2017)

strong genomic platform upon which studies of FIV in wildlife have built. This model system has enjoyed few genomic limitations due to the extremely small size of the FIV genome but will continue to benefit from increasing host genomic resources and genomic approaches for characterizing intra-host FIV diversity.

## *2.5 Tasmanian Devil Facial Tumor Disease*

In 1996, it was noted that increasing numbers of Tasmanian devils (*Sarcophilus harrisii*) were being found with strange facial tumors. Soon after, the once localized tumors began to be detected across a larger geographic area, and devils that had visible, ulcerated tumors almost invariably died. A combination of karyotyping and DNA sequencing showed that the tumors were a clonal, transmissible lineage (Pearse and Swift 2006; Murchison et al. 2010). Comparative transcriptomics of tumor tissue and a variety of devil tissues and expression of periaxin suggested that this infectious cell line originated from a mutated Schwann cell in a female devil (Murchison et al. 2010). The transmissible cancer, named devil facial tumor disease (DFTD) has since spread about 98% of the way across Tasmania, causing over 80% localized population declines and a species' wide decline of approximately 80% (McCallum et al. 2009; Storfer et al. 2018a). A second facial tumor disease (i.e., DFT2) was discovered in 2014 and is also determined to be of Schwann cell origin,

but this tumor arrived from a male and thus far remains more localized than DFTD in its geographic distribution (Pye et al. 2016; James et al. 2019). The emergence of a second transmissible cancer within 20 years of the initial discovery of DFTD begs the question as to whether such cancers are part of the devil's evolutionary history. Nonetheless, Tasmanian devils are now listed as endangered, with DFTD considered the primary cause and paramount management concern (Storfer et al. 2018a; Patchett and Woods 2019; Hohenlohe et al. 2019).

The extensive mark-recapture database of over 28,000 devil records combined with over 10,000 devil tissue samples and several thousand tumor samples make the Tasmanian devil-DFTD system somewhat unparalleled for wildlife disease studies. Further, genomic resources range from reference genomes and transcriptomes to several tumor isolates maintained in cell culture, providing researchers with ample genomic tools to study evolutionary and population dynamics of both the pathogen and its host.

In terms of the tumor, further karyotypic studies have provided more information about the likely origin of DFTD and DFT2. Both tumors share fragmentation of chromosome 1, which both changes gene order on this chromosome and results in formation of double minutes (Deakin et al. 2012; Taylor et al. 2017; Fig. 3). Genomic approaches helped solve another mystery – why devils are apparently universally susceptible to DFTD, which is an allograft. DFTD does not express MHC class I $\beta_2$m and epigenetically downregulates $\beta_2$m in neighboring host cells, essentially making tumor cells invisible to the devil host (Siddle et al. 2013). Comparative genomic sequencing of different DFTD isolates shows somewhere between 15,000 and 17,000 SNPs, representing a mutation rate higher than most human cancers but slower than that of lung cancer or melanomas (Murchison et al. 2012).

Recent drug screens have been used to test the susceptibility of DFTD and DFT2 to possible treatments. Both tumor types are sensitive to receptor tyrosine kinases, with DFTD being particularly sensitive to afatinib, likely mediated by ERBB2 inhibition (Stammnitz et al. 2018). DFT2 is sensitive to axitinib, which has activity against PDGFR, KIT, and VEGFR (vascular endothelial growth factor; Stammnitz et al. 2018). Additionally, both DFTD and DFT2 show sensitivity to dasatinib, which inhibits PDGFR, ABL, SRC, and KIT. Further testing suggests that the ERBB-STAT3 axis is perhaps more important in determining the pathogenicity of DFTD (Kosack et al. 2019). A combination of drug screens showed that the activity of ERBB2 and ERBB3 was most commonly inhibited by various drug candidates (kinase inhibitors) and that blocking ERBB expression leads to restoration of MHC Class I gene expression in DFTD (Kosack et al. 2019). As a result, it is now hypothesized that evolution of DFTD transmissibility progressed in the following stages: (1) a Schwann cell became tumorous with MHC expression and low concentrations of ERBB3; (2) some mutants of this initial tumor later had high ERBB3 expression and consequent downregulation of MHC I; and (3) these tumors evaded the immune responses of devils and were able to be spread by biting (Kosack et al. 2019; Patchett and Woods 2019). Treatment of DFTD by IFN-γ can potentially

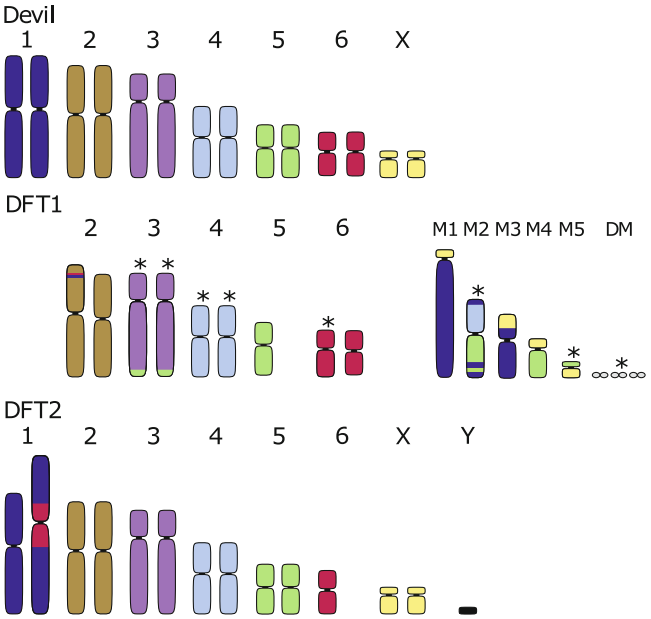


**Fig. 3** Normal female Tasmanian devil karyotype (top) with a representative karyotype of DFTD (aka "DFT1"; middle) and DFT2 (bottom). Colors in rearranged/fragmented chromosomes show the chromosome of origin in the normal devil. M1–M5 indicate marker chromosomes resulting from fragmentation of chromosome 1 and fusion with chromosomes 5 and 1 in DFT1. Asterisks show chromosomes/markers that have been observed to vary among karyotyped strains of DFT1. DFT2 originated from a male devil and thus possesses a Y chromosome. While one copy of chromosome 6 has been inserted into the pericentric region of chromosome 1, the limited available karyotypes of DFT2 thus far suggest it is otherwise relatively undifferentiated from the normal devil karyotype. DFT, devil facial tumor. Figure and legend reproduced from Storfer et al. (2018a)

restore tumor MHC expression and thus facilitate detection by the devil host (Kosack et al. 2019). In addition, prophylactic treatment of devils with irradiated DFTD cells has shown initiation of some adaptive immunity (i.e., antibody production; Pye et al. 2018), but it is unclear the extent to which this is protective against exposure to live DFTD cells. Taken together, genomic and cytogenetic studies provided clues as to how DFTD originated, as well as possible treatment strategies moving forward.

A number of population genomic studies in the Tasmanian devil host have suggested evolution of possible resistance to DFTD. A time-series analysis using RADseq of 276 devils at 90,000 loci from three different populations collected both pre- and post-DFTD emergence showed parallel signatures of changes in allele frequencies in chromosomes 2 and 3 (Epstein et al. 2016). Indeed, evolution occurred rapidly – in as little as 4–6 generations – across a small number of candidate genes involved in cell adhesion and cancer recognition (Epstein et al. 2016). Moreover, a genome-wide association study of nearly 700 devils using a RAD-capture panel showed that a small number of loci explained variation in

case-control and survival of infected females (Margres et al. 2018a). Field evidence corroborates this result by showing that females may have evolved to be more tolerant of DFTD than males, which suffer a significantly greater decline of body condition when infected (Ruiz et al. 2018). A landscape genomic study of over 3,200 devils using a RAD-capture panel of 6,886 SNPs showed evidence of local adaptation to abiotic factors pre-disease (Fraik et al. 2019). However, when sampled post-disease, signatures of local adaptation to abiotic factors appeared to be swamped out by selection due to DFTD (Fraik et al. 2019). Taken together, population genomic studies of Tasmanian devils suggest rapid evolutionary responses to the extremely strong selective pressure generated by DFTD.

DFTD has also undergone spontaneous regression in a handful of cases, and one comparative genomic study of devils showing evidence of tumor regression versus those in which tumors did not regress showed that two candidate genes (TLL1 and NBAS1) are involved (Wright et al. 2017). An additional study suggested there were no detectable non-synonymous substitutions in the genome and that regression was likely driven by changes in regulatory regions that influence gene expression patterns (Margres et al. 2018b). Tumor regression, rapid evolutionary responses in devils, and recent ecological models that suggest extinction of devils is unlikely (Wells et al. 2017, 2019) suggest hope for the future of the Tasmanian devil.

**Lessons Learned** The Tasmanian devil-DFTD system is somewhat unique in terms of the amount of long-term mark-recapture data that exist, combined with the extensive sample archive and genomic resources. As such, better long-term monitoring studies of other species and cryobanking of samples (which is occurring on a larger scale than ever before) can allow pre- and post-disease analyses in a wider variety of disease systems. One advantage of working on a transmissible cancer is that the pathogen genome is a mutating clone that originated from the host genome, thereby essentially limiting the need to develop resources for one genome instead of two as in other host-pathogen systems. Genomic studies have yielded important clues as to how DFTD emerged and has spread through devil populations and most recently some potential vulnerabilities that may be treatment targets. In the devils, genomic analyses of time series have allowed researchers to track the evolutionary responses to DFTD across populations. Similarly, the opportunity exists for analysis of a large number of tumor samples to track mutations and tumor evolution through time.

## 3 Future Directions

Case studies of both epidemic and endemic pathogens presented herein demonstrate that genomics is at the forefront of our growing understanding of wildlife diseases. In already well-characterized systems such as WNS and FIV where intensive, targeted sampling is possible, phylogenomic and population genomic tools can identify and trace pathogen lineages spatially and temporally. These patterns can

reflect recent transmission among and within host species and populations, as well as deeper host and pathogen coevolutionary histories through which causative host and environmental factors can be identified. Further, functional and comparative genomics can identify genomic characteristics driving factors such as host susceptibility, pathogen fitness, pathogen virulence, and the likelihood of pathogen emergence and spillover. Our understanding of wildlife disease systems can be hindered by a poorly characterized or difficult to isolate causative agent; however, advances in metagenomics and single-cell sequencing are helping to overcome these challenges.

The lack of an annotated host reference genome is frequently a limiting factor in non-model organisms, an obstacle in systems such as chytridiomycosis due to multiple amphibian hosts with large genomes. However, some hope has emerged in a recent study that showed genome quality has minimal impact on our ability to reconstruct species' demographic histories and historic effective population sizes (Patton et al. 2019). Additionally, more tractable approaches such as transcriptomics and reduced representation libraries can prove useful for understanding host responses to infection where genomic resources are limited.

Notwithstanding the diversity and utility of the genomic tools demonstrated in the above examples, we perceive several emerging avenues through which genomics will increasingly benefit the study of wildlife disease. A key aim of wildlife disease ecology is to be able to identify and predict the circumstances leading to disease outbreaks and consequent geographic spread. Contact networks are an increasingly popular method in wildlife epidemiology, whereby the simulation of pathogens along networks representing host contact rates can help us understand how pathogen transmission can be predicted by host behavior (Craft and Caillaud 2011; Craft 2015). Epidemiological networks are traditionally devoid of genomic data, but integrating genomics into network approaches, such as by inferring pathogen transmission networks or host contact networks using measures of pathogen relatedness, can enhance the predictive power of such models (Gilbertson et al. 2018).

A similarly predictive discipline is landscape genomics, which investigates how landscape heterogeneity influences patterns of genetic and genomic variation (Manel et al. 2003; Manel and Holderegger 2013; Storfer et al. 2018b). Traditionally used to predict gene flow and identify environmental adaptation in free-living organisms, landscape genomics has also been used to understand how pathogen spread is shaped by environmental and geographic variation, but examples remain relatively few (Hemming-Schroeder et al. 2018; Kozakiewicz et al. 2018). The incorporation of spatially explicit biotic factors such as host or pathogen distributions or movement models into landscape genomic models will enhance predictions of both host connectivity and pathogen transmission. Recent methods for integrating landscape data and multiple genetic datasets from hosts and pathogens into a single analytical framework are particularly promising but have not been widely implemented (Leo et al. 2016; Schwabl et al. 2017).

Another relatively unexplored direction in wildlife disease genomics is epigenetics (Gómez-Díaz et al. 2012; Fitak et al. 2019). Distinct from the traditional understanding of adaptation as arising through selection on standing genetic variation, epigenetics is a process through which environmentally mediated changes

in gene regulation lead to heritable phenotypic change independently of the underlying DNA sequence. Many pathogen traits are thought to be under epigenetic control, leading to phenotypic plasticity that enables pathogens to adapt to varying host traits and environmental conditions (Gómez-Díaz et al. 2012). Conversely, pathogens are also likely to induce or manipulate epigenetic responses in their hosts that may impact host fitness and epidemiological dynamics over multiple generations (Paschos and Allday 2010; Gómez-Díaz et al. 2012). Many high-throughput sequencing approaches exist for identifying epigenetic markers (Meaburn and Schulz 2012) but have scarcely been implemented for wildlife disease research. Epigenetic studies of wildlife diseases will be an important component for our understanding of host-pathogen coevolution.

## References

Altizer S, Ostfeld RS, Johnson PT, Kutz S, Harvell CD. Climate change and infectious diseases: from evidence to a predictive framework. Science. 2013;341(6145):514–9.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17(2):81.

Antunes A, Troyer JL, Roelke ME, Pecon-Slattery J, Packer C, Winterbach C, Winterbach H, Hemson G, Frank L, Stander P, Siefert L, Driciru M, Funston PJ, Alexander KA, Prager KC, Mills G, Wildt D, Bush M, O'Brien SJ, Johnson WE. The evolutionary dynamics of the lion Panthera leo revealed by host and viral population genomics. PLoS Genet. 2008;4(11): e1000251.

Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B. Has the Earth's sixth mass extinction already arrived? Nature. 2011;471(7336):51.

Barr MC, Calle PP, Roelke ME, Scott FW. Feline immunodeficiency virus infection in nondomestic felids. J Zoo Wildl Med. 1989;20:265–72.

Berger L, et al. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. Proc Natl Acad Sci U S A. 1998;95:9031–6.

Biek R, Rodrigo AG, Holley D, Drummond AJ, Anderson CR Jr, Ross HA, Poss M. Epidemiology, genetic diversity, and evolution of endemic feline immunodeficiency virus in a population of wild cougars. J Virol. 2003;77:9578–89.

Biek R, Drummond AJ, Poss M. A virus reveals population structure and recent demographic history of its carnivore host. Science. 2006;311:538–41.

Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB. Landscape genetics and the spatial distribution of chronic wasting disease. Biol Lett. 2008;4:130–3.

Blehert DS, Hicks AC, Behr MJ, Meteyer CU, Berlowski-Zier BM, et al. Bat white-nose syndrome: an emerging fungal pathogen? Science. 2009;323:227.

Byrne AQ, Vredenburg VT, Martel A, Pasmans F, Bell RC, Blackburn DC, Bletz MC, Bosch J, Briggs CJ, Brown RM, Catenazzi A. Cryptic diversity of a widespread global pathogen reveals expanded threats to amphibian conservation. Proc Natl Acad Sci U S A. 2019;116(41):20382–7.

Carpenter MA, Brown EW, Culver M, Johnson WE, Pecon-Slattery J, Brousset D, O'Brien SJ. Genetic and phylogenetic divergence of feline immunodeficiency virus in the puma (*Puma concolor*). J Virol. 1996;70:6682–93.

Chaturvedi V, Springer DJ, Behr MJ, Ramani R, Li X, et al. Morphological and molecular characterizations of psychrophilic fungus *Geomyces destructans* from New York bats with white-nose syndrome (WNS). PLoS One. 2010;5:e10783.

Cornman RS, Schatz MC, Spencer Johnston J, Chen Y-P, Pettis J, Hunt G, Bourgeois L, et al. Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. BMC Genomics. 2010;11(1):602.

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan P-L, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. Science. 2007;318(5848):283–7.

Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140107.

Craft ME, Caillaud D. Network models: an underutilized tool in wildlife epidemiology? Interdiscip Perspect Infect Dis. 2011;2011:676949.

Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife: threats to biodiversity and human health. Science. 2000;287:443–9.

Davy CM, Donaldson ME, Willis CKR, Saville BJ, McGuire LP, Mayberry H, Wilcox A, et al. The other white-nose syndrome transcriptome: tolerant and susceptible hosts respond differently to the pathogen *Pseudogymnoascus destructans*. Ecol Evol. 2017;7(18):7161–70.

De Castro F, Bolker B. Mechanisms of disease-induced extinction. Ecol Lett. 2005;8:117–26.

Deakin JE, Bender HS, Pearse AM, Rens W, O'Brien PC, Ferguson-Smith MA, Cheng Y, Morris K, Taylor R, Stuart A, Belov K, Amemiya CT, Murchison EP, Papenfuss AT, Graves JA. Genomic restructuring in the Tasmanian devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. PLoS Genet. 2012;8(2):e1002483.

Di Prisco G, Desiderato A, Margiotta M, Ferrara R, Varricchio P, Zanni V, Caprio E, Nazzi F, Pennacchio F. A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and health. Proc Natl Acad Sci U S A. 2016;113(12):3203–8.

DiRenzo GV, Langhammer PF, Zamudio KR, Lips KR. Fungal infection intensity and zoospore output of *Atelopus zeteki*, a potential acute chytrid supershedder. PLoS One. 2014;9(3):e93356.

Eaton B, Broder C, Middleton D, et al. Hendra and Nipah viruses: different and dangerous. Nat Rev Microbiol. 2006;4:23–35.

Ellis JD, Evans JD, Pettis J. Colony losses, managed colony population decline, and colony collapse disorder in the United States. J Apic Res. 2010;49(1):134–6.

Ellison AR, et al. Fighting a losing battle: vigorous immune response countered by pathogen suppression of host defenses in the chytridiomycosis-susceptible frog *Atelopus zeteki*. G3. 2014;4:1275–89.

Ellison AR, et al. More than skin deep: functional genomic basis for resistance to amphibian chytridiomycosis. Genome Biol Evol. 2015;7:286–98.

Ellison AR, DiRenzo GV, McDonald CA, Lips KR, Zamudio KR. First in vivo *Batrachochytrium dendrobatidis* transcriptomes reveal mechanisms of host exploitation, host-specific gene expression, and expressed genotype shifts. G3. 2017;7:269–78.

Enguita FJ, Costa MC, Fusco-Almeida AM, Mendes-Giannini MJ, Leitão AL. Transcriptomic crosstalk between fungal invasive pathogens and their host cells: opportunities and challenges for next-generation sequencing methods. J Fungi (Basel). 2016;2(1):7.

Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, Schönfeld B, Wiench CG, Hohenlohe PA, Storfer A. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Commun. 2016;7:12684.

Evans JD, Saegerman C, Mullin C, Haubruge E, Nguyen BK, Frazier M, Frazier J, et al. Colony collapse disorder: a descriptive study. PLoS One. 2009;4(8):e6481.

Farrer RA, et al. Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. Proc Natl Acad Sci U S A. 2011;108(46):18732–6.

Farrer RA, Martel A, Verbrugghe E, Abouelleil A, Ducatelle R, Longcore JE, James TY, Pasmans F, Fisher MC, Cuomo CA. Genomic innovations linked to infection strategies across emerging pathogenic chytrid fungi. Nat Commun. 2017;8:14742.

Fèvre EM, Bronsvoort BMDC, Hamilton KA, Cleaveland S. Animal movements and the spread of infectious diseases. Trends Microbiol. 2006;14(3):125–31.

Fisher MC. Molecular toolkit unlocks life-cycle of the panzootic amphibian pathogen *Batrachochytrium dendrobatidis*. Proc Natl Acad Sci U S A. 2008;105:17209–10.

Fitak RR, Antonides JD, Baitchman EJ, Bonaccorso E, Braun J, Kubiski S, Chiu E, Fagre AC, Gagne RB, Lee JS, Malmberg JL, Stenglein MD, Dusek RJ, Forgacs D, Fountain-Jones NM, MLJ G, KEL W-T, Funk WC, Trumbo DR, Ghersi BM, Grimaldi W, Heisel SE, Jardine CM, Kamath PL, Karmacharya D, Kozakiewicz CP, Kraberger S, Loisel DA, McDonald C, Miller S, O'Rourke D, Ott-Conn CN, Páez-Vacas M, Peel AJ, Turner WC, Van Acker MC, VandeWoude S, Pecon-Slattery J. The expectations and challenges of wildlife disease research in the era of genomics: forecasting with a horizon scan-like exercise. J Hered. 2019;110:261–74.

Fountain-Jones NM, Craft ME, Funk WC, Kozakiewicz C, Trumbo DR, Boydston EE, Lyren LM, Crooks K, Lee JS, VandeWoude S, Carver S. Urban landscapes can change virus gene flow and evolution in a fragmentation-sensitive carnivore. Mol Ecol. 2017a;26(22):6487–98.

Fountain-Jones NM, Packer C, Troyer JL, Vander Waal K, Robinson S, Jacquot M, Craft ME. Linking social and spatial networks to viral community phylogenetics reveals subtype-specific transmission dynamics in African lions. J Anim Ecol. 2017b;86:1469–82.

Fraik AK, Margres MJ, Epstein B, Barbosa S, Jones M, et al. Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. bioRxiv. 2019:780122. https://doi.org/10.1101/780122.

Franklin SP, Troyer JL, Terwee JA, Lyren LM, Boyce WM, Riley SP, Roelke ME, Crooks KR, VandeWoude S. Frequent transmission of immunodeficiency viruses among bobcats and pumas. J Virol. 2007;81:10961–9.

Garroway CJ, Radersma R, Sepil I, Santure AW, Cauwer I, Slate J, Sheldon BC. Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. Evolution. 2013;67:3488–500.

Gilbertson MLJ, Fountain-Jones NM, Craft ME. Incorporating genomic methods into contact networks to reveal new insights into animal behaviour and infectious disease dynamics. Behaviour. 2018;155:759–91.

Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JL, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheiffelin JS, Lander ES, Happi C, Gevao SM, Gnirke A, Rambaut A, Garry RF, Khan SH, Sabeti PC. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345(6202):1369–72.

Gómez-Díaz E, Jordà M, Peinado MA, Rivero A. Epigenetics of host–pathogen interactions: the road ahead and the road behind. PLoS Pathog. 2012;8:e1003007.

Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G. Landscape genetics: a toolbox for studying vector-borne diseases. Front Ecol Evol. 2018;6:1–11.

Hewson I, Button JB, Gudenkauf BM, Miner B, Newton AL, Gaydos JK, Wynne J, Groves CL, Hendler G, Murray M, Fradkin S. Densovirus associated with sea-star wasting disease and mass mortality. Proc Natl Acad Sci U S A. 2014;111(48):17278–83.

Hohenlohe PA, McCallum HI, Jones ME, Lawrance MF, Hamede RK, Storfer A. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Conserv Genet. 2019;20:81–7.

Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature. 2006;443(7114):931.

James S, Jennings G, Kwon Y-M, Stammnitz M, Fraik A, Storfer A, Comte S, Pemberton D, Fox S, Brown B, Pye R, Woods G, Lyons B, Hohenlohe P, Siddle HV, Thomas F, Ujvari B, Murchison E, Jones M, Hamede R. Tracing the rise of malignant cell lines: distribution,

epidemiology and evolutionary interactions of two transmissible cancers in Tasmanian devils. Evol Appl. 2019;12:1772–80. https://doi.org/10.1111/eva.12831.
Johnson RM, Evans JD, Robinson GE, Berenbaum MR. Changes in transcript abundance relating to colony collapse disorder in honey bees (*Apis mellifera*). Proc Natl Acad Sci U S A. 2009;106(35):14790–5.
Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. Nature. 2008;451:990–4.
Kerr TJ, Matthee S, Govender D, Tromp G, Engelbrecht S, Matthee CA. Viruses as indicators of contemporary host dispersal and phylogeography: an example of feline immunodeficiency virus ($FIV_{Ple}$) in free-ranging African lion (*Panthera leo*). J Evol Biol. 2018;31:1529–43.
Kohno H, Suenami S, Takeuchi H, Sasaki T, Kubo T. Production of knockout mutants by CRISPR/Cas9 in the European honeybee, *Apis mellifera* L. Zoolog Sci. 2016;33(5):505–12.
Kosack L, Wingelhofer B, Popa A, Orlova A, Agerer B, et al. The ERBB-STAT3 axis drives Tasmanian devil facial tumor disease. Cancer Cell. 2019;35:125–39.
Kozakiewicz CP, Burridge CP, Funk WC, VandeWoude S, Craft ME, Crooks KR, Ernest HB, Fountain-Jones NM, Carver S. Pathogens in space: advancing understanding of pathogen dynamics and disease ecology through landscape genetics. Evol Appl. 2018;11:1763–78.
Krakoff E, Gagne RB, VandeWoude S, Carver S. Variation in intra-individual lentiviral evolution rates: a systematic review of human, non-human primate and felid species. J Virol. 2019;93:e00538-19. https://doi.org/10.1128/JVI.00538-19.
Lafferty KD. The ecology of climate change and infectious diseases. Ecology. 2009;90:888–900.
Langley RJ, Hirsch VM, O'Brien SJ, Adger-Johnson D, Goeken RM, Olmsted RA. Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic organization and relationship to other lentiviruses. Virology. 1994;202:853–64.
Lee JS, Ruell EW, Boydston EE, Lyren LM, Alonso RS, Troyer JL, Crooks KR, VandeWoude S. Gene flow and pathogen transmission among bobcats (*Lynx rufus*) in a fragmented urban landscape. Mol Ecol. 2012;21:1617–31.
Lee JS, Bevins SN, Serieys LE, Vickers W, Logan KA, Aldredge M, Boydston EE, Lyren LM, McBride R, Roelke-Parker M, Pecon-Slattery J, Troyer JL, Riley SP, Boyce WM, Crooks KR, VandeWoude S. Evolution of puma lentivirus in bobcats (*Lynx rufus*) and mountain lions (*Puma concolor*) in North America. J Virol. 2014;88:7727–37.
Lee J, Malmberg JL, Wood BA, Hladky S, Troyer R, Roelke M, Cunningham M, McBride R, Vickers W, Boyce W, Boydston E, Serieys L, Riley S, Crooks K, VandeWoude S. Feline immunodeficiency virus cross-species transmission: implications for emergence of new lentiviral infections. J Virol. 2017;91:e02134–16.
Leo SST, Gonzalez A, Millien V, Cristescu ME. Multi-taxa integrated landscape genetics for zoonotic infectious diseases: deciphering variables influencing disease emergence. Genome. 2016;59:349–61.
Leopardi S, Blake D, Puechmaille SJ. White-nose syndrome fungus introduced from Europe to North America. Curr Biol. 2015;25(6):R217–9.
Lips KR, et al. Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. Proc Natl Acad Sci U S A. 2006;103:3165–70.
Lips KR, Diffendorfer J, Mendelson JR III, Sears MW. Riding the wave: reconciling the roles of disease and climate change in amphibian declines. PLoS Biol. 2008;6(3):e72.
Longcore JE, Pessier AP, Nichols DK. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. Mycologia. 1999;91:219–27.
Looi LM, Chua KB. Lessons from the Nipah virus outbreak in Malaysia. Malays J Pathol. 2007;29:63–7.
Lorch JM, Meteyer CU, Behr MJ, Boyles JG, Cryan PM, et al. Experimental infection of bats with *Geomyces destructans* causes white-nose syndrome. Nature. 2011;480:376–8.
Manel S, Holderegger R. Ten years of landscape genetics. Trends Ecol Evol. 2013;28:614–21.
Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 2003;18:189–97.

Margres M, Jones M, Epstein B, Kerlin DH, Comte S, Fox S, Fraik AK, Hendricks SA, Huxtable S, Lachish S, Lazenby B, O'Rourke SM, Stahlke AR, Wiench CG, Hamede R, Schönfeld B, McCallum H, Miller MR, Hohenlohe PA, Storfer A. Large-effect loci affect survival and tumor growth in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. Mol Ecol. 2018a;27:4189–99.

Margres MJ, Ruiz-Aravena M, Hamede R, Jones ME, Lawrance MF, Hendricks SA, Patton A, Davis BW, Ostrander EA, McCallum H, Hohenlohe PA, Storfer A. The genomic basis of tumor regression in Tasmanian devils (*Sarcophilus harrisii*). Genome Biol Evol. 2018b;10:3012–25.

Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, Münk C, Nymark-McMahon H, Landau NR. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell. 2003;114:21–31.

Martel A, Spitzen-van der Sluijs A, Blooi M, Bert W, Ducatelle R, Fisher MC, Woeltjes A, et al. *Batrachochytrium salamandrivorans* sp. nov. causes lethal chytridiomycosis in amphibians. Proc Natl Acad Sci U S A. 2013;110(38):15325–9.

Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, Powell M, Nikaido S, Schroeder DC. Global honey bee viral landscape altered by a parasitic mite. Science. 2012;336(6086):1304–6.

McCallum H, Jones M, Hawkins C, Hamede R, Lachish S, Sinn DL, Beeton N, Lazenby B. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology. 2009;90(12):3379–92.

McDonald CA, Ellison AR, Toledo LF, James TY, Zamudio KR. Gene expression varies within and between enzootic and epizootic lineages of *Batrachochytrium dendrobatidis* (Bd) in the Americas. Fungal Biol. 2020;124(1):34–43.

Meaburn E, Schulz R. Next generation sequencing in epigenetics: insights and challenges. Semin Cell Dev Biol. 2012;23:192–9.

Minnis AM, Lindner DL. Phylogenetic evaluation of Geomyces and allies reveals no close relatives of *Pseudogymnoascus destructans*, comb. nov., in bat hibernacula of eastern North America. Fungal Biol. 2013;117(9):638–49.

Morehouse EA, et al. Multilocus sequence typing suggests the chytrid pathogen of amphibians is a recently emerged clone. Mol Ecol. 2003;12:395–403.

Morgan JAT, et al. Population genetics of the frog-killing fungus *Batrachochytrium dendrobatidis*. Proc Natl Acad Sci U S A. 2007;104:13845–50.

Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, Rebbeck C, et al. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science. 2010;327(5961):84–7.

Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell. 2012;148(4):780–91.

Nazzi F, Brown SP, Annoscia D, Del Piccolo F, Di Prisco G, Varricchio P, Della Vedova G, Cattonaro F, Caprio E, Pennacchio F. Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. PLoS Pathog. 2012;8(6):e1002735.

O'Hanlon SJ, Rieux A, Farrer RA, Rosa GM, Waldman B, Bataille A, Kosch TA, Murray KA, Brankovics B, Fumagalli M, Martin MD. Recent Asian origin of chytrid fungi causing global amphibian declines. Science. 2018;360(6389):621–7.

Ochoa A, Onorato DP, Fitak RR, Roelke-Parker ME, Culver M. *De novo* assembly and annotation from parental and F1 puma genomes of the Florida panther genetic restoration program. G3. 2019;9(11):3531–6.

Olmsted RA, Langley R, Roelke ME, Goeken RM, Adger-Johnson D, Goff JP, Albert JP, Packer C, Laurenson MK, Caro TM, et al. Worldwide prevalence of lentivirus infection in wild feline species: epidemiologic and phylogenetic aspects. J Virol. 1992;66:6008–18.

Parejo M, Wragg D, Henriques D, Vignal A, Neuditschko M. Genome-wide scans between two honeybee populations reveal putative signatures of human-mediated selection. Anim Genet. 2017;48:704–7.

Paschos K, Allday MJ. Epigenetic reprogramming of host genes in viral and microbial pathogenesis. Trends Microbiol. 2010;18:439–47.
Patchett A, Woods G. Targeting transmissible cancers in animals. Science. 2019;365:438–40.
Patton AH, Margres MJ, Jones ME, Hendricks S, Stahlke A, Lewallen K, Hamede RK, Ruiz-Aravena M, McCallum HI, Ryder O, Hohenlohe PA, Storfer A. Contemporary demographic reconstruction methods are robust to genome quality: a case study in Tasmanian devils. Mol Biol Evol. 2019;36(12):2906–21.
Pearse AM, Swift K. Allograft theory: transmission of devil facial-tumour disease. Nature. 2006;439(7076):549.
Pecon-Slattery J, Troyer JL, Johnson WE, O'Brien SJ. Evolution of feline immunodeficiency virus in Felidae: implications for human health and wildlife ecology. Vet Immunol Immunopathol. 2008;123:32–44.
Pedersen NC, Ho EW, Brown ML, Yamamoto JK. Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. Science. 1987;235:790–3.
Pounds JA, et al. Widespread amphibian extinctions from epidemic disease driven by global warming. Nature. 2006;439:161–7.
Puechmaille SJ, Wibbelt G, Korn V, Fuller H, Forget F, et al. Pan-European distribution of white-nose syndrome fungus (*Geomyces destructans*) not associated with mass mortality. PLoS One. 2011;6:e19167.
Pye RJ, Pemberton D, Tovar C, Tubio JM, Dun KA, Fox S, Darby J, Hayes D, Knowles GW, Kreiss A, Siddle HV, Swift K, Lyons AB, Murchison EP, Woods GM. A second transmissible cancer in Tasmanian devils. Proc Natl Acad Sci U S A. 2016;113(2):374–9.
Pye R, Patchett A, McLennan E, Thompson R, Carver S, et al. Immunization strategies producing a humoral IgG immune response against devil facial tumor disease in the majority of Tasmanian devils destined for wild release. Front Immunol. 2018;9:259.
Rachowicz LJ, HERO J, Alford RA, Taylor JW, Morgan JA, Vredenburg VT, Collins JP, Briggs CJ. The novel and endemic pathogen hypotheses: competing explanations for the origin of emerging infectious diseases of wildlife. Conserv Biol. 2005;19(5):1441–8.
Rajkumar SS, Li X, Rudd RJ, Okoniewski JC, Xu J, et al. Clonal genotype of *Geomyces destructans* among bats with white-nose syndrome, New York, USA. Emerg Infect Dis. 2011;17:1273–6.
Reeder SM, Sophia M, Palmer JM, Prokkola JM, Lilley TM, Reeder DAM, Field KA. *Pseudogymnoascus destructans* transcriptome changes during white-nose syndrome infections. Virulence. 2017;8(8):1695–707.
Reynolds HT, Barton HA, Slot JC. Phylogenomic analysis supports a recent change in nitrate assimilation in the white-nose syndrome pathogen, *Pseudogymnoascus destructans*. Fungal Ecol. 2016;23:20–9.
Ribas L, et al. Expression profiling the temperature-dependent amphibian response to infection by *Batrachochytrium dendrobatidis*. PLoS One. 2009;4:e8408.
Rosenblum EB, Stajich JE, Maddox N, Eisen MB. Global gene-expression profiles for life stages of the deadly amphibian pathogen *Batrachochytrium dendrobatidis*. Proc Natl Acad Sci U S A. 2008;105:17034–3.
Rosenblum EB, et al. Genome-wide transcriptional response of *Silurana (Xenopus) tropicalis* to infection with the deadly chytrid fungus. PLoS One. 2009;4:e6494.
Rosenblum EB, Poorten TJ, Settles M, Murdoch GK. Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. Mol Ecol. 2012;21:3110–20.
Rosenblum EB, James TY, Zamudio KR, Poorten TJ, Ilut D, Rodriguez D, Eastman JM, et al. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proc Natl Acad Sci U S A. 2013;110(23):9385–90.
Rosenkranz P, Aumeier P, Ziegelmann B. Biology and control of *Varroa destructor*. J Invertebr Pathol. 2010;103:S96–S119.
Ruiz M, Jones ME, Carver S, Estay S, Espejo C, Storfer A, Hamede RK. Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease. Proc R Soc B. 2018;285:20182239.
Ryabov EV, Wood GR, Fannon JM, Moore JD, Bull JC, Chandler D, Mead A, Burroughs N, Evans DJ. A virulent strain of deformed wing virus (DWV) of honeybees (*Apis mellifera*)

prevails after *Varroa destructor*-mediated, or in vitro, transmission. PLoS Pathog. 2014;10(6):e1004230.

Saremi NF, Supple MA, Byrne A, Cahill JA, Coutinho LL, Dalén L, Figueiró HV, Johnson WE, Milne HJ, O'Brien SJ, O'Connell B, Onorato DP, Riley SPD, Sikich JA, Stahler DR, Villela PMS, Vollmers C, Wayne RK, Eizirik E, Corbett-Detig RB, Green RE, Wilmers CC, Shapiro B. Mountain lion genomes provide insights into genetic rescue of inbred populations. BioRxiv. 2018; https://doi.org/10.1101/482315.

Savage AE, Zamudio KR. MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci U S A. 2011;108:16705–10.

Savage AE, Zamudio KR. Adaptive tolerance to a pathogenic fungus drives major histocompatibility complex evolution in natural amphibian populations. Proc Biol Sci. 2016;283(1827):20153115.

Schloegel LM, Toledo LF, Longcore JE, Greenspan SE, Vieira CA, Lee M, Zhao S, Wangen C, Ferreira CM, Hipolito M, Davies AJ, Cuomo CA, Daszak P, James TY. Novel, panzootic and hybrid genotypes of amphibian chytridiomycosis associated with the bullfrog trade. Mol Ecol. 2012;21(21):5162–77.

Schwabl P, Llewellyn MS, Landguth EL, Andersson B, Kitron U, Costales JA, Ocaña S, Grijalva MJ. Prediction and prevention of parasitic diseases using a landscape genomics framework. Trends Parasitol. 2017;33:264–75.

Siddle HV, Kreiss A, Tovar C, Yuen CK, Cheng Y, Belov K, Swift K, Pearse AM, Hamede R, Jones ME, Skjødt K, Woods GM, Kaufman J. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. Proc Natl Acad Sci U S A. 2013;110(13):5103–8.

Skerratt LF, et al. Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. Ecohealth. 2007;4:125–34.

Smith KF, Sax DF, Lafferty KD. Evidence for the role of infectious disease in species extinction and endangerment. Conserv Biol. 2006;20:1349–57.

Stammnitz MR, Coorens THH, Gorl K, Hayes D, Fu B, et al. The origins and vulnerabilities of two transmissible cancers in Tasmanian devils. Cancer Cell. 2018;33(4):607–19.

Storfer A, Hohenlohe PA, Margres MJ, McCallum HI, Patton A, Fraik A, Lawrance M, Stahlke A, Jones ME, Ricci L. The devil is in the details: genomics of transmissible cancers in Tasmanian devils. PLoS Pathog. 2018a;14:e1007098.

Storfer A, Patton A, Fraik AK. Navigating the interface between landscape genetics and landscape genomics. Front Genet. 2018b;9:68.

Taylor RL, Zhang Y, Schoning JP, Deakin JE. Identification of candidate genes for devil facial tumour disease tumourigenesis. Sci Rep. 2017;7:8761.

Tham H-W, Balasubramaniam VMK, Chew M-F. Viral determinants and vector competence of Zika virus transmission. Front Microbiol. 2018;9:1040. https://doi.org/10.3389/fmicb.2018.01040.

Trapp J, McAfee A, Foster LJ. Genomics, transcriptomics and proteomics: enabling insights into social evolution and disease challenges for managed and wild bees. Mol Ecol. 2017;26(3):718–39.

Trivedi J, Lachapelle J, Vanderwolf KJ, Misra V, Willis CKR, Ratcliffe JM, Ness RW, Anderson JB, Kohn LM. Fungus causing white-nose syndrome in bats accumulates genetic variability in North America with no sign of recombination. mSphere. 2017;2(4):e00271–17.

Trovão NS, Suchard MA, Baele G, Gilbert M, Lemey P. Bayesian inference reveals host-specific contributions to the epidemic expansion of influenza A H5N1. Mol Biol Evol. 2015;32(12):3264–75.

Troyer JL, Pecon-Slattery J, Roelke ME, Johnson W, VandeWoude S, Vazquez-Salat N, Brown M, Frank L, Woodroffe R, Winterbach C, Winterbach H, Hemson G, Bush M, Alexander KA, Revilla E, O'Brien SJ. Seroprevalence and genomic divergence of circulating strains of feline immunodeficiency virus among Felidae and Hyaenidae species. J Virol. 2005;79:8282–94.

Underwood RM, Vanengelsdorp D. Colony collapse disorder: have we seen this before? Bee Cult. 2007;35:13–8.

VandeWoude S, Apetrei C. Going wild: lessons from naturally occurring T-lymphotropic lentiviruses. Clin Microbiol Rev. 2006;19:728–62.

VandeWoude S, Troyer J, Poss M. Restrictions to cross species transmission of lentiviral infection gleaned from studies of FIV. Vert Immunol Immunopathol. 2010;134:25–32.

Vega Thurber RL, Barott KL, Hall D, Liu H, Rodriguez-Mueller B, Desnues C, Edwards RA, et al. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites compressa*. Proc Natl Acad Sci U S A. 2008;105(47):18413–8.

Wake DB, Vredenburg VT. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proc Natl Acad Sci U S A. 2008;105(Supplement 1):11466–73.

Webster NS, Negri AP, Botté ES, Laffy PW, Flores F, Noonan S, Schmidt C, Uthicke S. Host-associated coral reef microbes respond to the cumulative pressures of ocean warming and ocean acidification. Sci Rep. 2016;6:19324.

Wells K, Hamede RK, Kerlin DH, Storfer A, Hohenlohe PA, Jones ME, McCallum H. Infection of the fittest: devil facial tumor disease has greatest effect on individuals with highest reproductive output. Ecol Lett. 2017;20:770–8.

Wells K, Hamede Ross R, Jones M, Hohenlohe PA, Storfer A, McCallum H. Individual and temporal variation in pathogen load predicts long-term impacts of an emerging infectious disease. Ecology. 2019;100(3):e02613.

Welsh L, Maleszka R, Foret S. Detecting rare asymmetrically methylated cytosines and decoding methylation patterns in the honeybee genome. R Soc Open Sci. 2017;4(9):170248.

Woodhams DC, et al. Resistance to chytridiomycosis varies among amphibian species and is correlated with skin peptide defenses. Anim Conserv. 2007;10:409–17.

Wright B, Willet CE, Hamede R, Jones M, Belov K, Wade CM. Variants in the host genome may inhibit tumor growth in devil facial tumors: evidence from genome-wide association. Sci Rep. 2017;7:423.

Yang X, Cox-Foster DL. Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. Proc Natl Acad Sci U S A. 2005;102(21):7470–5.

Yang X, Cox-Foster D. Effects of parasitization by *Varroa destructor* on survivorship and physiological traits of *Apis mellifera* in correlation with viral incidence and microbial challenge. Parasitology. 2007;134(3):405–12.

Zielonka J, Marino D, Hofmann H, Yuhki N, Löchelt M, Münk C. Vif of feline immunodeficiency virus from domestic cats protects against APOBEC3 restriction factors from many felids. J Virol. 2010;84:7312–24.

# Population Genomics of Wildlife Cancer


**Sarah A. Hendricks, Andrew Storfer, and Paul A. Hohenlohe**

**Abstract** Studies of cancer in wildlife species present unique challenges, but research is beginning to uncover causes of cancer and its impact on wildlife populations. Causes of cancer in wildlife include environmental carcinogens, viruses and other pathogens, hereditary factors, and direct transmission of tumor cells. Here, we review progress and potential for population genomics to address issues such as genetic variation for susceptibility, comparative genomics of tumor suppressor genes, and evolutionary response to cancers. We also address the implications of cancer, and the potential of population genomics research, to inform conservation and management of wildlife populations. As an illustrative case study, we focus on the unique case of a transmissible cancer, devil facial tumor disease (DFTD), which has had a dramatic impact on demography and life history of Tasmanian devils (*Sarcophilus harrisii*). Recent population genomics research has revealed genetic variation underlying DFTD-related phenotypes and signatures of rapid evolution at candidate loci associated with cancer and immune function. The DFTD-devil system illustrates how genomics tools can be applied to an epizootic cancer in a wildlife population, providing insights into basic cancer biology as well as lessons for potential conservation strategies.

**Keywords** Conservation genomics · Devil facial tumor disease · Tasmanian devils · Transmissible cancer · Tumor evolution · Wildlife disease

S. A. Hendricks · P. A. Hohenlohe (✉)
Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA
e-mail: hohenlohe@uidaho.edu

A. Storfer
School of Biological Sciences, Washington State University, Pullman, WA, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_81, 

## 1 Introduction: Cancer in Wildlife

Ongoing research is elucidating the role of cancer in natural populations of wildlife species of ecological, cultural, and conservation importance (McAloose and Newton 2009; Pesavento et al. 2018; Hamede et al. 2020). Cancer, a collection of diseases characterized by abnormal and uncontrolled cell proliferation caused by somatic mutations, affects nearly every known multicellular organism. Rather than an accumulation of genomic aberrations in a single-cell lineage, it is now clear from numerous studies that cancers are heterogeneous collections of cells (Nowell 2002; Maley et al. 2006; Campbell et al. 2008; Merlo and Maley 2010; Park et al. 2010) that evolve in tumor microenvironments with complex ecological interactions (Bissell and Radisky 2001; Ujvari et al. 2019). Cancer can affect wild populations by reducing reproductive success and survival, altering interactions with predators or other species, and directly or indirectly leading to population declines (Dawson et al. 2018; Ujvari et al. 2019). Anthropogenic influences, including direct impacts, such as pollution and the reduction of genetic diversity in natural populations that are fragmented or reduced in size, can increase the prevalence of cancer in wildlife (McAloose and Newton 2009; Giraudeau et al. 2018; Pesavento et al. 2018). Aspects of wildlife behavior, life history, and genetic factors have been shaped by an evolutionary history with cancer as a selective force (Ujvari et al. 2019; Thomas et al. 2018, 2020). Understanding these many impacts of cancer on wildlife populations can help inform management and conservation efforts. Additionally, cancers in wildlife species may provide new biological models for understanding the complex causes of cancer, with the potential for biomedical benefits.

Studying cancer in wildlife species is difficult because of the ethical, logistical, and legal limits on invasive sampling and experimentation. Relatively few studies have estimated cancer prevalence in wild populations. Madsen et al. (2017) found that estimates of cancer prevalence in mammal populations range from 2% (sea otter, *Enhydra lutris*; Williams and Pulley 1981) to 64% (Baltic gray seal, *Halichoerus grypus*; Bäcklin et al. 2016). Low-prevalence cancers may go undetected and are likely to affect many more species than observed. In addition, estimates of prevalence in natural populations may be down-biased due to several factors (Hamede et al. 2020). First, most wildlife cancers lack diagnostic tools, particularly for detection in the absence of obvious clinical signs. Further, the cancer may lead to other conditions, including secondary parasite or pathogen infections, reduced body condition, and an increased level of predation resulting from a compromised ability to avoid predators. Thus, individuals may die from these secondary factors before the cancer exhibits obvious clinical signs (Vittecoq et al. 2013; Ujvari et al. 2019; Perret et al. 2020). Despite these difficulties, genomics research on wildlife cancer is beginning to reveal the underpinnings of cancer in natural populations. Here we review some of these developments and discuss how population genomics tools in cancer research can inform wildlife conservation.

## 2 Causes of Cancer

The uncontrolled cell proliferation of cancer is caused by somatic mutations and epigenetic alterations in a population of cells (Box 1). The probability of an individual developing cancer is influenced by several fundamental factors that can be considered extrinsic (e.g., environmental conditions) or intrinsic to an individual (e.g., genetic factors). In wildlife species, these factors can be the result of human influence, either on the surrounding environment or on the genetic diversity and evolutionary processes of wildlife populations. Other underlying causes of cancer, such as viruses or direct transmission of tumor cells, can also result in epizootic spread of cancer as an infectious disease across wildlife populations. Factors may interact as well, such as environmental factors or infectious agents interacting with variation in genetic susceptibility.

**Box 1 Key Mutation Types in Cancer**

Progression to cancer involves somatic mutations or epigenetic changes that remove the constraints to uncontrolled proliferation in a population of cells (Hanahan and Weinberg 2000, 2011). Normal control of cell proliferation involves a number of pathways, and genes whose inactivation can allow tumorigenesis are called tumor suppressor genes (Vogelstein et al. 2013). The "gatekeepers" are the genes directly involved in preventing unregulated cell division by inhibiting growth or promoting death of cells with chromosomal abnormalities, while the "caretakers" are involved in error-free DNA replication, effective DNA repair, and the maintenance of appropriate epigenetic patterning and chromosomal structure (Kinzler and Vogelstein 1997; Stoler et al. 1999; Shields and Harris 2000; Sarkies and Sale 2012). Loss of function at caretaker genes can increase the rate of mutation and chromosomal alteration, and inherited mutations in caretaker genes are often associated with hereditary cancers because they increase the likelihood of cancer from subsequent somatic mutation (Negrini et al. 2010). Oncogenes are those that promote tumorigenesis when increased in activation or expression level by somatic mutation or epigenetic change, and they often exhibit recurrent mutations at the same positions across tumors (Vogelstein et al. 2013).

Because tumor progression is associated with increased mutation rates, tumors typically show large numbers of genetic differences from their respective hosts. Tumors are heterogeneous populations of cells, with selection acting among cellular lineages. Mutations that increase relative fitness of a cell lineage are "driver" mutations, while those that are neutral are "passenger" mutations that increase in frequency solely because of hitchhiking in successful cell lineages (Vogelstein et al. 2013; Cannataro and Townsend 2018). Massive genomic sequencing efforts, including single-cell sequencing and subsampling of tumor cell populations, have revealed a number of genes that

(continued)

**Box 1** (continued)
are strongly associated with driver mutations in particular cancer types (Vogelstein et al. 2013; Heng 2017; ICGC/TCGA 2020; Rheinbay et al. 2020). Important types of mutations and genetic factors in cancer progression are:

1. *Single nucleotide polymorphisms (SNPs)* make up approximately 95% of mutations from cancer genomes (Vogelstein et al. 2013; Heng 2017). These mutations can result in nonsynonymous changes in proteins, as well as other functional consequences, such as changes in micro-RNA loci or regulatory binding sites that affect gene expression.
2. *Copy number variation (CNV)* is defined as the amplification or deletion of DNA fragments >50bp (Girirajan et al. 2011). Somatic copy number alterations (SCNAs) are common in cancer; however, distinguishing driver SNCAs from numerous SCNAs that randomly accumulate during tumorigenesis is not straightforward (Zack et al. 2013; Heng 2017).
3. *Chromosomal structural abnormalities* such as translocations or aneuploidy are extremely common for many cancer types and can have large effects on gene function and cellular phenotypes (Stephens et al. 2009; Heng et al. 2013). Chromothripsis, which literally means "chromosome shattering," is defined by a single, localized event within genomic regions in one or few chromosomes characterized by thousands of clustered chromosomal rearrangements. Similarly, chromoplexy is characterized by chromosomal rearrangements that involve segments of DNA from multiple chromosomes (e.g., five or more). These abnormalities have been implicated in cancer phenotypes, particularly metastasis and drug resistance (Heng et al. 2013).
4. *Telomere dynamics* are involved in many cancers as well as somatic maintenance, aging, and apoptosis. Progressive shortening of telomeres typically induces cellular senescence, which can provide a defense against cancer. Mutations that affect the function of telomerase or promote telomere lengthening have been associated with multiple types of cancer (Artandi and DePinho 2010; Vogelstein et al. 2013).
5. *Epigenetic factors*, heritable changes in gene expression that are not accompanied by changes in DNA sequence, can contribute to tumorigenesis, for instance, by increasing the expression of oncogenes (Jones and Baylin 2007; Vogelstein et al. 2013). Abnormalities in methylation, histone modification, nuclear topology, and noncoding RNA have been implicated in the silencing of key tumor suppressor, regulatory, and repair genes resulting in cancer (reviewed by Grunau 2017). Epigenetic modifications are not detected by DNA sequencing, but by other approaches such as transcriptomics or methylation profiling.

## 2.1 Environmental Conditions

Environmental contaminants and other external influences are often associated with human and wildlife cancer incidence. Ultraviolet and other radiation exposure, smoking, environmental pollution, and ingestion of certain foods or toxins influence cancer initiation and progression by increasing somatic mutation rates (Perera 1998; Irigaray et al. 2007; Soto and Sonnenschein 2010). Moreover, there is some evidence that environmental endocrine disruptors, stress, and trauma influence somatic mutation rates and cancer risk (Reiche et al. 2004; Antoni et al. 2006; Aktipis and Nesse 2013; Pesavento et al. 2018). In natural populations of wildlife species, stress and multiple types of environmental pollution may play important roles in cancer incidence (Pesavento et al. 2018). Aquatic and terrestrial wildlife in the Chernobyl area of Ukraine, subject to elevated levels of pollution and radioactivity following the 1986 nuclear disaster, have higher cancer prevalence than populations in less polluted areas (Yablokov 2009; Mousseau and Møller 2015). Similarly, the beluga whale (*Delphinapterus leucas*) population in the St. Laurence River estuary (Canada) has a higher rate of cancer than other populations (Martineau et al. 2002). This population shows evidence of contamination by agricultural and industrial chemicals, such as polycyclic aromatic hydrocarbons, polychlorinated biphenyls, dichlorodiphenyltrichloroethane, and their metabolites (Letcher et al. 2000; Martineau et al. 2002).

## 2.2 Viruses and Other Pathogens

Parasite-induced cancers can be due to a variety of subcellular, unicellular, or multicellular parasites and pathogens. Cancer associated with multicellular parasites may be the result of chronic inflammation (Pesavento et al. 2018), such as that caused by infection from ear mites in Santa Catalina Island foxes (*Urocyon littoralis catalinae*; Vickers et al. 2015) or from nematodes in ring-necked pheasants (*Phasianus colchicus*; Himmel and Cianciolo 2017). Many pathogen-induced cancers in wildlife are attributed specifically to viruses, which have direct mutagenic effects on host tissue (McAloose and Newton 2009; McCallum and Jones 2012; Pesavento et al. 2018). Evaluating the effects of pathogens, particularly viruses, on cancer development and establishing a causal link between cancer incidence and infection is challenging partially because of a lag between the presence of the parasite and cancer detection. However, there is growing evidence that viruses may be associated with a substantial proportion of cancers in humans as well as natural populations of wildlife (Ewald and Swain Ewald 2015, 2019). Viruses disrupt a variety of cellular barriers to oncogenesis; for instance, infected cells may lose the ability to control the total number of cellular divisions, apoptosis, adhesive properties to other cells, and/or cellular arrest (Ewald and Swain Ewald 2015). Examples of empirical evidence for virus-induced cancer in wildlife include

otarine herpesvirus-1 and genital carcinoma in California sea lions (*Zalophus californianus*; Lipscomb et al. 2016), deltapapillomavirus associated with fibropapillomas and fibromas in deer (subfamily Capreolinae), giraffe (*Giraffa* spp.), and zebra (*Equus* spp.) species (Pesavento et al. 2018), papillomas and carcinomas in western barred bandicoots (*Perameles bougainville*), and lymphomas in Attwater's prairie chickens (*Tympanuchus cupido attwateri*; Drechsler et al. 2009) (see McAloose and Newton 2009; Ewald and Swain Ewald 2017; Pesavento et al. 2018; Hamede et al. 2020 for further examples).

## 2.3 Transmissible Cancers

Known from only a handful of animal species, transmissible cancers are much rarer than virus-associated cancers. Transmissible cancers are spread directly by transfer of tumor cells between individuals; in other words, the etiologic agent is the neoplastic cells derived from an original host (Metzger and Goff 2016; Ostrander et al. 2016). The tumor cells are a set of clonal lineages, spreading from the original host to secondary hosts across a population as an infectious disease. Transmission occurs with direct contact during mating, biting, or feeding, or tumor cells may be spread and acquired through the environment in marine systems (Metzger and Goff 2016; Ostrander et al. 2016). A well-studied example is canine transmissible venereal tumor (CTVT), which affects dogs (*Canis lupus domesticus*) and is believed to have originated thousands of years ago, making it perhaps the oldest continuously propagated cell lineage (Murchison et al. 2014; Baez-Ortega et al. 2019). A group of transmissible cancers produces leukemia-like conditions such as disseminated neoplasia or hemic neoplasia, in at least 15 different bivalve species (Metzger et al. 2016). Some lineages within this group of transmissible cancers have spread across species and across wide geographic areas (Yonemitsu et al. 2019), and they are strongly associated with the integration of retrotransposons into the host genomes (Arriagada et al. 2014; Metzger et al. 2018). Devil facial tumor disease (DFTD) and DFT2 are two recent independent origins of transmissible cancer that infect Tasmanian devils (*Sarcophilus harrisii*; Pearse and Swift 2006; Pye et al. 2016); we discuss this case in detail below. In transmissible cancers, the genome of tumor cells descends from the original host, so that the highest degree of genomic similarity is expected among tumors across a host population, rather than between each tumor and its respective, contemporary host. Accordingly, transmissible cancers can be diagnosed by genomic similarities, such as shared chromosomal rearrangements and other mutations, across a set of tumors (Pearse and Swift 2006; Pye et al. 2016; Leathlobhair et al. 2017).

### 2.4 Hereditary Factors

Hereditary susceptibility to cancer is widely established in humans, with over 200 cases known, most of which are inherited as autosomal dominant alleles (Nagy et al. 2004). The following characteristics designate an inherited cancer susceptibility: "two or more relatives with the same type of cancer on the same side of the family; several generations affected; earlier ages of cancer diagnosis than what is typically seen for that cancer type; individuals with multiple primary cancers; the occurrence of cancers in one family, which are known to be genetically related; and the occurrence of nonmalignant conditions and cancer in the same person and/or family" (Nagy et al. 2004). Hereditary factors include genetic mutations that increase the susceptibility to cancer progression (Box 1). Many of these susceptibility syndromes are rare, but collectively hereditary cancers account for at least 1–10% of all cancers in humans (Fearon 1997; Nagy et al. 2004). Most research on wildlife focuses on a population rather than an individual or family level, so we know little about hereditary cancer in wildlife. In particular, rare variants that increase individual susceptibility are very difficult to detect in wildlife, even if they collectively impose a large cancer burden on the population. Genetic variants that increase cancer susceptibility are expected to behave at a population level much like other deleterious variants, such as those associated with other pathologies. As a result, we can predict that cancer may contribute to reduced population fitness in wildlife populations that are small or fragmented and subject to reduced genetic diversity and inbreeding (Ujvari et al. 2019).

## 3 Genomics and Evolution of Cancer in Wildlife

### 3.1 Evolution of Cancer Resistance

An evolutionary perspective is useful for understanding cancer at multiple scales, from the behavior of cellular lineages within a tumor, to understanding the genetics of resistance at the individual level, to population-level susceptibility (Frank 2004; Ujvari et al. 2019; Thomas et al. 2020). Natural selection is predicted to act against cancer susceptibility because of its effect on fitness, but this effect is reduced to the extent that timing of onset is later in an individual's life, after some proportion of reproduction has occurred (Leroi et al. 2003). As a result, early-onset cancers tend to be more attributable to a specific cause, while later onset may reflect multiple causes, such as rare alleles persisting in a population because of weak negative selection (Frank 2004). In general, natural selection favors mechanisms in the genome for resistance and tolerance to cancer (Seluanov et al. 2018; Thomas et al. 2020). These mechanisms in turn suppress the selection acting at the cellular level, which favors cell lineages that proliferate at the expense of the multicellular individual (Michod 2000). A variety of ecological and evolutionary processes also occur within the

population of cells that make up a tumor and its microenvironment. Genetically, there is greater evidence for positive selection than purifying selection in tumors, compared to evolution at the species level, meaning that multiple mutations can increase cell proliferation within a tumor (Martincorena et al. 2017). Nonetheless, the large majority of mutations still appear neutral within tumors, so that neutral theory from population genetics can be fruitfully applied within a tumor cell population as well (Cannataro and Townsend 2018).

Adaptations to prevent cancer may have trade-offs in the capacity for wound healing, growth, reproduction, and aging. This tension is present because somatic maintenance and growth require controlled cell division, while suppressing cell proliferation is central to cancer resistance (Guo and DiPietro 2010; Hofman and Vouret-Craviari 2012). Similarly, there may be trade-offs with reproductive effort, such as the relationships between early menarche or fertility and susceptibility to breast cancer in humans (Smith et al. 2012). There is evidence that wildlife populations in captivity with limited opportunities for reproduction, such as the white rhinoceros (*Ceratotherium simum*), may experience a higher rate of neoplasias because of the proliferative effects of increased estrous cycling (Pesavento et al. 2018). Wildlife species may be frequently exposed to chronic infection by parasites and wounds from predators or other ecological interactions, so the capacity for inflammatory responses and wound healing is important for fitness. However, the inflammatory response can foster neoplastic cell proliferation, cause DNA damage, and create a microenvironment conducive to tumor progression (de Visser et al. 2006; Pesavento et al. 2018; Ujvari et al. 2019).

Balancing among these selective forces in the evolutionary history of wildlife species has left a complex legacy of cancer susceptibility in many populations (Thomas et al. 2018). This balance may help explain Peto's paradox: the lack of correlation between body size, life span, and cancer risk (Abegglen et al. 2015). The general expectation should be that larger body size and life span should require more somatic cell divisions and somatic maintenance, resulting in greater opportunity for somatic mutations leading to cancer. However, the lineages of some long-lived species, such as elephants and naked mole rats, have evolved remarkable resistance to cancer (Seluanov et al. 2018; Tollis et al. 2019). Comparative and other genomics approaches have uncovered some of the mechanisms explaining this resistance, discussed below.

## *3.2 Genetics of Population Susceptibility*

Wildlife species are subject to reduced population size, fragmentation, and inbreeding, which can increase the frequency of slightly deleterious alleles and their presence in the homozygous state. Some of these alleles may increase susceptibility to cancer, thus increasing the overall genetic load and potentially increasing extinction risk of small wildlife populations (Gomulkiewicz and Holt 1995; Frankham 2005). The degree to which cancer susceptibility contributes to genetic load is very difficult to quantify, although population genomics studies could reveal the patterns

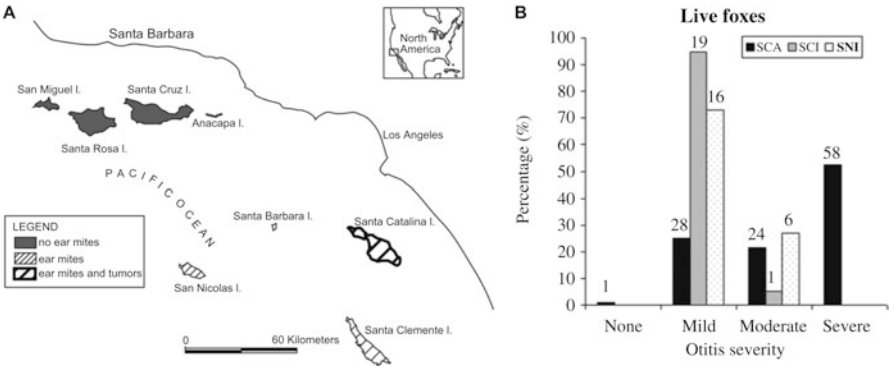


**Fig. 1** One population of island foxes on Santa Catalina Island (*Urocyon littoralis catalinae*) exhibits high incidence of inflammation-induced ear canal cancer associated with ear mite infection. (**a**) Map of the Channel Islands off the coast of California, showing the presence of fox populations without ear mites on three northern islands, foxes with ear mites on three southern islands, and high incidence of cancer on just one island. (**b**) Severity of otitis (inflammation of the ear canal) is much higher in foxes on Santa Catalina Island (SCA) compared to the other two islands with ear mites, San Nicolas Island (SNI) and San Clemente Island (SCI). Reproduced from Vickers et al. (2015)

of historic selection on cancer-associated genes and their potential effects on fitness by using functional genetic information from annotated genomes (Oh et al. 2019; Robinson et al. 2018).

One example of population susceptibility to cancer is found in Channel Island foxes (*Urocyon littoralis*), which are endemic to individual islands off the coast of California (Fig. 1). These populations have undergone severe genetic bottlenecks resulting in strong genetic differentiation among islands and the accumulation of deleterious mutations, although they retain enough genetic variation to facilitate local adaptation to different environmental conditions (Funk et al. 2016; Robinson et al. 2016). Populations among islands differ markedly in the incidence of cancer: on one island (Santa Catalina Island, SCA), foxes have a high prevalence of ear canal (ceruminous gland) carcinoma and adenoma that appear to be associated with inflammation from chronic infection by ear mites (*Otodectes* spp.; Vickers et al. 2015). Ceruminous gland tumors have not been documented on other islands (San Clemente Island and San Nicolas Island) despite similar levels of chronic mite infection, nor in the three island fox populations that do not have ear mites. Treatment of individual foxes with acaricide, which removes ear mite infection, significantly reduced inflammation and hyperplasia, and had a non-significant trend toward reducing tumor progression likely due to low sample size (Moriarty et al. 2015). Ear mite infection also induces changes in the ear canal microbiome (DeCandia et al. 2019). Other factors, such as a virus or environmental differences among islands, could play a role in the strikingly different level of cancer prevalence between SCA and the other populations. Nonetheless, a leading hypothesis is that the severe genetic drift in the SCA population has increased frequencies of alleles that contribute to inflammation-induced cancer susceptibility.

### 3.3 Population Genomics Studies of Wildlife Cancer

Powerful new genomics approaches in wildlife can be used to estimate the prevalence of cancer, its effects on population fitness and conservation, broad-scale evolutionary patterns, and the specific genetic mechanisms of cancer susceptibility. The large case-control or genome-wide association studies that have been critical to understanding the genetic basis of cancer in humans or other model organisms are often not feasible in wildlife species, but population genomics studies of wildlife cancer are still tractable. For example, in a case-control study of California sea lions (*Zalophus californianus*), urogenital carcinoma was significantly associated with homozygosity of a microsatellite loci within an intron of the heparanase 2 gene (HPSE2; Browning et al. 2014), which has been implicated in several human carcinomas. Two unusual rodents, the naked mole rat (*Heterocephalus glaber*) and the blind mole rat (*Spalax ehrenbergi*), are not closely related to each other but have independently evolved extremely low incidences of cancer, despite their long life spans. Cancer resistance in naked mole rats appears to involve multiple genes that control telomere dynamics (MacRae et al. 2015; Tollis et al. 2017; Seluanov et al. 2018). In blind mole rats, pre-cancerous hyperplasia triggers a strong concerted cell death response through interferon-β, suppressing tumor progression (Seluanov et al. 2018).

Transmissible cancers have been the focus of genomics studies to understand the genetic basis of susceptibility as well as the mechanisms that allow tumors to transmit among individual hosts and evade host immune and tumor suppression responses. In the cluster of transmissible cancers in bivalve molluscs, the *Steamer* retrotransposon exhibits extreme amplification in neoplastic cells as well as evidence of multiple cross-species transfers (Metzger et al. 2018). In the canine transmissible venereal tumor (CTVT), genomics work has dated the origin of the disease, roughly coincident with the domestication of dogs, and characterized the mutational signature of CTVT (Baez-Ortega et al. 2019). One particular mutational type, a C->T transition in the context of the five-nucleotide motif GTCCA, was prevalent until ~1,000 years ago but then subsided. Additionally, most driver mutations appear to have occurred relatively early in CTVT history, and more recent genetic evolution appears to be neutral (Baez-Ortega et al. 2019).

The ability to produce whole-genome sequence data and reference genome assemblies in wildlife species allows for comparative genomics studies of cancer, examining the genomes of a group of related taxa in a phylogenetic framework to reveal evolutionary history (Gorbunova et al. 2014; MacRae et al. 2015; Tollis et al. 2017; Seluanov et al. 2018). Comparative genomics has revealed evolutionary relationships among body size, life span, and cancer susceptibility and the genetic mechanisms that relate these factors (Tollis et al. 2017; Seluanov et al. 2018). Studies of genetic adaptations to reduce cancer susceptibility in long-lived and large-bodied animals have revealed cancer-related genomic evolution such as copy number variants of specific tumor suppressor and genome maintenance genes, and they help to resolve Peto's paradox (Caulin et al. 2015; Seluanov et al. 2018). Elephants, which have low cancer mortality, possess more copies of the TP53

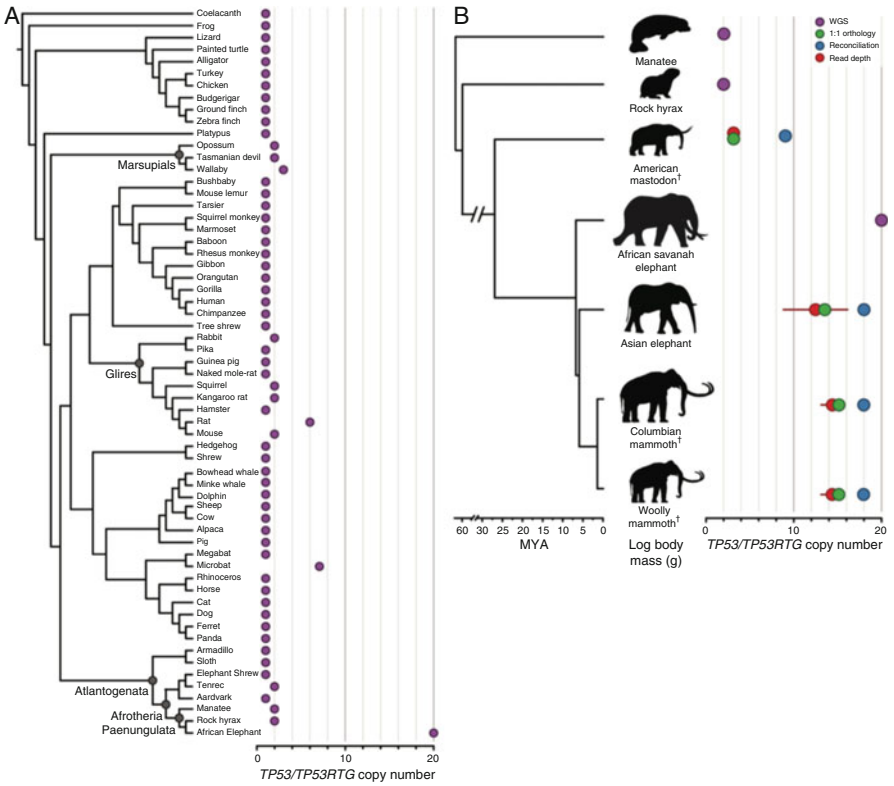


**Fig. 2** Elephants, despite large body sizes and long life spans, have evolved resistance to cancer by increasing copy number of the tumor suppressor gene TP53 as revealed by comparative genomics at two different evolutionary scales: (**a**) across a phylogeny of tetrapod vertebrates, and (**b**) among elephants and two closely related mammalian orders. Reproduced from Sulak et al. (2016)

(p53) tumor suppressor gene than 61 other vertebrate species (Fig. 2; Sulak et al. 2016). Elephant cells, as compared to human cells, demonstrate an increased p53-mediated apoptotic response following DNA damage (Abegglen et al. 2015), which may be due to the transcription and likely translation of several of the TP53 retrogenes (Sulak et al. 2016). Sulak et al. (2016) also found a positive association between body size and copy number of TP53 retrogenes (Fig. 2).

Tollis et al. (2019) compared the genomes of ten cetacean species – the largest-bodied animals – and found substantial evidence for selection for multiple mechanisms of cancer resistance that differ from elephants. These included segmental duplications of regions containing genes associated with apoptosis and evidence for positive selection in other loci linked to cell cycle checkpoints, cell signaling, and proliferation (Tollis et al. 2019). In another recent study on rodents, telomere maintenance strategies were found to differ depending on body mass and differential cancer risks (Tian et al. 2018). Larger species evolved repression of somatic telomerase activity and replicative senescence while longer-lived smaller species

evolved telomere-independent anticancer mechanisms that act to slow down cell proliferation and prevent premalignant hyperplasia. Patterns are similar across mammals: body size is related to telomere length and telomerase activity as a result of trade-offs among selection for cancer resistance and selection for protection against DNA damage and replicative senescence (Tollis et al. 2017; Risques and Promislow 2018). In contrast, animals with longer life spans tend to reduce cell proliferation rates and evolve toward early-acting tumor suppressor genes (Seluanov et al. 2018).

## 4 Tasmanian Devils and DFTD

### *4.1 An Epidemic Transmissible Cancer*

Devil facial tumor disease (DFTD) was first observed in 1996 by a wildlife photographer who documented ulcerative neoplasias on the face of Tasmanian devils (*Sarcophilus harrisii*; Hawkins et al. 2006). The disease has since spread across most of the island of Tasmania with only a few devil populations yet unaffected in the far western and northwestern parts of the species' range. The census population size has been reduced by ~80% due to these metastatic tumors that typically result in mortality within 6 months to 1 year of transmission (Hamede et al. 2012, 2015; Lazenby et al. 2018; Jones et al. 2019). DFTD cells are undifferentiated neoplasms with highly pleomorphic and anaplastic cells (Pyecroft et al. 2007). Tumors result in ulcerating proliferative masses that tend to occur around the face and jaw, and masses within the oral cavity can prevent feeding and are prone to secondary infection (Hawkins et al. 2006). Live cancer cells are the infectious agent and are transmitted to new hosts by biting during social interactions (Pearse and Swift 2006; Pyecroft et al. 2007; Hamilton et al. 2019). Uninfected, aggressive biters become infected after biting the tumors of infected, less aggressive bite recipients; therefore, more socially dominant devils appear more likely to get DFTD (Wells et al. 2017). Thus far, there is no evidence of vertical transmission from mothers to their offspring, and low levels of prevalence in juveniles could be associated with dramatic changes in immune capacity at sexual maturity (Cheng et al. 2017, 2019). The pattern of infection and mortality has effects on population age structure, with substantial shifts toward younger animals (Lachish et al. 2009; Hamede et al. 2012). Changes in life history strategies have also been observed, and age at first breeding has shifted from 2+ years to 14 months in some areas (Jones et al. 2008; Lachish et al. 2009). Early models predicted extinction of the species in the wild (McCallum et al. 2009), but growing evidence from multiple sources suggests extinction is unlikely (Hohenlohe et al. 2019; Wells et al. 2019), and local populations have not gone extinct in the wild (Lazenby et al. 2018; Storfer et al. 2018).

The etiology of DFTD and characterization of the cell of origin were largely determined through molecular cytogenetic, immunogenetic, and genomics methods. Clonality of DFTD was initially established by karyotypic data, which showed that tumors from different individuals contain the same complex chromosomal rearrangements (Pearse and Swift 2006; Deakin et al. 2012). Microsatellite and MHC analysis indicating a lack of diversity across tumors, consistent with clonal transmissibility (Siddle et al. 2007), and further genomic sequencing and genotyping of somatic mutations revealed details of the pattern of spread and mutational process in the DFTD tumor cell population (Murchison et al. 2012). Tumors were found to express diagnostic neuron-specific markers indicating that the ancestral cell type of DFTD was Schwann cell origin (Murchison et al. 2010; Loh et al. 2016). Antibody staining indicated that tumor cells produce a Schwann cell-specific protein, periaxin (Murchison et al. 2010), which is now considered a sensitive and specific diagnostic for DFTD tumors (Tovar et al. 2011).

A few hypotheses have been offered regarding host evasion leading to the rapid spread and near-universal susceptibility to DFTD. First, irregular tumor MHC expression and downregulation of host MHC by DFTD may help the tumor to escape host surveillance (Siddle et al. 2013). During the initial neoplastic transformation, epigenetic downregulation of multiple aspects within the antigen-presenting system occurs (Siddle et al. 2013). This leads to the inability of DFTD to display functional MHC class I molecules, in vivo or in vitro, thereby avoiding recognition by T cells. Additionally, devils may lack enough MHC diversity to recognize and destroy aberrant tumor cells (Siddle et al. 2007). Siddle et al. (2007) did not detect lymphocyte response when lymphocytes from devils were tested against each other as well as lymphocytes isolated from other parts of the island. However, MHC diversity is not linked to variation in disease susceptibility among individuals, and devils can reject tissue allografts (Kreiss et al. 2011; Lane et al. 2012). More recent work has implicated the ERBB-STAT3 signaling pathway in MHC expression and tumor transmissibility (Kosack et al. 2019). Rather than alternative explanations, it may be that all of these factors – reduced species-wide diversity in MHC, downregulation of MHC expression in tumor cells, tumor suppression of the host immune response, and alteration of other genetic pathways in tumor cells – act in combination to facilitate DFTD transmission.

Remarkably, a second transmissible cancer has arisen recently in Tasmanian devils, called DFT2, with multiple lines of evidence supporting an independent origin from the first DFTD (Pye et al. 2016). DFT2 appeared in a geographically distinct area (southern Tasmania, as opposed to northeast Tasmania for DFTD), and cytogenetic evidence suggests that DFT2 originated in a male devil, in contrast to a female devil for DFTD. While similar in cell type origin, mode of transmission, and gross appearance, these two transmissible cancers differ in histology, in the specific mutations characteristic to each, and in the way in which changes in MHC expression facilitate evasion of the host immune system (Pye et al. 2016; Caldwell et al. 2018; Stammnitz et al. 2018; James et al. 2019; Patchett et al. 2020). Nonetheless, DFTD and DFT2 appear to share some broad-scale chromosomal rearrangements that may point to genetic changes that play a role in transmissibility in both cancers

(Deakin et al. 2012; Taylor et al. 2017; Storfer et al. 2018). Both diseases spread in an epidemic fashion across the devil population, although DFT2 exhibits sex bias with males appearing to be more susceptible, perhaps due to rejection of tumor cells by females based on Y-chromosome-associated factors (James et al. 2019). The independent origin of two transmissible cancers in Tasmanian devils within just two decades raises the hypothesis that devils are uniquely susceptible to this type of disease, and similarities among them may point toward the specific mechanisms that allow transmissible cancers in this species (Stammnitz et al. 2018; Patchett et al. 2020). It may also be the case that transmissible cancers are more widespread across the animal kingdom than previously recognized.

## *4.2 Devil Genomics*

Population genomics tools have been used in Tasmanian devils to understand responses to the DFTD epidemic, as well as inform multiple aspects of management and conservation priorities in natural populations and management of the captive insurance population. First, we have an emerging view of the demographic history of devils and their current levels of genetic diversity and phylogeographic relationships among populations. These factors will strongly influence the ability of devils to adapt to DFTD, as well as other threats such as environmental change and anthropogenic disturbances (Hendricks et al. 2017). Previous studies have revealed that devils have low genetic diversity, based on data from microsatellite loci (Jones et al. 2004; Brüniche-Olsen et al. 2014; Storfer et al. 2017), MHC loci (Siddle et al. 2010; Cheng et al. 2012), SNPs (Hendricks et al. 2017; Fraik et al. 2020), and whole-genome sequencing (Miller et al. 2011; Murchison et al. 2012; Patton et al. 2019). Low genetic diversity in Tasmanian devils is potentially the result of historical fluctuations in population size and extinction of the species on mainland Australia and its restriction to the island of Tasmania (Guiler 1978; Hawkins et al. 2006; Brüniche-Olsen et al. 2018; Patton et al. 2019). The quality of the reference genome has been improved, and re-sequencing of 12 individuals robustly supports demographic reconstructions of a historic bottleneck using multiple genomic analyses (Patton et al. 2019). Nonetheless, devils show consistent evidence of population structure, particularly a large genetic cluster covering the eastern half of the island, another cluster in the northwest, and a broad zone of admixture between them (Hendricks et al. 2017).

Despite the overall low genetic diversity of the species, several lines of evidence suggest a rapid evolutionary response to the strong selection imposed by DFTD (Hohenlohe et al. 2019). First, three independent populations were found to show a parallel, rapid (4–6 generations) evolutionary response to the disease (Fig. 3; Epstein et al. 2016). This study scanned across 90K SNP loci, generated by high-density RAD sequencing, for signatures of selection and found two genomic regions, which contained genes with immunological and oncogenic functions. Second, using the data from Epstein et al. (2016), another study used a maximum likelihood approach

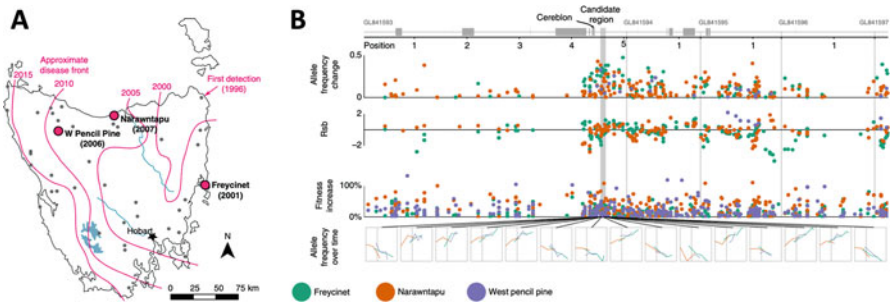


**Fig. 3** Genomic evidence of rapid evolution in response to transmissible cancer in Tasmanian devils. (**a**) Map of Tasmania showing spread of devil facial tumor disease (DFTD) across the species range, with three focal populations for which genomic samples were collected before and after the disease appeared. (**b**) A region of chromosome II showing multiple concordant signatures of selection based on SNPs derived from RAD sequencing. The gray bar highlights the candidate selected region based on three signatures of selection: allele frequency change, a metric of linkage disequilibrium, and estimated fitness effect of the increasing allele. SNP loci are colored by population, and allele frequency changes over time at individual SNPs are shown across the bottom; note the concordance in the direction of allele frequency change across the three populations. Reproduced from Epstein et al. (2016)

and improved functional annotations to find more signatures of selection in the devil genome (Hubert et al. 2018). In total, 97 genomic regions showed evidence of selection, most of which were population-specific with one region common to all three populations. These regions harbored 148 protein-coding genes (or human orthologues), nearly all of which have a link with cancer. Third, a genome-wide association study (GWAS) of ~600 individuals found that phenotypic variation in female survivorship (length of time after infection) could be explained by a few loci of large effect (~5 SNPs explained about >61% of the total variance; Margres et al. 2018a). Further, Margres et al. (2018a) found that female infection rates (female case-control) could be explained by more SNPs of smaller effect (~56 SNPs explained about >23% of the total variance). Given that DFTD has spread across multiple genetic clusters in the devil population, any allelic variation for resistance to DFTD may be able to spread across the devil population and increase in frequency because of selection (Hendricks et al. 2017).

Given early predictions of extinction of devils in the wild, a captive insurance metapopulation distributed across a number of locations was established in 2006 with the goal of maintaining a disease-free population that is "genetically representative of the species" (CBSG/DPIPWE/ARAZPA 2009). The insurance population has been managed using a combination of molecular and pedigree information geared to maximize genetic diversity across the genome (Hogg et al. 2015; Grueber et al. 2018). A panel of microsatellite markers (Wright et al. 2015) has been used to monitor genetic diversity in the insurance population, and genomic information can be more informative than pedigree relationships for assessing diversity and inbreeding (Kardos et al. 2015; Hogg et al. 2018; Brandies et al. 2019). Disease-free wild populations have been established on Maria Island and Forestier Peninsula, but long-

term genetic conservation using these isolated populations would require continued supplementation (McLennan et al. 2018). While genomics tools have been used, none of the captive or reintroduced populations are currently managed for variation at any specific cancer-related loci. As our understanding of the genetic basis of DFTD susceptibility continues to improve, it would be possible for management of the insurance populations to consider maintaining overall diversity at the growing list of genes associated with DFTD (Hohenlohe et al. 2019).

### *4.3 Tumor Genomics*

A number of different karyotypic strains of DFTD have been discovered (Pearse et al. 2012). These strains resemble the original DFTD karyotype reported by Pearse and Swift (2006), designated strain 1, but are characterized by additional cytogenetic rearrangements consistent with ongoing tumor evolution as the disease continues to spread through the population (Deakin et al. 2012). It appears from both cytogenetic and sequencing analysis that DFTD strains are continuing to accumulate karyotypic, copy number, and sequence variants, but compared with most human cancers, DFTD strains are remarkably stable (Deakin et al. 2012; Murchison et al. 2012; Taylor et al. 2017). Selection may be working to maintain the tumorigenic properties of the DFTD genome while permitting genomic instability and sequence substitutions in regions not critical for the survival of the DFTD cell (Deakin et al. 2012; Taylor et al. 2017). The number of somatic point mutations varies widely in humans, yet the mutation rate in DFTD is likely to be less than some human cancers, such as lung or skin cancer (Martincorena and Campbell 2015). As compared to the reference devil genome, two sequenced DFTD genomes identified approximately 17,000 somatic mutations that are present in the tumor (Murchison et al. 2012).

A candidate gene approach has identified rearrangements in tumor genomes at several genes known to be associated with cancer in other species, providing a possible list of driver mutations for transmissibility of DFTD (Taylor et al. 2017). While evolution in the tumor cell population of DFT2 has had much less time to proceed, genomic comparisons of DFT2 and DFTD are already revealing similarities between the two, for instance, in frequencies of single-nucleotide mutations (Stammnitz et al. 2018). DFT2 still appears to express MHC class I molecules, demonstrating that complete suppression is not necessary for transmissibility, but the most highly expressed molecules are either common or non-polymorphic among hosts, and MHC expression in DFT2 appears to be evolving (Caldwell et al. 2018). In both DFTD and DFT2, distinguishing somatic mutations from those found in the original or transient host is important for understanding what drives tumor growth and how the tumor evades immune detection by accumulating mutations in pathways related to recognition of self versus non-self. Genomics approaches, such as those used in canids (Decker et al. 2015; Baez-Ortega et al. 2019), would involve including large catalogs of variation found in modern devils, which are critical for identifying these somatic mutations.

Genomics is beginning to reveal mechanisms leading to spontaneous tumor regression or even complete recovery from the disease in some devils (fewer than 20), and it appears that features of both the host and tumor genomes may play a role (Pye et al. 2016; Wright et al. 2017; Margres et al. 2018b). Using a comparative case-control approach, two key genomic regions in the tumor were identified to putatively be associated with tumor regression and, therefore, the ability of the host to survive DFTD (Wright et al. 2017). Using targeted genotyping in additional samples, the authors were able to confirm that three genes may be involved in slowing tumor growth and allowing additional time for the effected individual to mount an immune response (Wright et al. 2017). Another study compared the genomes of devils that showed evidence of tumor regression to those that succumbed to DFTD and found a different set of three highly differentiated regions containing several genes with immunological or oncogenetic functions (Margres et al. 2018b). Putative regulatory variation in candidate genes suggests that changes in gene expression may drive natural tumor regression. Despite the small number of animals that have recovered from the disease, strong selection pressure from the disease may cause the frequency of these variants to increase over time in the devil population.

## *4.4 Conservation of Tasmanian Devils*

One application of population genomics tools for devil conservation in the face of DFTD is the development of vaccines or other intervention techniques to promote population-level resistance (Owen and Siddle 2019; Patchett and Woods 2019). Several studies have explored immune-stimulatory agents and vaccines against DFTD (Tovar et al. 2017, 2018; Patchett et al. 2017; Pye et al. 2018). For example, heat shock proteins (HSPs) derived from tumor cells have been used as a source of antigens for cancer immunotherapy in humans (see review by Murshid et al. 2008). A recent study by (Tovar et al. 2018) found that DFTD cancer cells express inducible HSP, which supports that a HSP-based vaccine against DFTD could be developed. A promising target could be the ERBB-STAT3 pathway, with therapies that could potentially recover MHC expression and arrest tumor growth (Kosack et al. 2019; Patchett and Woods 2019). Despite this progress, work remains to be done to show that an immune stimulation or vaccine protocol could confer sufficient immunity or resistance to treated individuals for a sufficient length of time to be effective in natural populations and to confirm that it would not have unintended consequences for DFTD epizootic behavior.

A population genomics understanding of variation in devil susceptibility to DFTD is also important for conservation and management of both captive and wild populations. As described above, multiple studies have established that devils have genetic variation for disease-related traits, even including tumor regression, and that populations are responding to selection by DFTD. Demographic modeling also predicts devil persistence under most scenarios, allowing time for an evolutionary response in nature (Wells et al. 2017, 2019). In contrast, captive populations have

not been exposed to the disease and are not managed for any disease-related variation. As a result, supplementing wild populations with devils from captive populations that have not been exposed to the disease could increase the severity of the disease by increasing transmission rates and population-level susceptibility (Hohenlohe et al. 2019). In other words, attempts at demographic rescue – increasing population size with supplementation in areas where the disease has greatly reduced devil density – could be counterproductive by impeding evolutionary rescue, the ability of populations to evolve higher fitness in the face of the disease. Additionally, the discovery of DFT2 favors the view that conservation strategies for devils consider not just genetic variation relevant to DFTD but also genetic variation relevant to immune function and cancer in general that could provide adaptive potential for the future (Hohenlohe et al. 2019). Genetic monitoring of both captive and wild populations should target allelic variation at both DFTD-specific and broader functional categories of loci associated with both transmissible cancers. The devil-DFTD system illustrates how population genomics tools can allow detection of adaptive and functionally significant loci associated with threats to species persistence, and this knowledge can guide conservation efforts.

## 5 Future Directions in Population Genomics of Wildlife Cancer

Many wildlife species are the focus of conservation efforts because of historic population declines, fragmentation and loss of genetic diversity, and social and economic importance. Population genomics tools have wide applications to management of natural and ex situ wildlife populations (Walters and Schwartz 2020), and cancer may be an important challenge that some wildlife species face and that can be incorporated into population genomics-based conservation strategies (Box 2; Hamede et al. 2020).

**Box 2 Management and Conservation of Wildlife Using Population Genomic Data**

Cancer may be one of many factors creating concern for conservation of wildlife populations, and genomics can provide powerful tools for assessing its impact. High-throughput genomic technologies have increased our ability to assess inbreeding coefficients, gene flow, demography including effective population size, epidemiology, adaptive potential, and population viability (Kardos et al. 2016; Flanagan et al. 2018; Hoelzel et al. 2019; Hohenlohe et al. 2020; Storfer et al. 2020). These sources of information have been used to guide wildlife management efforts in natural and captive populations (Walters and Schwartz 2020). When populations are small, both inbreeding

(continued)

**Box 2** (continued)
and genetic drift can increase homozygosity at loci with deleterious alleles, reducing fitness (i.e., increasing "genetic load") and contributing significantly to extinction risk (Frankham 2005; Díez-del-Molino et al. 2018). Hereditary cancer susceptibility due to the accumulation of oncogenic mutations could be a source of genetic load in wildlife populations.

If a population suffers from genetic load or inbreeding, genetic rescue through mediated migration, translocation, and reintroduction via captive breeding programs can increase population fitness due to an increase in heterozygosity, which can mask deleterious mutations, and facilitate adaptive evolution (Bell et al. 2019). Population genomics tools can be used to inform genetic rescue, for instance, by identifying source populations or assessing the risk of outbreeding depression (Fitzpatrick and Funk 2019). Alternatively, evolutionary rescue, evolution from standing genetic variation without migration (Hufbauer et al. 2015), may be possible particularly when there is evidence that a population has adaptive genetic variation. In wildlife populations, hereditary cancer may be caused by relatively rare, deleterious variants. If so, genetic rescue or evolutionary rescue may be highly effective in reducing cancer susceptibility, although there may be trade-offs between the two (Hohenlohe et al. 2019).

## *5.1 Monitoring and Population Management*

With the expanding set of tools for designing and genotyping panels of genetic variants in wildlife species, cancer-related marker panels could be informative for monitoring. Particularly in wildlife populations with high prevalence of a specific hereditary or environmental contaminant-caused cancer, or in the case of transmissible cancers, the disease may have a substantial impact on population fitness and viability. Genetic marker panels targeting the host genome could be used to predict population-level susceptibility and disease impacts, screen individuals or family groups with particularly high cancer susceptibility, or track evolution of genetic variation at loci associated with cancer incidence (Leroy et al. 2018). Genetics tools can also be used to detect the disease itself, for instance, in transmissible, pathogen-driven, or environmental cancers, where it can be important to detect cancer in individuals before obvious clinical signs (McAloose and Newton 2009; Kwon et al. 2018). For instance, assessing population-level cancer incidence would provide information on overall population health, consequences of reduced genetic diversity or inbreeding, and the effects of exposure to environmental contaminants or viral pathogens (Leroy et al. 2018; Pesavento et al. 2018). Finally, genetic marker panels designed for tumor samples could also be used to track tumor evolution.

All of these sources of information could inform the targeting of conservation efforts toward natural populations in which cancer may have strong effects on population persistence (Box 2). For example, small or isolated populations in

which hereditary cancer contributes to genetic load could be targets for genetic rescue (Bell et al. 2019). Conversely, evidence of evolution in natural populations in response to disease could argue against translocations for genetic rescue (Hohenlohe et al. 2019). Individual-level metrics for cancer susceptibility, as well as genetic diagnostic tools for preclinical cancer screening, could be applied in selecting individuals for reintroduction or translocation (Fitzpatrick and Funk 2019).

## 5.2 Captive Breeding Programs

Genomics can support the identification of candidate loci responsible for heritable disorders, which can inform breeding decisions in captive populations of wildlife species. Genome-wide association studies (GWAS) have found large numbers of variants associated with complex human traits and diseases such as cancer, leading to genetic panels for preventive and personalized medicine (Vazquez et al. 2012; Vogelstein et al. 2013). Despite the limits on statistical power from feasible sample sizes in wildlife studies, GWAS, along with a range of other population genomics tools, is increasingly being applied in wildlife species, including studies of cancer (Leathlobhair et al. 2017; Margres et al. 2018a; Baez-Ortega et al. 2019). In a captive wildlife population, genetic information on cancer-associated loci could be combined with pedigree information and used for strategic breeding. For example, a similar method was used in the case of the critically endangered California condor (*Gymnogyps californianus*), which suffered from high incidence of the lethal disease chondrodystrophy (Romanov et al. 2009; Grueber 2015). Through the pedigrees obtained in the captive breeding program, researchers found this disease to show Mendelian segregation (Ralls et al. 2000). Genomic resources were developed to identify causal polymorphisms linked to the disease, with the aim of informing the captive breeding protocols to reduce the frequency of chondrodystrophy while maintaining genetic diversity at other loci (Romanov et al. 2009; Walters et al. 2010). Overall, this approach would help to safeguard against inbreeding depression to avoid further decreases in individual fitness (Frankham 2010). Finding a single locus of major effect on inbreeding depression may be unusual, and cancer susceptibility may be more often highly polygenic. However, the ability of genomics tools to screen many thousands of loci and lead to genotyping marker panels of hundreds to thousands of markers opens the door to managing captive populations with genetic metrics targeting multi-locus traits.

## 5.3 Interventions in Wildlife Cancer

As our understanding of cancer treatment options in humans improves, these advances may be translated to wildlife in the context of direct interventions at the individual level. For example, the recognition of tumor cell populations as

heterogeneous, evolving systems can be used in designing treatment strategies (Gatenby et al. 2009; Hanahan and Weinberg 2011), and this is particularly relevant in the case of transmissible cancers (Caldwell et al. 2018; Stammnitz et al. 2018). In natural wildlife populations, ongoing invasive treatment of individuals is typically not possible. However, advancements in oral vaccine development and delivery for infectious diseases have led to successful infectious disease control as seen in the case of sylvatic plague affecting prairie dogs (*Cynomys* spp.) and the endangered black-footed ferret (*Mustela nigripes*; Salkeld 2017). These approaches to reduce infectious pathogens could reduce cancer impacts, for instance, in the case of virus-associated cancers or the ear mites and island foxes discussed above. It may also be possible to develop vaccines against transmissible cancers that could suppress epidemic spread (Owen and Siddle 2019; Patchett and Woods 2019).

## *5.4 Advances in Wildlife Cancer Genomics Research*

As the taxonomic scope of genomic data continues to increase, comparative genomics approaches will continue to increase our understanding of the genetic basis of cancer susceptibility and mechanisms of cancer suppression in wildlife (Caulin and Maley 2011; Tollis et al. 2017). In addition to the mammal studies discussed above, the long lives, slow developmental rates, probable low cancer rates, and the rapid development of genomic resources for large reptiles will provide ample opportunity to study genomic mechanisms of cancer suppression in these ectothermic amniotes (Tollis et al. 2015). Birds may also have relatively low incidence of cancer in most species (Madsen et al. 2017), which suggests that the numerous avian genomes available (Zhang et al. 2014) could provide more comparative genomics information about cancer suppression.

Many general types of cancer are shared across species (Schiffman and Breen 2015; Madsen et al. 2017; Pesavento et al. 2018), as evidenced by the widespread use of non-human mammal species used as models for human cancer. Investigation of shared cancers using a multi-species approach will highlight genes associated with carcinogenesis in the context of both genetics and environmental exposure. Important insights can also be gained from studying lineages that have a high prevalence for cancer, such as the marine mammals, Santa Catalina Island foxes, and Tasmanian devils discussed above. The wide diversity of causes associated with cancer in these taxa means that conclusions may not be applicable across systems, but wildlife can provide a broad view of multiple types of cancer susceptibility.

Technical advances in sequencing and bioinformatics will benefit the study of cancer in wildlife. For instance, the high levels of diversity and gene duplication (Nei et al. 1997; Temperley et al. 2008) that make immunity highly adaptable also make immune-gene regions challenging to assemble. Therefore, it is difficult to determine how many copy number variants of genes exist in a species or individual genome (Cheng et al. 2012; Alcaide et al. 2014). However, technological advances, such as Oxford NanoPore and Pacific Biosciences (PacBio) sequencing, continue to increase

the length of single DNA molecules that can be directly sequenced. Additionally, continued development and assessment of computational approaches (e.g., Putnam et al. 2016) may aid in resolving the challenges presented by gene duplications and repetitive regions. Long sequence reads can also help identify runs of homozygosity, a measure of inbreeding and critical for identifying deleterious loci in small populations (Hohenlohe et al. 2020). These advances will help identify candidate loci associated with disease susceptibility and inbreeding depression in wildlife populations.

## 6 Conclusions

Cancer affects nearly all multicellular organisms, yet our understanding of the role of cancer in wildlife populations remains limited. In reduced or fragmented wildlife populations with reduced genetic diversity or inbreeding, cancer may contribute to genetic load and reduced population fitness (McAloose and Newton 2009; Pesavento et al. 2018). Additionally, this may have impacts on interacting species and ecosystem function (Vittecoq et al. 2013; Ujvari et al. 2019). Population genomics approaches can inform multiple aspects of wildlife cancer. As genomic data continues to accumulate across taxa, our understanding of how evolutionary forces have shaped cancer suppression mechanisms will improve, providing new models for biomedical cancer research and a clearer view of the genetic susceptibility of wildlife populations to cancer (Seluanov et al. 2018). Intensive genomic studies of wildlife populations can reveal the specific genetic mechanisms of cancer susceptibility. The ability to rapidly identify putative functional loci and design marker genotyping panels opens the door to high-throughput genetic monitoring and management tools for wildlife populations (Leroy et al. 2018). As population genomics tools continue to develop across wildlife applications, as detailed throughout the chapters of this volume, cancer and other diseases will be an important component of wildlife genomics.

**Acknowledgments** Support was provided to SAH by the National Institute of Health grant P30GM103324; NIH grant R01GM126563 and National Science Foundation grant DEB-1316549 as part of the joint NIH-NSF-USDA Ecology and Evolution of Infectious Diseases Program to PAH and AS; and the Bioinformatics and Computational Biology program at the University of Idaho.

## References

Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, et al. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. JAMA. 2015;314:1850–60. https://doi.org/10.1001/jama.2015.13134.

Aktipis CA, Nesse RM. Evolutionary foundations for cancer biology. Evol Appl. 2013;6:144–59. https://doi.org/10.1111/eva.12034.

Alcaide M, Muñoz J, Martínez-de la Puente J, Soriguer R, Figuerola J. Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian Coot *Fulica atra* (Aves: Rallidae). Ecol Evol. 2014;4:688–98. https://doi.org/10.1002/ece3.974.

Antoni MH, Lutgendorf SK, Cole SW, Dhabhar FS, Sephton SE, McDonald PG, et al. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. Nat Rev Cancer. 2006;6:240–8. https://doi.org/10.1038/nrc1820.

Arriagada G, Metzger MJ, Muttray AF, Sherry J, Reinisch C, Street C, et al. Activation of transcription and retrotransposition of a novel retroelement, *Steamer*, in neoplastic hemocytes of the mollusk *Mya arenaria*. Proc Natl Acad Sci U S A. 2014;111:14175–80. https://doi.org/10.1073/pnas.1409945111.

Artandi SE, DePinho RA. Telomeres and telomerase in cancer. Carcinogenesis. 2010;31:9–18. https://doi.org/10.1093/carcin/bgp268.

Bäcklin BM, Eriksson L, Olovsson M. Histology of uterine leiomyoma and occurrence in relation to reproductive activity in the Baltic gray seal (*Halichoerus grypus*). Vet Pathol. 2016;40:175–80. https://doi.org/10.1354/vp.40-2-175.

Baez-Ortega A, Gori K, Strakova A, Allen JL, Allum KM, Bansse-Issa L, et al. Somatic evolution and global expansion of an ancient transmissible cancer lineage. Science. 2019;365:6452. https://doi.org/10.1126/science.aau9923.

Bell DA, Robinson ZL, Funk WC, Fitzpatrick SW, Allendorf FW, Tallmon DA, et al. The exciting potential and remaining uncertainties of genetic rescue. Trends Ecol Evol. 2019;34(12):1070–9.

Bissell MJ, Radisky D. Putting tumours in context. Nat Rev Cancer. 2001;1:46–54. https://doi.org/10.1038/35094059.

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. Genes (Basel). 2019;10(11). pii: E846.

Browning HM, Acevedo-Whitehouse K, Gulland FMD, Hall AJ, Finlayson J, Dagleish MP, et al. Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. Proc Biol Sci. 2014;281:20140240. https://doi.org/10.1098/rspb.2014.0240.

Brüniche-Olsen A, Jones ME, Austin JJ, Burridge CP, Holland BR. Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease. Biol Lett. 2014;10:20140619. https://doi.org/10.1098/rsbl.2014.0619.

Brüniche-Olsen A, Jones ME, Burridge CP, Murchison EP, Holland BR, Austin JJ. Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. J Biogeogr. 2018;45(5):963–76.

Caldwell A, Coleby R, Tovar C, Stammnitz MR, Kwon YM, Owen RS, et al. The newly-arisen devil facial tumour disease 2 (DFT2) reveals a mechanism for the emergence of a contagious cancer. Elife. 2018;7:e35314. https://doi.org/10.7554/eLife.35314.

Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proc Natl Acad Sci U S A. 2008;105:13081–6. https://doi.org/10.1073/pnas.0801523105.

Cannataro VL, Townsend JP. Neutral theory and the somatic evolution of cancer. Mol Biol Evol. 2018;35:1308–15. https://doi.org/10.1093/molbev/msy079.

Caulin AF, Maley CC. Peto's paradox: evolution's prescription for cancer prevention. Trends Ecol Evol. 2011;26:175–82.

Caulin AF, Graham TA, Wang L-S, Maley CC. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140222–10. https://doi.org/10.1098/rstb.2014.0222.

CBSG/DPIPWE/ARAZPA. The Save the Tasmanian Devil Program: strategic framework for an insurance meta-population. Hobart, Australia: Department of Primary Industries, Parks, Water and Environment; 2009.

Cheng Y, Stuart A, Morris K, Taylor R, Siddle H, Deakin J, et al. Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. BMC Genomics. 2012;13:87. https://doi.org/10.1186/1471-2164-13-87.

Cheng Y, Heasman K, Peck S, Peel E, Gooley RM, Papenfuss AT, et al. Significant decline in anticancer immune capacity during puberty in the Tasmanian devil. Sci Rep. 2017;7:44716. https://doi.org/10.1038/srep44716.

Cheng Y, Makara M, Peel E, Fox S, Papenfuss AT, Belov K. Tasmanian devils with contagious cancer exhibit a constricted T-cell repertoire diversity. Commun Biol. 2019;2:99. https://doi.org/10.1038/s42003-019-0342-5.

Dawson EH, Bailly TPM, Dos Santos J, Moreno C, Devilliers M, Maroni B, et al. Social environment mediates cancer progression in Drosophila. Nat Comms. 2018;9:3574. https://doi.org/10.1038/s41467-018-05737-w.

de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer. 2006;6:24–37. https://doi.org/10.1038/nrc1782.

Deakin JE, Bender HS, Pearse A-M, Rens W, O'Brien PCM, Ferguson-Smith MA, et al. Genomic restructuring in the Tasmanian devil facial tumour: chromosome painting and gene mapping provide clues to evolution of a transmissible tumour. PLoS Genet. 2012;8:e1002483. https://doi.org/10.1371/journal.pgen.1002483.

DeCandia AL, Brenner LJ, King JL, vonHoldt BM. Ear mite infection is associated with altered microbial communities in genetically depauperate Santa Catalina Island foxes (*Urocyon littoralis catalinae*). Mol Ecol. 2019; https://doi.org/10.1111/mec.15325.

Decker B, Davis BW, Rimbault M, Long AH, Karlins E, Jagannathan V, et al. Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. Genome Res. 2015;25:1646–55. https://doi.org/10.1101/gr.190314.115.

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. Quantifying temporal genomic erosion in endangered species. Trends Ecol Evol. 2018;33:176–85. https://doi.org/10.1016/j.tree.2017.12.002.

Drechsler Y, Bohls RL, Smith R, Silvy N, Lillehoj H, Collisson EE. An avian, oncogenic retrovirus replicates *in vivo* in more than 50% of CD4+ and CD8+ T lymphocytes from an endangered grouse. Virology. 2009;386:380–6.

Epstein B, Jones ME, Hamede R, Hendricks S, McCallum H, Murchison EP, et al. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nat Commun. 2016;7:12684. https://doi.org/10.1038/ncomms12684.

Ewald PW, Swain Ewald HA. Infection and cancer in multicellular organisms. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140224. https://doi.org/10.1098/rstb.2014.0224.

Ewald PW, Swain Ewald HA. Infection and cancer in nature. In: Ujvari B, Roche B, Thomas F, editors. Ecology and evolution of cancer. London: Academic; 2017. p. 47–56.

Ewald PW, Swain Ewald HA. The scope of viral causation of human cancers: interpreting virus density from an evolutionary perspective. Philos Trans R Soc B. 2019;374:20180304. https://doi.org/10.1098/rstb.2018.0304.

Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. Science. 1997;278:1043–50.

Fitzpatrick SW, Funk WC. Genomics for genetic rescue. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/13836_2019_64.

Flanagan SP, Forester BR, Latch EK, Aitken SN, Hoban S. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. Evol Appl. 2018;119:1267–18. https://doi.org/10.1111/eva.12569.

Fraik AK, Margres MJ, Epstein B, Barbosa S, Jones M, Hendricks S, et al. Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. Evolution. 2020 (in press).

Frank SA. Genetic predisposition to cancer – insights from population genetics. Nat Rev Genet. 2004;5:764–72. https://doi.org/10.1038/nrg1450.

Frankham R. Genetics and extinction. Biol Conserv. 2005;126:131–40. https://doi.org/10.1016/j.biocon.2005.05.002.

Frankham R. Challenges and opportunities of genetic approaches to biological conservation. Biol Conserv. 2010;143:1919–27. https://doi.org/10.1016/j.biocon.2010.05.011.

Funk WC, Lovich RE, Hohenlohe PA, Hofman CA, Morrison SA, Sillett TS, et al. Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol Ecol. 2016;25:2176–94. https://doi.org/10.1111/mec.13605.

Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009;69:4894–903. https://doi.org/10.1158/0008-5472.CAN-08-3658.

Giraudeau M, Sepp T, Ujvari B, Ewald PW, Thomas F. Human activities might influence oncogenic processes in wild animal populations. Nat Ecol Evol. 2018;2:1065–70. https://doi.org/10.1038/s41559-018-0558-7.

Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. Annu Rev Genet. 2011;45:203–26. https://doi.org/10.1146/annurev-genet-102209-163544.

Gomulkiewicz R, Holt RD. When does evolution by natural selection prevent extinction? Evolution. 1995;49:201.

Gorbunova V, Seluanov A, Zhang Z, Gladyshev VN, Vijg J. Comparative genetics of longevity and cancer: insights from long-lived rodents. Nat Rev Genet. 2014;15:531–40. https://doi.org/10.1038/nrg3728.

Grueber CE. Comparative genomics for biodiversity conservation. Comput Struct Biotechnol J. 2015;13:370–5. https://doi.org/10.1016/j.csbj.2015.05.003.

Grueber CE, Peel E, Wright B, Hogg CJ, Belov K. A Tasmanian devil breeding program to support wild recovery. Reprod Fertil Dev. 2018; https://doi.org/10.1071/RD18152.

Grunau C. The epigenetic component in cancer evolution. In: Ujvari B, Roche B, Thomas F, editors. Ecology and evolution of cancer. Amsterdam: Elsevier; 2017. p. 87–98.

Guiler ER. Observations on the Tasmanian devil, *Sarcophilus harrisii* (Dasyuridae: Marsupiala) at Granville Harbour. Papers Proc R Soc Tasmania. 1978;112:161–88.

Guo S, DiPietro LA. Factors affecting wound healing. J Dent Res. 2010;89:219–29. https://doi.org/10.1177/0022034509359125.

Hamede R, Lachish S, Belov K, Woods G, Kreiss A, Pearse A-M, et al. Reduced effect of Tasmanian devil facial tumor disease at the disease front. Conserv Biol. 2012;26:124–34. https://doi.org/10.1111/j.1523-1739.2011.01747.x.

Hamede RK, Pearse A-M, Swift K, Barmuta LA, Murchison EP, Jones ME. Transmissible cancer in Tasmanian devils: localized lineage replacement and host population response. Proc Biol Sci. 2015;2015282:20151468. https://doi.org/10.1098/rspb.2015.1468.

Hamede R, Owen R, Siddle H, Peck A, Jones M, Dujon AM, et al. The ecology and evolution of wildlife cancers: applications for management and conservation. Evol Appl. 2020; https://doi.org/10.1111/eva.12948.

Hamilton DG, Jones ME, Cameron EZ, McCallum H, Storfer A, Hohenlohe PA, et al. Rate of intersex interactions affects injury likelihood in Tasmanian devil contact networks. Behav Ecol. 2019;30:1087–95.

Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70. https://doi.org/10.1016/S0092-8674(00)81683-9.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74. https://doi.org/10.1016/j.cell.2011.02.013.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, et al. Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. Biol Conserv. 2006;131:307–24. https://doi.org/10.1016/j.biocon.2006.04.010.

Hendricks S, Epstein B, Schönfeld B, Wiench C, Hamede R, Jones M, et al. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (*Sarcophilus harrisii*). Conserv Genet. 2017;18:977–82. https://doi.org/10.1007/s10592-017-0939-5.

Heng HH. The genomic landscape of cancers. In: Ujvari B, Roche B, Thomas F, editors. Ecology and evolution of cancer. Amsterdam: Elsevier; 2017. p. 69–86.

Heng HH, Bremer SW, Stevens JB, Horne SD, Liu G, Abdallah BY, et al. Chromosomal instability (CIN): what it is and why it is crucial to cancer evolution. Cancer Metastasis Rev. 2013;32:325–40. https://doi.org/10.1007/s10555-013-9427-7.

Himmel L, Cianciolo R. Nodular typhlocolitis, heterakiasis, and mesenchymal neoplasia in a ring-necked pheasant (*Phasianus colchicus*) with immunohistochemical characterization of visceral metastases. J Vet Diagn Invest. 2017;29:561–5.

Hoelzel AR, Bruford MW, Fleischer RC. Conservation of adaptive potential and functional diversity. Conserv Genet. 2019;20:1–5.

Hofman P, Vouret-Craviari V. Microbes-induced EMT at the crossroad of inflammation and cancer. Gut Microbes. 2012;3:176–85. https://doi.org/10.4161/gmic.20288.

Hogg CJ, Ivy JA, Srb C, Hockley J, Lees C, Hibbard C, et al. Influence of genetic provenance and birth origin on productivity of the Tasmanian devil insurance population. Conserv Genet. 2015;16:1465–73.

Hogg CJ, Wright B, Morris KM, Lee AV, Ivy JA, Grueber CE, et al. Founder relationships and conservation management: empirical kinships reveal the effect on breeding programmes when founders are assumed to be unrelated. Anim Conserv. 2018; https://doi.org/10.1111/acv.12463.

Hohenlohe PA, McCallum HI, Jones ME, Lawrance MF, Hamede RK, Storfer A. Conserving adaptive potential: lessons from Tasmanian devils and their transmissible cancer. Conserv Genet. 2019;20:81–7. https://doi.org/10.1007/s10592-019-01157-5.

Hohenlohe PA, Funk WC, Rajora OP. Population genomics for wildlife conservation and management. Authorea. 2020; https://doi.org/10.22541/au.158480040.06912807.

Hubert J-N, Zerjal T, Hospital F. Cancer- and behavior-related genes are targeted by selection in the Tasmanian devil (*Sarcophilus harrisii*). PLoS One. 2018;13:e0201838. https://doi.org/10.1371/journal.pone.0201838.

Hufbauer RA, Szűcs M, Kasyon E, Youngberg C, Koontz MJ, Richards C, et al. Three types of rescue can avert extinction in a changing environment. Proc Natl Acad Sci U S A. 2015;112:10557–62. https://doi.org/10.1073/pnas.1504732112.

ICGC/TGCA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93. https://doi.org/10.1038/s41586-020-1969-6.

Irigaray P, Newby JA, Clapp R, Hardell L, Howard V, Montagnier L, et al. Lifestyle-related factors and environmental agents causing cancer: an overview. Biomed Pharmacother. 2007;61:640–58. https://doi.org/10.1016/j.biopha.2007.10.006.

James S, Jennings G, Kwon YM, Stammnitz M, Fraik A, Storfer A, et al. Tracing the rise of malignant cell lines: distribution, epidemiology and evolutionary interactions of two transmissible cancers in Tasmanian devils. Evol Appl. 2019;12:1772–80. https://doi.org/10.1111/eva.12831.

Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128:683–92. https://doi.org/10.1016/j.cell.2007.01.029.

Jones ME, Paetkau D, Geffen E, Moritz C. Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. Mol Ecol. 2004;13:2197–209. https://doi.org/10.1111/j.1365-294X.2004.02239.x.

Jones ME, Cockburn A, Hamede R, Hawkins C, Hesterman H, Lachish S, et al. Life-history change in disease-ravaged Tasmanian devil populations. Proc Natl Acad Sci U S A. 2008;105:10023–7.

Jones ME, Hamede RK, Storfer A, Hohenlohe P, Murchison EP, McCallum H. Emergence, transmission and evolution of an uncommon enemy: Tasmanian devil facial tumour disease. In: Wilson K, Fenton A, Tompkins D, editors. Wildlife disease ecology: linking theory to data and application. Cambridge: Cambridge University Press; 2019. p. 321–41.

Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity (Edinb). 2015;115:63–72. https://doi.org/10.1038/hdy.2015.17.

Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9:1205–18. https://doi.org/10.1111/eva.12414.

Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature. 1997;386:761–3. https://doi.org/10.1038/386761a0.

Kosack L, Wingelhofer B, Popa A, Orlova A, Agerer B, Vilagos B, et al. The ERBB-STAT3 axis drives Tasmanian devil facial tumor disease. Cancer Cell. 2019;35:125–39. https://doi.org/10.1016/j.ccell.2018.11.018.

Kreiss A, Cheng Y, Kimble F, Wells B, Donovan S, Belov K, et al. Allorecognition in the Tasmanian devil (*Sarcophilus harrisii*), an endangered marsupial species with limited genetic diversity. PLoS One. 2011;6:e22402. https://doi.org/10.1371/journal.pone.0022402.

Kwon YM, Stammnitz MR, Wang J, Swift K, Knowles GW, Pye RJ, et al. Tasman-PCR: a genetic diagnostic assay for Tasmanian devil facial tumour diseases. R Soc Open Sci. 2018;5:180870. https://doi.org/10.1098/rsos.180870.

Lachish S, McCallum HI, Jones ME. Demography, disease and the devil: life-history changes in a disease-affected population of Tasmanian devils (*Sarcophilus harrisii*). J Anim Ecol. 2009;78:427–36. https://doi.org/10.1111/j.1365-2656.2008.01494.x.

Lane A, Cheng Y, Wright B, Hamede R, Levan L, Jones M, et al. New insights into the role of MHC diversity in devil facial tumor disease. PLoS One. 2012;7:e36955. https://doi.org/10.1371/journal.pone.0036955.

Lazenby BT, Tobler MW, Brown WE, Hawkins CE, Hocking GJ, Hume F, et al. Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. J Appl Ecol. 2018;55:1368–79. https://doi.org/10.1111/1365-2664.13088.

Leathlobhair MN, Gulland FMD, Murchison EP. No evidence for clonal transmission of urogenital carcinoma in California sea lions (*Zalophus californianus*). Wellcome Open Res. 2017;2:46. https://doi.org/10.12688/wellcomeopenres.11483.1.

Leroi AM, Koufopanou V, Burt A. Cancer selection. Nat Rev Cancer. 2003;3:226–31. https://doi.org/10.1038/nrc1016.

Leroy G, Carroll EL, Bruford MW, DeWoody JA, Strand A, Waits L, et al. Next-generation metrics for monitoring genetic erosion within populations of conservation concern. Evol Appl. 2018;11:1066–83. https://doi.org/10.1111/eva.12564.

Letcher RJ, Norstrom RJ, Muir D, Sandau K. Methylsulfone PCB and DDE metabolites in beluga whale (*Delphinapterus leucas*) from the St. Lawrence River estuary and western Hudson Bay. Environ Toxicol Chem. 2000;19:1378–88.

Lipscomb TP, Scott DP, Garber RL, Krafft AE, Tsai MM, Lichy JH, et al. Common metastatic carcinoma of California sea lions (*Zalophus californianus*): evidence of genital origin and association with novel gammaherpesvirus. Vet Pathol. 2016;37:609–17. https://doi.org/10.1354/vp.37-6-609.

Loh R, Hayes D, Mahjoor A, O'Hara A, Pyecroft S, Raidal S. The immunohistochemical characterization of devil facial tumor disease (DFTD) in the Tasmanian devil (*Sarcophilus harrisii*). Vet Pathol. 2016;43:896–903. https://doi.org/10.1354/vp.43-6-896.

MacRae SL, Zhang Q, Lemetre C, Seim I, Calder RB, Hoeijmakers J, et al. Comparative analysis of genome maintenance genes in naked mole rat, mouse, and human. Aging Cell. 2015;14:288–91. https://doi.org/10.1111/acel.12314.

Madsen T, Arnal A, Vittecoq M, Bernex F, Abadie J, Labrut S, et al. Cancer prevalence and etiology in wild and captive animals. In: Ujvari B, Roche B, Thomas F, editors. Ecology and evolution of cancer. Amsterdam: Elsevier; 2017. p. 11–46. https://doi.org/10.1016/B978-0-12-804310-3.00002-8.

Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet. 2006;38:468–73. https://doi.org/10.1038/ng1768.

Margres MJ, Jones ME, Epstein B, Kerlin DH, Comte S, Fox S, et al. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. Mol Ecol. 2018a;27:4189–99. https://doi.org/10.1111/mec.14853.

Margres MJ, Ruiz-Aravena M, Hamede R, Jones ME, Lawrance MF, Hendricks SA, et al. The genomic basis of tumor regression in Tasmanian devils (*Sarcophilus harrisii*). Genome Biol Evol. 2018b;10:3012–25. https://doi.org/10.1093/gbe/evy229.

Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349:1483–9. https://doi.org/10.1126/science.aab4082.

Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171:1029–41. https://doi.org/10.1016/j.cell.2017.09.042.

Martineau D, Lemberger K, Dallaire A, Labelle P, Lipscomb TP, Michel P, et al. Cancer in wildlife, a case study: beluga from the St. Lawrence estuary, Québec, Canada. Environ Health Perspect. 2002;110:285–92.

McAloose D, Newton AL. Wildlife cancer: a conservation perspective. Nat Rev Genet. 2009;9:517–26. https://doi.org/10.1038/nrc2665.

McCallum HI, Jones ME. Infectious cancers in wildlife. In: Aguire AA, Ostfeld RS, Daszak P, editors. Emerging infectious diseases and conservation medicine. New York: Oxford University Press; 2012. p. 270–83.

McCallum HI, Jones ME, Hawkins C, Hamede R, Lachish S, Sinn DL, et al. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology. 2009;90:3379–92.

McLennan EA, Gooley RM, Wise P, Belov K, Hogg CJ, Grueber CE. Pedigree reconstruction using molecular data reveals an early warning sign of gene diversity loss in an island population of Tasmanian devils (*Sarcophilus harrisii*). Conserv Genet. 2018;19:439–50.

Merlo LMF, Maley CC. The role of genetic diversity in cancer. J Clin Invest. 2010;120:401–3. https://doi.org/10.1172/JCI42088.

Metzger MJ, Goff SP. A sixth modality of infectious disease: contagious cancer from devils to clams and beyond. PLoS Pathog. 2016;12:e1005904. https://doi.org/10.1371/journal.ppat.1005904.

Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, et al. Widespread transmission of independent cancer lineages within multiple bivalve species. Nature. 2016;534:705–9.

Metzger MJ, Paynter AN, Siddall ME, Goff SP. Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. Proc Nat Acad Sci U S A. 2018;115:E4227–35. https://doi.org/10.1073/pnas.1717227115.

Michod RE. Darwinian dynamics: evolutionary transitions in fitness and individuality. Princeton: Princeton University Press; 2000.

Miller W, Hayes VM, Ratan A. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). Proc Natl Acad Sci. 2011;108:12348–53. https://doi.org/10.1073/pnas.1102838108.

Moriarty ME, Vickers TW, Clifford DL, Garcelon DK, Gaffney PM, Lee KW, et al. Ear mite removal in the Santa Catalina Island fox (*Urocyon littoralis catalinae*): controlling risk factors for cancer development. PLoS One. 2015;10:e0144271. https://doi.org/10.1371/journal.pone.0144271.

Mousseau TA, Møller AP. Landscape portrait: a look at the impacts of radioactive contaminants on Chernobyl's wildlife. Bull At Sci. 2015;67:38–46. https://doi.org/10.1177/0096340211399747.

Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, Rebbeck CA, et al. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science. 2010;327:84–7. https://doi.org/10.1126/science.1180616.

Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell. 2012;148:780–91. https://doi.org/10.1016/j.cell.2011.11.065.

Murchison EP, Wedge DC, Alexandrov LB, Fu B, Martincorena I, Neng Z, et al. Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage. Science. 2014;343:437–40. https://doi.org/10.1126/science.1247167.

Murshid A, Gong J, Calderwood SK. Heat-shock proteins in cancer vaccines: agents of antigen cross-presentation. Expert Rev Vaccines. 2008;7:1019–30. https://doi.org/10.1586/14760584.7.7.1019.

Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. Oncogene. 2004;23:6445–70. https://doi.org/10.1038/sj.onc.1207714.

Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability – an evolving hallmark of cancer. Nat Rev Mol Cell Biol. 2010;11:220–8. https://doi.org/10.1038/nrm2858.

Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A. 1997;94:7799–806.

Nowell PC. Tumor progression: a brief historical perspective. Semin Cancer Biol. 2002;12:261–6.

Oh KP, Aldridge CL, Forbey JS, Dadabay CY, Oyler-McCance SJ. Conservation genomics in the sagebrush sea: population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). Genome Biol Evol. 2019;11:2023–34. https://doi.org/10.1093/gbe/evz112.

Ostrander EA, Davis BW, Ostrander GK. Transmissible tumors: breaking the cancer paradigm. Trends Genet. 2016;32:1–15. https://doi.org/10.1016/j.tig.2015.10.001.

Owen RS, Siddle HV. Devil facial tumours: towards a vaccine. Immunol Invest. 2019; https://doi.org/10.1080/08820139.2019.1624770.

Park SY, Gönen M, Kim HJ, Michor F, Polyak K. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. J Clin Invest. 2010;120:636–44. https://doi.org/10.1172/JCI40724.

Patchett A, Woods G. Targeting transmissible cancers in animals. Science. 2019;365:438–41. https://doi.org/10.1126/science.aau8936.

Patchett AL, Tovar C, Corcoran LM, Lyons AB, Woods GM. The toll-like receptor ligands Hiltonol (polyICLC) and imiquimod effectively activate antigen-specific immune responses in Tasmanian devils (*Sarcophilus harrisii*). Devel Comp Immun. 2017;76:352–60. https://doi.org/10.1016/j.dci.2017.07.004.

Patchett AL, Flies AS, Lyons AB, Woods GM. Curse of the devil: molecular insights into the emergence of transmissible cancers in the Tasmanian devil (*Sarcophilus harrisii*). Cell Mol Life Sci. 2020; https://doi.org/10.1007/s00018-019-03435-4.

Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, et al. Contemporary demographic reconstruction methods are robust to genome assembly quality: a case study in Tasmanian devils. Mol Biol Evol. 2019;36:2906–21. https://doi.org/10.1093/molbev/msz1919.

Pearse AM, Swift K. Allograft theory: transmission of devil facial-tumour disease. Nature. 2006;439:549. https://doi.org/10.1038/439549a.

Pearse A-M, Swift K, Hodson P, Hua B, McCallum H, Pyecroft S, et al. Evolution in a transmissible cancer: a study of the chromosomal changes in devil facial tumor (DFT) as it spreads through the wild Tasmanian devil population. Cancer Genet. 2012;205:101–12. https://doi.org/10.1016/j.cancergen.2011.12.001.

Perera FP. Molecular epidemiology of environmental carcinogenesis. In: Schwab M, Rabes HM, Munk K, Hofschneider P, editors. Recent results in cancer research. Berlin: Springer International Publishing AG, part of Springer Nature; 1998. p. 39–46.

Perret C, Gidoin C, Ujvari B, Thomas F, Roche B. Predation shapes the impact of cancer on population dynamics and the evolution of cancer resistance. Evol Appl. 2020; https://doi.org/10.1111/eva.12951.

Pesavento PA, Agnew D, Keel MK, Woolard KD. Cancer in wildlife: patterns of emergence. Nat Rev Cancer. 2018;18:646–61. https://doi.org/10.1038/s41568-018-0045-0.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016;26:342–50. https://doi.org/10.1101/gr.193474.115.

Pye RJ, Pemberton D, Tovar C, Tubio JMC, Dun KA, Fox S, et al. A second transmissible cancer in Tasmanian devils. Proc Natl Acad Sci U S A. 2016;113:374–9. https://doi.org/10.1073/pnas.1519691113.

Pye R, Patchett A, McLennan E, Thompson R, Carver S, Fox S, et al. Immunization strategies producing a humoral IgG immune response against devil facial tumor disease in the majority of

Tasmanian devils destined for wild release. Front Immunol. 2018;9:259. https://doi.org/10.3389/fimmu.2018.00259.

Pyecroft SB, Pearse A-M, Loh R, Swift K, Belov K, Fox N, et al. Towards a case definition for devil facial tumour disease: what is it? Ecohealth. 2007;4:346–51. https://doi.org/10.1007/s10393-007-0126-0.

Ralls K, Ballou JD, Rideout B, Frankham R. Genetic management of chondrodystrophy in California condors. Anim Conserv Forum. 2000;3:145–53.

Reiche EMV, Nunes SOV, Morimoto HK. Stress, depression, the immune system, and cancer. Lancet Oncol. 2004;5:617–25. https://doi.org/10.1016/S1470-2045(04)01597-9.

Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020;578:102–11. https://doi.org/10.1038/s41586-020-1965-x.

Risques RA, Promislow DEL. All's well that ends well: why large species have short telomeres. Philos Trans R Soc B. 2018;373:20160448. https://doi.org/10.1098/rstb.2016.0448.

Robinson JA, Vecchyo DO-D, Fan Z, Kim BY, vonHoldt BM, Marsden CM, et al. Genomic flatlining in the endangered island fox. Curr Biol. 2016;26:1183–9. https://doi.org/10.1016/j.cub.2016.02.062.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in Island Foxes. Curr Biol. 2018;28:3487–94. https://doi.org/10.1016/j.cub.2018.08.066.

Romanov MN, Tuttle EM, Houck ML, Modi WS, Chemnick LG, Korody ML, et al. The value of avian genomics to the conservation of wildlife. BMC Genomics. 2009;10:S10. https://doi.org/10.1186/1471-2164-10-S2-S10.

Salkeld DJ. Vaccines for conservation: plague, prairie dogs & black-footed ferrets as a case study. Ecohealth. 2017;14:432–7. https://doi.org/10.1007/s10393-017-1273-6.

Sarkies P, Sale JE. Cellular epigenetic stability and cancer. Trends Genet. 2012;28:118–27.

Schiffman JD, Breen M. Comparative oncology: what dogs and other species can teach us about humans with cancer. Philos Trans R Soc B. 2015;370:20140231. https://doi.org/10.1098/rstb.2014.0231.

Seluanov A, Gladyshev VN, Vijg J, Gorbunova V. Mechanisms of cancer resistance in long-lived animals. Nat Rev Cancer. 2018;18:433–41. https://doi.org/10.1038/s41568-018-0004-9.

Shields PG, Harris CC. Cancer risk and low-penetrance susceptibility genes in gene-environment interactions. J Clin Oncol. 2000;18:2309–15. https://doi.org/10.1200/JCO.2000.18.11.2309.

Siddle HV, Kreiss A, Eldridge MDB, Noonan E, Clarke CJ, Pyecroft S, et al. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proc Natl Acad Sci U S A. 2007;104:16221–6. https://doi.org/10.1073/pnas.0704580104.

Siddle HV, Marzec J, Cheng Y, Jones M, Belov K. MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. Proc R Soc Lond B Biol Sci. 2010;277:2001–6. https://doi.org/10.1098/rspb.2009.2362.

Siddle HV, Kreiss A, Tovar C, Yuen CK, Cheng Y, Belov K, et al. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. Proc Natl Acad Sci U S A. 2013;110:5103–8. https://doi.org/10.1073/pnas.1219920110.

Smith KR, Hanson HA, Mineau GP, Buys SS. Effects of *BRCA1* and *BRCA2* mutations on female fertility. Proc R Soc Lond B Biol Sci. 2012;279:1389–95. https://doi.org/10.1098/rspb.2011.1697.

Soto AM, Sonnenschein C. Environmental causes of cancer: endocrine disruptors as carcinogens. Nat Rev Endocrinol. 2010;6:363–70. https://doi.org/10.1038/nrendo.2010.87.

Stammnitz MR, Coorens THH, Gori KC, Hayes D, Fu B, Wang J, et al. The origins and vulnerabilities of two transmissible cancers in Tasmanian devils. Cancer Cell. 2018;33:607–619.e15. https://doi.org/10.1016/j.ccell.2018.03.013.

Stephens PJ, McBride DJ, Lin M-L, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009;462:1005–10. https://doi.org/10.1038/nature08645.

Stoler DL, Chen N, Basik M, Kahlenberg MS, Rodriguez-Bigas MA, Petrelli NJ, et al. The onset and extent of genomic instability in sporadic colorectal tumor progression. Proc Natl Acad Sci U S A. 1999;96:15121–6. https://doi.org/10.1073/pnas.96.26.15121.

Storfer A, Epstein B, Jones M, Micheletti S, Spear SF, Lachish S, et al. Landscape genetics of the Tasmanian devil: implications for spread of an infectious cancer. Conserv Genet. 2017;18:1287–97. https://doi.org/10.1007/s10592-017-0980-4.

Storfer A, Hohenlohe PA, Margres MJ, Patton A, Fraik AK, Lawrance M, et al. The devil is in the details: genomics of transmissible cancers in Tasmanian devils. PLoS Pathog. 2018;14:e1007098. https://doi.org/10.1371/journal.ppat.1007098.

Storfer A, Kozakiewicz CP, Beer MA, Savage AE. Applications of population genomics for understanding and mitigating wildlife disease. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2020. https://doi.org/10.1007/13836_2020_73.

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, et al. TP53copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. Elife. 2016;5:1850–30. https://doi.org/10.7554/eLife.11994.

Taylor RL, Zhang Y, Schöning JP, Deakin JE. Identification of candidate genes for devil facial tumour disease tumourigenesis. Sci Rep. 2017;7:8761. https://doi.org/10.1038/s41598-017-08908-9.

Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW. Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss. BMC Genomics. 2008;9:62. https://doi.org/10.1186/1471-2164-9-62.

Thomas R, Kareva I, Raven N, Hamede R, Pujol P, Roche B, et al. Evolved dependence in response to cancer. Trends Ecol Evol. 2018;33:269–76. https://doi.org/10.1016/j.tree.2018.01.012.

Thomas F, Giraudeau M, Gouzerh F, Boutry J, Renaud F, Pujol P, et al. The evolution of resistance and tolerance as cancer defenses. Parasitology. 2020;147:255–62. https://doi.org/10.1017/S0031182019001501.

Tian X, Doerig K, Park R, Qin ACR, Hwang C, Neary A, et al. Evolution of telomere maintenance and tumour suppressor mechanisms across mammals. Philos Trans R Soc B. 2018;373:20160443–10. https://doi.org/10.1098/rstb.2016.0443.

Tollis M, Hutchins ED, Kusumi K. Reptile genomes open the frontier for comparative analysis of amniote development and regeneration. Int J Dev Biol. 2015;58:863–71. https://doi.org/10.1387/ijdb.140316kk.

Tollis M, Schiffman JD, Boddy AM. Evolution of cancer suppression as revealed by mammalian comparative genomics. Curr Opin Genet Dev. 2017;42:40–7. https://doi.org/10.1016/j.gde.2016.12.004.

Tollis M, Robbins J, Webb AE, Kuderna LFK, Caulin AF, Garcia JD, et al. Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (*Megaptera novaeangliae*). Mol Biol Evol. 2019;36:1746–63. https://doi.org/10.1093/molbev/msz099.

Tovar C, Obendorf D, Murchison EP, Papenfuss AT, Kreiss A, Woods GM. Tumor-specific diagnostic marker for transmissible facial tumors of Tasmanian devils. Vet Pathol. 2011;48:1195–203. https://doi.org/10.1177/0300985811400447.

Tovar C, Pye RJ, Kreiss A, Cheng Y, Brown GK, Darby J, et al. Regression of devil facial tumour disease following immunotherapy in immunised Tasmanian devils. Sci Rep. 2017;7:43827. https://doi.org/10.1038/srep43827.

Tovar C, Patchett AL, Kim V, Wilson R, Darby J, Lyons AB, et al. Heat shock proteins expressed in the marsupial Tasmanian devil are potential antigenic candidates in a vaccine against devil facial tumour disease. PLoS One. 2018;13:e0196469. https://doi.org/10.1371/journal.pone.0196469.

Ujvari B, Fitzpatrick J, Raven N, Osterkamp J, Thomas F. The ecology of cancer. In: Bernicker EH, editor. Cancer and society. Cham: Springer International Publishing AG, part of Springer Nature; 2019. p. 153–74.

Vazquez AI, de los Campos G, Klimentidis YC, Rosa GJM, Gianola D, Yi N, et al. A comprehensive genetic approach for improving prediction of skin cancer risk in humans. Genetics. 2012;192:1493–502. https://doi.org/10.1534/genetics.112.141705.

Vickers TW, Clifford DL, Garcelon DK, King JL, Duncan CL, Gaffney PM, et al. Pathology and epidemiology of ceruminous gland tumors among endangered Santa Catalina Island Foxes (*Urocyon littoralis catalinae*) in the Channel Islands, USA. PLoS One. 2015;10:e0143211–8. https://doi.org/10.1371/journal.pone.0143211.

Vittecoq M, Roche B, Daoust SP, Ducasse H, Missé D, Abadie J, et al. Cancer: a missing link in ecosystem functioning? Trends Ecol Evol. 2013;28:628–35.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science. 2013;339:1546–58. https://doi.org/10.1126/science.1235122.

Walters AD, Schwartz MK. Population genomics for the management of wild vertebrate populations. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply; 2020. https://doi.org/10.1007/13836_2020_72.

Walters JR, Derrickson SR, Fry DM, Haig SM, Marzluff JM, Wunderle JM. Status of the California condor (*Gymnogyps californianus*) and efforts to achieve its recovery. Auk. 2010;127:969–1001. https://doi.org/10.1525/auk.2010.127.4.969.

Wells K, Hamede RK, Kerlin DH, Storfer A, Hohenlohe PA, Jones ME, et al. Infection of the fittest: devil facial tumour disease has greatest effect on individuals with highest reproductive output. Ecol Lett. 2017;20:770–8. https://doi.org/10.1111/ele.12776.

Wells K, Hamede RK, Jones ME, Hohenlohe PA, Storfer A, McCallum HI. Individual and temporal variation in pathogen load predicts long-term impacts of an emerging infectious disease. Ecology. 2019;100:e02613. https://doi.org/10.1002/ecy.2613.

Williams TD, Pulley LT. Leiomyomas in two sea otters, *Enhydra lutris*. J Wildl Dis. 1981;17:401–4.

Wright B, Morris K, Grueber CE, Willet CE, Gooley R, Hogg CJ, et al. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. BMC Genomics. 2015;16:791–11. https://doi.org/10.1186/s12864-015-2020-4.

Wright B, Willet CE, Hamede R, Jones M, Belov K, Wade CM. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. Sci Rep. 2017;7:423. https://doi.org/10.1038/s41598-017-00439-7.

Yablokov AV. Chernobyl's radioactive impact on fauna. Ann N Y Acad Sci. 2009;1181:255–80. https://doi.org/10.1111/j.1749-6632.2009.04833.x.

Yonemitsu MA, Giersch RM, Polo-Prieto M, Hammel M, Simon A, Cremonte F, et al. A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. Elife. 2019;8:e47788. https://doi.org/10.7554/elife.47788.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45:1134–40. https://doi.org/10.1038/ng.2760.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346:1311–20. https://doi.org/10.1126/science.1251385.

# Part V
# Wildlife Conservation and Management

# Population Genomics for the Management of Wild Vertebrate Populations


**Ashley D. Walters and Michael K. Schwartz**

**Abstract** Management of genetic variation in natural populations is necessary to mitigate the effects of environmental change and biodiversity loss. While traditional genetics techniques have aided management of biodiversity, the rapid advancement of genomics technology has provided advancements for the field of biodiversity conservation including increased resolution and identification of adaptive loci corresponding to ecologically relevant phenotypes. In this chapter, we explore how population genomics has been implemented into wildlife management via a literature review and discussions with management and conservation agencies. We discuss the future prospects of population genomics applications in biodiversity conservation and management of wild populations. Overall, we see the potential for population genomics to improve our understanding of wild populations of fish and wildlife and have several important examples that are paving the way for the adoption of these technologies into management. However, there has been a severe lag in the implementation of population genomics data into management decisions, likely due to the length of the research cycle and the slow absorption into regulation and policy.

**Keywords** Biodiversity conservation · Evolutionary significant unit · Management unit · Population genomics · Wildlife management

## 1 Introduction and General Overview

Genetic diversity is a fundamental component of population fitness and adaptive potential (Reed and Frankham 2003; Hoffman et al. 2017); therefore, management of genetic variation in wild populations of wildlife and fish is necessary to mitigate

A. D. Walters (✉) · M. K. Schwartz
US Department of Agriculture, Forest Service, Rocky Mountain Research Station, National Genomics Center for Wildlife and Fish Conservation, Missoula, MT, USA
e-mail: ashley.walters@usda.gov

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)], https://doi.org/10.1007/13836_2020_72, 

the effects of a changing environment. Conservation genetics provides both a theoretical framework and an effective tool for addressing contemporary conservation and wildlife management issues. Traditional genetic techniques have aided management and conservation efforts on multiple fronts including biodiversity monitoring, resolving taxonomic uncertainty, wildlife forensics, and designation of conservation units. However, the rapid advancement of genomics technologies, such as the generation of large genetic marker panels and whole genome-sequencing for non-model organisms, in recent decades has provided important advancements in the field and improved traditional genetic inferences in two key areas: increased resolution and identification of ecologically important loci (Fig. 1). The population genomics approach offers a dramatic increase in the number of variable genetic markers used compared to traditional genetic methods, therefore, improving the precision of estimating diversity and population demographic parameters, connectivity, and designation of conservation units.

While genetics has provided information regarding neutral evolutionary processes for management efforts, the application of genomics increases the power to detect signatures of divergent selection and local adaptation. Preserving adaptive potential has been recognized as an important component of ensuring long-term persistence, especially in the face of global climate change (Sgró et al. 2011; Harrisson et al. 2014). From the perspective of wildlife and fisheries management, understanding how wild populations are adapted to their ecosystem can aid in a multitude of traditional wildlife management activities including translocations, reintroductions, and population augmentations as well as identification of units of conservation concern. For example, one of the first questions addressed prior to most reintroduction efforts is the assessment of the most suitable source population. Using population genomics to evaluate functional differences among populations will ultimately lead to better translocation decisions and hopefully more successful management efforts.

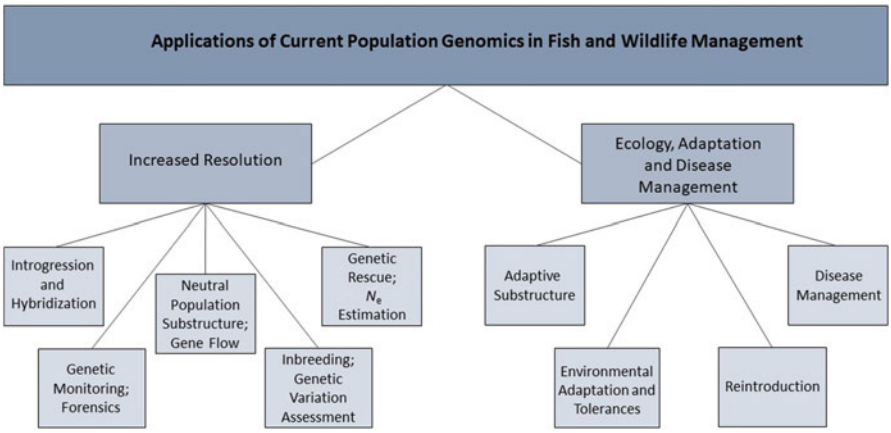


**Fig. 1** Current applications for population genomics in fish and wildlife management

Currently, one of the most important uses of population genomics is in preserving evolutionary processes (Allendorf et al. 2010; Funk et al. 2012; Hohenlohe et al. 2019) which support future biodiversity under changing environmental conditions (Mortiz 2002). The genomics approach helps accomplish this goal by providing the enhanced power of many more loci, or by aiding in the understanding of how populations are adapted to local environments to designate conservation units (Allendorf et al. 2010; Funk et al. 2012), However, there is ongoing debate about the relative importance of neutral versus adaptive variation in the designation of conservation units (e.g., Crandall et al. 2000; Fraser and Bernatchez 2001; Pearse 2016). In a conservation context, two types of units have been defined: Evolutionary Significant Units (ESUs) and Management Units (MUs; Mortiz 1994). An ESU is generally defined as a population or group of populations with high genetic and ecological distinctiveness, maximizing evolutionary potential (Fraser and Bernatchez 2001). However, MUs were originally defined as populations with significant divergence of allele frequencies at either nuclear or mitochondrial loci, regardless of their uniqueness phylogeographically. This concept thus focuses on demographic independence among populations, therefore, preserving genetic diversity of intraspecific units. Funk et al. (2012) proposed a method for identifying appropriate units for management and conservation involving three steps: (1) delineating ESUs using all loci, (2) recognizing MUs via non-outlier loci, and (3) identifying adaptive variation among MUs through the use of outlier loci.

The identification of adaptive variation may not alter conservation and management strategies if concordant with patterns in neutral variation; however, there have been instances of incongruence between the two marker types (i.e., Pacific salmon; Prince et al. 2017; Fig. 2). The underlying controversy of a gene-targeted approach is based on the knowledge of correlative genomic variation responsible for adaptive variation and the trade-off and preservation of future adaptive potential (i.e., genomic diversity). There has been substantial argument on the prioritization of populations based on a specific allele(s) because the resulting conservation effort may reduce total genomic diversity and ultimately limit future adaptive potential (Luikart et al. 2003, 2019; Allendorf 2017; Kardos and Shafer 2018). Conservation geneticists do not advocate for managing only outlier loci, consistent with historical conservation debates over conserving allelic diversity for a set of disease resistance genes (major histocompatibility complex) known to be under selection (Vrijenhoek and Leberg 1991). In cases where only putative genetic contributions to adaptive variation can be identified without a causal mechanism, management frameworks advocate for the incorporation of additional information including neutral genetic diversity and/or important phenotypic and ecological variation (Funk et al. 2012, 2019; Shafer et al. 2015; Benestan et al. 2016; Allendorf 2017). Therefore, genomic signatures of selection are most useful in a management framework when combining variation in conservation-relevant phenotypes and loci of known function with signatures of selection (Epstein et al. 2016; Wright et al. 2017).

While the promise of genomics for conservation has been looming for a decade and has been evident in some areas of wildlife management (Schwartz et al. 2010), we explored how population genomics has been implemented in wildlife

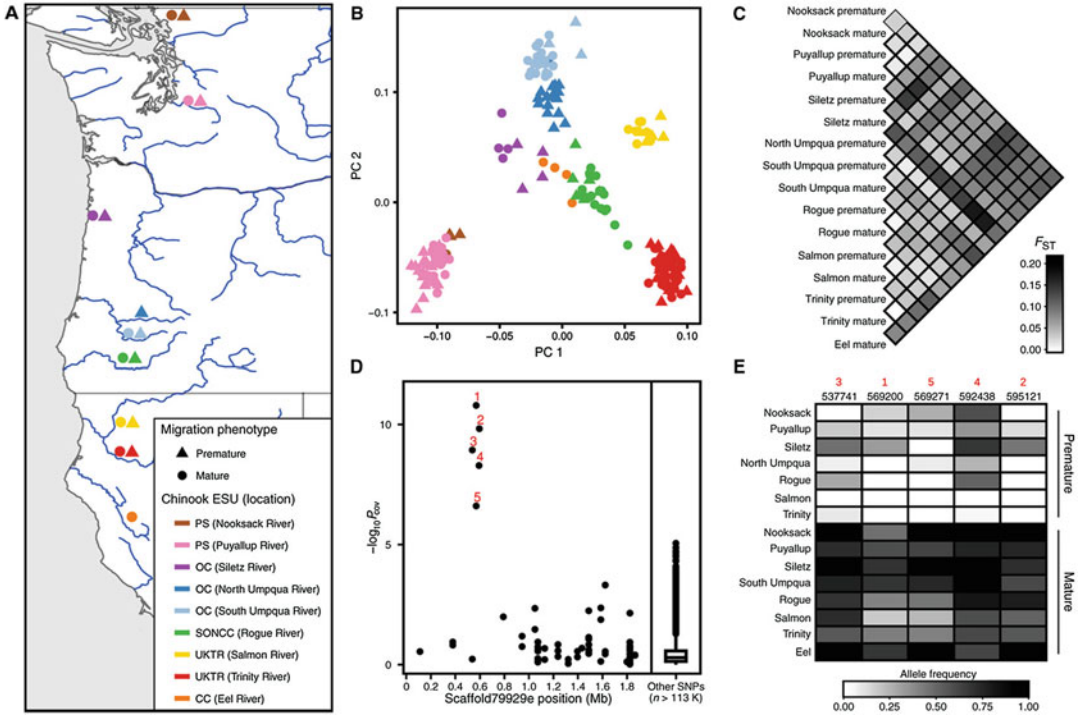


**Fig. 2** Genetic and evolutionary basis of premature migration in Chinook salmon. (**a**) Map of sample locations and migration phenotypes; color indicates location and shape indicates migration category. (**b**) PCA and (**c**) pairwise $F_{ST}$ estimates using genome-wide SNP data. (**d**) Association mapping of migration category in Chinook; red numbers indicate significant SNPs. (**e**) Allele frequency shift at significant SNPs between premature and mature migrating populations. Black numbers indicate SNP position on scaffold79929e. Reprinted with permission of AAAS from Prince et al. (2017). 

management. We searched the literature for uses of population genomic data in wildlife and fish management. We also sent out a survey on The Ecological Society of America (ECOLOG-L) listserv to see if there were forthcoming uses of population genomic data that we missed in the literature. This allowed us to examine the extent that population genomics has been integrated into wildlife and fisheries management.

## 2 Applications of Population Genomics in Fish and Wildlife Management

Below we describe the common uses of population genomics data in wildlife and fisheries management.

### *2.1 Increased Resolution*

#### 2.1.1 Introgression/Hybridization

As exploitation of wild living resources becomes increasingly unsustainable (Hutchings 2000; Myers and Worm 2003), domestication and captive production provide an opportunity to supplement economically important populations. However, release or escape of captively bred individuals into the wild population has the potential to negatively impact genetics of wild populations. Hybridization between wild species and their domesticated relatives has been acknowledged in a variety of organisms (Randi 2008) but has been particularly common in fishes.

Management efforts may include the deliberate supplementation of wild populations with the stocking of hatchery-reared offspring (Laikre et al. 2010). Current steelhead, *Oncorhynchus mykiss*, hatchery programs in northern Puget Sound are designed as segregated programs where hatchery and wild populations are deliberately kept separate. The hatchery broodstock is restricted to hatchery-origin individuals, and the reproductive interaction between hatchery-origin and wild fish is limited. Segregated program efforts are aimed at producing a sufficient number of hatchery-origin individuals to accommodate the next generation of hatchery broodstock and support recreational harvest opportunities. Regulations require monitoring for the presence of naturally spawning hatchery fish and the reproductive interaction between hatchery-origin individuals and the wild populations. Steelhead spawning is difficult to monitor in Puget Sound because hatchery-origin and hybrid fish are morphologically indistinguishable from wild fish. Therefore, the Washington Department of Fish and Wildlife (WDFW) adopted the use of genetic tools for monitoring reproductive interaction between hatchery-origin and wild populations; however, the effectiveness of genetic monitoring programs is dependent on the ability to differentiate hatchery-origin,

hybrid, and wild fish. As a consequence of continual introductions of hatchery-origin steelhead and the cumulative effects of gene flow over time, increased resolution through a population genomics approach was required to differentiate between the origins in Puget Sound. Warheit (2014) used expressed sequence tags (ESTs) and restriction site-associated DNA sequencing (RADseq) to develop SNP panels to monitor introgression from hatchery-origin into wild steelhead populations. A total of 192 SNPs were used to genotype individuals throughout Washington state as baseline samples for genetic stock identification and genotyped in thousands of individuals annually to monitor for and quantify introgression of hatchery-origin fish into wild populations. Results suggest that segregated steelhead hatchery programs have affected the genetic structure of wild populations; however, the effects vary among river systems (Warheit 2014).

Alternatively, wild populations may accidently be exposed to escapees from farming operations with the expansion of commercial aquaculture practices, including salmon farming. Norwegian Atlantic salmon farms dominate global production with stock originating from 40 Norwegian rivers but have undergone multiple generations of domestication selection (Gjedrem 2010). As a result of captive breeding and directional selection for economically important traits, farmed and wild salmon differ in many characters. Farmed salmon display lower levels of genetic variation compared to wild populations (Skaala et al. 2004), as well as genetic differences for traits such as growth (Glover et al. 2009), physiology (Fleming et al. 2002), and behavior (Fleming and Einum 1997). In 2011, the Institute of Marine Research initiated a risk assessment of Norwegian salmon farming to evaluate the potential hazard of introgression of farmed escapees on wild salmon populations. Researchers used next-generation sequencing to discover a panel of 60 SNPs distributed across 27 of 29 chromosomes (Karlsson et al. 2011). Because of the genetic similarity between wild and farmed salmon, the genomics approach provided the resolution required to identify hatchery-informative markers. The detection of farmed escapees and the quantification of cumulative introgression are being used to address risk of continued and future genetic changes in wild Atlantic salmon populations (Taranger et al. 2014). A study of 20 Norwegian rivers utilizing the SNP panel from Karlsson et al. (2011) revealed less introgression of farmed Atlantic salmon than may be expected based upon the reported numbers of escapees observed in the populations, suggesting limited reproductive success of farmed escaped salmon in Norwegian rivers (Glover et al. 2013). These results are consistent with previous controlled experiments examining spawning success (Fleming et al. 1996, 2000).

### 2.1.2 Genetic Rescue

A few conservation and management efforts have applied genome-wide markers and technology to address genetic factors affecting endangered populations, such as the loss of the genetic diversity. Genetic rescue is a conservation tool used to increase the fitness of at-risk populations by introducing new genetic variation into a

population. Genomic tools can aid the identification of target populations, as well as guide management and conservation efforts (Fitzpatrick and Funk 2019). With growing knowledge of the negative effects of inbreeding depression, there is increasing interest in using genetic translocations which is expected to be most useful for small, isolated populations (Frankham 2015).

The mountain pygmy possum, *Burramys parvus*, is one of Australia's most threatened marsupials. *Burramys parvus* is restricted to alpine regions of Australia, and the remaining populations are geographically isolated and genetically distinct. In 2009, the southern-most population experienced a rapid population decline. Because the population was vulnerable to extinction, a recovery program was implemented that aimed at restoration of habitat, predator control, and the introduction of males from healthy and genetically variable populations of *B. parvus*. To assess genetic diversity, individuals were genotyped using eight previously isolated microsatellite markers (Mitrovski et al. 2005), and 16 additional markers were generated with the 454 sequencing platform. Males from two source populations were introduced to the southern-most population, and alleles from introduced males became integrated into the gene pool. As a result of the introduction efforts, genetic diversity (allelic richness, heterozygosity) increased along with population size (Weeks et al. 2017).

A second example of the application of population genomics for informing reintroduction is the Burmese roofed turtle, *Batagur trivittata*, one of the world's most endangered turtles with a single wild population in Myanmar (Çilingir et al. 2017). The use of double-digest restriction site-associated DNA sequencing (ddRAD-seq) produced 1,500 genome-wide single nucleotide polymorphisms, revealing a low effective population size ($\leq$10 individuals). The most genetically diverse individuals from the captive pool were reintroduced into the wild resulting in an increase in the number of fertile eggs. Additionally, genetically diverse individuals were retained for the captive-breeding program. The use of genome-wide markers informed the genetic and demographic reinforcement of the wild population and management of captive populations.

### 2.1.3 Neutral Population Substructure: Population/Stock Genetic Differentiation and Conservation Units

Delineation of neutral substructure is the most common population genomics application in management and conservation. Managers have often used population genomics to track the movement and harvest of fisheries resources. One common use of the application of population genomics into resource management is genetic stock identification (GSI), an important tool for fisheries management. The efficacy of this tool is dependent upon the ability to differentiate stocks of interest; however, this can be difficult when populations are closely related. GSI uses observed allele frequencies of baseline populations sampled on the spawning grounds to infer the natal origin of fish captured in mixed-stock fisheries (Utter and Ryman 1993; Beacham et al. 2012). Researchers used RADseq to identify 38 population informative SNPs to monitor stock composition in sockeye salmon, *Oncorhynchus*

*nerka* (Dann et al. 2013). The use of SNPs improved accuracy in stock identification over microsatellites. The efficacy of SNPS to identify stock origin allows managers to shift fishing efforts based on return rates to stock of origin, benefiting the commercial fishery and local economy while reducing the risk of overharvest (Dann et al. 2013). Similar work has been done in Chinook salmon, *Oncorhynchus tshawytscha* (Larson et al. 2014).

Additionally, population-informative SNP panels for both chum and sockeye salmon have been used to genotype >150,000 individuals over a 3-year period along the Alaska coastline (Habicht et al. 2012; Munro et al. 2012). The Alaska Department of Fish and Game's Western Alaska Salmon Stock Identification Program (WASSIP) used this data to directly inform management decisions including allocation of resources and protecting weak stocks.

Management using a parentage-based tagging approach (PBT; Anderson and Garza 2005) involves the annual genotyping of hatchery broodstock to create a database of parental genotypes from each hatchery. Parentage analysis allows for offspring to be traced back to their parents, therefore, identifying the hatchery of origin and age. A PBT program has been implemented in the Snake River basin using <100 SNPs developed from next-generation sequencing to reconstruct pedigrees and genetically tag offspring of hatchery steelhead trout (Steele et al. 2013). The accuracy and application of SNPs for parentage analysis have resulted in the genetic tagging of approximately 95% of the steelhead and Chinook salmon in the region (Steele et al. 2013). Management agencies are continuing to genotype broodstock throughout the Snake River basin, with future efforts targeting the entire Columbia River basin.

A notable example of using population genomics to delineate fisheries conservation units is of rockfish in Puget Sound, WA, USA. In 2010, yelloweye rockfish, *Sebastes ruberrimus*, and canary rockfish, *S. pinniger*, in the inland waterways of Puget Sound were listed as threatened under the US Endangered Species Act (ESA). The definition of listable under the ESA includes taxonomically identified species and subspecies, as well as distinct population segments (DPS; USFWS-NMFS 1996). In order for a vertebrate population to be designated as a DPS, it must be "discrete" from other populations of the same species as well as "significant" to the remainder of the species (USFWS-NMFS 1996). These two rockfish species were designated as DPSs. Because little information was available for the rockfish species of interest in Puget Sound, the decision to designate DPSs relied heavily on evidence from other species that these populations were "discrete" taxonomic units (Drake et al. 2010). To investigate the conservation designations, researchers collaborated with recreational fishing communities to collect tissue samples from the outer coasts of the USA and Canada. Multiple analyses using RADseq (>7,000 SNPs) suggest that yelloweye rockfish in Puget Sound and British Columbia, Canada, were genetically different from coastal populations, while canary rockfish showed no genetic differentiation between the two areas (Andrews et al. 2018). For yelloweye rockfish, the genomic data support the original ESA designation with an expansion of protected areas. Interestingly, the data for canary rockfish in Puget Sound do not

support a "discrete" population, therefore, not meeting the first criterion of the ESA. Consequently, in January 2017, the National Marine Fisheries Service removed the canary rockfish from the federal list of threatened and endangered species.

## *2.2 Ecological Adaptation and Disease Management*

### 2.2.1 Adaptive Population Substructure

Population genomics approaches can be used to identify adaptively differentiated populations (Hoban 2018; Funk et al. 2019; Razgour et al. 2018) which can be prioritized for conservation (Funk et al. 2012). One prominent example of the utility of a genomics approach for identification of adaptive substructure is life history variation in Chinook salmon. The seasonal timing of migration by adult salmonids into freshwater is a key life history trait. Chinook salmon exhibit two seasonal migration strategies, spring and fall, to spawn in freshwater. Mature migrating Chinook enter freshwater in a sexually mature state in the fall and migrate directly to spawning grounds (Quinn et al. 2015). In contrast, premature migrating Chinook enter freshwater sexually immature in the spring, migrate to headwaters and oversummer while their gonads develop before spawning in the fall (Quinn et al. 2015). The storage of excess fat allows the uncoupling of migration and spawning (Hearsey and Kinziger 2015). Historically, both spring and fall types inhabited many rivers. Because premature migrating salmon spend an extended period of time in freshwater, populations have experienced a dramatic decline as a result of anthropogenic activities that affect river conditions such as logging, mining, dam construction, and water diversion (Quinn et al. 2015). In the Klamath River basin, the spring Chinook population experienced a severe decline compared to the fall Chinook. The Salmon River hosts the last remaining population of spring salmon in the Klamath River basin comprising approximately 100 individuals. However, the rapid population decline has been met with limited conservation concern. Genetic analyses using neutral markers find little differentiation between premature and mature migrating populations and, therefore, are generally grouped into the same ESU or DPS (Myers et al. 1998). Population genomic analysis of 301,562 SNPs identified through RADseq suggests overall genetic variation reflects geography as opposed to migration type, corresponding to the ESU designations (Prince et al. 2017; Fig. 2). However, a single genetic locus (Greb1-Like) was associated with premature migration (Prince et al. 2017; Fig. 2) which has been shown to modulate diverse behavior and metabolic processes in mice (Henry et al. 2015). Further investigation of Chinook salmon suggests that less than 1% of the remaining fish carry a copy of the early migration version of the gene (Thompson et al. 2019). The results generated from the application of the genomics approach have prompted the Karuk Tribe to petition the listing of Klamath's spring Chinook as threatened or endangered under the Endangered Species Act. This proposal has been controversial

because some individuals advocate for the need to protect adaptive variation, while others are concerned about the implications of conservation decisions based on a single gene (Langin 2018).

### 2.2.2 Environmental Adaptation and Tolerances

Although not a vertebrate (i.e., chapter topic), researchers used RADseq to identify SNPs (>9,000) to delineate locally adapted populations of commercially important greenlip abalone, *Haliotis laevigata* (Sandoval-Castillo et al. 2018). This study incorporated genotype-environment association analyses and outlier tests to identify 8,786 putatively neutral and 323 candidate adaptive loci. Neutral genetic structure was low across the study region; however, candidate adaptive loci suggested five distinct population clusters congruent with oceanic regions with associations corresponding to differences in oxygen concentration and temperature between the regions. Additionally, 80 of the 323 candidate loci were annotated to genes whose functions putatively affect high temperature and/or low oxygen tolerance (Sandoval-Castillo et al. 2018). The results have provided a genetic baseline to develop a genetic management plan which includes a risk assessment of impacts of the commercial sea ranching and restocking activities, including policies that promote the maintenance of patterns of adaptive divergence (DoF 2016; Daume et al. 2017; Hart et al. 2017). In particular, the Abalone Aquaculture Policy's "progeny diversity strategy" recommends the use of broodstock collection and captive breeding programs based on adaptive clusters to ensure that only genetically appropriate progenies are released into the marine environment (Hart et al. 2017).

### 2.2.3 Disease Management

The application of population genomics can provide knowledge for management and treatment of disease in wildlife species. One example is chondrodystrophy in the California condor, *Gymnogyps californianus*. Historically, the California condor was widespread throughout southern North America. However, anthropogenic activity caused a rapid decline in the population. To prevent extinction, the last wild bird was brought into captivity in 1987, resulting in a human-induced bottleneck (D'Elia et al. 2016), and the contemporary population has descended from 14 individuals (USFWS 1996). The specific goal for management of the California condor population focuses on maximizing genetic diversity in captive and reintroduced populations (Ralls and Ballou 2004) which has been supported by the application of genomics methodologies. Thirty-six complete California condor genomes, representing the entire gene pool of the species, have been sequenced, which identified four million SNPs (Ryder et al. 2016). This information has provided insight into the genetic similarities among founding birds and allowed reassessment of kinship, which is now being applied to the captive propagation effort in the selection of breeding pairs. However, the captive propagation effort is

hindered by the expression of chondrodystrophy, which has an autosomal recessive mode of transmission in condors (Ralls et al. 2006). Whole genome sequencing suggests that this lethal form of dwarfism is correlated to a series of linked markers localized in a 1 Mb region of the genome, which is currently being used to detect carrier condors heterozygous for the lethal mutation (Ryder et al. 2016). This information is being incorporated into population management to reduce the risk of reproductive failure.

The Tasmanian devil '*Sarcophilus harrisii*' is the largest extant marsupial carnivore and is endemic to the island of Tasmania. In the last 20 years, the emergence of an infectious cancer, devil facial tumor disease (DFTD), has resulted in an 85% decline in the population (Hawkins et al. 2006; McCallum et al. 2007). DFTD is a transmissible cancer that spreads through physical contact during feeding and mating (Pearse and Swift 2006). In 2006, a disease-free insurance population of devils was established (Conservation Breeding Specialist Group 2008). The main goals of this program are to preserve genetic diversity, minimize inbreeding, prevent adaptation to captivity, and prevent disease (Conservation Breeding Specialist Group 2008; Lees and Andrew 2012). Whole genome sequencing has been used to understand the origins, transmission, and diversity of DFTD in Tasmanian devils (Murchison et al. 2012). Interestingly, in a single population of infected devils, a small number of individuals spontaneously recovered from the typically lethal disease. To determine if resistance to DFTD has a genetic basis, researchers sequenced the genomes of the seven individuals that recovered from DFTD infection and compared them to the genomes of six devils that succumbed to the disease (Wright et al. 2017). This genome-wide association approach identified two genomic regions where genotypes were strongly associated with disease status and associated with angiogenesis (Wright et al. 2017). The information generated from genomics techniques is being used to inform disease management, breeding programs, and reintroductions to manage genetic diversity of this endangered species.

## 3 Future Perspectives

The current application of population genomics to wildlife and fish management focuses on monitoring and managing existing diversity; however, there is growing interest in the application of functional genomics approaches to gain a better understanding of adaptive potential in wild populations (Corlett 2017; Connon et al. 2018; Rey et al. 2019). With the recent advances in RNA-sequencing technology allowing the application in non-model organisms (Todd et al. 2016), a population transcriptomics approach can be applied in a conservation context to identify biomarkers associated with environmental stress (Connon et al. 2018). Additionally, genomics technology can also be used to study the epigenome (i.e., epigenomics). Mutations in the epigenome are molecular changes in gene expression without altering the DNA sequence and can be involved in an organism's adaptive response to environmental change, representing the short-term interaction between organisms

and their environment. The integration of epigenetics can provide insight into the physiological, biological, and ecological status of organisms and aid in the designation of management units (Rey et al. 2019). The recent advances in genomics technology that allow the quantification of functional diversity and structure among wild populations will improve our understanding of organismal response to changing environmental conditions.

Furthermore, there is an opportunity for new genomics technology to provide the potential for genome manipulation in conservation applications (Phelps et al. 2019). Genome-editing technology refers to methods that can insert, delete, and/or replace specific regions of the genome allowing for sequence modification that can knock out the function of a gene or insertion of a specific sequence with genome edits. The most well-known genome-editing technique is CRISPR-CAS9, an RNA-guided molecule that identifies and binds to a specific region within the genome and subsequently cuts and replaces DNA (Mei et al. 2016). The promise of this tool is enhanced by its simplicity, precision, and relatively low cost (Mei et al. 2016; Phelps et al. 2019).

Genome-editing technology has the potential to aid biodiversity conservation in several ways by identifying and altering regions of the genome that may impact fitness and limit survival in wild populations. Genomic data and gene-editing technology can be used in the replacement of alleles which can assist evolution of disease resistance, remove deleterious genetic variants such as chondrodystrophy in California condors, and enhance adaptation to changing climates (i.e., facilitated adaptation; Harrisson et al. 2014). Gene-editing technology can aid in the eradication of invasive species, which are the main driver of vertebrate extinctions (e.g., Bellard et al. 2016), through gene drive methodologies. Gene drives alter Mendelian inheritance by increasing inheritability of an engineered allele to spread the desired trait throughout the population (Esvelt et al. 2014; Webber et al. 2015). For example, genome-editing techniques that alter the sex determination pathway of invasive rodents on islands aimed at producing all male offspring ultimately lead to a reduction in reproduction (Campbell et al. 2015). Additionally, gene-editing technology has the potential to aid in the resurrection of ecologically important traits (i.e., trait resurrection, modification of phenotypically relevant genes from extinct species into closely related extant taxa; Johnson et al. 2016). One notable example of trait resurrection is the alteration of genes from the extinct wooly mammoth, *Mammuthus primigenius*, in the Asian elephant, *Elephas maximum*, associated with cold temperature adaptations (Lynch et al. 2015; Shapiro 2015).

While gene-editing technology has the potential to aid in conservation and management efforts, this methodology does not come without controversy. Agencies currently lack a clear framework for implementing gene-editing approaches as conservation tools (Sandler 2019). Before implementation into conservation and management efforts of natural populations, it is imperative to evaluate practical, ethical, and legal considerations (Johnson et al. 2016; Webber et al. 2015) to reduce any unintended, irreversible consequences.

## 4 Conclusions

We conclude from our search of the literature and our survey that population genomics is still only partially embraced with few examples of genomic data being implemented in a management or conservation framework. The field of fisheries has been an early adopter of some genomics technologies. However, most of the advances in natural resource management have been in the use of genomics for more power and precision instead of looking at the functionality or interactions of genes and how they help organisms adapt to their environment. We had initially hoped that the lack of examples of management decisions being affected by population genomics was a function of a lag in the process. It can take time from results being obtained, to having them accepted through a peer review process, to having them adopted by management agencies to change how natural populations are conserved or managed. This may still be the case; however, our literature review and survey through ECOLOG-L listserv has led us to believe that while population genomics has an important role for wildlife management, it is still an underused tool with an important future. We predict that the next decade will see an explosion of uses of population genomic data in the management of wild populations, although as the technical details of the science becomes more specialized it will require active engagement between managers and researchers to ensure the interpretation of the science for policy is done responsibly.

## References

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

Anderson EC, Garza JC. A description of full parentage genotyping (online). Report submitted to the Pacific Salmon Commission, Vancouver, BC; 2005. Retrieved from http://swfsc.noaa.gov/publications/FED/00675.pdf.

Andrews KS, Nichols KM, Elz A, Tolimieri N, Harvey CJ, Pacunski R, Lowry D, Yamanaka KL, Tonnes DM. Cooperative research sheds light on population structure and listing status of threatened and endangered rockfish species. Conserv Genet. 2018;19:865–78.

Beacham TD, Candy JR, Wallace C, Wetklo M, Deng L, MacConnachie C. Microsatellite mixed-stock identification of Coho Salmon in British Columbia. Mar Coast Fish. 2012;4:85–100.

Bellard C, Cassey P, Blackburn TM. Alien species as a driver of recent extinctions. Biol Lett. 2016;12:20150623.

Benestan LM, Ferchaud A, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Schwartz MK, Kelley JL, Luikart G. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25:2967–77.

Campbell KJ, Beek J, Eason CT, Glen AS, Godwin J, Gould F, Holmes ND, Howald GR, Madden FM, Ponder JB, Threadgill DW, Wegmann AS, Baxter GS. The next generation of rodent eradications: innovative technologies and tools to improve species specificity and increase their feasibility on islands. Biol Conserv. 2015;185:47–58.

Çilingir FG, Rheindt FE, Garg KM, Platt K, Platt SG, Bickford DP. Conservation genomics of the endangered Burmese roofed turtle. Conserv Biol. 2017;31:1469–76.
Connon RE, Jeffries KM, Komoroske LM, Todgham AE, Fangue NA. The utility of transcriptomics in fish conservation. J Exp Biol. 2018;221:jeb148833.
Conservation Breeding Specialist Group. Tasmanian devil PHVA final report. Apple Valley, MN: IUCN/SSC Conservation Breeding Specialist Group; 2008.
Corlett RT. A bigger toolbox: biotechnology in biodiversity conservation. Trends Biotechnol. 2017;35:55–65.
Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK. Considering evolutionary processes in conservation biology. Trends Ecol Evol. 2000;15:290–5.
D'Elia J, Haig SM, Mullins TD, Miller MP. Ancient DNA reveals substantial genetic diversity in the California condor (*Gymnogyps californianus*) prior to a population bottleneck. Condor. 2016;11:703–14.
Dann TH, Habicht C, Baker TT, Seeb JE. Exploiting genetic diversity to balance conservation and harvest migratory salmon. Can J Fish Aquat Sci. 2013;70:785–93.
Daume S, Gardner C, Leporati S, Trott P. Western Australia abalone fishery MSC full assessment public certification report. SCS Global Services; 2017. p. 254.
DoF. Abalone aquaculture in Western Australia: principles and considerations relating to management of abalone aquaculture in Western Australia. Fisheries Occasional Publication 132; 2016. p. 18.
Drake JS, Berntson EA, Gustafson RG, Holmes EE, Levin PS, Tolimieri N, Waples RS, Sogard SM, Williams GD, Cope JM. Status review of five rockfish species in Puget Sound, Washington: bocaccio (*Sebastes paucispinis*), canary rockfish (*S. pinniger*), yelloweye rockfish (*S. ruberrimus*), greenstriped rockfish (*S. elongates*), and redstripe rockfish (*S. proriger*). US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service; 2010.
Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, Schönfeld B, Wiench C, Hohenlohe P, Storfer A. Rapid evolutionary response to transmissible cancer in Tasmanian devils. Nat Commun. 2016;7:12684.
Esvelt KM, Smidler AL, Catteruccia F, Church GM. Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. Elife. 2014;3:e03401.
Fitzpatrick SW, Funk WC. Genomics for genetic rescue. In: Population genomics. Cham: Springer; 2019.
Fleming IA, Einum S. Experimental tests of genetic divergence of farmed from wild Atlantic salmon due to domestication. ICES J Mar Sci. 1997;54:1051–63.
Fleming IA, Jonsson B, Gross MR, Lamberg A. An experimental study of the reproductive behavior and success of farmed and wild Atlantic salmon (*Salmo salar*). J Appl Ecol. 1996;33:893–905.
Fleming IA, Hindar K, Mjølnerød IB, Jonsson B, Balstad T, Lamberg A. Lifetime success and interactions of farm salmon invading a native population. Proc Roy Soc Lond B Biol Sci. 2000;267:1517–23.
Fleming IA, Agustsson T, Finstad B, Johnsson JI, Bjornsson BT. Effects of domestication on growth physiology and endocrinology of Atlantic salmon (*Salmo salar*). Can J Fish Aquat Sci. 2002;59:1323–30.
Frankham R. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Mol Ecol. 2015;24:2610–8.
Fraser DJ, Bernatchez L. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol Ecol. 2001;10:2741–52.
Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.
Funk WC, Forester BR, Converse SJ, Darst C, Morey S. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet. 2019;20:115–34.
Gjedrem T. The first family based breeding program in aquaculture. Rev Aquac. 2010;2:2–15.

Glover KA, Ottera H, Olsen RE, Slinde E, Taranger GL, Skaala Ø. A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. Aquaculture. 2009;286:203–10.

Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala Ø. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. BMC Genet. 2013;14:4.

Habicht C, Munro AR, Dann TH, Eggers DM, Templin WD, Witteveen MJ, Baker TT, Howard KG, Rogers Olive SD, Liller HL, Chenoweth EL, Volk EC. Harvest and harvest rates of sockeye salmon stocks in fisheries of the Western Alaska Salmon Stock Identification Program (WASSIP), 2006–2008. Alaska Department of Fish and Game, Special Publication No. 12-24. Anchorage, AK; 2012.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. Using genomics to characterize evolutionary potential for conservation of wild populations. Evol Appl. 2014;7:1008–25.

Hart A, Strain L, Hesp A, Fisher E, Webster F, Brand-Gardner S, Walters S. Marine Stewardship Council Full Assessment Report Western Australian Abalone Managed Fishery. Perth, WA: Department of Fisheries, Western Australia; 2017.

Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, Mann D, Mooney N, Pemberton D, Pyecroft S, Restani M, Wiersma J. Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. Biol Conserv. 2006;131:307–24.

Hearsey JW, Kinziger AP. Diversity in sympatric Chinook salmon runs: timing, relative fat content and maturation. Environ Biol Fishes. 2015;98:413–23.

Henry FE, Sugino K, Tozer A, Branco T, Sternson SM. Cell type-specific transcriptomics of hypothalamic energy-sensing neuron responses to weight loss. Elife. 2015;4:e09800.

Hoban S. Integrative conservation genetics: prioritizing populations using climate predictions, adaptive potential and habitat connectivity. Mol Ecol Resour. 2018;18:14–7.

Hoffman AA, Sgrò CM, Kristensen TN. Revisiting adaptive potential, population size, and conservation. Trends Ecol Evol. 2017;32:506–17.

Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population genomics provides key insights in ecology and evolution. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 483–510.

Hutchings JA. Collapse and recovery of marine fishes. Nature. 2000;406:882–5.

Johnson JA, Altwegg R, Evans DM, Ewen JG, Gordon IJ. Is there a future for genome-editing technologies in conservation? Anim Conserv. 2016;19:97–101.

Kardos M, Shafer ABA. The perils of gene-targeted conservation. Trends Ecol Evol. 2018;33:827–39.

Karlsson S, Moen T, Lien S, Glover KA, Hindar K. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol Ecol Resour. 2011;11:377–87.

Laikre L, Schwartz MK, Waples RS, Ryman N, GeM Working Group. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. Trends Ecol Evol. 2010;25:520–9.

Langin K. Salmon spawn fierce debate over protecting endangered species, thanks to a single gene. Sci News. 2018. http://www.sciencemag.org/news/2018/05/salmon-spawn-fierce-debate-over-protecting-endangered-species-thanks-single-gene.

Larson WA, Seeb JE, Pascal CE, Templin WD, Seeb LW. Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Can J Fish Aquat Sci. 2014;71:698–708.

Lees C, Andrew P. The Tasmanian devil insurance meta-population: 2012 evaluation and review. Apple Valley, MN: IUCN/SSC Conservation Breeding Specialist Group; 2012.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94.

Luikart G, Kardos M, Hand B, Rajora OP, Aitken S, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Spring Nature Switzerland AG; 2019. p. 3–79.

Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, Perry GH, Miller W, Schuster SC. Elephantid genomes reveal the molecular bases of wooly mammoth adaptations to the arctic. Cell Rep. 2015;12:217–8.

McCallum H, Tompkins DM, Jones M, Lachish S, Marvanek S, Lazenby B, Jocking G, Wiersma J, Hawkins CE. Distribution and impacts of Tasmanian devil facial tumor disease. Ecohealth. 2007;4:318–25.

Mei Y, Wang Y, Chen H, Sun ZS, Ju XD. Recent progress in CRISPR/Cas9 technology. J Genet Genomics. 2016;43:63–75.

Mitrovski P, Heinze DA, Guthridge K, Weeks AR. Isolation and characterization of microsatellite loci from the Australian endemic mountain pygmy-possum, *Burramys parvus* Broom. Mol Ecol Notes. 2005;5:395–7.

Moritz C. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst Biol. 2002;51:238–54.

Mortiz C. Defining 'Evolutionary Significant Units' for conservation. Trends Ecol Evol. 1994;9:373–5.

Munro AR, Habicht C, Dann TH, Eggers DM, Templin WD, Witteveen MJ, Baker TT, Howard KG, Jasper JR, Rogers Olive SD, Liller HL, Chenoweth EL, Volk EC. Harvest and harvest rates of Chum salmon stocks in fisheries of the Western Alaska Salmon Stock Identification Program (WASSIP), 2007–2009. Alaska Department of Fish and Game, Special Publication No. 12-25, Anchorage, AK; 2012.

Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, Hims M, Ding Z, Ivakhno S, Stewart C, Ng BL, Wong W, Aken B, White S, Alsop A, Becq J, Bignell GR, Cheetham RK, Cheng W, Connor TR, Cox AJ, Feng ZP, Gu Y, Grocock RJ, Harris SR, Khrebtukova I, Kingsbury Z, Kowarsky M, Kreiss A, Luo S, Marshall J, McBride DJ, Murray L, Pearse AM, Raine K, Rasolonjatovo I, Shaw R, Tedder P, Tregidgo C, Vilella AJ, Wedge DC, Woods GM, Gormley N, Humphray S, Schroth G, Smith G, Hall K, Searle SM, Carter NP, Papenfuss AT, Futreal PA, Campbell PJ, Yang F, Bentley DR, Evers DJ, Stratton MR. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell. 2012;148:780–91.

Myers RA, Worm B. Rapid worldwide depletion of predatory fish communities. Nature. 2003;423:280–3.

Myers JM, Kope RG, Bryant GJ, Teel D, Lierheimer LJ, Wainwright TC, Grant WS, Waknitz FW, Neely K, Lindley ST, Waples RS. Status Review of Chinook salmon from Washington, Idaho, Oregon, and California. U.S. Department of Commerce, NOAA Technical Memorandum. NMFS-NWFSC-35; 1998.

Pearse DE. Saving the spandrels? Adaptive genomic variation in conservation and fisheries management. J Fish Biol. 2016;89:2697–716.

Pearse AM, Swift K. Allograft theory: transmission of devil facial-tumour disease. Nature. 2006;439:549.

Phelps M, Seeb L, Seeb J. Transforming ecology and conservation biology through genome editing. Conserv Biol. 2019; https://doi.org/10.1111/cobi.13292.

Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller MR. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Evol Genet. 2017;3:e1603198.

Quinn TP, McGinnity P, Reed TE. The paradox of "premature migration" by adult anadromous salmonid fishes: patterns and hypotheses. Can J Fish Aquat Sci. 2015;73:1015–30.

Ralls K, Ballou JD. Genetic status and management of California condors. Condor. 2004;106:215–28.

Ralls K, Ballou JD, Rideout BA, Frankham R. Genetic management of chondrodystrophy in the California condor. Anim Conserv. 2006;3:145–53.

Randi E. Detecting hybridization between wild species and their domesticated relatives. Mol Ecol. 2008;17:285–93.

Razgour O, Taggart JB, Manel S, Juste J, Ibánez C, Rebelo H, Alberdi A, Jones G, Park K. An integrated framework to identify wildlife populations under threat from climate change. Mol Ecol Resour. 2018;18:18–31.

Reed DH, Frankham R. Correlation between fitness and genetic diversity. Conserv Biol. 2003;17:230–7.

Rey O, Eizaguirre C, Angers B, Baltazar-Soares M, Sagonas K, Prunier JG, Blanchet S. Linking epigenetics and biological conservation: towards a conservation epigenetics perspective. Funct Ecol. 2019:1–14.

Ryder O, Miller W, Ralls K, Ballou JD, Steiner CC, Mitelberg A, Romanov M, Chemnick LG, Mace M, Schuster S. Whole genome sequencing of California condors is now utilized for guiding genetic management. International plant and animal genome XXIV conference, San Diego, CA; 2016.

Sandler R. The ethics of genetic engineering and gene drives in conservation. Conservat Genet. 2019; https://doi.org/10.1111/cobi.13407.

Sandoval-Castillo J, Robinson NA, Hart AM, Strain LWS, Beheregaray LB. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusk, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. Mol Ecol. 2018;27:1603–20.

Schwartz MK, McKelvey KS, Cushman SA, Luikart G. Landscape genomics: a brief perspective. In: Spatial complexity, informatics, and wildlife conservation. Tokyo: Springer; 2010. p. 165–74.

Sgró CM, Lowe AJ, Hoffmann AA. Building evolutionary resilience for conserving biodiversity under climate change. Evol Appl. 2011;4:326–37.

Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De Meester L, Ekblom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoezel AR, Höglund J, Jensen EL, Krause J, Kristensen TN, Krützen M, JK MK, Norman AJ, Ogden R, Österling EM, Ouborg NJ, Piccolo J, Popović D, Primmer CR, Reed FA, Roumet M, Salmona J, Schenekar T, Schwartz MK, Segelbacher G, Senn H, Thaulow J, Valtonen M, Veale A, Vergeer P, Vijay N, Vilà C, Weissensteiner M, Wennerström L, Wheat CW, Zieliński P. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30:78–87.

Shapiro B. Mammoth 2.0: will genome engineering resurrect extinct species. Genome Biol. 2015;16:228.

Skaala Ø, Hoyheim B, Glover K, Dahle G. Microsatellite analysis in domesticated and wild Atlantic salmon (*Salmo salar* L.): allelic diversity and identification of individuals. Aquaculture. 2004;240:131–43.

Steele CA, Anderson EC, Ackerman MW, Hess MA, Campbell NR, Narum SR, Campbell MR. A validation of parentage-based tagging using hatchery Steelhead in the Snake River basin. Can J Fish Aquat Sci. 2013;70:1046–54.

Taranger GL, Karlsen Ø, Bannister RJ, Glover KA, Husa V, Karlsbakk E, Kvamme BO, Boxaspen KK, Bjørn PA, Finstad B, Madhun AS, Morton HC, Svåsand T. Risk assessment of the environmental impact of Norwegian Atlantic salmon farming. ICES J Mar Sci. 2014;72:997–1021.

Thompson TQ, Bellinger RM, O'Rourke SM, Prince DJ, Stevenson AE, Rodriques AT, Sloat MR, Speller CF, Yang DY, Butler VL, Banks MA, Miller MR. Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. Proc Natl Acad Sci U S A. 2019;116:177–86.

Todd EV, Black MA, Gemmell NJ. The power and promise of RNA-seq in ecology and evolution. Mol Ecol. 2016;25:1224–41.

USFWS. California condor recovery plan, third revision, US Fish and Wildlife Service; 1996.

USFWS-NMFS. Policy regarding the recognition of distinct vertebrate population segments under the Endangered Species Act. Fed Regist. 1996;61:4721–5.

Utter F, Ryman N. Genetic markers and mixed stock fisheries. Fisheries. 1993;18:11–21.

Vrijenhoek RC, Leberg PL. Let's not throw the baby out with the bathwater: a comment on management for MHC diversity in captive populations. Conserv Biol. 1991;5:252–4.

Warheit KI. Measuring reproductive interactions between hatchery-origin and wild Steelhead (*Oncorhynchus mykiss*) from northern Puget Sound populations potentially affected by segregated hatchery programs. Washington Department of Fish and Wildlife, Unpublished Report, Olympia, WA; 2014.

Webber BL, Raghu S, Edwards OR. Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? Proc Natl Acad Sci. 2015;112:10565–7.

Weeks AR, Heinze D, Perrin L, Stoklosa J, Hoffmann AA, van Rooyen A, Kelly T, Mansergh I. Genetic rescue increases fitness and aids rapid recovery of an endangered marsupial population. Nat Commun. 2017;8:1071.

Wright B, Willet CE, Hamede R, Jones M, Belov K, Wade CM. Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genome-wide association. Sci Rep. 2017;7:1–6.

# Genomics for Genetic Rescue


**Sarah W. Fitzpatrick and W. Chris Funk**

**Abstract** Genetic rescue, where new alleles cause increased population growth, has successfully reversed population declines in several iconic species. Concerns over outbreeding depression and genomic swamping limit this technique's application in wildlife management. New genomic approaches can improve the implementation and monitoring of genetic rescue, making it an even more effective management strategy. In planning stages, genomics can help identify populations that would benefit most from augmented gene flow and populations and individuals that would be the best sources. After augmented gene flow, genomics can be used to monitor the outcome of genetic rescue and determine if and when additional gene flow is needed. Here, we outline specific ways in which genomics can be used to (1) test for inbreeding and inbreeding depression; (2) predict the probability that gene flow will cause outbreeding depression versus genetic rescue; (3) monitor the results of assisted gene flow; and (4) determine the genetic architecture underlying genetic rescue to improve future attempts. We conclude by outlining recommendations on how genomic data can be used to improve the effectiveness of genetic rescue and reduce uncertainty about its outcome.

**Keywords** Fitness · Gene flow · Genetic drift · Genetic rescue · Inbreeding depression · Population genomics · Small populations

S. W. Fitzpatrick (✉)
W.K. Kellogg Biological Station, Department of Integrative Biology, Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, Hickory Corners, MI, USA
e-mail: sfitz@msu.edu

W. C. Funk
Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2019_64, 

# 1 Introduction

Many species with historically continuous distributions are now restricted to small and isolated fragments. In the absence of gene flow, small populations are subject to increased probability of mating among relatives, which can result in accumulation and fixation of deleterious alleles, and reduced fitness (Keller and Waller 2002). Reduced gene flow can also decrease genetic variation available to selection and limit the spread of adaptive alleles, thus constraining adaptive potential (Garant et al. 2007). Genetic rescue (GR) has potential to become a powerful means to conserve small and declining populations (Tallmon et al. 2004; Whiteley et al. 2015). The ability to genotype thousands of loci or sequence entire genomes of many individuals in any natural population is rapidly improving our understanding of long-standing questions in evolution, ecology, and conservation, and the subfield of genetic rescue is no exception. This chapter considers how genomics can be used to improve successful implementation and monitoring of genetic rescue for conservation purposes.

## *1.1 Definitions*

*Genetic rescue* (GR) is an increase in population fitness caused by gene flow inferred from some demographic vital rate or phenotypic trait, by more than can be attributed to the demographic contribution of immigrants (Tallmon et al. 2004; Whiteley et al. 2015). We prefer this definition to others previously proposed (Hedrick et al. 2011) because it provides the most useful and desired metric for management and conservation practitioners in requiring a population-level response to the introduction of new alleles. Populations in need of GR tend to be small, declining populations with low genetic and phenotypic variation and low fitness. Adding individuals to a dwindling population can cause *demographic rescue* (see Glossary for terms in italics) through nongenetic benefits such as alleviation of Allee effects (Brown and Kodric-Brown 1977). However, successful GR involves an increase in abundance caused by the introduction of beneficial genetic variation, whether through masking of deleterious alleles, heterozygote advantage, adaptive evolution, or a combination of these processes. Throughout this chapter, we refer to GR as the outcome of a management strategy involving augmented gene flow, rather than the strategy itself. For example, the outcome of adding one or several migrants to a small, declining population could result in genetic rescue if population growth rates increase due to the addition of beneficial alleles. Alternatively, this action could result in *outbreeding depression* if population growth rates decrease due to intrinsic genetic incompatibilities, the addition of maladaptive alleles, or by breaking apart coadapted gene complexes. If population growth rates remain the same as before gene flow, even if genetic diversity has increased, the action resulted in neither genetic rescue nor outbreeding depression.

GR is related to, but distinct from, a suite of similar terms such as *genetic restoration*, which is when gene flow causes an increase in genetic variation that does not necessarily result in an increase in *absolute fitness* (Whiteley et al. 2015). GR is within the umbrella of *evolutionary rescue* (Carlson et al. 2014), which also requires a demographic response, but evolutionary rescue can be achieved through de novo mutation (Bell and Gonzalez 2009, 2011), standing genetic variation (Zuk et al. 2006), or gene flow (i.e., genetic rescue). GR is therefore one specific type of evolutionary rescue. GR is also related to *assisted gene flow* (Aitken and Whitlock 2013), which has been defined as the managed movement of individuals into populations specifically to reduce local maladaptation to climate or other environmental change. However, assisted gene flow pertains to the action of moving alleles and does not by definition result in a positive demographic response.

This chapter focuses on the consequences of gene flow from other wild populations, rather than supportive breeding (release of individuals from captive populations). However, many of the issues and recommendations discussed in this chapter may also apply to captive breeding and wild release strategies.

## 1.2 Genetic Rescue in Conservation and Management

Prescribed gene flow intended for GR is typically viewed as a last-ditch attempt to prevent extinction of small populations. Despite widespread evidence for the reduction in fitness caused by inbreeding (Frankham 2005; Frankham et al. 2017), and well-documented evidence for declines and extinctions of small populations (Newman and Pilson 1997; Saccheri et al. 1998), there have been fewer than 30 known attempts globally at genetic rescue for conservation purposes (Frankham et al. 2017; Whiteley et al. 2015). Iconic examples of successful GR include increases in population fitness of European adders (Madsen et al. 2004), bighorn sheep (Hogg et al. 2006), greater prairie chickens (Westemeier 1998), and Florida panthers (Johnson et al. 2010). A recent literature search revealed that in studies which rigorously tested for absolute fitness effects of low levels of gene flow, most resulted in positive ($N = 10$) or a mix of positive or neutral fitness effects ($N = 4$) and few studies showed negative fitness effects ($N = 4$; Whiteley et al. 2015). However, GR is not widely implemented due to concerns over outbreeding depression, genetic swamping of the native genotypes, and risks of disease spread (Frankham et al. 2011). Uncertainty over the outcomes of augmented gene flow has limited widespread use of this strategy for conservation purposes. Yet, substantial *inbreeding depression* is likely to occur in the majority of recently fragmented small populations, especially under changing environments or stressful conditions (Keller and Waller 2002). Successful genetic rescue may be more likely when informed selection of candidate source and recipient populations remain possible, rather than waiting until the last minute when fewer options exist because multiple populations have gone extinct. Accumulating evidence for successful genetic rescue and harmful effects of inbreeding depression suggest this strategy should be more widely considered.

### *1.3 Outline of the Chapter*

Genomic approaches have not been widely utilized for genetic rescue [with the exception of Miller et al. (2012)], but there are many opportunities for the application of genomic tools to inform and improve the design, implementation, and monitoring of GR. Ideas for how genomics can and does contribute to conservation have been reviewed in Allendorf et al. (2010), Benestan et al. (2016), Funk et al. (2018), Kohn et al. (2006), and Ouborg et al. (2010). The purpose of this chapter is specifically to address how genomic tools can inform and improve GR and to provide recommendations for conservation biologists and practitioners interested in GR.

We first discuss how genomic tools can be used to identify populations that are most in need of gene flow augmentation. Genomics can increase effectiveness of GR by helping identify which potential source populations and even donor individuals are most likely to have the desired effect of increasing fitness and population growth rates in declining target populations (Whiteley et al. 2015). In many ways, predicting whether a declining population will benefit from genetic rescue involves evaluating and mitigating inbreeding and inbreeding depression, processes that are becoming better understood as large next-generation sequencing (NGS) datasets from non-model organisms accumulate. In this section, we also discuss screening candidate target populations for loci that may be indicative of inbreeding depression or cause negative effects on fitness. For imperiled taxa, risks of outbreeding depression, genetic swamping, and disease must be weighed against possible extinction due to inaction. In Sect. 3, we discuss how genomic tools can be used to identify variables associated with outbreeding depression and therefore used to predict the risk that it will occur. We base our suggestions in this section on the (Frankham et al. 2011) flowchart for predicting probability of outbreeding depression (Fig. 1). We then discuss how genomics can be used to monitor the outcomes of augmented gene flow as well as the underlying genetic mechanisms giving rise to those outcomes. As more studies document genomic patterns underlying genetic rescue and outbreeding depression, the uncertainty in predicting fitness effects of gene flow will diminish. Finally, we provide specific recommendations for design, implementation, and monitoring gene flow augmentations with the goal of improving understanding and likelihood of successful genetic rescue outcomes in conservation and management.

## 2 Evaluating the Potential Benefits of Genetic Rescue

Determining whether a population will benefit from augmented gene flow is a major challenge for wildlife managers. Not all small populations that have lost substantial genetic variation are necessarily headed for extinction, and there are cases of populations having recovered from extreme bottlenecks without the need for

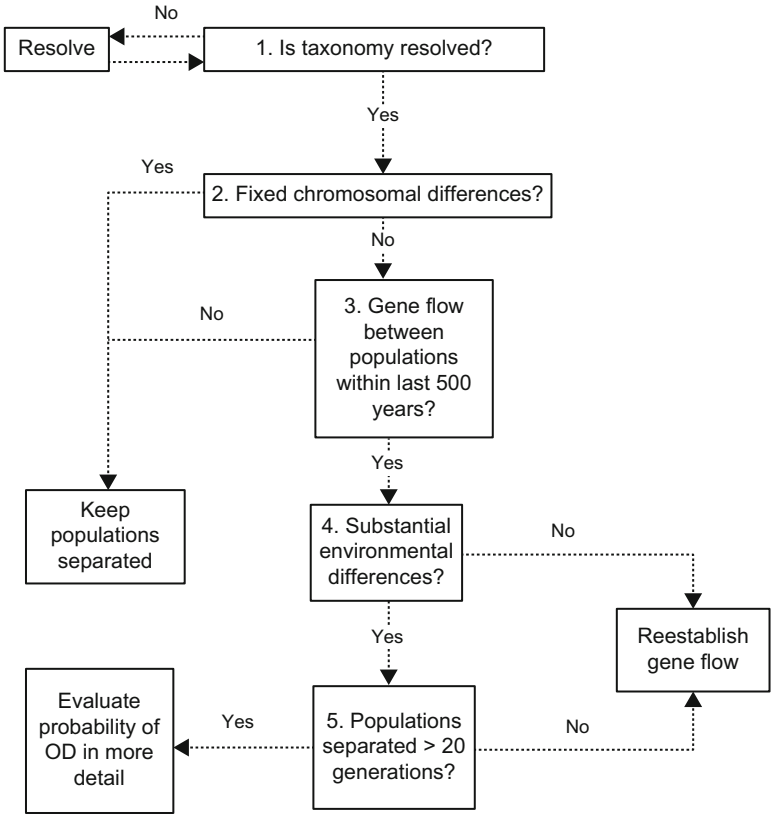


**Fig. 1** Decision tree for determining the probability of outbreeding depression (OD) between two populations reproduced from Frankham et al. (2011). Genomic data can be used to inform the five questions in this decision tree

augmented gene flow (Allendorf et al. 1982; McCullough et al. 1996). However, small and fragmented populations are undoubtedly more vulnerable to extinction due to a combination of demographic and genetic factors, and these have contributed to staggering losses of populations worldwide (Ceballos et al. 2017). Additionally, evidence for inbreeding depression in small populations is ubiquitous, especially under changing environments and/or stressful conditions (Frankham et al. 2017; Hoffmann et al. 2017). Given the uncertainty associated with augmentation, the decision to carry out this management strategy should involve an informed risk assessment, including the risks associated with inaction (Meek et al. 2015). When paired with a strong understanding of the natural history of the species, genomic data can provide invaluable insights as to whether a population will benefit from genetic rescue as well as how to prioritize target recipient populations given limited resources.

## 2.1 Inbreeding and Inbreeding Depression

Large-scale molecular genetic datasets have improved the ability to measure and understand *inbreeding* and inbreeding depression and their importance in evolution and conservation (Hedrick and Garcia-Dorado 2016; Hoffman et al. 2014; Kardos et al. 2015, 2016; Kristensen et al. 2010). Small populations, where most or all individuals are closely related, are particularly vulnerable to negative fitness consequences of inbreeding (Keller and Waller 2002). Inbreeding depression is caused by increased homozygosity for either (1) recessive or partially recessive deleterious alleles or (2) alleles at loci with heterozygote advantage (Charlesworth and Charlesworth 1999; Charlesworth and Willis 2009). Cumulative effects of inbreeding depression on individuals in small populations can cause population declines and increase the probability of extinction (Saccheri et al. 1998). Given all that is known about detrimental effects of inbreeding depression, managers should be confident that populations suffering from inbreeding depression will stand to benefit from augmented gene flow. One of the most promising ways that genomics can improve implementation and success of genetic rescue (GR) is by taking advantage of the advances in detection and understanding of inbreeding and inbreeding depression. Improved estimation of the *genetic load* caused by drift and inbreeding in small populations will allow researchers to screen candidate populations for those at highest risk of inbreeding depression and in greatest need of genetic variation from gene flow.

Inbreeding is typically measured with Wright's inbreeding coefficient ($F$), which is the probability that a diploid individual carries two alleles that are *identical by descent* (IBD) at a given locus (Wright 1931). Traditionally, the use of a detailed pedigree was the best approach for estimating $F$ (Pemberton 2004, 2008; Slate et al. 2004), wherein the inbreeding coefficient ($F_p$) predicts the probability of a locus being identical by descent based on a known pedigree where founders are assumed to be unrelated and noninbred (Wright 1922). However, this assumption is often not met in small, declining populations, and reliable pedigrees are unavailable for most organisms. In addition, $F_p$ is an estimate of the expected probability of IBD, rather than the realized proportion of the genome that is IDB. In contrast, genomic datasets can provide precise estimates of individual inbreeding through the analysis of genetic variation across individual genomes and can be used to study inbreeding depression without needing to conduct parentage analysis over many generations.

Recently fragmented populations that experience the highest levels of inbreeding should stand to benefit the most from immigration and be prioritized over those with lower inbreeding to receive prescribed gene flow. Therefore, identifying populations where mean $F$ is high should be a top goal for identifying candidate populations for GR. However, the extent of inbreeding depression depends on the evolutionary history of the population (Garcia-Dorado 2012). For example, if a population has been small for many generations, purifying selection facilitated by inbreeding may purge some of the genetic load, even if the population has very high inbreeding coefficients (Garcia-Dorado 2012). It is important to point out that although purging

has received a large amount of theoretical attention, little empirical evidence has been documented for purging in animal and plant populations (Gulisija and Crow 2007; Leberg and Firmin 2008). Ideally, we would want to know the relationship between fitness and inbreeding at the individual level to directly test for evidence of inbreeding depression in the target population (see Sect. 2.4). However, if fitness data are unavailable, there are several possible ways to use genomic data to indirectly assess whether a population has been recently isolated and subject to inbreeding or has undergone recent declines (see Sects. 2.2 and 2.3).

## 2.2 Estimating Inbreeding with a Reference Genome

If the target species or a close relative has a reference genome, availability of thousands of typed loci with known physical or genetic positions in the genome enables estimation of $F$ based on discrete chromosome segments in an individual's genome that are identical by descent (IBD). Contiguous regions of the genome where an individual is homozygous across all sites are known as a *run of homozygosity* (ROH) (Ceballos et al. 2018). The inbreeding coefficient $F_{\mathrm{ROH}}$ is estimated as the proportion of the genome that is in ROH:

$$F_{\mathrm{ROH}} = \Sigma L_{\mathrm{ROH}} / \Sigma L_{\mathrm{AUT}}$$

where the numerator is the sum of the base pair lengths of all the ROH of a given size or larger (long run of homozygosity) and the denominator is the sum of the overall length of the autosomal genome covered by the SNPs in the dataset (Keller et al. 2011). $F_{\mathrm{ROH}}$ ranges from 0 to 1 and provides a direct estimate of the proportion of the genome in an individual that is IBD due to recent inbreeding (Hedrick and Garcia-Dorado 2016; Keller et al. 2011).

The distribution and lengths of ROH can also be analyzed to infer population history, which can be used to infer how long a population has experienced inbreeding (Kirin et al. 2010). Many long ROH suggest recent small effective population size ($N_e$) with insufficient generations for recombination to break up ROH. Many short ROH, in contrast, suggest a longer history of small $N_e$. Recently small populations are more likely to experience strong inbreeding depression than those that have remained small for multiple generations because selection has not had time to purge the genetic load (Charlesworth and Willis 2009). Therefore, populations with a history of more recent reductions in $N_e$ should be prioritized as candidates for GR.

Recent fluctuations in $N_e$ (i.e., over the past few hundred generations) can be identified using large-scale SNP or whole-genome resequencing data. For example, a method called *IBDNe* uses the distribution of lengths of chromosome segments with shared IBD between pairs of individuals to estimate a time series of $N_e$ from a few generations before sampling to a few hundred generations back in time (Browning and Browning 2015). Using this method, Kardos et al. (2017) uncovered

population variation in effective population size within flycatcher species and, importantly, documented substantial reduction in $N_e$ within the last 50 generations. Populations that have undergone recent and/or human-caused bottlenecks should be considered high-priority candidates for GR over populations that have historically persisted with $N_e$.

## 2.3 Estimating Inbreeding Without a Reference Genome

Until recently, multiple-locus heterozygosity (MLH) and other marker-based measures of $F$ had low precision because most studies used small numbers of loci (e.g., usually a few dozen or fewer microsatellites; Kardos et al. 2016). Large numbers of SNPs can now readily be analyzed for any organism, providing greater precision in marker-based measures of $F$, even for species where linkage maps are not available. Heterozygosity-based measures (such as MLH) are indirect estimates of individual inbreeding based on the idea that individuals whose parents are more closely related will have lower heterozygosity on average across the genome due to the presence of IBD chromosome segments (Crow and Kimura 1970; Szulkin et al. 2010). The usefulness of indirect measures of $F$ depends on the number of loci and their expected heterozygosity and the variance of $F$ (Kardos et al. 2015; Miller et al. 2014; Slate et al. 2004). Studies of simulated populations show that mean genome-wide heterozygosity within a population can be estimated precisely by genotyping a modest number of loci spread throughout the genome on a surprisingly small number of individuals. For example, fewer than 1,000 SNPs typed in 30 individuals with a mean $H_e = 0.3$ is sufficient to clearly differentiate populations with high versus low mean $F$ (Kardos et al. 2016). Populations with high mean $F$ should be prioritized as candidates for genetic rescue. The *inbreedR* package in R (Stoffel et al. 2016) provides a useful framework to estimate variation in individual inbreeding; correlations between inbreeding, MLH, and fitness; and the ability to evaluate the precision and magnitude of estimates based on simulations of different numbers of genetic markers (Table 1).

A limitation of using unmapped markers is that it is not possible to explicitly identify IBD chromosome segments, and mean heterozygosity will be reduced both in populations with recently reduced effective population size ($N_e$) and in populations that have been small for many generations. However, detection of population bottlenecks based on heterozygosity excess, as implemented in the program Bottleneck (Cornuet and Luikart 1996), can be used with genome-wide SNP data, as in Funk et al. (2016).

**Table 1** Suggested data types and non-exhaustive list of statistical approaches for addressing problems associated with characterizing inbreeding, predicting probability of outbreeding depression (OD), and monitoring the spread of immigrant alleles using genomic data

| How it informs genetic rescue | Specific problem to address | Type of genomic data | Need reference genome? | Suggested software for analysis | Platform | Citation |
|---|---|---|---|---|---|---|
| Identify candidate recipient populations; screen donor individuals | Estimate inbreeding using ROH | RR, WGS | Y | Plink | Command line | Purcell et al. (2007) |
| Identify candidate recipient populations; screen donor individuals | Estimate inbreeding using heterozygosity-based measures | RR, WGS | N | inbreedR | R package | Stoffel et al. (2016) |
| Identify candidate recipient populations | Estimate effective population sizes | RR | N | NeEstimator | Java (GUI) | Do et al. (2013) |
| Identify candidate recipient populations | Test for recent population bottleneck | RR, WGS | Y | IBDNe | Command line | Browning and Browning (2015) |
| Identify candidate recipient populations | | RR | N | Bottleneck | GUI | Cornuet and Luikart (1996) |
| Identify candidate recipient populations; screen donor individuals | Test for reduced heterozygosity in MHC loci | Targeted DNA sequencing | N | jMHC | Java (GUI) | Stuglik et al. (2011) |
| Choose source population with low probability of OD | Clarify taxonomic status | Targeted DNA sequencing | N | Anchored hybrid enrichment | Command line | Lemmon et al. (2012) |
| Choose source population with low probability of OD | Test for evidence of fixed inversions | RR, WGS | N | LDna | R package | Kemppainen et al. (2015) |
| Choose source population with low probability of OD | Estimate demographic history and divergence times using IBD tracts | RR, WGS | Y | Shared haplotype methods | Command line | Gravel (2012), Harris and Nielsen (2013), and Pool et al. (2010) |
| Choose source population with low probability of OD | Estimate demographic history and divergence times using site-frequency spectrum | RR, WGS | N | δaδi | Command line | Gutenkunst et al. (2009) |

(continued)

**Table 1** (continued)

| How it informs genetic rescue | Specific problem to address | Type of genomic data | Need reference genome? | Suggested software for analysis | Platform | Citation |
|---|---|---|---|---|---|---|
| Choose source population with low probability of OD | Characterize population structure and recent admixture using neutral or outlier loci or both | RR, WGS | N | Structure | GUI | Pritchard et al. (2000) |
| | | RR, WGS | N | Adegenet | R package | Jombart and Ahmed (2011) |
| | | RR, WGS | N | BAPS | GUI | Corander et al. (2003) |
| | | RR, WGS | N | TreeMix | Command line | Pickrell and Pritchard (2012) |
| | | RR, WGS | N | SpaceMix | Command line | Bradburd et al. (2016) |
| Monitor outcome of gene flow augmentation | Test for non-neutral introgression patterns | RR, WGS | Y | ABC methods | Command line | Foll et al. (2015) |
| Monitor outcome of gene flow augmentation | | RR, WGS | Y | bgc | Command line | Gompert and Buerkle (2011) |

*WGS* whole-genome sequencing or resequencing, *RR* reduced-representation DNA sequencing

## 2.4 Inbreeding and Fitness

The extent of inbreeding depression resulting from recent reductions in population size is difficult to predict for a given population (Bouzat 2010). Multiple factors such as selection, history, environment, and genetic stochasticity contribute to fitness consequences of a demographic bottleneck. Linking estimates of individual fitness and inbreeding is therefore the best possible way to determine whether a population would benefit from introduced variation caused by gene flow. Reliable estimates of fitness in the wild therefore remain a top priority and limiting factor. Although it is becoming clear that non-pedigree-based estimates of inbreeding coefficient $F$ outperform pedigree-based estimates (Kardos et al. 2015, 2016), multigenerational wild pedigrees do provide the gold standard for fitness estimates in natural populations. In fact, multigenerational SNP data used for pedigree reconstruction and subsequent fitness estimates could also be used to estimate both $F_{\mathrm{ROH}}$ and lifetime reproductive success as exemplified by work done on federally endangered Florida scrub jays (Chen et al. 2016). Decreasing immigration rates in this highly fragmented species was shown to be associated with increased inbreeding and consequent inbreeding depression (Chen et al. 2016), providing strong evidence that populations may benefit from additional genetic variation supplied by gene flow.

A large body of literature seeks to address whether individuals that are more heterozygous tend to have greater fitness (Chapman et al. 2009; Coltman and Slate 2003; Szulkin et al. 2010). The theory behind heterozygosity-fitness correlations (HFCs) is that an increased proportion of loci in the genome that are identical by descent may lead to the unmasking of deleterious recessive alleles and a reduction in heterozygote advantage by decreasing genome-wide heterozygosity (Hansson and Westerberg 2002). Positive heterozygosity-fitness correlations (HFCs) have been documented across many species and traits (Hansson and Westerberg 2002; Szulkin et al. 2010). However, effect sizes are typically small, and in most cases heterozygosity estimates have been based on 10–20 microsatellite markers, providing little power to estimate the portion of loci in the genome that are IBD. In contrast, dense genetic marker panels provide much higher power than a small set of microsatellites to quantify the effects of inbreeding on fitness. If fitness or phenotypic trait-as-a-proxy-for-fitness data exist for a target species, individual estimates of heterozygosity can be used in testing for HFCs. The strength of the relationship between heterozygosity and fitness can provide evidence as to whether a population would benefit from added variation provided by gene flow.

## 2.5 Quantifying Functional Variation

Understanding intraspecific variation at functional loci may also inform which populations should be prioritized for gene flow augmentation. For example, populations and individuals could be screened for low MHC variation or a high

preponderance of putative loss-of-function or deleterious mutations. Although relatively little is known about the genomic architecture of inbreeding depression, the emerging consensus is that inbreeding depression is primarily caused by many loci of varying effect sizes (Paige 2010). If large effect loci are identified, populations could be screened for alleles known to have detrimental effects on fitness. In a recent study, Robinson et al. (2016) screened Channel Islands foxes for putative deleterious mutations and found twice the number of homozygous loss-of-function genotypes compared to more genetically diverse mainland gray foxes, implying a substantial genetic load in small, recently bottlenecked island fox populations. This approach could be used to identify populations with high genetic load to prioritize for assisted gene flow.

Targeted capture is an appealing approach for high-throughput sequencing of preselected genomic regions of interest (Jones and Good 2016). Although capture approaches rely on prior sequence knowledge, there are several options to overcome this barrier in non-model species without reference genomes. For example, panels of targeted genes can be selected a priori based on knowledge of gene function or genomic regions that have been identified from closely related species as potentially functionally relevant (Peñalba et al. 2014). De novo transcriptomic approaches (i.e., RNA-seq) may be used to design probes corresponding to exonic regions and may also be useful for investigating loss of variation in the expressed functional genome. There is some evidence that reduced variation in expressed genes underlies the commonly observed pattern that inbreeding depression is exacerbated in stressful environments (Schou et al. 2018). Importantly, RNA-sequencing technology can be used for functional genomic comparisons among individuals and/or populations without prior molecular information. Combining transcriptomics with stress-tolerance assays can help identify mechanisms underlying stress-sensitivity differences among individuals and populations (Jeffries et al. 2016). Although these approaches have not yet been used to inform gene flow augmentation, decreasing costs and increasing genomic resources and analytical tools for non-model organisms will facilitate integration of multiple types of "omics" data that will ultimately help identify and prioritize populations.

## 3 Predicting Outbreeding Depression with Genomics

Gene flow augmentation remains a controversial and perhaps underutilized management strategy largely due to concerns that outbreeding depression will cause reduced fitness of offspring between genetically divergent parents (Edmands 2006). The probability of outbreeding depression is generally determined by the time since isolation of immigrant and recipient populations, the magnitude of environmental differences and resulting level of adaptive divergence between populations, and the level of inbreeding in the recipient population (Frankham et al. 2011). For example, crossing populations with fixed chromosomal differences or those that have been geographically isolated for millions of years is likely to result in outbreeding

depression caused by the evolution of postzygotic reproductive barriers such as Dobzhansky-Muller incompatibilities (Coyne and Orr 2004; Edmands 1999). But the extent to which gene flow between differentiated populations reduces overall fitness when the recipient population is small, declining, and inbred remains uncertain. Given this gray area, managers are often understandably hesitant to move forward with augmentation. Frankham et al. (2011, 2017) provide useful guidelines and decision trees for predicting the probability of outbreeding depression (Fig. 1). However, the information needed to proceed along the tree is typically not available for non-model organisms such as those of conservation concern. Below we discuss how each of these decisions can be informed using genetic or genomic data. Suggested approaches for sequencing and analyses are provided in Table 1.

### *3.1 Resolving Taxonomic Uncertainty*

Taxonomic status of potential recipient and source populations should be resolved before establishing gene flow between them. In most cases, establishing gene flow between distinct species is not recommended because crosses between different species are expected to be more likely to result in outbreeding depression (Muhlfeld et al. 2009). On the other hand, species delineations that are split excessively may preclude genetic rescue of small inbred populations with low genetic variation (Frankham et al. 2017). In extreme cases, such as Australian honeyeaters (Harrisson et al. 2016), the only option may be to cross populations of closely related species or subspecies. Phylogenomic approaches, which take advantage of genome-wide data, advances in high-performance computing, and more realistic models of molecular evolution can be applied to resolve species-level taxonomy with greater accuracy than previous single marker approaches (Delsuc et al. 2005). Large genomic datasets combined with coalescent species delimitation methods can be used to detect distinct lineages in non-model organisms (Pante et al. 2014; Zhang et al. 2014). For example, a study that sequenced 40,000 bp of the nuclear genome in all living elephant taxa contributed to the recognition of the forest elephant as a distinct taxon from the savannah elephant (Rohland et al. 2010). However, multispecies coalescent methods risk oversplitting due to misidentification of population structure as putative species (Sukumaran and Knowles 2017). Implementation of beneficial gene flow augmentation and subsequent GR could be impeded if isolated populations are incorrectly being treated as distinct species. Resolving taxonomic uncertainties prior to augmentation of gene flow will be especially important in groups with high levels of cryptic diversity, where two or more distinct species are incorrectly classified as a single species. Multiple cryptic lineages of lizards endemic to the Brazilian Cerrado were identified using the anchored hybrid enrichment (AHE) method (Domingos et al. 2017; Lemmon et al. 2012). This is a cost-effective and relatively rapid method that uses enrichment probes in highly conserved regions of vertebrate genomes to capture more rapidly evolving adjacent regions, resulting in sequence fragments that

are informative at both deep and shallow phylogenetic levels (Lemmon et al. 2012). Protocols, probe sets, and bioinformatic pipelines for this method can be found at www.anchoredphylogeny.com.

## 3.2 Determining Potential for Fixed Chromosomal Differences

Fixation of chromosomal rearrangements in different populations can be caused by genetic drift or natural selection (Rieseberg 2001). Outbreeding depression is more likely between crosses of populations that have fixed chromosomal differences (Hoffmann and Rieseberg 2008; Rieseberg 2001). Crosses between populations with different numbers of chromosomes (e.g., due to polyploidy) cause the most adverse fitness effects, followed by translocations, centric fusions, and inversions, and the extent of negative fitness effects increases as the number of fixed differences increase (White 1978). Determining chromosomal information is particularly important when considering population crosses in groups with varying ploidy levels (e.g., plants) and/or high rates of chromosomal evolution (Frankham et al. 2017). Detecting chromosomal rearrangements is greatly aided by the availability of a reference genome. However, patterns of multilocus linkage disequilibrium (LD) may be used to detect signatures of chromosomal rearrangement polymorphisms such as inversions without a full genome (Kemppainen et al. 2015). The R package *LDna* can be used to identify loci that are putatively associated with inversions even without mapping information (Kemppainen et al. 2015). In order to evaluate concerns about potential fixed inversions among populations, a researcher could run *LDna* on a reduced-representation DNA sequencing (i.e., RADseq) dataset and screen for clusters of loci that share high LD and are therefore putatively associated with inversions. Pairs of populations without clusters of high LD regions would make better candidates for subsequent crosses than those with putatively fixed inversions.

## 3.3 Estimating Time and Number of Generations Since Most Recent Gene Flow

The goal of augmented gene flow in the context of genetic rescue is to reestablish connectivity between *recently* isolated populations. This is because recently isolated populations will have had less time to accumulate genetic incompatibilities that may cause outbreeding depression such as mitonuclear incompatibilities (Box 1). Furthermore, populations that have become fragmented and isolated as a result of human activity are likely the best candidates for genetic rescue. The 500-year time frame suggested by Frankham et al. (2011) encompasses the period of increased

fragmentation associated with exponential increases in human population size. In many cases of recent fragmentation, managers will have previous knowledge about when distinct populations became isolated. However, genomic data can inform cases where population divergence history is not known. Many methods have been developed for characterizing patterns of demographic history and population divergence times with multilocus data. Here we highlight two classes of methods that have been shown to perform well with large genomic datasets. The first class of methods requires a reference genome and is based on identical-by-descent (IBD) tracts of DNA that are shared within and between populations (Gravel 2012; Pool et al. 2010). Given that recombination breaks apart ancestry tracts through successive generations, timing of admixture can be inferred from tract length distribution. Tract length distribution is sensitive to recent migration events (i.e., tens of generations). A complementary second class of methods that does not necessarily require a reference genome is based on the joint site-frequency spectrum (SFS), which is the distribution of allele frequencies across polymorphic sites, and is more sensitive at time scales of hundreds to thousands of generations (Gutenkunst et al. 2009). A common method, diffusion approximation for demographic inference (δaδi), generates a SFS under a specific demographic scenario and maximizes the similarity between the expected allele frequency distribution and the observed SFS over the parameter values space (Gutenkunst et al. 2009). Liu et al. (2014) used both method classes to infer the joint demographic history of polar bears and brown bears, finding similar results with regard to divergence time, relative effective population sizes, and direction of gene flow.

**Box 1 Mitonuclear Interactions, Mismatch, and Females**

Coadaptation between mitochondrial and nuclear genotypes within populations is thought to be caused by compensatory evolution in the nuclear genome in response to mutation accumulation in the non-recombining, maternally inherited mitochondrial genome (Hill 2015). A concern of using females with non-native mitochondria for gene flow augmentation is the potential for mitonuclear incompatibilities in F2 or later generations (Fig. 2). These incompatibilities may result in negative fitness effects such as reduced fecundity, decreased longevity, metabolic deficiencies, lowered stress tolerance, and developmental abnormalities (Burton et al. 2013). This concern has not yet received much attention in the genetic rescue literature (Havird et al. 2016). However, mitochondrial DNA (mtDNA) is abundant in most eukaryotic cell types, and whole genome, exome, or transcriptome data from high-throughput sequencing can be mined for mitochondrial sequences (Ekblom et al. 2014). Phylogenetic analyses based on mitochondrial sequences from different populations can be used to determine how divergent populations are across their mitochondria. Populations with more similar mitochondrial haplotypes should be prioritized for crosses in order to avoid potential negative fitness

(continued)

**Box 1** (continued)

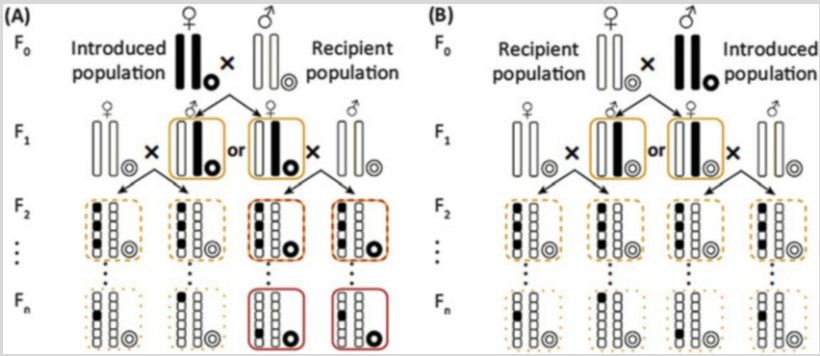


**Fig. 2** Mitonuclear incompatibility during genetic rescue reproduced from Havird et al. (2016). (**a**) When females are introduced during genetic rescue, mitonuclear incompatibilities can be propagated through generations (colored boxes), resulting in reduced fitness, especially in later generations, because females will bring in a novel mitochondrial genome that might be maladapted to the local nuclear genome. (**b**) Introducing males is less likely to cause mismatch, as their mitochondrial genomes are not transmitted. Each pedigree shows two nuclear chromosomes (linear) and a mitochondrial chromosome (circular). Orange boxes indicate genotypes that are at risk of mitonuclear incompatibilities if a single "mismatched" nuclear allele is sufficient to have a deleterious effect. Red boxes indicate genotypes that are at risk even if both nuclear alleles must be mismatched with the mitochondrial genome to have a harmful effect. Dashed and dotted lines indicate lower risks associated with a lower overall frequency of mismatched alleles

consequences associated with mitonuclear incompatibilities. If divergent mitochondrial haplotypes are the only option for crossing, managers should consider prioritizing males for translocation.

In addition to gene flow timing estimates, many population genomic methods are available for inferring how much historical and/or contemporary gene flow occurs among populations. The accuracy and precision of these methods has greatly improved with large genomic datasets. Common methods for assessing population structure and recent admixture include ADMIXTURE (Alexander et al. 2009), STRUCTURE (Pritchard et al. 2000), BAPS (Corander et al. 2003), principal component analysis (PCA) (Patterson et al. 2006), and *TreeMix* (Pickrell and Pritchard 2012). Spatial data can be included in programs such as SpaceMix (Bradburd et al. 2016) in order to summarize both genetic and geographic distances as well as admixture events.

### *3.4 Evaluating the Extent of Adaptive Differentiation*

The probability of outbreeding depression is generally thought to increase as crossed populations originate from increasingly divergent environments (Edmands and Timmerman 2003; Frankham et al. 2011; Storfer 1999). Additionally, strong selection against migrants may entirely prevent gene flow from occurring in the new environment. For example, an experiment in which Trinidadian guppies adapted to a low-predation environment were introduced to a high-predation site resulted in little to no gene flow because immigrants suffered very high mortality (Weese et al. 2011). On the other hand, the occurrence of adaptive phenotypic divergence should not automatically rule out the possibility of a successful genetic rescue. Experimental gene flow manipulation studies with guppies have shown clear positive benefits of gene flow when guppies are moved in the opposite direction, namely, when immigrants adapted to high-predation environments were added to populations adapted to low-predation environments (Fitzpatrick et al. 2016; Kronenberger et al. 2017a; Kronenberger et al. 2018; Box 2). Determining the extent to which adaptive differentiation predicts the probability of outbreeding depression will likely often be species- or trait-dependent. However, identifying genomic regions or specific loci underlying local adaptation can inform putative crosses and improve the design of translocation efforts.

**Box 2 Playing "God" with Guppies: A Model System for Genetic Rescue Experiments**

A recurring problem in conservation biology is the lack of replication, randomization, and experimental controls needed to conclusively document applied treatments such as augmented gene flow to observed outcomes, such as genetic rescue (Tallmon 2017). Trinidadian guppies have provided a powerful model system in which to conduct gene flow augmentation experiments and test predictions about subsequent outcomes on traits and fitness over multiple generations (Fitzpatrick et al. 2016, 2017; Kronenberger et al. 2017a, 2018). In Trinidad, translocation experiments facilitated testing the effects of gene flow from a divergent source into two small and isolated headwater guppy populations. Genetic rescue was documented in both streams, evidenced by sustained increases in population growth caused by high hybrid fitness (Fitzpatrick et al. 2016; Fig. 3). Furthermore, locally adaptive trait variation was maintained in the wild, despite high levels of homogenizing gene flow (Fitzpatrick et al. 2015), and common garden experiments revealed that gene flow caused genetically based evolution in many traits, some of which shifted in the adaptive direction (Fitzpatrick et al. 2017).

Mesocosm experiments, with the advantage of controls and replication, mimicked the above-described field scenario but also allowed manipulation of the source of gene flow. Immigrants from source populations with varying

(continued)

**Box 2** (continued)

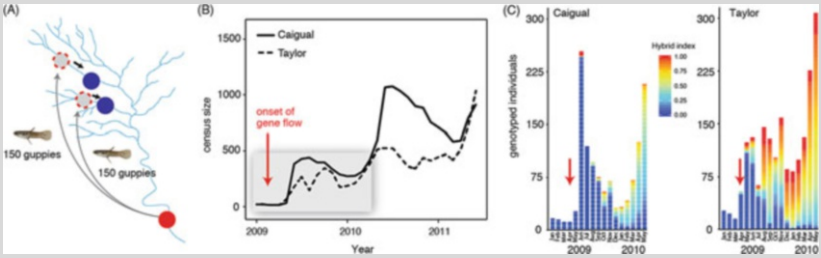


**Fig. 3** Gene flow manipulation experiments in Trinidad. (**a**) Map of the Guanapo River drainage. In 2009, guppies were translocated from a downstream high-predation locality (red) into two headwater sites (dashed red) that were upstream of native recipient populations in low-predation environments (dark blue). Unidirectional, downstream gene flow began shortly after the introductions, indicated by black arrows. (**b**) Census sizes in Caigual (solid) and Taylor (dashed) following the onset of gene flow from the upstream introduction sites. Gray box indicates the time span in which all captured individuals were genotyped at 12 microsatellite loci. (**c**) Temporal patterns of continuous hybrid index assignments throughout the first 17 months of the study (~4–6 guppy generations). Recipient populations prior to gene flow had a hybrid index = 0, and pure immigrant individuals had a hybrid index = 1. Hybrid indices were assigned using data from 12 microsatellite loci. Red arrows indicate the onset of gene flow

levels of adaptive and neutral genetic differentiation were added to replicated small populations of lab guppies (Kronenberger et al. 2017a, 2018). Gene flow generally increased population sizes relative to no gene flow controls, especially when recipient populations started with lower levels of genetic variation (Kronenberger et al. 2018), but in one experiment, gene flow from the adaptively similar source was most beneficial (Kronenberger et al. 2017a).

This work has provided important experimental support to the increasing consensus that gene flow often has a positive effect on fitness, especially when recipient populations are small, isolated, and inbred. Importantly, positive effects were documented even when immigrants originate from an adaptively divergent source, suggesting that adaptive differentiation should not automatically rule out translocation.

Detecting local adaptation is often difficult for rare and endangered species, but genomic tools provide several opportunities to achieve this (Hoban et al. 2016). Methods for identifying loci that are adaptive or linked to adaptive loci can be used to estimate how adaptively divergent various potential source populations are from the target population (Funk et al. 2012, 2018). Two general approaches for identifying local adaptation throughout the genome include (1) identifying loci with unusually high genetic differentiation among populations (i.e., outlier methods) and (2) scanning for correlations between local population allele frequencies or individual genotypes and local environments [i.e., genotype-environment

association (GEA) methods]. Both approaches benefit from the availability of a reference genome, but neither requires it. Outlier methods use average genome-wide differentiation between populations for detecting outlier loci that show unusually high divergence (measured by $F_{ST}$ or other measures of differentiation) and are presumably under divergent selection (Nielsen 2005). High $F_{ST}$ outliers can be identified that are putatively adaptive or linked to adaptive loci, and these loci can be used to estimate how adaptively divergent various potential source populations are from the target recipient population. One of the problems with this method is that processes other than local adaptation (i.e., stochastic effects in declining or expanding populations, cryptic hybrid zones, variation in recombination rates across the genome) can also lead to $F_{ST}$ outliers (Lotterhos and Whitlock 2014). Genotype-environment association methods test for associations between SNP frequency or genotype and environmental variables while controlling for population structure (Rellstab et al. 2015). Populations on a landscape typically exhibit some degree of isolation by distance such that spatial autocorrelation in allele frequencies can cause associations between gene frequencies and the environment by chance (Meirmans 2012), but some GEA methods control for population structure (Forester et al. 2018).

Findings of outlier loci or genotype-environment associations should not necessarily preclude population crosses. Rather, these patterns should be evaluated in the context of optimizing the choice of source population: one that is moderately differentiated from the recipient population at neutrally behaving loci and minimally differentiated at putatively adaptive loci. If a range of source populations does not exist, Frankham et al. (2011) advise that if populations have been separated for more than 20 generations in divergent environments, then they should be subjected to a detailed analysis and potentially crossed in an experimental context before large-scale translocations for augmenting gene flow are undertaken. On the other hand, if environments are more or less similar, emerging consensus suggests that mating between populations that have been isolated for up to 50–100 generations can cause fitness increases with minor risk of outbreeding depression (Aitken and Whitlock 2013; Allendorf et al. 2013; Frankham et al. 2011). Ideally, evaluations of historic and contemporary gene flow patterns among populations (Sect. 3.3) should be carried out in addition to the analyses outlined in this section on characterizing adaptive variation. Large-scale genomic datasets that contain neutral- and non-neutral-behaving loci can therefore be used to evaluate both adaptive and neutral differentiation among potential crosses (Funk et al. 2012).

## 4 Monitoring and Mechanisms

### *4.1 Monitoring the Outcome of Genetic Rescue*

Genomics will not only play a key role in determining if gene flow augmentation should be attempted for the purpose of genetic rescue, and if so, which source populations and individuals should be used, but also in monitoring the outcome of

genetic rescue attempts. In order for genetic rescue to be considered successful, the first prerequisite is that immigrants interbreed with local residents to produce viable offspring. Genomics will provide high power to determine if this occurs. Genomic approaches involving hundreds or thousands of SNPs can accurately identify different categories of hybrids, including first-generation crosses (F1), second-generation crosses (F2), and backcrosses (F1 × immigrant or F1 × resident) (Hohenlohe et al. 2013). To date, most genomic studies of hybridization have focused on interspecific hybridization, but genomics should also provide higher power to detect intraspecific introgression due to the large number of loci available. For example, Funk et al. (2016) detected evidence for a low level of historical introgression from one island fox population to another using 4,858 SNPs, suggesting movement of foxes by humans, which was supported by shared mitogenomic haplotypes (Hofman et al. 2015). Genomic monitoring of the augmented population will also provide important information about the persistence of the benefits of gene flow over multiple generations. A concern of crossing divergent populations is that initial heterosis will give way to a reduction in fitness as recombination breaks apart coadapted gene complexes (Edmands 1999). However, empirical studies suggest that genetic rescue can produce benefits that persist for many generations in outbred species (Frankham 2016). Genomic monitoring can clear up this question by documenting the frequency of different hybrid classes following gene flow. This task can be implemented in software such as NewHybrids (Anderson and Thompson 2002).

Second, once it is determined that immigrants are interbreeding with residents, the next key question concerning the outcome of genetic rescue is which individuals and alleles have the greatest effect on fitness. This information is key for fine-tuning any future genetic rescue efforts. If specific alleles have a disproportionate effect on fitness, then future genetic rescue attempts can identify individuals that carry the combination of alleles at multiple loci that are predicted to have the greatest positive effect. Genomic approaches has already been used for this purpose for a population of bighorn sheep at the National Bison Range in Montana, USA, which was the recipient of immigrants as part of a successful genetic rescue management strategy (Miller et al. 2012). Miller et al. (2012) found that 30 loci had effects on fitness above and beyond that predicted by overall levels of introgression. Targeted DNA sequencing approaches could also be used to determine allele frequency changes in functional regions of the genome. For example, Grueber et al. (2017) found that a small number of migrants into isolated New Zealand South Island robin populations restored genetic diversity of immunogenetic toll-like receptor and MHC loci. This type of information could be used in future gene flow augmentation efforts to choose source individuals with beneficial alleles at these loci.

Lastly, genomics should be used to monitor changes in genetic variation and inbreeding after release of immigrants. If natural connectivity is not reestablished to the recipient population, then, eventually, genetic variation will likely decline, and inbreeding will increase after genetic rescue. Ultimately, genetic rescue is a stop-gap measure. A single genetic rescue event is not expected to prevent the ongoing effects of small population size and isolation, as seen in Isle Royale wolves (Hedrick et al.

2014). The only way to prevent repeated loss of genetic variation would be to reverse the ultimate causes of population declines, whatever they may be for a particular population, and restore natural connectivity. Thus, repeated augmentation may be necessary. In order to determine if and when new immigrants should be added, genomic monitoring can be used to assess if genetic variation declines or inbreeding increases again after an initial genetic rescue effort (Schwartz et al. 2007). Any genetic rescue attempt should explicitly plan for ongoing genomic monitoring and decide a priori at what point additional genetic rescue efforts are required.

## *4.2 Determining the Genomic Mechanisms of Genetic Rescue*

There are three primary mechanisms that can contribute to genetic rescue (Whiteley et al. 2015). Two of these depend on the genetically distinct mechanisms of inbreeding depression (Charlesworth and Willis 2009). For example, if inbreeding depression is primarily caused by the expression of deleterious alleles in the homozygous state – referred to as dominance – then genetic rescue is caused by the masking of these deleterious alleles by beneficial, dominant alleles from the source population. On the other hand, if inbreeding depression is caused by increased homozygosity for alleles at loci where there is heterozygote advantage (i.e., overdominance), then genetic rescue is caused by the increase in heterozygosity due to gene flow from the source population. The third mechanism that can result in genetic rescue does not invoke inbreeding depression and is the introduction of novel, adaptive alleles that perhaps never previously occurred in, or have been lost from, the recipient population. To complicate matters more, these mechanisms are not mutually exclusive, and they can all conceivably occur simultaneously at different loci.

Understanding the mechanisms causing fitness declines and subsequent genetic rescue can inform how best to implement genetic rescue by guiding choice of donor population and individual immigrant selection. There are multiple analytical frameworks that can harness genomic data to determine which of the above mechanisms contribute to genetic rescue at specific loci and be used to inform management. Perhaps the most powerful approach is to directly link estimates of fitness to genotypes (homozygous for recipient alleles, homozygous for immigrant alleles, or heterozygous) across the genome (Miller et al. 2012). Fitness is best estimated as lifetime reproductive success, which can be inferred using pedigree analysis (Chen et al. 2016). General linear models can be used to test for associations between fitness and overall levels of introgression, as well as genotypes at specific loci (Zhu et al. 2013). For example, if the mechanism of genetic rescue is dominance, then individuals that are homozygous for the immigrant allele can be targeted and should have the highest fitness. In contrast, if the mechanism is overdominance, heterozygotes will have the highest fitness. If high fitness immigrant alleles are identified,

then source individuals with these alleles can preferentially be used for genetic rescue. Genomic approaches such as these promise not only to help inform which individuals to use for genetic rescue but also will greatly advance our understanding of the genetic basis of inbreeding depression and genetic rescue.

## 5 Future Perspectives and Conclusions

We echo the paradigm shift suggested by Frankham et al. (2017), whereby evidence of genetic differentiation among populations should trigger questions of whether any population segments are suffering genetic problems and if they can be rescued by restoring gene flow, rather than routinely recommending that segments be managed separately. Many, if not most species of conservation concern are declining in part due to problems associated with habitat loss, fragmentation, and isolation. Genomic data can now provide a wealth of information on key demographic and evolutionary processes (e.g., population size, gene flow, adaptive variation and potential) linked to viability.

### *5.1 Proposed Guidelines for Using Genomics to Improve Genetic Rescue*

In this section, we provide recommendations for steps to take following the collection of population genomic data from a species of conservation concern. In Box 3 we provide a hypothetical scenario in which gene flow augmentation would be useful and how genomics can inform difficult choices. Sampling design and marker choice should be based on the set of scientific questions needing answers (Benestan et al. 2016; Funk et al. 2012). The cheapest and most comprehensive approach will likely be one of the reduced-representation DNA sequencing methods such as RADseq or GBS (Andrews et al. 2016). Increasingly, whole-genome sequences are becoming available for species of conservation concern with smaller genome sizes, providing even more opportunities for pinpointing the specific genomic regions underlying reduced fitness and declines. If resources permit, whole-genome sequencing is preferred over reduced-representation methods. However, it is important to weigh the trade-offs between whole-genome sequencing of fewer individuals from fewer populations versus the capacity to evaluate a larger number of individuals and populations using a reduced-representation method (Catchen et al. 2017; Lowry et al. 2016). As a general rule, it has been shown that 25–30 individuals per population are required to accurately estimate allele frequencies, expected heterozygosity, and genetic distances between populations (Hale et al. 2012).

**Box 3 Hypothetical Scenario**

Figure 4 represents a hypothetical management scenario in which genomic data could inform gene flow augmentation and lead to successful genetic rescue. Seven extant and fragmented populations of varying census sizes (black circles) are found throughout a previously continuous distribution (gray background) along an environmental gradient, and three populations are known to have recently gone extinct (white circles with dashed red outline). We use this hypothetical scenario to illustrate the recommendations from Sect. 5.1 of this chapter.

*Step 1*: Sample tissue and collect genome-wide SNP data (e.g., RADseq or whole-genome resequencing data) from 25 to 30 individuals from each of the seven extant populations.

*Step 2*: Assess variation in inbreeding coefficients and genome-wide heterozygosity, and determine candidate populations to receive gene flow

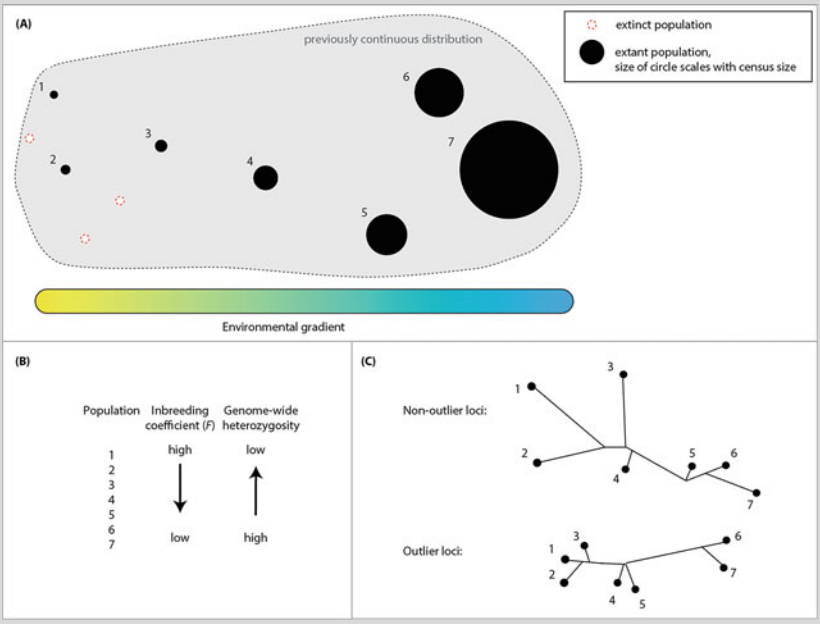


**Fig. 4** Hypothetical scenario of a fragmented species that would likely benefit from gene flow augmentation (**a**) in which a species from a previously continuous distribution (outlined in gray) now exists in isolated populations (black circles) along an environmental gradient. Several small populations outlined by red dashes have already gone extinct. Extant populations range in inbreeding coefficient (*F*) and genome-wide heterozygosity (**b**). Neighbor-joining trees (**c**) using non-outlier versus outlier marker sets show different patterns of population similarity

(continued)

**Box 3** (continued)

augmentation. In this scenario, population 1 has the highest inbreeding coefficient and lowest levels of genomic variation. The added evidence that other nearby populations of similar size have gone extinct suggest that this edge population is vulnerable to extinction and could benefit from gene flow.

*Step 3*: Quantify population genomic structure based on non-outlier or "neutral" loci. Neighbor-joining tree based on these loci suggest that populations 1–3 have experienced high genetic drift and divergence from each other and all other populations.

*Step 4*: Quantify population genomic structure based on outlier or "putatively adaptive" loci. Neighbor-joining tree based on this subset of loci suggest that there are three adaptively similar clusters (1–3, 4–5, 5–7).

The ideal source population for gene flow augmentation would be genetically diverse, somewhat divergent at neutral loci, and adaptively similar. In this scenario, populations 2 and 3 are good candidates in that they are found in a similar environment and are moderately divergent at neutral loci. However, these populations are quite small themselves and potentially vulnerable to extinction. There may not be individuals to spare for translocation. The largest populations (6 and 7), while they may be the most diverse, are the most dissimilar to the recipient population at both neutral and outlier loci, suggesting they pose the greatest risk for outbreeding depression. Given these constraints, populations 4 or 5 would be optimal source populations. Note that the recommendations in this scenario are based solely on population genomic results. In reality, translocation planning will also involve ecological, behavioral, logistical, financial, or other factors and constraints that may contribute to the final decision-making.

Proposed guidelines after population genomic data has been acquired are as follows:

1. Test whether a population will benefit from augmented gene flow.
   (a) *Ideal scenario*: Directly test for evidence of inbreeding depression by testing for a relationship between individual inbreeding coefficient ($F$) and fitness (i.e., through collection of pedigree data). Direct evidence of inbreeding depression is the most convincing evidence that a population would benefit from augmented gene flow.
   (b) *Most likely scenario*: Identify isolated populations with low mean heterozygosity, high inbreeding coefficients, and/or a history of recent reduction in effective population size. Populations with these characteristics would likely benefit from augmented gene flow.

2. Identify source populations for translocations.
   (a) *Ideal scenario*: Identify "Goldilocks" populations with intermediate genomic divergence at neutral loci and low levels of divergence at putatively adaptive (i.e., outlier) loci (Box 4).
   (b) If there is strong evidence that a population is declining due to inbreeding depression, a less than ideal scenario should be pursued in order to prevent extinction. For example, source populations that are either highly similar or highly divergent could be used, given that we know that even when source populations are divergent, successful GR can still work (Fitzpatrick et al. 2016; Kronenberger et al. 2017a, 2018).
3. Screen source populations for donor individuals.
   (a) Many context-dependent nongenetic factors (e.g., sex, stage, disease) should be considered in the choice of donor individuals. For example, female panthers from Texas were chosen for gene flow augmentation of the Florida panther population due to concerns about male-male aggression (Johnson et al. 2010).
   (b) In addition to nongenetic factors, genomic screening of a set of candidate donor individuals could be used to help identify which individuals are most likely to reduce inbreeding depression and most likely to result in successful genetic rescue (GR). For example, candidate individuals could be screened for highest levels of genome-wide heterozygosity or fewest and shortest runs of homozygosity. In most cases, fitness-related and life history traits are multigenic, and thus estimates of genome-wide statistics will be most informative for screening for "genetic health." However, if alleles associated with inbreeding depression have been previously identified, individuals could be pre-screened for absence of these alleles. Finally, candidate individuals could be screened for alleles that have certain known desirable effects in a given environment (e.g., alleles conferring disease resistance or heat tolerance). Risks of causing genetic bottlenecks or selective sweeps should be considered if only particular variants are targeted for management. Similar risks are associated with the idea of gene editing for conservation (Box 5).
4. Monitor outcomes of gene flow augmentation.
   (a) *Before translocation*: Sample tissue from all donor individuals and as many individuals in the recipient population as possible for genotyping.
   (b) *After translocation*: Sample tissue from as many individuals as possible for multiple generations following the translocation in order to track the allele frequency changes. Ideally, pedigrees can be inferred for estimating individual fitness to conclusively document genetic rescue as being caused by an infusion of new alleles. Genome-wide genotypic data also allow inference of the genomic architecture underlying population-level responses to gene flow.

**Box 4 How Much Introgression and Immigrant Divergence Is Too Much?**

The goal of genetic rescue is to increase population fitness without swamping out locally adapted alleles or causing outbreeding depression. To achieve this goal, it is important to consider the amount of gene flow and the level of divergence of the source population. Too little gene flow will result in continued loss of genetic variation, while too much can swamp out locally adaptive alleles (Fig. 5). Based on early population genetics theory of Wright (1931), the rule of thumb of one migrant per generation emerged as an appropriate level of gene flow to maintain within population genetic variation without the loss of important allelic differences among populations (Spieth 1974). However, Mills and Allendorf (1996) argued that this amount of gene flow may be inadequate for many populations. They conclude that 1–10 migrants is a better rule of thumb and that the species' natural history and particular situation must be taken into account (Mills and Allendorf 1996). For example, if migrants are likely to have low survival or breeding success, more may be needed. On the other hand, if a small number of immigrants are introduced, and they have much greater fitness than residents, then most individuals will be offspring of the immigrants within a few generations, potentially resulting in inbreeding depression as has happened in Isle Royale wolves (Hedrick et al. 2014).

Another important consideration is the level of divergence of immigrants. If immigrants are too genetically similar to the recipient population, then they will not contribute many new alleles, and their contribution will primarily be demographic. On the other hand, if immigrants are too divergent, then outbreeding depression can occur (Frankham et al. 2011). The most divergent

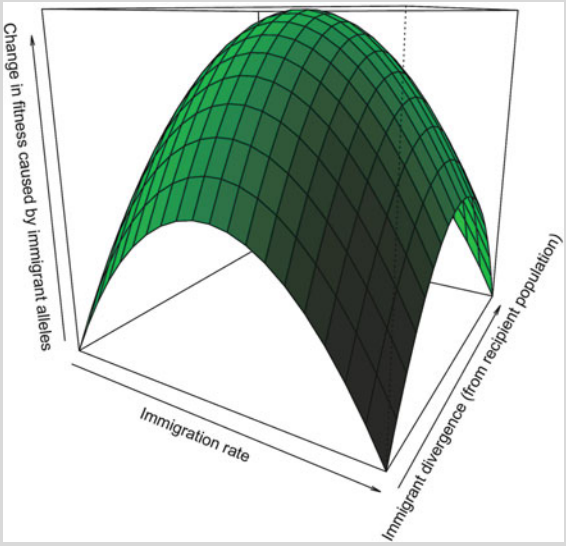


**Fig. 5** Conceptual illustration of the predicted change in fitness caused by immigrant alleles as a function of immigration rate and immigrant divergence

(continued)

**Box 4** (continued)
type of immigrant is a different species, resulting in interspecific hybridization, which is well recognized to generally result in reduced fitness (Allendorf et al. 2001; Kovach et al. 2016a). Some have suggested purposely using different species for introduction of adaptive alleles (Hamilton and Miller 2016), but others have warned of the danger of intentional interspecific hybridization (Kovach et al. 2016b; Miller and Hamilton 2016). Frankham et al.'s (2011) analysis concluded that outbreeding depression is most likely when the individuals that are crossed are different species, have fixed chromosomal differences, have not exchanged genes in the last 500 years, or are adapted to different environments. These types of crosses should, therefore, generally be avoided. However, experimental gene flow manipulations documented genetic rescue in field and mesocosm guppy populations that received adaptively and neutrally divergent immigrants. They conclude that if the only source population available is genetically or adaptively divergent, genetic rescue can still work and may be a better option than no action.

**Box 5 Possibilities and Pitfalls of Genetic Rescue Through Transgenic Augmentation**
In a bold, high-profile opinion piece in *Nature*, Thomas et al. (2013) argued that genetic engineering could be used to genetically improve endangered species by introducing adaptive variants. The idea of a technological fix to endangered species conservation is tantalizing, but several complications make genetic "tweaking" problematic (Hedrick et al. 2013).

First, low fitness in isolated populations is usually caused by inbreeding depression, rather than a lack of adaptive variants (Hedrick et al. 2013). Inbreeding depression is typically caused by alleles at many loci (Paige 2010). Thus, it would be necessary to replace many different alleles in order to alleviate low fitness caused by inbreeding depression. Second, perhaps an even more fundamental issue is that in most cases, the loci that contribute to inbreeding depression will not be known for the non-model species that are typically endangered. Third, alleviating inbreeding depression through gene editing may have the unintended consequence of adding genetic variation at a small subset of loci identified as causing inbreeding depression at the cost of losing genetic variation throughout the rest of the genome. This is because gene editing would likely only be performed on a small subset of individuals in the population, which could result in the loss of those alleles harbored by individuals not included in the gene-editing effort. Finally, despite advances in

(continued)

**Box 5** (continued)
gene editing due to new tools, such as the CRISPR-Cas system (Barrangou et al. 2007), there are many technological hurdles to jump before they are readily available for most species.

Fortunately, "traditional" genetic rescue (introduce individual immigrants into the target population) circumvents all of the above problems. Although there are potential risks of introducing immigrants into a target population, including disease and disrupting social interactions, increasing evidence indicates that traditional genetic rescue works. Although "gene tweaking" appeals to a technology-focused world, it is unlikely to advance conservation of imperiled populations suffering from inbreeding depression anytime in the near future.

## *5.2 Conclusions*

The decline and extirpation of small populations represents a substantial fraction of global biodiversity loss (Ceballos et al. 2017; Frankham et al. 2017; Willi et al. 2006). Despite overwhelming evidence for inbreeding depression and reduced adaptive potential in small and recently fragmented populations, these factors are rarely incorporated into conservation planning and management (Pierson et al. 2016). The good news is that restoring gene flow between recently isolated populations can reliably reverse the negative fitness consequences associated with small populations. Genomic tools will be invaluable for identification of target populations and to guide the design, implementation, and monitoring of genetic rescue. Ideally, information provided by these tools will help shift the genetic management of populations toward earlier evaluation and implementation when informed selection of candidate source and recipient populations remains possible, rather than waiting until the last minute when fewer options exist. Ultimately, when applied to genetic rescue, genomic tools will increase our fundamental understanding of how contemporary evolution shapes population dynamics as well as hopefully buy time for threatened populations that would otherwise be lost.

## Glossary

**Absolute fitness** Mean number of offspring per capita, measured as population growth rate or abundance.

**Assisted gene flow** Managed movement of individuals into populations to reduce local maladaptation to climate or other environmental change.

**Demographic rescue** An increase in population growth caused by the addition of new individuals.

**Evolutionary rescue** An increase in population growth resulting from adaptation to otherwise extinction-causing environmental stress from standing genetic variation, de novo mutation, or gene flow.

**Genetic load** The reduction in mean fitness of members of a population owing to deleterious genes, or gene combinations, in the population.

**Genetic rescue** An increase in population growth owing to immigration of new alleles beyond the demographic contribution of immigrants themselves.

**Genetic restoration** An increase in genetic variation and relative, but not absolute, fitness owing to immigration of new alleles.

**Identical by descent (IBD)** A matching segment of DNA shared by two or more individuals that has been inherited from a common ancestor without any intervening recombination.

**Inbreeding** Mating between related individuals which results in an increase of homozygosity in the progeny because they possess alleles that are identical by descent.

**Inbreeding depression** The relative reduction in fitness of progeny from matings between related individuals compared with progeny from unrelated individuals.

**Outbreeding depression** Reduced fitness of offspring from matings between genetically divergent individuals.

**Runs of homozygosity (ROH)** Stretches of homozygous genotypes at mapped SNPs.

## References

Aitken SN, Whitlock MC. Assisted gene flow to facilitate local adaptation to climate change. Annu Rev Ecol Evol Syst. 2013;44:367–88.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.

Allendorf FW, Knudsen KL, Blake GM. Frequencies of null alleles at enzyme loci in natural populations of ponderosa and red pine. Genetics. 1982;100:497–504.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. Trends Ecol Evol. 2001;16:613–22.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

Allendorf FW, Luikart G, Aitken SN. Conservation and the genetics of populations. 2nd ed. Oxford: Wiley-Blackwell; 2013.

Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 2002;160:1217–29.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–12.

Bell G, Gonzalez A. Evolutionary rescue can prevent extinction following environmental change. Ecol Lett. 2009;12:942–8.

Bell G, Gonzalez A. Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. Science. 2011;332:1327–30.
Benestan LM, Ferchaud AL, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25:2967–77.
Bouzat JL. Conservation genetics of population bottlenecks: the role of chance, selection, and history. Conserv Genet. 2010;11:463–78.
Bradburd GS, Ralph PL, Coop GM. A spatial framework for understanding population structure and admixture. PLoS Genet. 2016;12:e1005703.
Brown JH, Kodric-Brown A. Turnover rates in insular biogeography: effect of immigration on extinction. Ecology. 1977;58:445.
Browning SR, Browning BL. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. Am J Hum Genet. 2015;97:404–18.
Burton RS, Pereira RJ, Barreto FS. Cytonuclear genomic interactions and hybrid breakdown. Annu Rev Ecol Evol Syst. 2013;44:281–302.
Carlson SM, Cunningham CJ, Westley PAH. Evolutionary rescue in a changing world. Trends Ecol Evol. 2014;29:521–30.
Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362–5.
Ceballos G, Ehrlich PR, Dirzo R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. PNAS. 2017;114:E6089–96.
Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19:220–34.
Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC. A quantitative review of heterozygosity-fitness correlations in animal populations. Mol Ecol. 2009;18:2746–65.
Charlesworth B, Charlesworth D. The genetic basis of inbreeding depression. Genet Res. 1999;74:329–40.
Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10:783–96.
Chen N, Cosgrove EJ, Bowman R, Fitzpatrick JW, Clark AG. Genomic consequences of population decline in the endangered Florida scrub-jay. Curr Biol. 2016;26:2974–9.
Coltman DW, Slate J. Microsatellite measures of inbreeding: a meta-analysis. Evolution. 2003;57:971–83.
Corander J, Waldmann P, Sillanpää MJ. Bayesian analysis of genetic differentiation between populations. Genetics. 2003;163:367–74.
Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 1996;144:2001–14.
Coyne JA, Orr HA. Speciation. Sunderland: Sinauer Associates; 2004.
Crow JF, Kimura M. An introduction to population genetics theory. New York: Harper & Row; 1970.
Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 2005;6:361–75.
Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimatorv2: re-implementation of software for the estimation of contemporary effective population size (*Ne*) from genetic data. Mol Ecol Resour. 2013;14:209–14.
Domingos FMCB, Colli GR, Lemmon A, Lemmon EM, Beheregaray LB. In the shadows: phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: *Tropidurus*). Mol Phylogenet Evol. 2017;107:455–65.
Edmands S. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution. 1999;53:1757–68.
Edmands S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol. 2006;16:463–75.

Edmands S, Timmerman CC. Modeling factors affecting the severity of outbreeding depression. Conserv Biol. 2003;17:883–92.
Ekblom R, Smeds L, Ellegren H. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. BMC Genomics. 2014;15:467.
Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC. Locally adapted traits maintained in the face of high gene flow. Ecol Lett. 2015;18:37–47.
Fitzpatrick SW, Gerberich JC, Angeloni LM, Bailey LL, Broder ED, Torres Dowdall J, et al. Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. Evol Appl. 2016;9:879–91.
Fitzpatrick SW, Handelsman C, Torres-Dowdall J, Ruell E, Broder ED, Kronenberger JA, Reznick DN, Ghalambor CK, Angeloni LM, Funk WC. Gene flow constrains and facilitates genetically based divergence in quantitative traits. Copeia. 2017;105:462–74.
Foll M, Shim H, Jensen JD. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol Ecol Resour. 2015;15:87–98.
Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. Mol Ecol. 2018;27:2215–33.
Frankham R. Genetics and extinction. Biol Conserv. 2005;126:131–40.
Frankham R. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. Biol Conserv. 2016;195:33–6.
Frankham R, Ballou JD, Eldridge MDB, Lacy RC, Ralls K, Dudash MR, et al. Predicting the probability of outbreeding depression. Conserv Biol. 2011;25:465–75.
Frankham R, Ballou JD, Ralls K, Eldridge M, Dudash MR, Fenster CB, et al. Genetic management of fragmented animal and plant populations. Oxford: Oxford University Press; 2017.
Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.
Funk WC, Lovich RE, Hohenlohe PA, Hofman CA, Morrison SA, Sillett TS, et al. Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol Ecol. 2016;25:2176–94.
Funk WC, Forester BR, Converse SJ, Darst C, Morey S. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet. 2018;65:2481.
Garant D, Forde SE, Hendry AP. The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct Ecol. 2007;21:434–43.
Garcia-Dorado A. Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. Genetics. 2012;190:1461–76.
Gompert Z, Buerkle CA. Bayesian estimation of genomic clines. Mol Ecol. 2011;20:2111–27.
Gravel S. Population genetics models of local ancestry. Genetics. 2012;191:607–19.
Grueber CE, Sutton JT, Heber S, Briskie JV, Jamieson IG, Robertson BC. Reciprocal translocation of small numbers of inbred individuals rescues immunogenetic diversity. Mol Ecol. 2017;26:2660–73.
Gulisija D, Crow JF. Inferring purging from pedigree data. Evolution. 2007;61:1043–51.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5:e1000695.
Hale ML, Burg TM, Steeves TE. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS One. 2012;7:e45170.
Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016;30:33–41.
Hansson B, Westerberg L. On the correlation between heterozygosity and fitness in natural populations. Mol Ecol. 2002;11:2467–74.

Harris K, Nielsen R. Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet. 2013;9:e1003521.
Harrisson KA, Pavlova A, Gonçalves da Silva A, Rose R, Bull JK, Lancaster ML, et al. Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. Mol Ecol. 2016;25:1242–58.
Havird JC, Fitzpatrick SW, Kronenberger J, Funk WC, Angeloni LM, Sloan DB. Sex, mitochondria, and genetic rescue. Trends Ecol Evol. 2016;31:96–9.
Hedrick PW, Garcia-Dorado A. Understanding inbreeding depression, purging, and genetic rescue. Trends Ecol Evol. 2016;31:940–52.
Hedrick PW, Adams JR, Vucetich JA. Reevaluating and broadening the definition of genetic rescue. Conserv Biol. 2011;25:1069–70.
Hedrick PW, Allendorf FW, Waples RS. Genetic engineering in conservation. Nature. 2013;502:303.
Hedrick PW, Peterson RO, Vucetich LM, Adams JR, Vucetich JA. Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. Conserv Genet. 2014;15:1111–21.
Hill GE. Mitonuclear ecology. Mol Biol Evol. 2015;32:1917–27.
Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 2016;188:379–97.
Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, et al. High-throughput sequencing reveals inbreeding depression in a natural population. PNAS. 2014;111:3775–80.
Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 2008;39:21–42.
Hoffmann AA, Sgro CM, Kristensen TN. Revisiting adaptive potential, population size, and conservation. Trends Ecol Evol. 2017;32:506–17.
Hofman CA, Rick TC, Hawkins MTR, Funk WC, Ralls K, Boser CL, et al. Mitochondrial genomes suggest rapid evolution of dwarf California Channel Islands foxes (*Urocyon littoralis*). PLoS One. 2015;10:e0118240.
Hogg JT, Forbes SH, Steele BM, Luikart G. Genetic rescue of an insular population of large mammals. Proc R Soc B Biol Sci. 2006;273:1491–9.
Hohenlohe PA, Day MD, Amish SJ, Miller MR, Kamps-Hughes N, Boyer MC, et al. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Mol Ecol. 2013;22:3002–13.
Jeffries KM, Connon RE, Davis BE, Komoroske LM, Britton MT, Sommer T, et al. Effects of high temperatures on threatened estuarine fishes during periods of extreme drought. J Exp Biol. 2016;219:1705–16.
Johnson WE, Onorato DP, Roelke ME, Land ED, Cunningham M, Belden RC, et al. Genetic restoration of the Florida panther. Science. 2010;329:1641–5.
Jombart T, Ahmed I. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27:3070–1.
Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25:185–202.
Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. 2015;115:63–72.
Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9:1205–18.
Kardos M, Qvarnstrom A, Ellegren H. Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. Genetics. 2017;205:1319–34.
Keller LF, Waller DM. Inbreeding effects in wild populations. Trends Ecol Evol. 2002;17:230–41.

Keller MC, Visscher PM, Goddard ME, Rosenberg NA. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 2011;189:237–49.

Kemppainen P, Knight CG, Sarma DK, Hlaing T, Prakash A, Maung Maung YN, et al. Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. Mol Ecol Resour. 2015;15:1031–45.

Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 2010;5:e13996.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. Genomics and conservation genetics. Trends Ecol Evol. 2006;21:629–37.

Kovach RP, Hand BK, Hohenlohe PA, Cosart TF, Boyer MC, Neville HH, et al. Vive la résistance: genome-wide selection against introduced alleles in invasive hybrid zones. Proc R Soc B Biol Sci. 2016a;283:20161380.

Kovach RP, Luikart G, Lowe WH, Boyer MC. Risk and efficacy of human-enabled interspecific hybridization for climate-change adaptation: response to Hamilton and Miller (2016). Conserv Biol. 2016b;30:428–30.

Kristensen TN, Pedersen KS, Vermeulen CJ, Loeschcke V. Research on inbreeding in the "omic" era. Trends Ecol Evol. 2010;25:44–52.

Kronenberger JA, Funk WC, Smith JW, Fitzpatrick SW, Angeloni LM, Broder ED, et al. Testing the demographic effects of divergent immigrants on small populations of Trinidadian guppies. Anim Conserv. 2017a;20:3–11.

Kronenberger JA, Fitzpatrick SW, Angeloni LM, Broder ED, Ruell EW, Funk WC. Playing God with guppies – informing tough conservation decisions using a model experimental system. Anim Conserv. 2017b;20:18–9.

Kronenberger JA, Gerberich JC, Fitzpatrick SW, Broder ED, Angeloni LM, Funk WC. An experimental test of alternative population augmentation scenarios. Conserv Biol. 2018;32:838–48.

Leberg PL, Firmin BD. Role of inbreeding depression and purging in captive breeding and restoration programmes. Mol Ecol. 2008;17:334–43.

Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol. 2012;61:727–44.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Curr Biol. 2014;157:785–94.

Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of *F*ST outlier tests. Mol Ecol. 2014;23:2178–92.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2016.

Madsen T, Ujvari B, Olsson M. Novel genes continue to enhance population growth in adders (*Vipera berus*). Biol Conserv. 2004;120:145–7.

McCullough DR, Fischer JK, Ballou JD. From bottleneck to metapopulation: recovery of the tule elk in California. In: Metapopulations and wildlife conservation. Washington: Island Press; 1996.

Meek MH, Wells C, Tomalty KM, Ashander J, Cole EM, Gille DA, et al. Fear of failure in conservation: the problem and potential solutions to aid conservation of extremely small populations. Biol Conserv. 2015;184:209–17.

Meirmans PG. The trouble with isolation by distance. Mol Ecol. 2012;21:2839–46.

Miller JM, Hamilton JA. Interspecies hybridization in the conservation toolbox: response to Kovach et al. (2016). Conserv Biol. 2016;30:431–3.

Miller JM, Poissant J, Hogg JT, Coltman DW. Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). Mol Ecol. 2012;21:1583–96.

Miller JM, Malenfant RM, David P, Davis CS, Poissant J, Hogg JT, et al. Estimating genome-wide heterozygosity: effects of demographic history and marker type. Heredity. 2014;112:240–7.

Mills LS, Allendorf FW. The one-migrant-per-generation rule in conservation and management. Conserv Biol. 1996;10:1509–18.
Muhlfeld CC, Kalinowski ST, McMahon TE, Taper ML, Painter S, Leary RF, et al. Hybridization rapidly reduces fitness of a native trout in the wild. Biol Lett. 2009;5:328–31.
Newman D, Pilson D. Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella*. Evolution. 1997;51:354–62.
Nielsen R. Molecular signatures of natural selection. Annu Rev Genet. 2005;39:197–218.
Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. Conservation genetics in transition to conservation genomics. Trends Genet. 2010;26:177–87.
Paige KN. The functional genomics of inbreeding depression: a new approach to an old problem. Bioscience. 2010;60:267–77.
Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC, et al. Use of RAD sequencing for delimiting species. Heredity. 2014;114:450–9.
Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2:e190.
Pemberton J. Measuring inbreeding depression in the wild: the old ways are the best. Trends Ecol Evol. 2004;19:613–5.
Pemberton JM. Wild pedigrees: the way forward. Proc R Soc B Biol Sci. 2008;275:613–21.
Peñalba JV, Smith LL, Tonione MA, Sass C, Hykin SM, Skipwith PL, et al. Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. Mol Ecol Resour. 2014;14:1000–10.
Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8:e1002967.
Pierson JC, Coates DJ, Oostermeijer JGB, Beissinger SR, Bragg JG, Sunnucks P, et al. Genetic factors in threatened species recovery plans on three continents. Front Ecol Environ. 2016;14:433–40.
Pool JE, Hellmann I, Jensen JD, Nielsen R. Population genetic inference from genomic sequence variation. Genome Res. 2010;20:291–300.
Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24:4348–70.
Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001;16:351–8.
Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, et al. Genomic flatlining in the endangered Island fox. Curr Biol. 2016;26:1183–9.
Rohland N, Reich D, Mallick S, Meyer M, Green RE, Georgiadis NJ, et al. Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. PLoS Biol. 2010;8:e1000564.
Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I. Inbreeding and extinction in a butterfly metapopulation. Nature. 1998;392:491–4.
Schou MF, Bechsgaard J, Muñoz J, Kristensen TN. Genome-wide regulatory deterioration impedes adaptive responses to stress in inbred populations of *Drosophila melanogaster*. Evolution. 2018;72:1614–28.
Schwartz MK, Luikart G, Waples RS. Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol. 2007;22:25–33.
Slate J, David P, Dodds KG, Veenvliet BA, Glass BC. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Mol Ecol. 2004;93:255–65.
Spieth PT. Gene flow and genetic differentiation. Genetics. 1974;78:961–5.
Stoffel MA, Esser M, Kardos M, Humble E, Nichols H, David P, et al. inbreedR: an R package for the analysis of inbreeding based on genetic markers. Methods Ecol Evol. 2016;7:1331–9.

Storfer A. Gene flow and endangered species translocations: a topic revisited. Biol Conserv. 1999;87:173–80.
Stuglik MT, Radwan J, Babik W. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Mol Ecol Resour. 2011;11:739–42.
Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. PNAS. 2017;114:1607–12.
Szulkin M, Bierne N, David P. Heterozygosity-fitness correlations: a time for reappraisal. Evolution. 2010;64:1202–17.
Tallmon DA. Get a move on: the value of rescue. Anim Conserv. 2017;20:16–7.
Tallmon DA, Luikart G, Waples RS. The alluring simplicity and complex reality of genetic rescue. Trends Ecol Evol. 2004;19:489–96.
Thomas MA, Roemer GW, Donlan CJ, Dickson BG, Matocq M, Malaney J. Gene tweaking for conservation. Nature. 2013;501:485–6.
Weese DJ, Schwartz AK, Bentzen P, Hendry AP, Kinnison MT. Eco-evolutionary effects on population recovery following catastrophic disturbance. Evol Appl. 2011;4:354–66.
Westemeier RL. Tracking the long-term decline and recovery of an isolated population. Science. 1998;282:1695–8.
White M. Modes of speciation. San Francisco: W.H. Freeman; 1978.
Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA. Genetic rescue to the rescue. Trends Ecol Evol. 2015;30:42–9.
Willi Y, Van Buskirk J, Hoffmann AA. Limits to the adaptive potential of small populations. Annu Rev Ecol Evol Syst. 2006;37:433–58.
Wright S. Coefficients of inbreeding and relationship. Am Nat. 1922;56:330–8.
Wright S. Evolution in Mendelian populations. Genetics. 1931;16:0097–159.
Zhang C, Rannala B, Yang Z. Bayesian species delimitation can be robust to guide-tree inference errors. Syst Biol. 2014;63:993–1004.
Zhu B, Ashley-Koch AE, Dunson DB. Generalized admixture mapping for complex traits. G3: Genes Genomes Genetics. 2013;3:1165–75.
Zuk M, Rotenberry JT, Tinghitella RM. Silent night: adaptive disappearance of a sexual signal in a parasitized population of field crickets. Biol Lett. 2006;2:521–4.

# Ex Situ Wildlife Conservation in the Age of Population Genomics


**Michael A. Russello and Evelyn L. Jensen**

**Abstract** As the loss of biodiversity accelerates, there is general recognition that managing species outside of their native range (ex situ) will become increasingly important as populations continue to decline. Well-grounded in population genetic theory, ex situ conservation strategies, such as captive breeding, have largely relied on pedigree-based management out of both necessity and preference, despite known violations of important assumptions. Since the advent of molecular markers, many studies have successfully used empirical genetic data for informing ex situ conservation, yet their utility has been questioned due to competing priorities and resources as well as concerns related to potential biases associated with estimating individual- and population-level parameters based on traditional suites of loci. Paired with modern genotyping-by-sequencing approaches, population genomics holds great promise for overcoming past limitations associated with the use of empirical genetic data in ex situ conservation, allowing for highly precise estimates of population genetic parameters and identification of specific loci underlying traits of interest. Here, we review available literature and discuss the clear advantages and ultimate potential of using genome-wide data when managing species outside of their native range, from refining breeding decisions and assessing lineage integrity to minimizing adaptation to the captive environment and informing interactive in situ/ex situ conservation strategies. With resource-driven and capacity-related barriers to adoption falling away, our ability to harness leading-edge technologies to mine the genomes of wildlife species will enable more effective and efficient planning, implementation and monitoring of ex situ conservation strategies moving forward.

**Keywords** Ex situ population management · Hybridization · Introgression · Reintroduction · SNPs · Species restoration

M. A. Russello (✉) · E. L. Jensen
Department of Biology, The University of British Columbia, Kelowna, BC, Canada
e-mail: michael.russello@ubc.ca

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2018_44,

## 1 Introduction

As the magnitude and rate of biodiversity loss were first being recognized within and outside of the scientific community (Myers 1979), preserving and managing species outside of their native range (ex situ) was proposed as one strategy for avoiding species extinction. The so-called ark paradigm soon followed, recommending that long-term captive breeding programs be established to maintain species ex situ, with the goal of eventual reintroduction to the wild once the danger of extinction had passed (Foose et al. 1992; Soulé and Simberloff 1986; Tudge 1992). An official International Union for Conservation of Nature (IUCN) policy statement on captive breeding validated the approach, highlighting that such "...programmes need to be established before species are reduced to critically low numbers (IUCN 1987)." The IUCN's Conservation Breeding Specialist Group went as far as to recommend captive breeding for 1,192 of 3,550 vertebrate taxa considered at the time to be at risk (Seal et al. 1994). Captive breeding as an ex situ conservation strategy also featured prominently in recovery plans for wildlife species listed under the US Endangered Species Act (Tear et al. 1993). Yet, there was a growing concern that too much emphasis was being placed on ex situ conservation strategies, given the substantial resource investment (Conway 1986) and relatively limited applicability across the vast number of species at risk of extinction (Balmford et al. 1995; Rahbek 1993; Snyder et al. 1996). However, with the biodiversity crisis accelerating (Pimm et al. 2014), there is a general recognition that ex situ conservation strategies will become increasingly important as populations continue to decline but they will be most effective within the context of an integrated species recovery plan (Bowkett 2009; Russello and Amato 2007).

Despite shifting attitudes towards the role ex situ strategies should take in conservation, captive breeding, in particular, has grown in prominence and, along with it, so has the recognition of the importance of integrating population genetic theory into the management of captive populations. Initial schemes for animal breeding were based on the "maximum avoidance of inbreeding" proposed by Wright (1921). Translated into management, this is simply to mate the least-related individuals in each succeeding generation; under an ideal breeding scheme with a sufficient number of founders ($n = 50–100$), this strategy would result in preserving half of the genetic diversity of the species over 100 generations (Flesness 1977). Over time, a basic strategy for genetic management of captive populations emerged that highlighted the importance of acquiring an adequate number of founders, expanding the population size to a predetermined carrying capacity and subdividing the populations in a way that also allows regulated exchange of genetic material across institutions (Ralls and Ballou 1983). Within each subpopulation and across the entire captive population, management strategies were designed to maximize effective population size, equalize founder representation and minimize inbreeding (Ralls and Ballou 1983). It was later shown by Ballou and Lacy (1995) that a strategy for minimizing mean kinship, the average of the kinship coefficients between an individual and all living individuals (including itself), was the most

effective for maintaining gene diversity and minimizing inbreeding, a paradigm that is still recommended and followed today (Willoughby et al. 2017).

Well-grounded in population genetic theory, the first two decades of captive breeding largely relied on pedigree-based management out of both necessity and preference, despite the known violations of important assumptions. One of the key assumptions of pedigree-based management is that all founders of a captive population are considered equally unrelated and non-inbred ("founder assumption"; Ballou 1983). This assumption is often violated, potentially leading to erroneous estimates of mean kinship and inbreeding coefficients upon which management decisions are based (Russello and Amato 2004). The founder assumption was necessary, however, as the ability to empirically estimate relatedness using molecular markers such as microsatellites and single nucleotide polymorphisms (SNPs) has only emerged in the last 20 years. Since the advent of molecular markers, many studies have successfully used empirical genetic data for addressing a wide range of questions of relevance to ex situ conservation (Table 1), including (1) monitoring genetic diversity and structure in ex situ populations (e.g. Henry et al. 2009; Hoeck et al. 2015; Ishtiaq et al. 2015; Kyle et al. 2016; Price et al. 2015; Urano et al. 2013); (2) quantifying cryptic lineage diversity, hybridization and introgression within captivity (e.g. Milián-García et al. 2015a, b; Ray et al. 2013; Russello et al. 2007, 2010a, b); (3) identifying genetically important individuals for captive breeding (e.g. Gonçalves da Silva et al. 2010; Russello and Amato 2004); and (4) evaluating/informing reintroduction (e.g. Strzala et al. 2015) and head-starting initiatives (e.g. Jensen et al. 2015).

Despite the proliferation in use of molecular markers and the ever-decreasing costs of data collection, the utility of empirical genetic data for ex situ conservation has been questioned. These have largely been "straw man" arguments related more to competing priorities and resources (e.g. diverting funds from in situ conservation) than to a critical examination of the past and potential utility of genetic data (Russello and Amato 2007). In other cases, however, important concerns were raised regarding potential biases introduced due to effect sizes and high sampling variances associated with estimating individual- and population-level parameters based on a limited set of hypervariable loci, such as microsatellites (Hoffman et al. 2014; Ritland 1996; Van de Casteele et al. 2001).

Modern genotyping-by-sequencing approaches, such as those that rely on reduced representation (e.g. restriction-site associated DNA (RAD) tags; Andrews et al. 2016; Baird et al. 2008; Etter et al. 2011) or whole-genome sequencing (Therkildsen and Palumbi 2017), are increasingly being employed to address questions of ecological, evolutionary and conservation relevance. Population genomics holds great promise for overcoming past limitations associated with incorporating empirical genetic data within ex situ conservation. The use of genome-wide data allows for highly precise estimates of population genetic parameters and, importantly, permits the identification of specific loci underlying traits of interest (Luikart et al. 2003). Yet, to date, few studies have employed population genomic approaches for addressing questions in ex situ conservation. Here, we identify four broad areas where population genomic data can substantially contribute to better informing ex situ conservation including

**Table 1** Examples of studies using population genetics/genomics to inform ex situ conservation

| Application | Species | Common name | Type of markers | Recommendation | Reference |
|---|---|---|---|---|---|
| Refining breeding decisions | *Margaritifera margaritifera* | Freshwater pearl mussel | Microsatellites ($n = 8$) | Captive-bred individuals are highly inbred; incorporating new founders from the wild will help alleviate inbreeding | Kyle et al. (2016) |
| | *Gyps spp.* | Asian vultures | Microsatellites ($n = 13$) | Founders are largely unrelated, and genotypic data will be used to minimize the loss of genetic diversity and inbreeding in the captive population | Ishtiaq et al. (2015) |
| | *Calotriton arnoldi* | Montseny brook newt | Microsatellites ($n = 24$) | Maintain two captive breeding lines and incorporate new founders | Valbuena-Urena et al. (2017) |
| Assessing lineage integrity | *Gopherus agassizii* | Agassiz's desert tortoise | Microsatellites ($n = 25$), mtDNA | Captive individuals do not originate from nearby wild populations and should not be considered useful for management | Edwards and Berry (2013) |
| | *Felis silvestris silvestris* | European wildcat | Microsatellites ($n = 10$), mtDNA | Captive individuals are highly admixed with domestic cats and are not recommended for further breeding or reintroduction | Witzenberger and Hochkirch (2014) |
| | *Microtus californicus scirpensis* | Amargosa vole | SNPs ($n = 20{,}000$) | Only one of two wild subgroups are currently represented in captivity; need to balance risk of outbreeding depression with the possibility of a catastrophic event in the wild | Krohn et al. (2018) |
| Minimizing adaptation to captivity | *Aburria jacutinga* | Black-fronted piping guan | Microsatellites ($n = 9$) | Maintain separate captive lineage and crossbreed them in the pre-release generation to minimize adaptation to captivity and improve reintroduction success | Oliveira et al. (2016) |
| | *Oncorhynchus tshawytscha* | Chinook salmon | SNPs ($n = 9{,}410$) | The offspring of hatchery broodstock sourced from the captive lineage had evidence of domestication selection, whereas offspring of wild broodstock did not | Waters et al. (2015) |
| | *Peromyscus leucopus* | White-footed mice | SNPs ($n = 5{,}500$) | Minimizing inbreeding and selection in captive populations is critical as captive-estimated fitness correlations likely underestimate the effects in wild populations | Willoughby et al. (2017) |
| Informing interactive in situ/ex situ conservation | *Petrogale lateralis* | Warru | Microsatellites ($n = 15$), mtDNA | Manage the reintroduction of captive individuals as one genetic group; re-establish gene flow between captive populations | West et al. (2018) |
| | *Tragelaphus eurycerus isaaci* | Mountain Bongo | SNPs ($n = 1{,}052$) | Captive individuals should be used to restock declining wild populations to increase population sizes and genetic diversity | Svengren et al. (2017) |
| | *Capra ibex* | European Ibex | SNPs ($n = 101{,}822$) | Introduce individuals from genetically differentiated populations to increase genetic variation and reduce inbreeding | Grossen et al. (2018) |

*SNPs* single-nucleotide polymorphisms

(1) refining captive breeding decisions, (2) assessing lineage integrity and quantifying the extent of hybridization/introgression, (3) minimizing adaptation to the captive environment and (4) informing interactive in situ/ex situ conservation strategies.

## 2 Refining Breeding Decisions

Small, closed populations, such as those found in captive breeding programs, will lose genetic diversity due to random genetic drift (Allendorf 1986; Lacy 1987), with potentially deleterious fitness consequences (Ralls and Ballou 1983). In the captive environment, the effects of drift can be attenuated by maximizing the genetically effective population size through careful decisions of which individuals should be bred together, historically informed by pedigree analysis (Lacy 1995; MacCluer et al. 1986; Thompson 1986). Both retrospective and prospective pedigree analyses typically invoke the founder assumption, where each founder is considered equally unrelated and non-inbred to all non-descendent members in the pedigree (Ballou 1983). Individuals in subsequent generations have their ancestry mapped back to the founders, and the pedigree is used to determine an individual's relationship to potential mates and the captive population as a whole. These relatedness values can be used to design breeding pairs, plan transfers between facilities, and determine the ideal number of offspring from a given pair. However, errors in pedigrees can arise in a variety ways, from mistakes in data entry and misassignment of parentage to violations of the founder assumption. Such errors can result in the unintentional breeding of close relatives, which can lead to deleterious fitness consequences due to inbreeding depression (Ralls and Ballou 1983).

Genetic markers have been used to help identify errors in pedigrees by determining parentage and estimating relatedness. In one example, microsatellite genotypic data were used to identify errors in parental assignment that had compounded throughout the pedigree within a captive population of the critically endangered Attwater's prairie-chicken (*Tympanuchus cupido attwateri*; Hammerly et al. 2016). In this instance, >25% of breeding pairs chosen based on the pedigree had a mean relatedness of >0.125. After reassigning breeding pairs based on genetically determined relatedness values, the mean relatedness of parents significantly decreased, and the proportion of chicks surviving to 5 weeks post-hatch significantly increased (Hammerly et al. 2016). Scores of other studies have similarly used microsatellite genotypic data to inform breeding decisions either to augment pedigree-based management (e.g. Gonçalves da Silva et al. 2010; Henkel et al. 2012; Jones et al. 2002) or to provide information on parentage and relatedness in the absence of pedigree information (e.g. Mitchell et al. 2011; Russello and Amato 2004). Yet, there are limitations associated with using a small number of molecular markers for informing captive breeding decisions (Ivy et al. 2009). In particular, molecular estimates of relationship (kinship or relatedness) derived from a limited number of traditional markers (e.g. microsatellites) are known to have high sampling variances (Blouin et al. 1996; Csilléry et al. 2006), an effect that can be exacerbated in closed

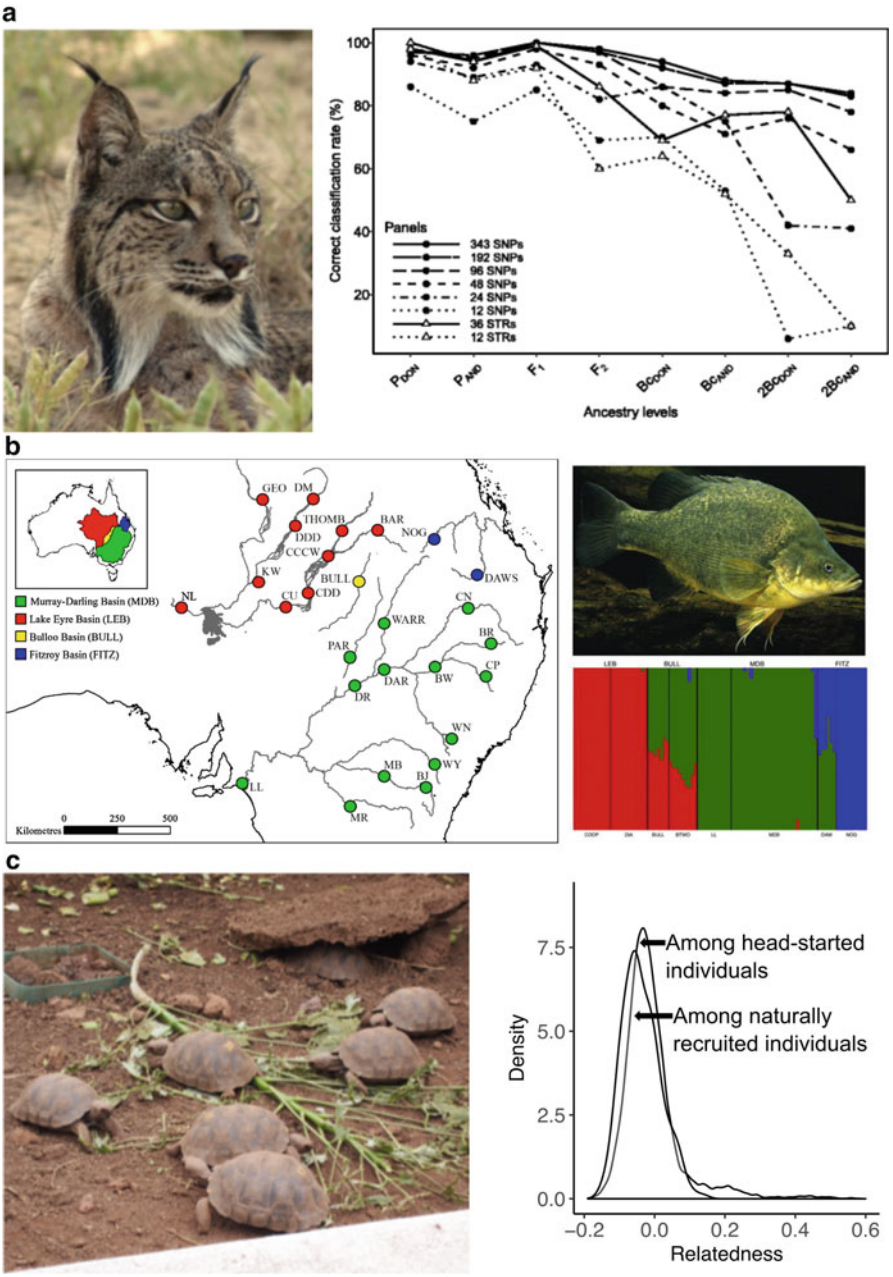


**Fig. 1** Examples of studies where genome-wide data improved upon insights from traditional molecular markers for informing in situ/ex situ population management. (**a**) Kleinman-Ruiz et al. (2017) demonstrated the enhanced power of SNPs for classifying unknown individuals into ancestry classes; (**b**) Beheregaray et al. (2017) used 18,979 SNPs to resolve taxonomic uncertainty that has persisted despite previous research using allozymes, microsatellites and mitochondrial

(De Bois et al. 1990) or genetically depauperate populations (Tokarska et al. 2009), as are often found in captivity.

Genome-wide SNP data hold great promise for refining estimates of pairwise kinship or relatedness to improve pedigrees and inform captive breeding decisions based on a criterion of minimizing mean kinship. Evaluating the use of SNP genotypic data for calculating molecular coancestry, Ivy et al. (2016) reported exceptionally low variances of $<0.0003$ when sampling $\geq 250$ loci derived from a SNP chip and $<0.0025$ when sampling $\geq 500$ loci genotyped via double digest RAD sequencing (Peterson et al. 2012). Likewise, in the endangered Iberian lynx (*Lynx pardinus*), targeted SNP panels were demonstrated to outperform microsatellites for relatedness estimation, individual identification and parentage assignment (Fig. 1a), all of which will inform ongoing captive breeding, translocation and reintroduction plans (Kleinman-Ruiz et al. 2017). In addition to the insights provided by panels of molecular markers, model development has also accelerated for inferring precise relatedness assignments from low-coverage whole-genome data (Martin et al. 2017). Yet another area of innovation is in exploring alternative conceptual definitions of relatedness. To date, most measures of relatedness have been based on probabilities of identity by descent, but these can be imprecise measures of genome sharing. For example, full siblings can vary in the actual amount of their genome shared due to the random nature of segregation and recombination during meiosis, which has been shown to range from 37 to 63% in humans (Speed and Balding 2015). Moving forward, the use of genome-wide SNP data will allow for more accurate estimates of genome sharing and the exploration of alternative approaches for inferring relatedness (e.g. genome-wide distributions of time since most recent common ancestor; Speed and Balding 2015). Accurate and precise estimates of relatedness that take into account the variance in inheritance can enable more specifically tailored breeding pairs to be matched and may be of particular utility when only a very small number of breeders are available, which is typically the case for critically endangered species.

In addition to using genome-wide neutral loci for improving pedigrees and minimizing mean kinship within captive populations, population genomics can be used to identify specific loci related to local adaptation or inbreeding depression that can inform management decisions (Allendorf et al. 2010). Genomic selection for specific traits has been successfully employed in livestock, particularly in the dairy

**Fig. 1** (continued) DNA. Bayesian clustering analyses and phylogenetic reconstruction revealed three independently evolving lineages associated with climate-determined boundaries of major river basins. (**c**) Jensen et al. (2018) paired RADseq and targeted capture to genotype 2,218 SNPs in population-level samples of the Pinzón Galapagos giant tortoise before (c.1906) and after (c.2014) a known bottleneck to investigate the genetic consequences of rapid population decline and to evaluate the performance of a head-start program implemented to support demographic recovery. The distributions of pairwise relatedness for naturally recruited and head-started individuals were largely overlapping, providing one line of evidence suggesting that genetic diversity has been captured consistently by the head-start program over a ~50 year period

industry, for improving milk production and other traits of interest (Hayes et al. 2013). Captive breeding for conservation purposes may use such information to select against particular genetic variants, as has been the case with the California condor (*Gymnogyps californianus*). Declared extinct in the wild in 1982, all living individuals descend from a captive population originally founded by 14 individuals. It was subsequently determined that the small founder population contained a relatively high frequency of a recessive lethal allele causing chondrodystrophy, a genetically based skeletal disorder (Ralls et al. 2000). Ongoing work has sequenced 36 complete genomes representing the entire extant gene pool and localizing the markers within a 1 Mb region to detect carrier condors heterozygous for the lethal mutation (Ryder et al. 2016). This information is now being directly incorporated into population management to minimize the frequency of the lethal allele while still maximizing genetic diversity elsewhere in the genome (Ryder et al. 2016). Ultimately, the appropriateness of genomic selection for conservation purposes will have to be determined on a case-by-case basis, as single-trait strategies could have unintended consequences, potentially resulting in loss of important adaptive variation at other genes and increased homozygosity across the genome (Hedrick 2001).

## 3 Assessing Lineage Integrity and Quantifying the Extent of Hybridization/Introgression

In some instances, ex situ populations are founded or augmented with opportunistically sourced individuals of uncertain origin and ancestry. This situation can be especially problematic for maintaining pure captive populations of conservation value when there is cryptic diversity or taxonomic uncertainty related to the target species and/or when there is a history of hybridization and introgression with a nontarget species. Conventional suites of molecular markers, such as mitochondrial DNA and microsatellites, have been successful in many cases in evaluating lineage integrity and detecting signals of admixture within ex situ populations. One example involves an ex situ population of a threatened subspecies of the Cuban parrot (*Amazona leucocephala palmarum*) that was initially founded by individuals confiscated from the illegal pet trade. However, a genetic study revealed that this captive population is not genetically distinct from one held at a separate facility believed to be restricted to another threatened subspecies of Cuban parrot (*A. l. leucocephala*), with some individuals exhibiting mtDNA haplotypes from yet a third formally described subspecies (*A. l. hesterna*; Milián-García et al. 2015a, b; Russello et al. 2010a, b). Taken together, these results suggest that current ex situ management strategies should be re-evaluated (Milián-García et al. 2015a, b). Similarly, a genetic assessment of ex situ populations of the European wildcat (*Felis silvestris silvestris*) revealed widespread introgression with domestic cats (*F. s. catus*), leading to a recommendation that the current captive populations should not be retained for conservation purposes (Witzenberger and Hochkirch 2014).

Population genomics provides opportunities for not just detecting evidence of cryptic diversity but also for resolving taxonomic uncertainty to inform ex situ management strategies. A study of golden perch (*Macquaria ambigua*) used 18,979 SNPs to resolve taxonomic uncertainty that has persisted despite previous research using allozymes, microsatellites and mitochondrial DNA (Beheregaray et al. 2017). In this case, three independently evolving lineages associated with climate-determined boundaries of major river basins were identified (Fig. 1b), leading the authors to recommend that each represents a cryptic species of golden perch (Beheregaray et al. 2017). From a management perspective, ex situ hatchery-based stocking initiatives should be conducted exclusively within each river basin using locally sourced fish to maintain lineage integrity.

Similarly, population genomics can be used to characterize the degree of admixture between target and nontarget species, even for more advanced introgression classes than were previously detectable with conventional suites of molecular markers (Abbott et al. 2016). From an ex situ management perspective, this will facilitate identification of pure-bred or minimally admixed individuals that can be prioritized for breeding. One notable example is provided by genomic research on the endangered Przewalski's horse (*Equus ferus przewalskii*), a species declared extinct in the wild in the 1960s. The species was saved from the brink of extinction by an intensive captive breeding and reintroduction program that has increased the current population to 2,109 individuals, all descended from 12 wild-caught Przewalski's horses and up to 4 domestic horses. A study by Der Sarkissian et al. (2015) sequenced the complete genomes of 11 Przewalski's horses, representing all founding lineages and 5 museum specimens collected prior to captive population establishment. The authors uncovered evidence for historical introgression occurring since the split of Przewalski's and domestic horses ~45,000 years ago. Within the modern-day Przewalski's horse gene pool, variable levels of introgression of domestic alleles were found, ranging from undetectable to >31% (Der Sarkissian et al. 2015). Moving forward, resulting markers from this study can be used to minimize domestic ancestry while maximizing current genetic diversity within the captive breeding program. This application of population genomics clearly has relevance for other programs, including the European wildcats discussed above, as pure lineages do still exist in the wild (Witzenberger and Hochkirch 2014). Genome-wide markers have been developed in this species (Oliveira et al. 2015) that can be used to more reliably determine the hybrid status of individuals and to identify potential founders of a genetically informed European wildcat captive breeding program. In more extreme cases where individuals are identified that carry ancestry from recently extinct species, genetic data are essential for identifying individuals that maximize the ancestry of interest while minimizing relatedness and inbreeding of resulting offspring, as demonstrated within the context of the Floreana Galapagos giant tortoise species restoration program (Miller et al. 2017).

## 4 Minimizing Adaptation to Captive Environment

A central purpose of captive breeding is to produce individuals for release to the wild or to house a population as insurance against extinction. In either case, it is important that captive individuals retain the ability to survive in the wild. However, the success rate of release programs can be startlingly low (Beck et al. 1994; Fischer and Lindenmayer 2000), which may partially be associated with genetic changes in the ex situ population due to adaptation to the captive environment. In benign captive conditions (e.g. no predators, few parasites, low environmental stress), selection for traits essential for survival in the wild may be relaxed, while at the same time there may also be positive selection for traits that allow individuals to thrive in captivity, such as tameness (Darwin 1868; Frankham 2008). Such traits, while maladaptive in the wild, may result in some individuals having enhanced breeding success in captivity, increasing their representation in subsequent captive generations.

Genetic adaptation to captivity has been documented across a broad array of taxa including mammals, fish, amphibians, insects, plants and bacteria (e.g. Ensslin et al. 2015; Frankham and Loebel 1992; Heath et al. 2003; Kraaijeveld-Smit et al. 2006; Lewis and Thomas 2001). This issue is considered to be one of the major concerns facing the genetic management of ex situ conservation programs (Frankham 2010), some of which now specifically try to prevent it from occurring (Oliveira et al. 2015). Recommendations to prevent adaptation to captivity include minimizing the number of generations spent in captivity before release, mitigating the strength of selection through managed breeding based on a criterion of minimizing mean kinship and maintaining replicate, fragmented captive populations that are crossed only when necessary to alleviate inbreeding (Frankham 2008; Willoughby et al. 2017). The constraints of ex situ conservation programs, particularly for highly endangered species, make achieving these recommendations challenging, but genomic tools can greatly assist these goals.

Depending on the nature of the threat to the wild population, it may not be possible to reduce the actual number of generations spent in captivity before ex situ populations can be reintroduced to the wild. However, there are a number of alternative means that can achieve the same ends. There has been increasing interest in cryoconservation in the form of "frozen zoos" (Benirschke 1984), which contain preserved, hopefully viable, germplasm collected from the founders or earlier generations of captive populations. In such cases, germplasm can be thawed and used to reintroduce the genetic diversity of those ancestral individuals into the current population, thereby helping to alleviate negative genetic changes that have occurred since the germplasm was archived (Lermen et al. 2009). A notable case where this technology has been used is in black-footed ferrets (*Mustela nigripes*), where spermatozoa frozen for 20 years were successfully used to produce new offspring from long dead founders of the population (Howard et al. 2016). For the black-footed ferrets, this intervention was undertaken to increase genetic diversity, not combat adaptation to captivity, but the case provides an important proof of concept in a conservation context. The use of archived germplasm will be most

effective when source individuals are specifically targeted to contribute needed components of genetic diversity. Recent advancements in single cell genomics (Linnarsson and Teichmann 2016) may allow genetic testing of archived gametes to match into breeding pairs. For example, such assisted reproductive technologies are being explored to attempt to reverse the seemly inevitable trajectory to extinction for the northern white rhinoceros (*Ceratotherium simum cottoni*; Saragusty et al. 2016), an endeavour made more urgent with the recent death of the last surviving male in March 2018 (Tunstall et al. 2018).

Another way to achieve a reduction in the impact of the number of generations spent in captivity is to introduce new founders sourced from wild populations every few generations. Empirical studies of Chinook salmon (*Oncorhynchus tshawytscha*; Waters et al. 2015) and Leon Springs pupfish (*Cyprinodon bovinus*; Black et al. 2017) both highlight the importance of integrating gene flow from the wild population into hatchery brood stock to help mitigate genetic and morphological divergence between the in situ and ex situ populations. There are, however, situations where introducing additional individuals from the wild into the captive population would not be appropriate. The fear of disease exposure, legal barriers to transport or causing further harm to the wild population by removing additional individuals are just some of the reasons why captive populations may be genetically isolated. One such example is the Tasmanian devil (*Sarcophilus harrisii*) insurance colony, which was founded using disease-free individuals in 2006 as an epidemic of a highly lethal and contagious facial tumour disease was causing drastic population declines (CBSG 2008).

Despite the promise of these approaches for decreasing the impact of the number of generations spent in captivity, there is increasing evidence, particularly in fish, that negative consequences of captivity can occur extremely rapidly (Araki et al. 2007). In steelhead trout (*Oncorhynchus mykiss*), adaptation to captivity, resulting in greater fitness in captivity and decreased fitness in the wild, has been documented after only a single generation (Christie et al. 2012). A follow-up transcriptomic study showed significant changes in the levels of expression at more than 700 genes after a single generation in captivity (Christie et al. 2016). Simply being reared in captivity for a period of time has also been shown to have negative impacts on fitness for Atlantic salmon (*Salmo salar*). An experimental study released fry at two developmental stages, one having spent an additional 5 months in captivity (Clarke et al. 2016). The offspring of individuals that spent longer in captivity had lower viability, suggesting that the effects of captivity are both increased by spending an extended period ex situ and transgenerational in Atlantic salmon (Clarke et al. 2016). In such cases, it is possible that spending early developmental stages in captivity results in epigenetic changes. The emerging field of epigenomics may be able to shed light on this phenomenon (Suzuki and Bird 2008), which may have a significant impact on the fitness of captive-reared individuals and their offspring (Fraser 2008).

In some situations, it may be possible to identify which genes are underlying the traits resulting in adaptation to captivity and maladaptation to the wild. By scanning for regions of the genome under selection in wild and captive populations, targeted breeding strategies could be devised to produce individuals with a higher chance of

survival upon reintroduction (Allendorf et al. 2010). Determining the genetic basis of traits that reduce fitness of captive individuals once reintroduced to the wild is a critical first step, and experimental studies are beginning to emerge. For example, after three generations in captivity, multiple genes were found to be outliers between captive and wild populations of Chinook salmon (*Oncorhynchus tshawytscha*; Waters et al. 2015), with the annotations of these genes suggesting that adaptation to the crowded conditions and feeding regimes in captivity as well as selection for early maturation may have been occurring. At present, however, it remains challenging to identify causal variants underlying complex traits, even in cases where quantitative trait loci (QTL) have been previously identified. Following up on a whole-genome linkage mapping study that identified several QTL for five morphological traits in an ex situ population of zebra finches (*Taeniopygia guttata*), Knief et al. (2017) conducted an association study to identify causal variants in both wild and captive populations. Several SNPs showed significant associations with the same trait across populations, but the direction and magnitude of the effects differed, possibly due to founder effects (Knief et al. 2017). Consequently, none of the identified SNPs were likely causative, yet captive populations, within which linkage disequilibrium (LD) is typically high, offer great potential moving forward as focal study systems to identify the genetic basis of traits of interest (Knief et al. 2017). In cases where causative SNPs are characterized, emerging genome-editing technologies, such as CRISPR-Cas9, may provide opportunities for integrating novel genetic variation without having to impact the in situ population or exposing the ex situ population to potential pathogens. There are clearly practical, ethical and legal considerations that must be thought through and addressed prior to adopting genome-editing technologies in conservation (Johnson et al. 2016), but ex situ populations may provide optimal targets for initial application.

## 5 Informing Interactive In Situ/Ex Situ Conservation Strategies

One critique of ex situ conservation is that it is costly and draws resources away from in situ conservation, thus often neglecting to address the cause of the species' decline (Derrickson and Snyder 1992; Frazer 1992; Meffe 1992; Philippart 1995). However, in situ and ex situ conservation strategies do not have to be mutually exclusive, and effective conservation can be achieved through interactive in situ/ex situ management.

A common interactive in situ/ex situ program is "restocking" by releasing captive-bred or raised individuals into the occupied range of the target species. Restocking is done to bolster population sizes and improve viability of endangered species (e.g. captive breeding with recurring release or head-starting) or to increase opportunities for harvesting of exploited species (e.g. stocking fish and waterfowl for recreational fishing and hunting). In either case, the large-scale release of captive-reared individuals can have demographic and genetic consequences for local wild populations, some of

which are only just beginning to be appreciated (Champagnon et al. 2012). One issue is that restocked individuals may be released to different areas from where their ancestors originated, resulting in artificial admixture among conspecific populations. These introductions may impact the genetic integrity of the recipient population and possibly fitness if outbreeding depression occurs (Frankham et al. 2011). These and other unintended consequences may be avoidable if careful genomic assessment for population differentiation and local adaptation is carried out prior to restocking.

Other concerns regarding captive breeding, outlined in the previous sections of this chapter, also apply to restocking. Head-starting can be an advantageous alternative to captive breeding that minimizes some of the negative consequences of ex situ conservation. Head-start programs involve collecting eggs or very young individuals and housing them in captivity until they have outgrown their most vulnerable life stage at which point they are repatriated within their native range. Since only young are housed in captivity, many more individuals can be included in the program than traditional ex situ captive breeding of adults. Moreover, individuals typically only spend a small portion of their life ex situ, decreasing the risk of adaptation to captivity as well as exposure to diseases. Similar to captive breeding and other types of restocking, it is critical that the number of "founders" (i.e. parents) of head-start generations is maximized so that the extent of genetic diversity in the wild population is adequately represented (Jensen et al. 2015, 2018).

Population genomics can play a critical role in all phases of an interactive in situ/ex situ strategy, including assessment of the genetic makeup of populations before, during and after captive breeding/head-starting and reintroduction to monitor outcomes and maximize success (Attard et al. 2016). As an example, a pre-captive breeding analysis of the wild population of green abalone (*Haliotis fulgens*) was used to inform the strategy for subsequent hatchery rearing and release (Gruenthal et al. 2014). By using population genomics to demonstrate panmixia across the proposed area for restocking, Gruenthal et al. (2014) determined that raising a single pool of individuals for release was appropriate. In this case, individuals for breeding in the proposed hatchery could be sourced from larger populations, with supplementation occurring at more depleted locations. In another example, Pinzón Island Galapagos giant tortoises (*Chelonoidis duncanensis*) have been the subject of a head-start program for more than 50 years to support demographic recovery following a population bottleneck; however, the degree to which genetic variation has been impacted by this long-term conservation intervention is unknown. To investigate this question, Jensen et al. (2018) genotyped 2,218 SNPs for population-level samples of *C. duncanensis* before (c.1906) and after (c.2014) the bottleneck. Reconstructed distributions of pairwise relatedness were largely overlapping for naturally recruited and head-started individuals. In tandem with consistent measures of within-population variation and lack of structure among population samples, these results suggest that genetic diversity has been captured consistently by the head-start program over time (Fig. 1c; Jensen et al. 2018).

Yet, even if diverse and representative cohorts of captive-bred or head-start individuals are released, the ultimate fitness consequences are unknown without sufficient monitoring, as demonstrated by a study of the endangered Burmese roofed

turtle (*Batagur trivittata*). In this instance, genotypic data collected at ~1,500 SNPs from ~40% of the remaining global population identified two groups of 30 individuals from the captive pool that were prioritized for reintroduction (Çilingir et al. 2017). Post-release monitoring identified 27 fertile eggs from the known wild nesting beaches attributed to the contribution of released males to the active breeding output, which had previously shrunk to zero (Çilingir et al. 2017). Moving forward, the application of population genomics within such a holistic framework for interactive in situ/ex situ management (Attard et al. 2016) will help ensure scientifically informed conservation action through all phases, from planning and implementation to post-release monitoring and impact assessment.

## 6 Conclusions

Just as ex situ management strategies are most effective as a component of an integrated species recovery plan, genetic information constitutes only one of several considerations in prioritizing, planning and implementing successful ex situ conservation programs. The days of using an economic argument to dismiss the value of genetic data for informing ex situ conservation are over; genomic approaches have greater statistical power than their predecessors and cost less for genotyping (Campbell et al. 2015; Davey et al. 2011; Narum et al. 2013). The bioinformatics associated with processing and analysing high-throughput DNA sequencing reads can be daunting for the uninitiated, but that too is not a barrier to entry; full-service facilities now exist where non-geneticists can send their tissue samples and receive genomic data, complete with summary statistics and results of more advanced analyses if desired. We cannot rely on technology to solve the biodiversity crisis, but we can harness it to mine the genomes of at-risk populations to more effectively and efficiently plan, implement and monitor conservation strategies, in situ and ex situ. Here, we have demonstrated the clear advantages of using genome-wide data when managing species outside of their native range, from refining breeding decisions and assessing lineage integrity to the potential for minimizing adaptation to the captive environment and informing interactive in situ/ex situ conservation strategies. Ultimately these applications will only be successful if the original causes of decline are ameliorated and the global community musters the will to take action to value and protect biodiversity.

**Acknowledgements** We thank G. Amato and P. Hohenlohe for helpful comments on the manuscript. M.A.R acknowledges the support of the National Science and Engineering Research Council of Canada Discovery program (grant # 2014-04736). E.L.J. was supported by an NSERC Postgraduate Scholarship.

## References


Abbott RJ, Barton NH, Good JM. Genomics of hybridization and its evolutionary consequences. Mol Ecol. 2016;25:2325–32.

Allendorf FW. Genetic drift and the loss of alleles versus heterozygosity. Zoo Biol. 1986;5:181–90.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81.

Araki H, Cooper B, Blouin MS. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science. 2007;318:100–3.

Attard C, Moller LM, Sasaki M, Hammer MP, Bice CM, Brauer CJ, et al. A novel holistic framework for genetic-based captive-breeding and reintroduction programs. Conserv Biol. 2016;30:1060–9.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3:e3376.

Ballou JD. Calculating inbreeding coefficients from pedigrees. In: Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL, editors. Genetics and conservation: a reference for managing wild animal and plant populations. Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc; 1983. p. 509–21.

Ballou JD, Lacy RC. Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou JD, Gilpin M, Foose TJ, editors. Population management for survival and recovery. New York, NY: Columbia University Press; 1995. p. 76–111.

Balmford A, Leader-Williams N, Green M. Parks or arks: where to conserve threatened mammals? Biodivers Conserv. 1995;4:595–607.

Beck BB, Rapaport LG, Price MRS, Wilson AC. Reintroduction of captive-born animals. In: Olney PJS, Mace GM, Feistner ATC, editors. Creative conservation: interactive management of wild and captive animals. Dordrecht, The Netherlands: Springer; 1994. p. 265–86.

Beheregaray LB, Pfeiffer LV, Attard CR, Sandoval-Castillo J, Domingos FM, Faulks LK, et al. Genome-wide data delimits multiple climate-determined species ranges in a widespread Australian fish, the golden perch (*Macquaria ambigua*). Mol Phylogenet Evol. 2017;111:65–75.

Benirschke K. The frozen zoo concept. Zoo Biol. 1984;3:325–8.

Black AN, Seears HA, Hollenbeck CM, Samollow PB. Rapid genetic and morphologic divergence between captive and wild populations of the endangered Leon Springs pupfish, *Cyprinodon bovinus*. Mol Ecol. 2017;26:2237–56.

Blouin M, Parsons M, Lacaille V, Lotz S. Use of microsatellite loci to classify individuals by relatedness. Mol Ecol. 1996;5:393–401.

Bowkett AE. Recent captive-breeding proposals and the return of the ark concept to global species conservation. Conserv Biol. 2009;23:773–6.

Campbell NR, Harmon SA, Narum SR. Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Resour. 2015;15:855–67.

CBSG. Tasmanian devil PHVA final report. Apple Valley: IUCN/SSC Conservation Breeding Specialist Group; 2008.

Champagnon J, Elmberg J, Guillemain M, Gauthier-Clerc M, Lebreton J-D. Conspecifics can be aliens too: a review of effects of restocking practices in vertebrates. J Nat Conserv. 2012;20:231–41.

Christie MR, Marine ML, French RA, Blouin MS. Genetic adaptation to captivity can occur in a single generation. Proc Natl Acad Sci U S A. 2012;109:238–42.

Christie MR, Marine ML, Fox SE, French RA, Blouin MS. A single generation of domestication heritably alters the expression of hundreds of genes. Nat Commun. 2016;7.

Çilingir FG, Rheindt FE, Garg KM, Platt K, Platt SG, Bickford DP. Conservation genomics of the endangered Burmese roofed turtle. Conserv Biol. 2017;31:1469–76.

Clarke CN, Fraser DJ, Purchase CF. Lifelong and carry-over effects of early captive exposure in a recovery program for Atlantic salmon (*Salmo salar*). Anim Conserv. 2016;19:350–9.
Conway WG. The practical difficulties and financial implications of endangered species breeding programmes. Int Zoo Yearb. 1986;24:210–9.
Csilléry K, Johnson T, Beraldi D, Clutton-Brock T, Coltman D, Hansson B, et al. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics. 2006;173:2091–101.
Darwin C. The variation of animals and plants under domestication. London: John Murray; 1868.
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510.
De Bois H, Dhondt A, Van Puijenbroeck B. Effects of inbreeding on juvenile survival of the okapi *Okapia johnstoni* in captivity. Biol Conserv. 1990;54:147–55.
Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol. 2015;25:2577–83.
Derrickson SR, Snyder NF. Potentials and limits of captive breeding in parrot conservation. New world parrots in crisis. Washington, DC: Smithsonian Institution Press; 1992. p. 133–63.
Edwards T, Berry KH. Are captive tortoises a reservoir for conservation? An assessment of genealogical affiliation of captive Gopherus agassizii to local, wild populations. Conserv Genet. 2013;14:649–59.
Ensslin A, Tschope O, Burkart M, Joshi J. Fitness decline and adaptation to novel environments in *ex situ* plant collections: current knowledge and future perspectives. Biol Conserv. 2015;192:394–401.
Etter P, Bassham S, Hohenlohe PA, Johnson E, Cresko W. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol. 2011;772:157–78.
Fischer J, Lindenmayer DB. An assessment of the published results of animal relocations. Biol Conserv. 2000;96:1–11.
Flesness NR. Gene pool conservation and computer analysis. Int Zoo Yearb. 1977;17:77–81.
Foose T, Flesness N, Seal U, De Boer B, Rabb G. Ark into the 21st century. Apple Valley, MN: International Union for the Conservation of Nature and Natural Resources/Captive Breeding Specialist Group; 1992.
Frankham R. Genetic adaptation to captivity in species conservation programs. Mol Ecol. 2008;17:325–33.
Frankham R. Challenges and opportunities of genetic approaches to biological conservation. Biol Conserv. 2010;143:1919–27.
Frankham R, Loebel DA. Modeling problems in conservation genetics using captive *Drosophila* populations: rapid genetic adaptation to captivity. Zoo Biol. 1992;11:333–42.
Frankham R, Ballou JD, Eldridge MD, Lacy RC, Ralls K, Dudash MR, et al. Predicting the probability of outbreeding depression. Conserv Biol. 2011;25:465–75.
Fraser DJ. How well can captive breeding programs conserve biodiversity? A review of salmonids. Evol Appl. 2008;1:535–86.
Frazer NB. Sea turtle conservation and halfway technology. Conserv Biol. 1992;6:179–84.
Gonçalves da Silva A, Lalonde DR, Quse V, Shoemaker A, Russello MA. Genetic approaches refine *ex situ* lowland tapir (*Tapirus terrestris*) conservation. J Hered. 2010;101:581–90.
Grossen C, Biebach I, Angelone-Alasaad S, Keller LF, Croll D. Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. Evol Appl. 2018;11:123–39.
Gruenthal KM, Witting DA, Ford T, Neuman MJ, Williams JP, Pondella DJ, et al. Development and application of genomic tools to the restoration of green abalone in southern California. Conserv Genet. 2014;15:109–21.
Hammerly SC, de la Cerda DA, Bailey H, Johnson JA. A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. Anim Conserv. 2016;19:296–303.
Hayes BJ, Lewin HA, Goddard ME. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet. 2013;29:206–14.

Heath DD, Heath JW, Bryden CA, Johnson RM, Fox CW. Rapid evolution of egg size in captive salmon. Science. 2003;299:1738–40.

Hedrick PW. Conservation genetics: where are we now? Trends Ecol Evol. 2001;16:629–36.

Henkel JR, Jones KL, Hereford SG, Savoie ML, Leibo S, Howard JJ. Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (*Grus canadensis pulla*). Zoo Biol. 2012;31:322–35.

Henry P, Miquelle D, Sugimoto T, McCullough DR, Caccone A, Russello MA. *In situ* population structure and *ex situ* representation of the endangered Amur tiger. Mol Ecol. 2009;18:3173–84.

Hoeck PEA, Wolak ME, Switzer RA, Kuehler CM, Lieberman AA. Effects of inbreeding and parental incubation on captive breeding success in Hawaiian crows. Biol Conserv. 2015;184:357–64.

Hoffman JI, Simpson F, David P, Rijks JM, Kuiken T, Thorne MA, et al. High-throughput sequencing reveals inbreeding depression in a natural population. Proc Natl Acad Sci U S A. 2014;111(10):3775–80.

Howard JG, Lynch C, Santymire RM, Marinari PE, Wildt DE. Recovery of gene diversity using long-term cryopreserved spermatozoa and artificial insemination in the endangered black-footed ferret. Anim Conserv. 2016;19:102–11.

Ishtiaq F, Prakash V, Green RE, Johnson J. Management implications of genetic studies for *ex situ* populations of three critically endangered Asian *Gyps* vultures. Anim Conserv. 2015;18:259–70.

IUCN. The IUCN policy statement on captive breeding. Gland, Switzerland: IUCN; 1987.

Ivy JA, Miller A, Lacy RC, DeWoody JA. Methods and prospects for using molecular data in captive breeding programs: an empirical example using parma wallabies (*Macropus parma*). J Hered. 2009;100:441–54.

Ivy JA, Putnam AS, Navarro AY, Gurr J, Ryder OA. Applying SNP-derived molecular coancestry estimates to captive breeding programs. J Hered. 2016;107:403–12.

Jensen EL, Tapia W, Caccone A, Russello MA. Genetics of a head-start program to guide conservation of an endangered Galápagos tortoise (*Chelonoidis ephippium*). Conserv Genet. 2015;16:823–32.

Jensen EL, Edwards DL, Garrick RC, Miller JM, Gibbs JP, Cayot LJ, et al. Population genomics through time provides insights into the consequences of decline and rapid demographic recovery through head-starting in a Galapagos giant tortoise. Evol Appl 2018 (in press). https://doi.org/10.1111/eva.12682

Johnson JA, Altwegg R, Evans DM, Ewen JG, Gordon IJ, Pettorelli N, et al. Is there a future for genome-editing technologies in conservation? Anim Conserv. 2016;19:97–101.

Jones KL, Glenn TC, Lacy RC, Pierce JR, Unruh N, Mirande CM, et al. Refining the whooping crane studbook by incorporating microsatellite DNA and leg-banding analyses. Conserv Biol. 2002;16:789–99.

Kleinman-Ruiz D, Martínez-Cruz B, Soriano L, Lucena-Perez M, Cruz F, Villanueva B, et al. Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. BMC Genomics. 2017;18:556.

Knief U, Schielzeth H, Backström N, Hemmrich-Stanisak G, Wittig M, Franke A, et al. Association mapping of morphological traits in wild and captive zebra finches: reliable within, but not between populations. Mol Ecol. 2017;26:1285–305.

Kraaijeveld-Smit FJL, Griffiths RA, Moore RD, Beebee TJC. Captive breeding and the fitness of reintroduced species: a test of the responses to predators in a threatened amphibian. J Appl Ecol. 2006;43:360–5.

Krohn AR, Conroy CJ, Pesapane R, Bi K, Foley JE, Rosenblum EB. Conservation genomics of desert dwelling California voles (*Microtus californicus*) and implications for management of endangered Amargosa voles (*Microtus californicus scirpensis*). Conserv Genet. 2018;19:383–95.

Kyle R, Beatty GE, Roberts D, Provan J. Using genetic monitoring to inform best practice in a captive breeding programme: inbreeding and potential genetic rescue in the freshwater pearl mussel *Margaritifera margaritifera*. Conserv Genet. 2016;17:1323–32.

Lacy RC. Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. Conserv Biol. 1987;1:143–58.
Lacy RC. Clarification of genetic terms and their use in the management of captive populations. Zoo Biol. 1995;14:565–77.
Lermen D, Bloemeke B, Browne R, Clarke A, Dyce PW, Fixemer T, et al. Cryobanking of viable biomaterials: implementation of new strategies for conservation purposes. Mol Ecol. 2009;18:1030–3.
Lewis OT, Thomas CD. Adaptations to captivity in the butterfly *Pieris brassicae* (L.) and the implications for *ex situ* conservation. J Insect Conserv. 2001;5:55–63.
Linnarsson S, Teichmann SA. Single-cell genomics: coming of age. Genome Biol. 2016;17:1–3.
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94.
MacCluer JW, VandeBerg JL, Read B, Ryder OA. Pedigree analysis by computer simulation. Zoo Biol. 1986;5:147–60.
Martin MD, Jay F, Castellano S, Slatkin M. Determination of genetic relatedness from low-coverage human genome sequences using pedigree simulations. Mol Ecol. 2017;26(16):4145–57.
Meffe GK. Techno-arrogance and halfway technologies: Salmon hatcheries on the Pacific coast of North America. Conserv Biol. 1992;6:350–4.
Milián-García Y, Jensen EL, Madsen J, Álvarez Alonso S, Serrano Rodríguez A, Espinosa López G, et al. Founded: genetic reconstruction of lineage diversity and kinship informs *ex situ* conservation of Cuban Amazon parrots (*Amazona leucocephala*). J Hered. 2015a;106:573–9.
Milián-García Y, Ramos-Targarona R, Perez-Fleitas E, Sosa-Rodriguez G, Guerra-Manchena L, Alonso-Tabet M, et al. Genetic evidence of hybridization between the critically endangered Cuban crocodile and the American crocodile: implications for population history and *in situ*/*ex situ* conservation. Heredity. 2015b;114:272–80.
Miller JM, Quinzin MC, Poulakakis N, Gibbs JP, Beheregaray LB, Garrick RC, et al. Identification of genetically important individuals of the rediscovered Floreana Galápagos giant tortoise (*Chelonoidis elephantopus*) provide founders for Species Restoration Program. Sci Rep. 2017;7:11471.
Mitchell AA, Lau J, Chemnick LG, Thompson EA, Alberts AC, Ryder OA, et al. Using microsatellite diversity in wild Anegada iguanas (*Cyclura pinguis*) to establish relatedness in a captive breeding group of this critically endangered species. Conserv Genet. 2011;12:771–81.
Myers N. The sinking ark: a new look at the problem of disappearing species. Oxford, UK: Pergamon Press; 1979. 307 p.
Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. Mol Ecol. 2013;22:2841–7.
Oliveira R, Randi E, Mattucci F, Kurushima JD, Lyons LA, Alves PC. Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). Heredity. 2015;115:195–205.
Oliveira PRR, Costa MC, Silveira LF, Francisco MR. Genetic guidelines for captive breeding and reintroductions of the endangered Black-fronted Piping Guan, *Aburria jacutinga* (galliformes, cracidae), an Atlantic Forest endemic. Zoo Biol. 2016;35:313–8.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7:e37135.
Philippart JC. Is captive breeding an effective solution for the preservation of endemic species? Biol Conserv. 1995;72:281–95.
Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, et al. The biodiversity of species and their rates of extinction, distribution, and protection. Science. 2014;344:1246752.
Price MR, Sischo D, Pascua MA, Hadfield MG. Demographic and genetic factors in the recovery or demise of *ex situ* populations following a severe bottleneck in fifteen species of Hawaiian tree snails. PeerJ. 2015;3:e1406.

Rahbek C. Captive breeding – a useful tool in the preservation of biodiversity? Biodivers Conserv. 1993;2:426–37.

Ralls K, Ballou J. Extinction: lessons from zoos. In: Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL, editors. Genetics and conservation: a reference for managing wild animal and plant populations. Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc.; 1983. p. 164–84.

Ralls K, Ballou JD, Rideout BA, Frankham R. Genetic management of chondrodystrophy in California condors. Anim Conserv. 2000;3:145–53.

Ray JW, King RB, Duvall MR, Robinson JW, Jaeger CP, Dreslik MJ, et al. Genetic analysis and captive breeding program design for the eastern massasauga *Sistrurus catenatus catenatus*. J Fish Wildl Manag. 2013;4:104–13.

Ritland K. Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res. 1996;67:175–85.

Russello MA, Amato G. *Ex situ* population management in the absence of pedigree information. Mol Ecol. 2004;13:2829–40.

Russello MA, Amato G. On the horns of a dilemma: molecular approaches refine *ex situ* conservation in crisis. Mol Ecol. 2007;16:2405–6.

Russello MA, Hyseni C, Gibbs JP, Cruz S, Marquez C, Tapia W, et al. Lineage identification of Galápagos tortoises in captivity worldwide. Anim Conserv. 2007;10:304–11.

Russello MA, Poulakakis N, Gibbs JP, Tapia W, Benavides E, Powell JR, et al. DNA from the past informs *ex situ* conservation for the future: an "extinct" species of Galápagos tortoise identified in captivity. PLoS One. 2010a;5:e8683.

Russello MA, Stahala C, Lalonde D, Schmidt KL, Amato G. Cryptic diversity and conservation units in the Bahama parrot. Conserv Genet. 2010b;11:1809–21.

Ryder O, Miller W, Ralls K, Ballou JD, Steiner CC, Mitelberg A, et al. Whole genome sequencing of California condors is now utilized for guiding genetic management. International Plant and Animal Genome XXIV Conference; 8–13 Jan 2016, San Diego, CA, USA.

Saragusty J, Diecke S, Drukker M, Durrant B, Friedrich Ben-Nun I, Galli C, et al. Rewinding the process of mammalian extinction. Zoo Biol. 2016;35:280–92.

Seal U, Foose T, Ellis S. Conservation assessment and management plans (CAMPs) and global captive action plans (GCAPs). Creative conservation: interactive management of wild and captive animals. Dordrecht, The Netherlands: Springer; 1994. p. 312–25.

Snyder NFR, Derrickson SR, Beissinger SR, Wiley JW, Smith TB, Toone WD, et al. Limitations of captive breeding in endangered species recovery. Conserv Biol. 1996;10:338–48.

Soulé ME, Simberloff D. What do genetics and ecology tell us about the design of nature reserves? Biol Conserv. 1986;35:19–40.

Speed D, Balding DJ. Relatedness in the post-genomic era: is it still useful? Nat Rev Genet. 2015;16:33–44.

Strzala T, Kowalczyk A, Lukaszewicz E. Reintroduction of the European capercaillie from the Capercaillie Breeding Centre in Wisla Forest District: genetic assessments of captive and reintroduced populations. PLoS One. 2015;10:13.

Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9:465–76.

Svengren H, Prettejohn M, Bunge D, Fundi P, Bjorklund M. Relatedness and genetic variation in wild and captive populations of Mountain Bongo in Kenya obtained from genome-wide single-nucleotide polymorphism (SNP) data. Glob Ecol Conserv. 2017;11:196–206.

Tear TH, Scott JM, Hayward PH, Griffith B. Status and prospects for success of the Endangered Species Act. Science. 1993;262:976–7.

Therkildsen NO, Palumbi SR. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Mol Ecol Resour. 2017;17:194–208.

Thompson EA. Pedigree analysis in human genetics. Baltimore: Johns Hopkins University Press; 1986.

Tokarska M, Marshall T, Kowalczyk R, Wójcik J, Pertoldi C, Kristensen T, et al. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity. 2009;103:326–32.

Tudge C. Last animals at the zoo: how mass extinction can be stopped. Washington, DC: Island Press; 1992. 265 p.

Tunstall T, Kock R, Vahala J, Diekhans M, Fiddes I, Armstrong J, et al. Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. Genome Res. 2018;28(6):780–8.

Urano K, Tsubono K, Taniguchi Y, Matsuda H, Yamada T, Sugiyama T, et al. Genetic diversity and structure in the Sado captive population of the Japanese crested ibis. Zoolog Sci. 2013;30:432–8.

Valbuena-Urena E, Soler-Membrives A, Steinfartz S, Alonso M, Carbonell F, Larios-Martin R, et al. Getting off to a good start? Genetic evaluation of the ex situ conservation project of the Critically Endangered Montseny brook newt (Calotriton arnoldi). PeerJ. 2017;5:25.

Van de Casteele T, Galbusera P, Matthysen E. A comparison of microsatellite-based pairwise relatedness estimators. Mol Ecol. 2001;10:1539–49.

Waters CD, Hard JJ, Brieuc MSO, Fast DE, Warheit KI, Waples RS, et al. Effectiveness of managed gene flow in reducing genetic divergence associated with captive breeding. Evol Appl. 2015;8:956–71.

West R, Potter S, Taggart D, Eldridge MDB. Looking back to go forward: genetics informs future management of captive and reintroduced populations of the black-footed rock-wallaby Petrogale lateralis. Conserv Genet. 2018;19:235–47.

Willoughby JR, Ivy JA, Lacy RC, Doyle JM, DeWoody JA. Inbreeding and selection shape genomic diversity in captive populations: implications for the conservation of endangered species. PLoS One. 2017;12.

Witzenberger KA, Hochkirch A. The genetic integrity of the *ex situ* population of the European wildcat (*Felis silvestris silvestris*) is seriously threatened by introgression from domestic cats (*Felis silvestris catus*). PLoS One. 2014;9(8):e106083.

Wright S. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics. 1921;6:124–43.

# Population Genomics Advances and Opportunities in Conservation of Kiwi (*Apteryx* spp.)


**Kristina M. Ramstad and Luke T. Dunning**

**Abstract** Kiwi (*Apteryx* spp.) are highly threatened flightless birds endemic to New Zealand. They are members of the most basal extant avian lineage, the paleognaths, and exhibit a suite of traits that are unusual in birds. Despite their iconic and imperiled status, there have been only four genomic studies of kiwi to date with only two of these aimed at improving conservation. There is, therefore, massive opportunity to use genomic techniques to elucidate the genetic basis and consequences of the strange ecology and evolution of kiwi and to inform their intensive management. In this chapter, we review genomic studies in paleognaths, assess prospects for the future of kiwi genomics, and define some lessons for population genomics and conservation of at-risk taxa generally. We also present an analysis of genomic signatures associated with the evolution of Apterygidae and the genes involved in diversification of kiwi via comparison of 3,774 orthologous protein coding genes among 28 avian species. We found strong signals of selection in genes associated with dwarfism, neurogenesis, retinal development, and temperature regulation. Our results provide clues as to why kiwi have such small body size (relative to other paleognaths), large egg size (relative to their body size), excellent olfaction, and poor vision. The data further suggest that coping with highly divergent temperature regimes may be a defining feature of the spotted kiwi clade which includes the only kiwi species that inhabits the alpine zone. Considerable genomic resources are now available for kiwi, including whole-genome sequences, transcriptome assemblies, thousands of SNP markers, and numerous candidate genes. There is also a myriad of outstanding questions about kiwi that genomic studies can inform. The challenge now is to bring these new genomic tools to bear on conservation and management of kiwi.

K. M. Ramstad (✉)
Department of Biology and Geology, University of South Carolina Aiken, Aiken, SC, USA
e-mail: kristinar@usca.edu

L. T. Dunning
Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK
e-mail: l.dunning@sheffield.ac.uk

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2019_71, 

**Keywords** *Apteryx* · Conservation · Genomics · Kiwi · Paleognath · Positive selection

## 1 Introduction

The fields of molecular ecology and conservation genetics have seen dramatic advances over the last decade as high-throughput sequencing has been widely adopted. The transition from genotyping a handful of loci to sequencing whole genomes has enabled researchers to shift their focus from neutral loci that capture patterns of population dynamics to identifying functional loci of adaptive significance (Allendorf 2017; Luikart et al. 2019; Shafer et al. 2015). Genomic approaches have also yielded an unprecedented ability to explore the genetic architecture of adaptation in non-model organisms by coupling reference genomes with resequencing data from multiple individuals (Berg et al. 2016; Hartmann et al. 2017; Hohenlohe et al. 2010a; Zhang et al. 2014a; Zhou et al. 2014). The increased number of markers used in whole-genome, and to a lesser extent reduced representation (e.g., RADseq), studies has greatly improved the power of traditional population genetic analyses aimed at estimating effective population size ($N_e$), inbreeding coefficients, population structure, dispersal, admixture, and introgression (Allendorf 2017; Berg et al. 2016; Bernatchez et al. 2017).

Conservation genomics is an extension of conservation genetics and seeks to apply genomic techniques to practical management of small populations of at-risk species. While there is widespread appreciation for the practical benefits of conservation genomics (Allendorf et al. 2010; Andrews et al. 2016; Funk et al. 2012; Hohenlohe et al. 2010b; Ryder 2005; Steiner et al. 2013), the real-world applications have been slow in coming (Funk et al. 2018), and there has been a persistent gap between theory and practice. Wildlife managers often view genomic research as too costly, confusing, and unreliable to add value to conservation programs. In addition, advancements are still needed, particularly in analytical techniques, as our ability to analyze data is being outpaced by data collection (Andrews and Luikart 2014; Benestan et al. 2016; Shafer et al. 2015, 2016; Steiner et al. 2013). Despite these challenges, an increasing number of case studies demonstrate the benefits of applying genomic techniques to species conservation (e.g., Tasmanian devil, *Sarcophilus harrisii*, Margres et al. 2018; greater sage grouse, *Centrocercus urophasianus*, Oyler-McCance et al. 2019; ungulates, Martchenko et al. 2018; Burmese roofed turtle, *Batagur trivittata*, Çilingir et al. 2017; Pacific salmon, *Oncorhynchus* spp., Prince et al. 2017; plateau deer mouse, *Peromyscus melanophrys*, Vega et al. 2017), and the gap between basic genomic research and applied conservation is narrowing.

Genomics of small populations, which we define here as those having an $N_e$ of less than 200 (and thus a census size of likely $<1{,}000$; Jamieson and Allendorf 2012; Palstra and Ruzzante 2008), differs in a number of fundamental ways from genomics of large populations. Small populations often exist in fragmented, isolated, inaccessible, and wild habitats; the distribution of such species is often patchy due to

extirpation of populations or whole lineages. Low census size and rarity often mean few individuals can be sampled, and tissue samples often must be collected using non-invasive techniques (Hoffman et al. 2013; Hoffman 2011; Ramstad et al. 2016). Thus, the quantity, type (e.g., lethal sampling for different tissues for RNA expression analysis), and quality of samples (e.g., feathers or hair rather than tissue or blood) for genomic analysis are limited for at-risk species. Applying traditional population genetic measures can be difficult for at-risk species as well. Long generation intervals and overlapping generations make estimating $N_e$ difficult (Waples et al. 2014), and low sample sizes make it difficult statistically to find loci under selection or associated with ecological traits (Margres et al. 2018; Shen et al. 2011). In addition, rare species with small populations are vulnerable to elevated levels of both interspecific hybridization and inbreeding depression (Edmands 2007). At-risk species are also often phylogenetically distinct, making it unlikely that a closely related reference genome will be available (Miller et al. 2012). All of these issues are a product of the biology and history of these endangered populations and still complicate matters despite the advent of genomics.

Genomics, and in particular whole-genome sequencing, initially focused on a few key model species and then expanded to include abundant and economically important species (Bernatchez et al. 2017). These early studies typically included a single individual or handful of individuals to maximize sequencing depth for de novo genome assembly. In recent years, however, genomic research has included more genome resequencing to assess the variation within and among populations, and there have been increasing numbers of projects aimed at sequencing the genomes of endangered species (e.g., koala, *Phascolarctos cinereus*, Johnson et al. 2018; giant panda, *Ailuropoda melanoleuca*, Li et al. 2010; Chinese alligator, *Alligator sinensis*, Wan et al. 2013). More reference genomes are becoming publicly available each year, and increased efficiency and decreased cost of collecting genomic data are allowing unprecedented depth and breadth of sampling of genomic variation at the population level. Techniques such as Rapture (Ali et al. 2016) and RADcap (Hoffberg et al. 2016) allow both to be accomplished simultaneously by assaying variation at the same suite of thousands of SNPs reliably across different genomic libraries.

In this chapter, we use kiwi (family Apterygidae) as a case study to illustrate the application of conservation genomic techniques to species of conservation concern. Avian taxa are excellent systems for comparative genomic and molecular evolution studies because they have relatively small genomes that show highly conserved synteny (Zhang et al. 2014a). We review how recent genomic research has significantly improved our understanding of the evolution, taxonomy, and ecology of kiwi. We combine existing genomic data for kiwi and present an analysis detecting genes under positive selection in the different lineages, illustrating that meaningful conservation genomic insights do not always require new data. Finally, we outline the conservation impact of this research and identify future genomic research that may aid conservation of this charismatic species. Although focused on kiwi, much of the research discussed is applicable to any taxon of conservation concern and so provides an example of how researchers might initially apply genomic techniques to non-model species.

## 1.1 Evolutionary and Ecological Significance of Kiwi

Kiwi are one of the most instantly recognizable bird species and a national icon of New Zealand. They represent the most basal extant avian lineage, the paleognaths, and possess a suite of traits that are rare or absent in other birds (Fig. 1). Paleognaths make up less than 1% of extant avian species and differ from all other birds (neognaths) by having a primitive and reptile-like palate (the paleognath palate), a distinct pelvic structure (a large and open ilioischiatic fenestra), and an unusual pattern of grooves on their bills (rhamphotheca; Cracraft 1974; Huxley 1867; Parkes and Clark 1966). The paleognaths include five extant and two extinct lineages of flightless ratites and the flighted tinamous (Harshman et al. 2008), all of which are restricted to the southern hemisphere (Fig. 1). Flightlessness is relatively rare (<1%) among extant birds (McCall et al. 1998), but is found in seven of eight paleognath lineages which together comprise the ratites. Not surprisingly, ratites share a number of traits associated with flightlessness that are not typically found in neognaths or tinamous. For example, they lack a keeled sternum and have reduced flight muscles, furcula (wishbones), and wings (Cracraft 1974; Fowler 1991). Ratites also tend to be large-bodied and herbivorous and have "hairlike" feathers (Fowler 1991).

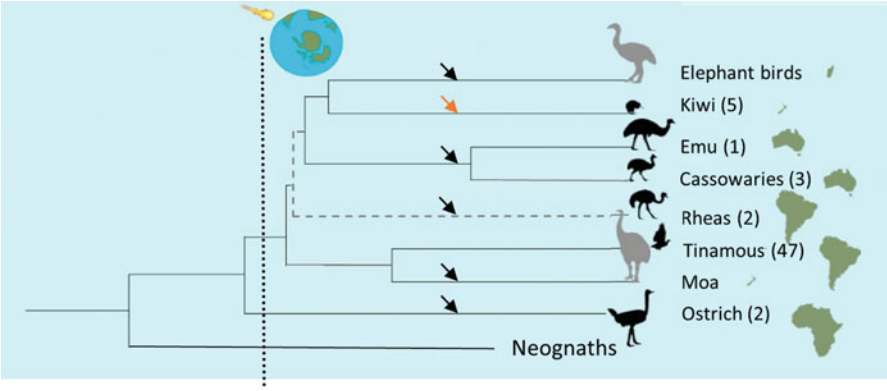


**Fig. 1** Phylogeny of the paleognaths, modified from Maderspacher (2017) and based on Yonezawa et al. (2017). More recent studies (Cloutier et al. 2019; Sackton et al. 2018) place rhea as sister taxa to kiwi/emu/cassowary (dashed lines) but do not change the position of kiwi. Paleognaths make up less than 1% of extant avian species and are the most basal lineage of extant birds. Extant paleognaths include five flightless lineages (the ratites) and one volant lineage (tinamous). The dotted line indicates the Cretaceous-Paleogene extinction event that occurred approximately 65 Mya, arrows indicate hypothesized loss of flight either coupled with gigantism (black) or not (orange; Mitchell et al. 2014), and silhouettes in gray and black indicate extinct and extant lineages, respectively. Geographic distribution of each taxon is indicated on the far right and number of extant species per lineage is in parentheses. Ostriches (family Struthionidae) are found in Africa, kiwi (family Apterygidae) and extinct moa (order Dinornithiformes) are found in New Zealand, emu (*Dromaius novaehollandiae*) and cassowary (*Casuarius* spp.) are native to Australia and New Guinea, rhea (family Rheidae) and tinamous (family Tinamidae) are found in South America, and the extinct elephant birds (family Aepyornithidae) were endemic to Madagascar (Harshman et al. 2008; Mitchell et al. 2014)

Kiwi diverged from other paleognaths approximately 60 million years ago (Mitchell et al. 2014; Phillips et al. 2010) and five species of kiwi comprising two clades are currently recognized (Fig. 2). The little spotted (LSK; *Apteryx owenii*) and great spotted kiwi (GSK; *A. haastii*) belong to the spotted clade, and North Island brown kiwi (NIB; *A. mantelli*), Okarito brown kiwi or rowi (*A. rowi*), and tokoeka (*A. australis*; Holzapfel et al. 2008) comprise the brown kiwi clade. The brown and spotted kiwi clades diverged from one another approximately 5 million years ago (Weir et al. 2016; Fig. 2) making kiwi an ancient lineage with deep evolutionary divergence between species.

Kiwi possess a set of traits not found in other paleognaths. These include paired functional ovaries (there is typically a single functional ovary in birds and non-avian reptiles) that produce extremely large eggs (up to >25% of the weight of the female; Kinsky 1971; Sales 2005). For example, NIB kiwi eggs are approximately 400% larger than the allometrically expected value for a bird of their body size (Prinzinger and Dietz 2002). These enormous eggs have unusually high yolk content (~65% compared to 30–40% on average) and antimicrobial properties relative to other birds (Sales 2005). Kiwi have small eyes and a small optic lobe and lack color vision (Le Duc et al. 2015; Sales 2005). In contrast, they have a highly enlarged olfactory

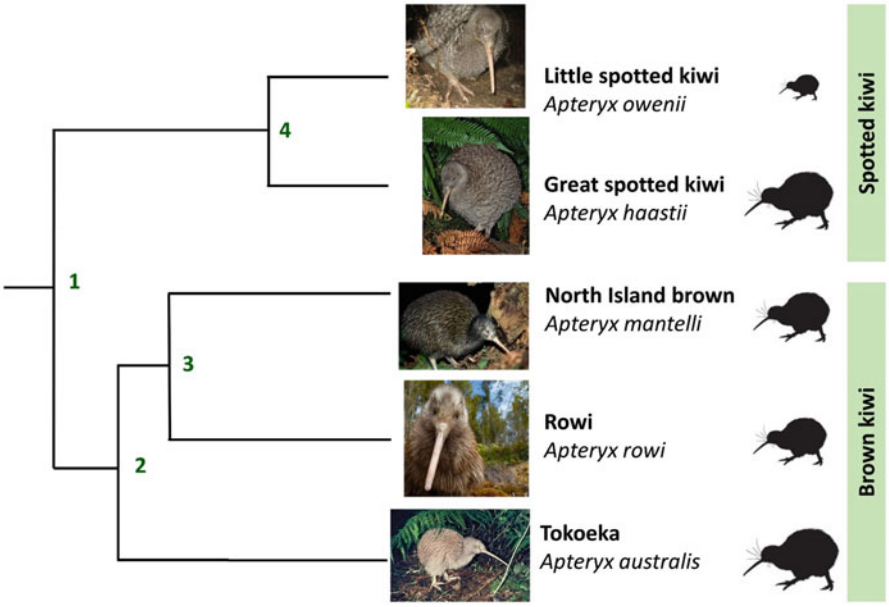


**Fig. 2** Phylogeny of family Apterygidae based on Weir et al.'s (2016) analysis of 1,710 mtDNA SNPs and 1,000 nuclear SNPs. Five species of kiwi in two clades (spotted and brown) are currently recognized with all divergence events having occurred within the last 8 million years. Node dates (95% CI) are (1) 3.85 Mya (1.87–7.00), (2) 1.56 Mya (0.76–2.83), (3) 1.12 Mya (0.54–2.02), and (4) 0.55 Mya (0.27–0.99; Weir et al. 2016). Bird silhouettes are scaled to mean body weight based on data in Table 1. Photo credits: *A. owenii* Andrew Digby, *A. haastii* and *A. mantelli* Tui De Roy, *A. rowi* Grant Maslowski, and *A. australis* John Kendrick

bulb and exhibit strong olfaction; they are the only bird in the world with nostrils on the end of their bill (Castro et al. 2010; Corfield et al. 2008). The enlarged olfactory bulb results in kiwi having unusually large brain size relative to body size, on par with that observed in parrots and songbirds (Corfield et al. 2008).

Kiwi behavior differs significantly from other paleognaths as well. They are almost entirely nocturnal (Heather and Robertson 2005), a behavior that is found in less than 3% of all avian species and none of the other paleognaths (Le Duc et al. 2015; Martin 1986). They are also fossorial, meaning they are adapted to digging and nest and shelter in underground burrows (Sales 2005). Kiwi form long-term monogamous pair bonds and have an extremely long incubation period (Fowler 1991; Sales 2005). They typically lay one to two eggs per clutch which are then incubated for 65–85 days either by both parents (rowi, tokoeka, GSK) or solely by the male (LSK and NIB kiwi; Sales 2005). The eggs have thin shells and are prone to cracking and microbial invasion during their long underground incubation period (Calder 1979). Chicks are precocial – they hatch fully feathered and reach maturity within 2–5 years (Sales 2005) – and kiwi can live to more than 50 years of age (life expectancy of a territorial adult LSK is 45 years (95% CI 27–83 years; Robertson and Colbourne 2004)). Age of reproductive senescence is unknown; the oldest known LSK was at least 32 years old in 2013 and was still producing chicks annually (Ramstad et al. 2013; Taylor et al. 2017).

Finally, while ratites are the largest birds in the world (Kummrow 2014), kiwi are the smallest of the ratites (Davies 2003; Sales 2005). It has been estimated that the elephant bird could weigh in excess of 400 kg (Amadon 1947) and moa up to 300 kg (Amadon 1947; Bunce et al. 2003). In contrast, kiwi mean weight is between 1 and 4 kg, a trait that is shared with the small-bodied and flighted tinamous (Davies 2003; Sales 2005; Table 1; Fig. 1). Along with small body size, kiwi display the lowest basal metabolic rate observed in birds, low body temperature (2–4°C lower than expected of similar-sized neognathous birds), and slow growth rate (Sales 2005).

**Table 1** Mean body weight of the five currently recognized kiwi species by sex (number of birds weighed)

| Species | Sex | Weight (kg) | Sources |
|---|---|---|---|
| Little spotted | M (51) | 1.14 | Jolly and Daugherty (2002) |
| *Apteryx owenii* | F (41) | 1.35 | |
| Great spotted | M (39) | 2.31 | McLennan and McCann (2002) |
| *A. haastii* | F (29) | 3.19 | |
| North Island brown | M (34) | 2.04 | McLennan et al. (2004) and Miles et al. (1997) |
| *A. mantelli* | F (22) | 2.66 | |
| Rowi | M (49) | 1.92 | Tennyson et al. (2003) |
| *A. rowi* | F (51) | 2.65 | |
| Tokoeka | M (71) | 2.6 | Edmonds (2015) |
| *A. australis* | F (27) | 3.3 | |

In summary, kiwi are unique among extant paleognaths in that they have larger eggs and brains relative to body size, are nocturnal and fossorial, form long-term monogamous pair bonds, and have an extremely long incubation period. In addition, they are the only extant paleognath that is both flightless and small-bodied. Their basal position in the avian phylogeny and unusual phenotypic traits make them an important taxon for studies of avian evolution.

## *1.2 Kiwi Are Highly Threatened and Intensively Managed*

Flightlessness, ground nesting, small body size, and large egg size make kiwi highly vulnerable to predation, while their life history (long generation interval, limited dispersal ability, low reproductive rate, intense parental investment) means adult mortality has a tremendous impact on their population growth and persistence (Sæther and Bakke 2000). Four of the five species of kiwi are listed as Threatened (Vulnerable) by the International Union for Conservation of Nature (IUCN 2019); the exception is LSK which are considered Near Threatened and are genetically imperiled (Ramstad et al. 2013; Taylor et al. 2017). Numerous diseases, parasites, and pathogens threaten kiwi (Sales 2005; White et al. 2016), but their principal threat is predation by introduced mammalian predators, including stoats, rats, possums, and dogs (Germano et al. 2018). Indeed, kiwi populations are declining at a rate of approximately 2% per year with chicks having a 5–6% chance of survival in wild and unmanaged landscapes (Germano et al. 2018; Holzapfel et al. 2008; McLennan et al. 1996). While there are approximately 70,000 kiwi at present, 76% are living in unmanaged areas, and kiwi numbers continue to decline (Germano et al. 2018).

Significant efforts have been made to conserve kiwi, including extensive mammal trapping and poisoning, establishing predator-free kiwi sanctuaries, translocation programs, and the captive incubation and rearing program called Operation Nest Egg (Colbourne et al. 2005). These efforts have been spearheaded by the New Zealand Department of Conservation (DOC) and involve countless community groups, trusts, zoos, sanctuaries, and volunteers throughout New Zealand (Germano et al. 2018). Over 100 Māori iwi (tribes) are also actively involved in kiwi conservation. Māori are the indigenous people of New Zealand and consider kiwi a tāonga (treasure) for which they are kaitiaki (guardians). Obtaining a permit to work with kiwi, or collect or use archived samples, requires consultation with iwi which can be time-consuming and sensitive. Conservation efforts have been successful; where managed, kiwi populations are growing at a rate of 2% or more. The current goal of kiwi management is to reverse the 2% decrease seen in unmanaged sites to a 2% annual increase across all kiwi species and to reach 100,000 kiwi by 2030 while also restoring the former distribution of kiwi and maintaining their genetic diversity (Germano et al. 2018).

Significant funding is put toward controlling predators at the landscape scale (100,000 of hectares) to conserve kiwi and other native species. For example, Predator Free New Zealand invests $5 million NZD annually toward their goal of

ridding New Zealand of its most damaging introduced predators by 2050 (Germano et al. 2018). The DOC Battle for Our Birds campaign put $21.3 million NZD toward controlling a boom in rat and stoat numbers associated with beech forest masting in 2017. In addition, the 2018 government budget included an additional $20 million NZD for predator control (Germano et al. 2018). The beloved and iconic status of kiwi has resulted in significant governmental funds being allocated to kiwi conservation specifically as well. For example, Operation Nest Egg costs approximately $75,000 NZD annually and the New Zealand government allocated $11.2 million NZD between 2015 and 2019 and another $6.8 million NZD annually thereafter for the Save Our Iconic Kiwi program (Germano et al. 2018).

## 2 Review of Relevant Genomic Research

### *2.1 Genomic Studies of Paleognaths*

Recent comparative genomic studies have transformed our understanding of paleognath and ratite evolution (Cloutier et al. 2019; Grealy et al. 2017; Mitchell et al. 2014; Sackton et al. 2018; Yonezawa et al. 2017). Prior to these studies, it was thought that ratites were monophyletic, kiwi were most closely related to moa, and ratites evolved via allopatric speciation associated with the breakup of Gondwana (Cracraft 1974). These conclusions were based on the geographic distribution of species and geological history, with continental rafting thought to be the driving force in their diversification (Cracraft 1974). Ratites were thought to have arisen from a common flightless and large-bodied ancestor that independently rafted to their current locations as Gondwana divided (Paton et al. 2002).

High-throughput sequencing was used to recover the mitochondrial genome and approximately 12,500 bp of nuclear exon sequence of the extinct elephant bird of Madagascar and for the first time showed they are sister taxa to kiwi (Grealy et al. 2017; Mitchell et al. 2014; Fig. 1). This result was contrary to expectations based on continental vicariance because Madagascar and New Zealand had never been directly connected and elephant birds and kiwi diverged approximately 50 Mya, which was well after the breakup of Gondwana. Mitchell et al. (2014) hypothesized that flighted dispersal, potentially from Antarctica which was warm and covered in trees at the time (Pross et al. 2012), gave rise to the current distribution of ratites. The presence of small, potentially volant paleognaths in New Zealand during the early Miocene and the fact that paleognaths in the early Tertiary were capable of long-distance flight support this hypothesis (Mayr 2009; Mitchell et al. 2014; Worthy et al. 2013). Following dispersal, every major ratite lineage would have lost flight independently (at least six times in total; Fig. 1). The revised paleognath phylogeny suggests also that gigantism evolved at least five times among ratites and that the large egg size to body size observed in kiwi was likely inherited from an ancestor kiwi shared with elephant birds (Mitchell et al. 2014; Yonezawa et al. 2017).

Parallel loss of flight and gigantism appear to have been associated with an explosive radiation event approximately 69–52 million years ago after the Cretaceous-Paleogene mass extinction (Grealy et al. 2017; Mitchell et al. 2014; Yonezawa et al. 2017). The extinction of dinosaurs would have provided an opportunity for large, flightless, herbivorous birds to evolve. Mitchell et al. (2014) further hypothesized that kiwi may have evolved to fill the alternative niche of a nocturnal, ground-dwelling, and small omnivore because the now extinct moa were already occupying the large herbivore niche when the ancestors of kiwi colonized New Zealand.

The most recent paleognath phylogenies are based on whole-genome datasets and challenge the typology of the previous trees based on concatenated data (Cloutier et al. 2019; Sackton et al. 2018). Cloutier et al. (2019) used over 41 million base pairs of aligned sequence data including over 20 thousand loci from three types of non-coding nuclear markers. Their tree places rheas, and not tinamous, as the sister taxa to the kiwi and emu + cassowary branch. Sackton et al. (2018) found the same result after constructing a phylogeny based on 284 thousand conserved non-exonic elements in 14 paleognath species, including 11 newly assembled genomes. The position of kiwi, however, remains unchanged. Collectively, these recent genomic studies overturned decades of morphological and genetics studies and revolutionized our view of avian evolution. Ratites have been used as an example of vicariance for decades, a scenario which is highly unlikely based on recent genomic studies, but is still in nearly all introductory biology textbooks. While there is general agreement on the order and timing of kiwi evolution, the tree typology and timing of divergence among paleognath lineages are still a matter of great debate.

## *2.2 Genomic Studies of Kiwi*

Despite their scientific and cultural value, a great deal remains unknown about kiwi due to their secretive nature, nocturnal habits, long generation interval, and highly protected status. Four genomic studies to date have focused on understanding the evolution of kiwi specifically.

The first kiwi transcriptome was published in 2010 and based on a single NIB individual (Subramanian et al. 2010). Over 1,543 conserved protein-coding regions were identified using the chicken genome as a scaffold, and 702 genes were used to estimate the divergence time between paleognaths and neognaths. Subramanian et al. (2010) estimated that the split occurred 132 million years which was consistent with previous estimates based on mitochondrial genomes (Brown et al. 2008; Pereira and Baker 2006). A mixture of tissues from a male embryo was sequenced for this study. Thus, differential expression between tissues and sex chromosomes (females are the heterogametic sex in birds) could not be assessed, and most of the genes identified were highly constrained and associated with protein synthesis, structural proteins, and developmental gene regulation.

The first kiwi genome was published in 2015 and based on three NIB kiwi (Le Duc et al. 2015). The authors annotated this genome with the NIB transcriptome of Subramanian et al. (2010) and other well-annotated avian genomes and tested for evidence of positive selection specific to NIB kiwi in 4,152 genes orthologs among 8 bird species. They found that several opsin genes are pseudogenized in kiwi and dated the loss of color vision to 30–38 million years ago after kiwi arrived in New Zealand. They also found that kiwi have a highly diverse suite of odorant receptor genes relative to all other birds, including other ratites, reflecting their greater reliance on smell than sight. A number of genes associated with metabolism and energy expenditure also show signs of selection in NIB. The authors attributed all of these findings to the kiwi's nocturnal lifestyle.

A more recent paper presented transcriptomes of 16 individual kiwi (8 rowi and 8 LSK) and showed gene expression differences between species and sexes (Ramstad et al. 2016). The study used non-lethal sampling of the two rarest kiwi species and focused on finding variation within and between species that could be useful for conservation genomic studies. More than 7,900 protein coding transcripts were identified based on homology with chicken. Though only whole blood was sequenced, the genes identified were related to diverse functions including growth, development, disease resistance, reproduction, and behavior. Importantly, a wealth of SNP markers within these transcripts differentiate between rowi and LSK (66,909) and between individuals within these rare species (LSK, 12,384; rowi, 29,313). Many of these SNPs are likely neutral as they are positioned in untranslated regions or reflect synonymous substitutions. The study defined a suite of functional genes potentially under selection in kiwi to aid in future studies of evolution and adaptation as well.

Ramstad et al. (2016) also found significant expression differences between male and female kiwi in 150 transcripts most of which showed male-biased expression ($n = 94$) and were syntenic with the Z chromosome ($n = 79$). Given incomplete dosage compensation in birds, male-to-female expression ratios and high SNP densities found in a minority of these chrZ genes suggest a large pseudoautosomal region in kiwi sex chromosomes. This pattern was previously known in emu and ostrich (Ogawa et al. 1998; Zhou et al. 2014), suggesting a common pattern of homomorphy, recombination, and gene dosage among living paleognaths.

Also in 2016, Weir and colleagues presented a genomic analysis of 1,710 bp of mtDNA sequence and 6,332 SNPs that resolved the taxonomy and provided evidence of historical demography of kiwi (Weir et al. 2016). They confirmed the long suspected presence of four extant genetic lineages within each of tokoeka and NIB kiwi, thus bringing the total number of extant kiwi lineages to 11. Divergence between some of these lineages may rise to the level of species ($F_{ST} = 0.21$–$0.63$) and provides evidence of greater divergence and diversity within extant kiwi than previously appreciated. Weir et al. (2016) also identified up to six additional extinct kiwi lineages (two tokoeka, two rowi, and one or two LSK) and showed there was an intense period of speciation associated with Pleistocene glaciation that accounts for 80% of kiwi diversification events. Habit fragmentation due to glacial ice resulted in numerous isolated refugia, particularly in the South Island of New Zealand to which

ten of these lineages were restricted. Kiwi diversification rates during the Pleistocene are greater than the most explosive radiations previously documented, including Galapagos finches and Tanganyika cichlids (Weir et al. 2016).

Fully assembled and annotated genomes of rowi, GSK, and LSK have become available since the analysis we present here (Sackton et al. 2018), as well as the complete mitochondrial genome of NIB (Liu et al. 2017) and the first nuclear genome assembly of a moa (Cloutier et al. 2018). Additional paleognath genomes have become available recently as well (Sackton et al. 2018), to add to the more than 50 avian genomes previously available (Zhang et al. 2014a). Analysis of these new genome assemblies has (1) shown that loss of flight in paleognaths is due to changes in regulatory sequences and not in protein coding sequences (Sackton et al. 2018), (2) allowed reevaluation of avian sex chromosome evolution (Xu et al. 2018), and (3) uncovered an anomaly zone in paleognaths that has been a likely the source of the difficulty in resolving their phylogeny (Cloutier et al. 2019). The focus, however, has remained almost entirely on evolution and not conservation per se. Of the papers reviewed above, only two discuss the utility of their results for conservation (Ramstad et al. 2016; Weir et al. 2016), one is focused on an extinct lineage (Cloutier et al. 2018), and none have been applied to paleognath management.

In the section below, we test for positive selection in each of three kiwi species, two kiwi clades and the kiwi lineage broadly. Previous work tested for positive selection in only NIB kiwi and so was not able to distinguish whether selection was acting before or after NIB diverged from the other kiwi species. There are numerous outstanding questions for kiwi genomics, including:

1. Is there a genomic signature associated with the evolution of Apterygidae?
2. What genes were involved in the diversification of kiwi species?
3. How can genomic data help with kiwi conservation?

## 3 Analysis of Genes Under Selection in Kiwi

### *3.1 Leveraging Publicly Available Data to Study Evolution in Kiwi*

We aligned protein coding sequences from the NIB kiwi genome (Le Duc et al. 2015) and rowi and LSK reference blood transcriptomes (Ramstad et al. 2016) with ortholog alignments from diverse avian species. High-quality alignments are essential to accurately infer positive selection (Jordan and Goldman 2011; Ramstad et al. 2016). We therefore utilized the 8,295 bird ortholog alignments from the Avian Genome Consortium (Zhang et al. 2014a). The initial alignments contained sequences from 48 species. To reduce computational time, we removed a subset of sequences from the overrepresented neonaves for a final data set of 25 species (Table 2). We used a previously published stringent phylogenetic annotation pipeline to identify orthologous kiwi genes and generate subsequent high-quality

**Table 2** Species included in the positive selection analysis

| Order | Family | Common group | Common species | Species |
|---|---|---|---|---|
| Passeriformes | Estrildidae | Finches | Zebra finch | *Taeniopygia guttata* |
| Psittaciformes | Strigopidae | Parrots | Kea | *Nestor notabilis* |
| Falconiformes | Falconidae | Falcons | Peregrine falcon | *Falco peregrinus* |
| Cariamiformes | Cariamidae | Seriemas | Red-legged seriema | *Cariama cristata* |
| Bucerotiformes | Bucerotidae | Hornbills | Rhinoceros hornbill | *Buceros rhinoceros* |
| Strigiformes | Tytonidae | Owls | Barn owl | *Tyto alba* |
| Accipitriformes | Accipitridae | Eagles | Bald eagle | *Haliaeetus leucocephalus* |
| Pelecaniformes | Pelecanidae | Pelicans | Dalmation pelican | *Pelecanus crispus* |
| Sphenisciformes | Spheniscidae | Penguins | Emperor penguin | *Aptenodytes forsteri* |
| Gaviiformes | Gaviidae | Loons | Red-throated loon | *Gavia stellata* |
| Phaethontiformes | Phaethontidae | Tropicbirds | White-tailed tropicbird | *Phaethon lepturus* |
| Gruiformes | Gruidae | Cranes | Grey crowned crane | *Balearica regulorum* |
| Opisthocomiformes | Opisthocomidae | Hoatzin | Hoatzin (aka Stinkbird) | *Opisthocomus hoazin* |
| Cuculiformes | Cuculidae | Cuckoos | Common cuckoo | *Cuculus canorus* |
| Columbiformes | Columbidae | Doves | Pigeon | *Columba livia* |
| Phoenicopteriformes | Phoenicopteridae | Flamingos | American flamingo | *Phoenicopterus ruber* |
| Galliformes | Phasianidae | Landfowl | Chicken | *Gallus gallus* |
| Anseriformes | Anatidae | Waterfowl | Pekin duck | *Anas peking* |
| Piciformes | Picidae | Woodpeckers | Downy woodpecker | *Dryobates pubescens* |
| Accipitriformes | Cathartidae | New World vultures | Turkey vulture | *Cathartes aura* |
| Pelecaniformes | Ardeidae | Herons | Little egret | *Egretta garzetta* |
| Pelecaniformes | Threskiornithidae | Ibis | Crested ibis | *Nipponia nippon* |
| Caprimulgiformes | Trochilidae | Hummingbirds | Anna's hummingbird | *Calypte anna* |
| Tinamiformes | Tinamidae | Tinamous | White-throated tinamou | *Tinamus guttatus* |
| Struthioniformes | Struthionidae | Ostrich | Common ostrich | *Struthio camelus* |

(continued)

**Table 2** (continued)

| Order | Family | Common group | Common species | Species |
|---|---|---|---|---|
| Apterygiformes | Apterygidae | Kiwi | North Island brown kiwi | *Apteryx mantelli* |
| Apterygiformes | Apterygidae | Kiwi | Little spotted kiwi | *Apteryx owenii* |
| Apterygiformes | Apterygidae | Kiwi | Rowi | *Apteryx rowi* |

Protein coding sequences were retrieved from the North Island brown kiwi genome (Le Duc et al. 2015) and reference blood transcriptomes for rowi and little spotted kiwi (Ramstad et al. 2016). Ortholog alignments for all other species listed are from the Avian Genome Consortium (Zhang et al. 2014a) available at http://avian.genomics.cn/en/jsp/database.shtml

alignments (Dunning et al. 2017). Only bird orthologs with all three kiwi species were considered, and if a kiwi ortholog was represented by more than one sequence for a species, then the longest sequences were retained for downstream analysis. Finally, the alignment procedure was repeated, with an additional TCS residue filtering (Chang et al. 2014) step so that only the highest confidence residues were retained, before Gblocks trimming and phylogeny inference (Dunning et al. 2017).

We tested for positive selection in each of the bird orthologs by optimizing several codon models (site and branch-site models) using codeml (Yang 1997). For each gene, the best-fit model was identified by comparing the corrected Akaike information criterion (AICc) of the null model which assumes no positive selection (M1a), a branch-site model that assumes a shift to relaxed selection on a defined foreground branch (model BSA), and a branch-site model that assumes a shift to positive selection on the foreground branch (BSA1). We defined several different foreground branches to test (base of kiwi, NIB, and LSK; NIB and rowi; LSK and rowi; NIB; LSK; rowi) and included a kiwi clade model to allow for bouts of recurrent selection during the diversification of this lineage. We rejected the null model if the ΔAICc score for the best fit model was more than 5.22 units less than the M1a model. A ΔAICc threshold of 5.22 is equivalent to a *P*-value threshold of 0.01 for a likelihood ratio test comparing the models using 2 degrees of freedom (df). Finally, we tested for a significant shift to positive selection if the null model could be rejected by preforming a likelihood ratio test (df = 1) between the two branch-site models, with *P*-values <0.01 considered significant. We preformed gene ontology over representation analysis using the PANTHER webserver (Mi et al. 2009) with the complete biological process set for the chicken.

## 3.2 Genes Under Selection in the Kiwi Lineage

Out of the 8,295 bird orthologs from the Avian Genome Consortium, 3,774 had orthologs in all 3 kiwi datasets and were used in subsequent analyses (Fig. 3a). The null model was not rejected for 3,548 of these orthologs, while 149 rejected the null

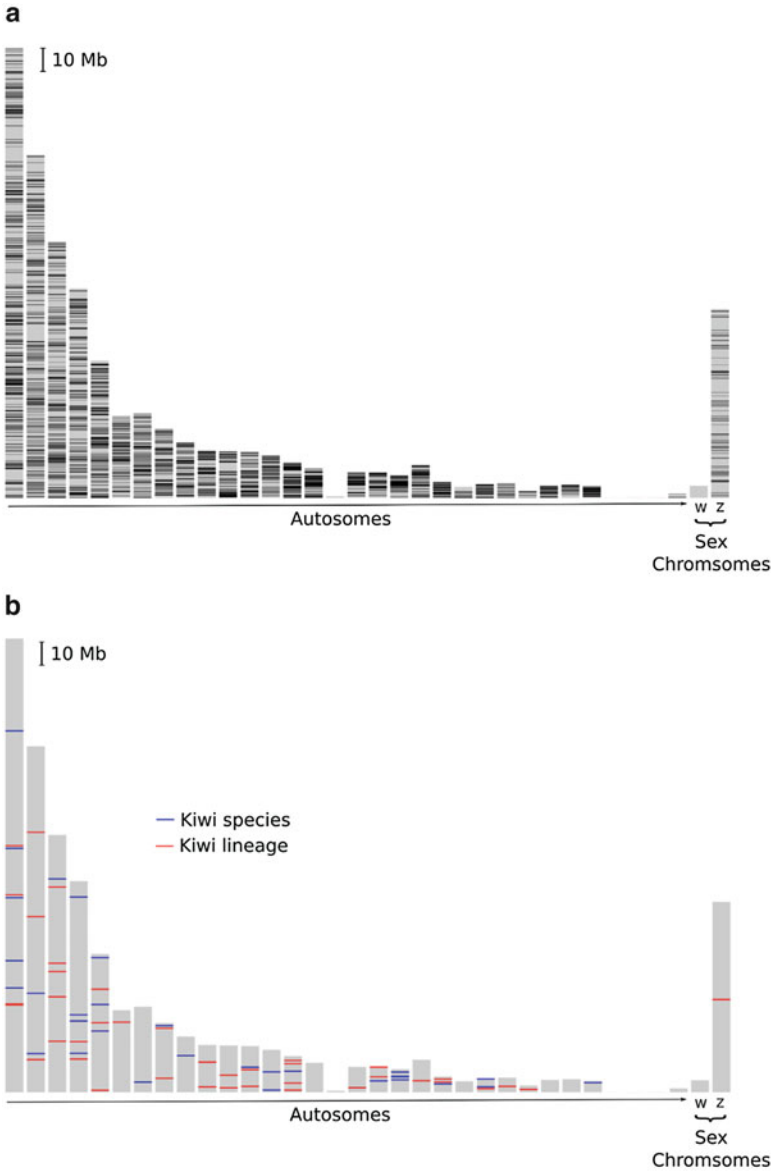


**Fig. 3** Highly conserved genome synteny among birds means that we can use the well-annotated chicken genome as a reference. (**a**) Positive selection signatures were assessed at 3,774 orthologous protein coding sequences (black lines) spread throughout the genome and identified in 28 avian species that include 3 species of kiwi (NIB, rowi, and LSK) and represent 23 avian orders. (**b**) A total of 77 genes show signatures of positive selection in nearly every chromosome both among kiwi species and between kiwi and other paleognath lineages

**Table 3** Number and functions of genes under positive selection in kiwi

| Lineage | Number of orthologs | Overrepresented gene ontologies |
|---|---|---|
| Kiwi branch | 30 | Endochondral bone morphogenesis, inflammatory response, response to stress, neurogenesis |
| Kiwi clade (i.e., recurrent selection) | 12 | T cell proliferation, negative regulation of gene expression (epigenetic), Notch receptor processing |
| LSK and NIB | 4 | Regulation of cell morphogenesis |
| LSK and rowi | 5 | Nucleotide-excision repair, DNA damage recognition |
| Brown kiwi clade (NIB and rowi) | 6 | Response to UV, cardiac muscle tissue morphogenesis |
| Spotted kiwi clade (LSK) | 5 | Diet-induced thermogenesis, cerebral cortex development |
| NIB | 11 | Positive regulation of glycolytic process, ATP metabolic process |
| Rowi | 4 | Epithelial cell morphogenesis, activation of innate immune response |

Branch-site codon models were used for 3,774 genes to identify those evolving under positive selection, and a comparison of corrected Akaike information criterion (AICc) values was used to determine the branch of the phylogeny on which this occurred

model but did not show significant positive selection. A signal of positive selection was detected in the remaining 77 genes, which are found on nearly every kiwi chromosome and spread apparently randomly throughout the genome (Fig. 3b). The majority of orthologs showing significant positive selection occurred on the branch separating kiwi from other species ($n = 42$), and 12 of these genes exhibited recurrent selection in the kiwi clade (signal present in both the lineage leading to kiwi and among kiwi species; Table 3). Positive selection was found in six genes in brown kiwi (NIB and rowi) and another five kiwi genes in LSK suggesting divergent selection between the brown and spotted kiwi clades. However, a similar number of genes showed positive selection in both LSK and rowi but not in NIB (five genes) and in both LSK and NIB but not in rowi (four genes). Another 11 and 4 genes showed positive selection exclusively within NIB and rowi, respectively.

The genes under selection are functionally diverse and include those associated with bone morphogenesis, stress and immune response, neurogenesis, metabolism, retinal development, and longevity (Table 3). Among the 42 genes showing positive selection in the branch leading to kiwi, we found several genes associated with small body size, short limbs, and skeletal abnormalities. For example, the thyroid hormone receptor (*TRIP11*) gene causes achondrogenesis type IA (ACG), an autosomal recessive disorder that can result in congenital chondrodysplasia (malformation of bones and cartilage; Parwanto 2017). Individuals with ACG are characterized by small body size, short limbs, and skeletal abnormalities. The protein encoded by the aggrecan (*ACAN*) gene is an integral part of the extracellular matrix in cartilaginous tissue. Mutations in this gene are associated with disturbed chondroskeletal development resulting in skeletal dysplasia or dwarfism (Dateki 2017).

The evolution of Apterygidae is also associated with positive selection in genes associated with stress resistance. The eukaryotic translation initiation factor (*EIF2AK1*) gene is involved in downregulating protein synthesis in response to oxidative stress, heme deficiency, osmotic shock, and heat shock (Krishna and Kumar 2018). The telomere maintenance 2 (*TELO2*, also known as *Tel2P* or *CLK2*) gene encodes a protein involved in cellular resistance to stressors that cause DNA damage, such as ionizing and ultraviolet radiation (Hurov et al. 2010). *TELO2* is also involved in telomere length regulation and, potentially, aging (Hurov et al. 2010; Lim et al. 2001). We also found signals of positive selection in genes that influence the effects of acetylcholine in the central and peripheral nervous system (cholinergic receptor, *CHMR5*; Hurov et al. 2010; Lim et al. 2001; Yamada et al. 2001) as well as genes associated with insulin regulation of metabolism (forkhead box K1, *FOXK1*; Sakaguchi et al. 2018), retinal degeneration (Bardet-Biedl syndrome 10, *BBS10*; Álvarez-Satta et al. 2017), and neurogenesis (nicastrin, *NCSTN*; Xie et al. 2014).

### 3.3 Genes Under Selection in Brown and Spotted Kiwi Clades

Selection signals found in six genes in the brown kiwi clade (NIB and rowi) are associated with immune response, metabolism, and neural development. The transactivator (*CIITA*) gene is essential for MHC class II gene expression. Mutations in this gene cause bare lymphocyte syndrome, where the immune system is severely compromised and cannot effectively fight infection (Steimle et al. 1993), as well as pathogenesis of several lymphomas (Steidl et al. 2011). Two of the genes identified are associated with metabolism of vitamin B (*THNSL1* B6, *ABCD4* B12) and thus normal brain development and proper functioning of nervous and immune systems. We also found evidence of positive selection among brown kiwi in the *WDR81* gene, which encodes a transmembrane protein predominantly expressed in the brain. Mutations in this gene are associated with severe progressive photoreceptor loss and the autosomal recessive Uner Tan syndrome, a degenerative disease of the nervous system (Traka et al. 2013). The *GEMIN4* and *UBE4B* (ubiquitination factor E4B) genes code for proteins involved in multiple pathological processes and several forms of cancer (Liu et al. 2012; Zhang et al. 2014b).

Signals of positive selection in the spotted kiwi clade suggest selection associated with temperature. The genes under selection in spotted kiwi were the monocarboxylate transporter *SLC16A3* gene, the *TTF2* transcription termination factor gene, the *TRPV1* (transient receptor potential cation channel) gene, and the *NIPAL3* and *TACC3* genes. *TRPV1* is also known as the capsaicin receptor gene and is involved in detecting scalding heat and pain and regulating body temperature (Xu et al. 2007). Also, one of the gene ontology terms overrepresented among these genes was diet-induced thermogenesis or the amount of energy expended above the basal metabolic rate due to the cost of processing food. The two spotted kiwi species inhabit locales with very different temperature regimes. Little spotted

kiwi have historically been restricted to lowland areas. In contrast, GSK are the only kiwi that occupy elevations up to 1,500 m above sea level and can survive in areas where soils can be frozen or covered in snow for days (McLennan and McCann 2002). Thus, our data suggests that selection for coping with highly divergent temperature regimes may be a defining feature of the spotted kiwi clade. This hypothesis should be tested in the future by broadly comparing LSK and GSK diversity at the *TRPV1* and other candidate genes associated with thermal tolerance.

### *3.4 What Makes a Kiwi a Kiwi?*

There is surely a great deal of important functional diversity that we did not detect in our tests as we only looked at a subset of protein coding genes. In particular, the use of blood samples limits us to genes expressed in this tissue; we could not test for selection in the opsin genes, for example. Finally, our data cannot differentiate between historical and recent selection and certainly cannot predict what genes will be important in the future. These data do, however, provide insight into the genomic signature associated with the evolution of Apterygidae and indicate candidate genes that could now be surveyed more broadly in kiwi to improve our understanding of kiwi evolution and management.

Recent genomic studies inform the ongoing debate of why kiwi have such a small body size relative to other ratites and such large eggs relative to their body size. The phyletic dwarfing hypothesis explained this pattern as being due to the kiwi ancestor having a large body size with a proportionately large egg and kiwi experiencing a subsequent reduction in body size, but not egg size, after arrival in New Zealand (Calder 1979; Cracraft 1974; Gould 1986). This made sense when kiwi were thought to be most closely related to moa or emu and cassowary (large-bodied ratites with proportionate eggs) and to have attained their present distribution via vicariance. The alternative view was that regardless of how kiwi arrived in New Zealand, they were small-bodied ancestrally and then evolved the large egg, presumably because there was an advantage to producing highly precocial young to avoid avian predators in New Zealand (e.g., aerial raptors, including the Haast eagle, and terrestrial *Aptornis* or adzebills, Worthy et al. 2013).

Genomic results suggest both hypotheses may be correct. These data show that kiwi's closest relatives are elephant birds (Mitchell et al. 2014) which, like kiwi, have disproportionately large egg size relative body size (Dickison 2007). Also, our data show positive selection among kiwi in genes associated with skeletal morphology and dwarfism. Together, these findings suggest that kiwi have disproportionately large eggs because they inherited the allometry from a common ancestor shared with elephant birds, but also experienced changes in body size after dispersal to New Zealand. Predation could have maintained a large egg size (precocial chicks) while also promoting small body size (to allow the use of burrows for protection). Competition with moa would likely also have promoted small body size to reduce niche overlap as suggested by Mitchell et al. (2014).

# 4 Conservation Impacts and Research Needed

## *4.1 Can Genomics Enhance Conservation of Kiwi?*

Kiwi exemplify a number of genomic difficulties often encountered when working with small populations. For example, they have experienced significant genetic bottleneck effects (Miller et al. 2011; Ramstad et al. 2010, 2013), hybridization (Ramstad, unpublished data), and extreme inbreeding (Taylor et al. 2017) and exhibit patterns of isolation by distance (White et al. 2018), sex-biased dispersal, and strong social organization (Ramstad, unpublished data). Below we describe how genomics can be applied to some of these issues in conservation of kiwi and small populations more broadly.

***Genetic Population Structure and Connectivity*** In small populations, drift overwhelms gene flow and selection in defining the genetic population structure of the species. Genomic data can provide more precise estimates of divergence between populations ($F_{ST}$) than limited genetic data from a handful of markers (typically microsatellites) and identify populations that have previously gone undetected (Luikart et al. 2019; Oyler-McCance et al. 2019). In kiwi, numerous small and extant subpopulations have recently been identified, some that almost certainly represent new species, and will require specific management in the future (Weir et al. 2016; White et al. 2018).

In addition, fine-scale genetic structuring is possible within populations (Beck et al. 2008; Kanno et al. 2011). Such structure can be sex, age, or stage biased and may indicate important social or geographic barriers to gene flow. There is now evidence of fine-scale genetic structure in the last remnant population of the rarest kiwi species, rowi, which exhibits sex-biased dispersal and isolation by distance within their 14 km spatial range (Ramstad, unpublished data). Genomic research will help refine these findings, allow us to look for similar patterns in other kiwi species, and significantly enhance our understanding of what population units and dispersal behaviors are important to conserve in kiwi.

***Genetic Bottleneck Effects and*** $\mathbf{N}_e$ Effective population size ($N_e$) is a critical parameter for understanding the magnitude and effects of drift, and therefore the rate of genetic erosion, in small populations. Many populations with large census size are functionally small from a conservation genomics perspective due to high reproductive skew, overlapping generations, differences in timing of reproduction between sexes, or an isolation by distance pattern of genetic structure (Neel et al. 2013; Waples et al. 2013, 2014). Genomics can be used in combination with advances in theory and statistical techniques, such as coalescent models and approximate Bayesian computation, to provide reliable estimates of current and historical $N_e$ and changes in population size (Salmona et al. 2019). For example, Nunziata et al. (2017) used RADseq data to estimate changes in population size in two salamander species (*Ambystoma* spp.) and found a pattern that agreed with 37 years of population size estimates based on mark-recapture data.

Likewise, genomic data and coalescent-based demographic models can be used to estimate the timing and magnitude of genetic bottleneck effects. In kiwi, genomic data resolved a significant genetic bottleneck effect and subsequent explosive diversification associated with glaciation (Weir et al. 2016). Microsatellite data suggest that genetic bottleneck effects and low $N_e$ are common and sometimes recent in kiwi (Ramstad, unpublished data; Ramstad et al. 2013; Taylor 2014). Genomic techniques would improve estimates of the timing and magnitude of bottlenecks, as well as provide more robust estimates of current $N_e$ for many small populations of kiwi. For example, the ratio between $N_e$ and census size ($N_C$) of Kapiti Island LSK was found to be exceptionally low for terrestrial vertebrates (0.03 using 15 microsatellite loci) and suggests that genetic diversity might still be eroding in this population, despite its large census size (Ramstad et al. 2013). The accuracy of this estimate is unclear, however; it may be downwardly biased due to residual gametic disequilibrium from the Kapiti Island founding event, overlapping generations, population subdivision, social structure, or simply an explosive increase in $N_C$ but not $N_e$ in this closed population (Ramstad et al. 2013).

***The Inbreeding Effect of Small Population Size*** When populations shrink, the likelihood of inbreeding, or mating with a close relative, increases (Crow and Kimura 1970). Many small populations would benefit from estimates of inbreeding coefficients ($F$) to aid in selecting unrelated founders for translocation and mating pairs for captive breeding programs. $F$ is typically estimated from a pedigree, but these are notoriously difficult to construct (Taylor et al. 2015), especially in small populations that have low genetic diversity. Unless all individuals in a population are sampled over many generations, then estimates of $F$ based on pedigrees ($F_P$) are often highly imprecise and downwardly biased (Taylor et al. 2015). Many populations cannot be sampled to near completion, making it essentially impossible to produce a robust pedigree or precise and unbiased estimates of inbreeding coefficients for small populations in the wild. The advent of genomics makes it possible to precisely measure $F$ of any individual without a pedigree (Kardos et al. 2016). $F$ can be estimated for both unmapped (using the diagonal elements of a genomic relatedness matrix) and mapped (the proportion of the genome that includes runs of homozygosity) loci (Kardos et al. 2016). The latter can measure $F$ virtually without error because an individual can reliably be scored as heterozygous or homozygous at nearly every position in the genome (Kardos et al. 2016). Indeed, marker-based estimates predict $F$ better than $F_P$ in recently bottlenecked and partially isolated small populations ($N_e = 75$, Kardos et al. 2015). In some cases, $F$ estimated with as few as 1,000 SNPs was more closely correlated with multilocus heterozygosity than $F_P$ estimated from a 20-generation pedigree (Kardos et al. 2015).

While strong inbreeding is known to occur in some kiwi populations (Taylor et al. 2017), no pedigrees have been built to date for any naturally occurring kiwi population. A pedigree based on 30 microsatellite loci is underway for rowi (Ramstad, unpublished data) and is possible only because nearly the entire extant population has been genotyped (Taylor 2014, 2015). Rowi are actively managed through captive incubation and translocation of founders to predator-free

sanctuaries. Reliable estimates of inbreeding coefficients and relatedness among birds will help managers choose founders for these new populations and avoid inbreeding in rowi. More precise estimates of $F$ based on runs of homozygosity would make it easier to detect inbreeding depression and to identify loci making a large contribution to inbreeding depression. This technique will no doubt become commonplace in conservation genomics once genome resequencing is occurring more routinely.

***Mating Systems and Social Organization*** Kiwi are thought to be monogamous, but a robust test of this assumption is lacking (but see Ziesemann 2011), and extra pair fertilization has been detected in many other bird species that were thought to be monogamous (Wink and Dyrcz 1999). Very little is known of natal dispersal in kiwi, how territories and pairs are established, and the potential for kin recognition. There is, however, clearly variation among kiwi in their mating and social systems. For example, LSK and NIB kiwi live solely in pairs, while small groups of tokoeka often occupy and fiercely defend a single territory. Robust genomic testing of relatedness among individual kiwi relative to their spatial distribution and behavior could elucidate the social and mating systems of the individual species. This information is critical for establishing new populations via translocation, selecting individuals for captive incubation and rearing programs, and effectively reintroducing captive-reared juveniles to the populations they were collected from as eggs.

***Hybridization and Outbreeding Depression*** Another potential threat to kiwi is hybridization, which can result in genomic extinction of species that have been reduced to one or few small populations. In some cases, however, hybridization is an important source of genetic diversity for what would otherwise be genetically depauperate species (Caniglia et al. 2018; Pimm et al. 2006). Genomic sequencing has greater power to differentiate between hybridization and incomplete lineage sorting than traditional genetic approaches (vonHoldt et al. 2018). For example, it is very difficult to determine the taxonomy of "species" with 0.5–2% genomic divergence (Roux et al. 2016). However, genomic data can provide a robust estimate of the fraction of the genome derived from admixture or hybridization between closely related species and the specific genes that have been introgressed (Wayne and Shaffer 2016).

Among kiwi, hybrids have been detected between GSK, LSK, and rowi based on mtDNA and nuclear DNA (Ramstad, unpublished data). To date, four F1 LSK x rowi hybrids have been found in the last remaining remnant population of rowi. As a precautionary measure, the hybrids were taken from Okarito sanctuary to an isolated kiwi-free island by the New Zealand Department of Conservation. The hybrids subsequently produced chicks, confirming that F1 kiwi hybrids are able to survive and successfully reproduce.

The historical frequency of hybridization among kiwi is unknown, and so it is unclear if current levels of hybridization are elevated due to anthropomorphic disturbance and small population size. The ~5 million years of divergence between the brown (rowi) and spotted (LSK) kiwi clades suggests that hybridization could

result in outbreeding depression (Allendorf et al. 2001; Frankham et al. 2011). Likewise, if hybridization were allowed to continue in Okarito sanctuary, the world's last remnant population of rowi could be lost to genomic extinction (Allendorf et al. 2001). However, kiwi hybrids have conservation value because they are the only extant source of otherwise extinct alleles (Allendorf et al. 2001; Chan et al. 2019; Frankham et al. 2011) and may aid in genetic rescue of highly inbred populations and improve our understanding of hybridization dynamics in kiwi.

***Adaptive and Functional Diversity*** Long-term persistence of populations relies on their having significant reservoirs of additive genetic variance (Jamieson and Allendorf 2012). This genetic diversity allows populations to survive challenges that impact their survival and reproduction, such as disease. Comparative genomic studies are elucidating signals of adaptive and functional diversity between taxa at various phylogenetic levels (Kober and Pogson 2017) and even within extinct species (Feigin et al. 2018). For example, a full genome sequence was obtained from a preserved specimen of Tasmanian tiger (*Thylacinus cynocephalus*) and used to place the species in a phylogeny of carnivorous marsupials and examine the genomic basis of its phenotypic convergence with other canids (Feigin et al. 2018).

It is not uncommon for populations that cannot be differentiated at neutral genetic loci to show adaptive divergent at functional loci due to local adaptation. For example, cryptic population structure was found in Atlantic cod (*Gadus morhua*; Berg et al. 2016) and sea scallop (*Placopecten magellanicus*; Van Wyngaarden et al. 2017) via genomic analysis of loci under selection. Indeed, significant adaptive genetic variation may not be protected if management units are based on overall or neutral genetic population structure. All that is currently known of potential adaptive genetic diversity in kiwi is outlined in this paper. Robust, hypothesis-driven studies of adaptive variation have the potential to greatly improve our understanding and management of kiwi.

## *4.2 Challenges and Opportunities*

Genomics is a young field and there are still numerous issues to be resolved in its application to species conservation. With the large number of markers now being tested simultaneously, it is unclear how best to correct for multiple tests and false positives are a concern (Johnson et al. 2010; Kober and Pogson 2017). It can also be difficult to assess independence among loci unless markers can be mapped to a reference genome. It remains important to detect outlier loci and deal with them separately based on the focal questions. Also, it is often difficult to assign meaning to or understand the effects of genes that show evidence of positive selection. Traditional statistical analyses can be challenging or useless with genomic datasets comprised of SNPs (e.g., tests for genetic bottleneck effects that rely on highly polymorphic loci; Cornuet and Luikart 1996; Luikart et al. 1998). However, new

statistical tests are being developed and the rich body of population genetics theory remains foundational. Effectively conveying the value of these techniques to the public broadly, including relevant findings and limitations, will also be a challenge.

Regardless of the difficulties listed above, genomic techniques are the new gold standard in conservation and should be taken up immediately to improve management of imperiled taxa. Critically small populations that are actively managed or manipulated should have a genomic management plan that it based on proper taxonomy and tracks individuals. Such plans help managers maintain genomic diversity via effective selection of founders for new populations and captive breeding programs, track reproductive success of and relatedness among individuals, monitor levels of inbreeding, and respond to demographic crises (e.g., disease outbreaks). Such a genomic management plan is urgently needed for kiwi. Five kiwi species are currently recognized but there could be as many as 11 (Weir et al. 2016; White et al. 2018). Understanding kiwi social structure will allow managers to maximize chick and sub-adult survival by defining windows for their capture and release. Identifying important functional variation among kiwi help will allow managers to weigh the costs and likelihood of outbreeding versus inbreeding depression and thus assess the conservation value of hybrids. A central repository for kiwi genomic data and samples would vastly improve monitoring, translocation, and captive incubation programs for this ancient avian lineage.

The shift from genetics to genomics will be a challenge for researchers and managers. However, reduced representation techniques, such as Rapture and RADcap, are excellent starting points for wading into the field. These techniques allow for discovery and reliable capture of thousands of SNPs in hundreds of individuals simultaneously and can be applied to numerous conservation questions. They are also relatively fast, straightforward, and cost-effective techniques and are being employed with increasing frequency to non-model species (Ali et al. 2016; Hoffberg et al. 2016). More genomic resources are forthcoming, including whole-genome sequences, transcriptome assemblies, new molecular techniques, and new analytical tools, and will provide a foundation for an unprecedented amount of highly powerful research. The challenge now is to bring the full weight of innovative genomics to the management and conservation of at-risk species broadly.

**Acknowledgments** Thank you to the editors for the invitation to contribute to this book and guidance in preparing this chapter. Many thanks to Hugh Robertson and Rogan Colbourne, collaborators, and Kiwi Whisperers extraordinaire, for collection of and access to archived kiwi samples and for sharing their invaluable insight into kiwi ecology and management. Thanks also to Helen Taylor and Duncan Kaye for many helpful kiwi conversations. Funding for this work was provided by the Allan Wilson Centre for Molecular Ecology and Evolution (KMR), the University of South Carolina (KMR), the Natural Environment Research Council (NE/M00208X/1, LTD), and an ERC grant (ERC-2014-STG-638333, LTD). Permission for sample collection and genetic research and cultural guidance in working with tāonga was generously provided by Kaitiaki o Kapiti Trust, Te Rūnanaga Ngāi Tahu, Port Nicholson Block Settlement Trust, Te Ātiawa Manawhenua Ki Te Tau Ihu Trust, Te Rūnanga o Makaawhio, and Waiorua Bay Trust.

## References


Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics. 2016;202:389–400.

Allendorf FW. Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol. 2017;26:420–30.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. Trends Ecol Evol. 2001;16:613–22. https://doi.org/10.1016/S0169-5347(01)02290-X.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

Álvarez-Satta M, Castro-Sánchez S, Valverde D. Bardet-Biedl syndrome as a chaperonopathy: dissecting the major role of chaperonin-like BBS proteins (BBS6-BBS10-BBS12). Front Mol Biosci. 2017;4:55.

Amadon D. An estimated weight of the largest known bird. Condor. 1947;49:159–64.

Andrews KR, Luikart G. Recent novel approaches for population genomics data analysis. Mol Ecol. 2014;23:1661–7. https://doi.org/10.1111/mec.12686.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81.

Beck N, Peakall R, Heinsohn R. Social constraint and an absence of sex-biased dispersal drive fine-scale genetic structure in white-winged choughs. Mol Ecol. 2008;17:4346–58.

Benestan LM, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol Ecol. 2016;25:2967–77.

Berg PR, et al. Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. Sci Rep. 2016;6:23246.

Bernatchez L, et al. Harnessing the power of genomics to secure the future of seafood. Trends Ecol Evol. 2017;32:665–80.

Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP. Strong mitochondrial DNA support for a cretaceous origin of modern avian lineages. BMC Biol. 2008;6:6.

Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, Drummond A, Cooper A. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. Nature. 2003;425:172–5.

Calder WA. The kiwi and egg design: evolution as a package deal. Bioscience. 1979;29:461–7.

Caniglia R, et al. Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfdog. BMC Genomics. 2018;19:533.

Castro I, Cunningham SJ, Gsell AC, Jaffe K, Cabrera A, Liendo C. Olfaction in birds: a closer look at the kiwi (Apterygidae). J Avian Biol. 2010;41:213–8.

Chan WY, Hoffmann AA, van Oppen MJ. Hybridization as a conservation management tool. Conserv Lett. 2019;12:e12652.

Chang J-M, Di Tommaso P, Notredame C. TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. Mol Biol Evol. 2014;31:1625–37.

Çilingir FG, Rheindt FE, Garg KM, Platt K, Platt SG, Bickford DP. Conservation genomics of the endangered Burmese roofed turtle. Conserv Biol. 2017;31:1469–76.

Cloutier A, Sackton TB, Grayson P, Edwards SV, Baker AJ. First nuclear genome assembly of an extinct moa species, the little bush moa (*Anomalopteryx didiformis*). bioRxiv. 2018; https://doi.org/10.1101/262816.

Cloutier A, Sackton TB, Grayson P, Clamp M, Baker AJ, Edwards SV. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. Syst Biol. 2019;68:937–55.

Colbourne R, Bassett S, Billing T, McCormick H, McLennan J, Nelson A, Robertson H. The development of Operation Nest Egg as a tool in the conservation management of kiwi. Wellington: Department of Conservation; 2005.

Corfield J, Gillman L, Parsons S. Vocalisations of the North Island brown kiwi (*Apteryx mantelli*). Auk. 2008;125:326–35.
Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 1996;144:2001–14.
Cracraft J. Phylogeny and evolution of the ratite birds. Ibis. 1974;116:494–521.
Crow JF, Kimura M. An introduction to population genetics theory. New York: Harper and Row; 1970.
Dateki S. ACAN mutations as a cause of familial short stature. Clin Pediatr Endocrinol. 2017;26:119–25.
Davies SJJF. Struthioniformes (Tinamous and Ratites). In: Hutchins M, Jackson JA, Bock WJ, Olendorf D, editors. Grzimek's animal life encyclopedia. Birds I Tinamous and Ratites to Hoatzins, vol. 8. 2nd ed. Farmington Hills: Gale Group; 2003. p. 56–105.
Dickison MR. The allometry of giant flightless birds. Durham: Duke University; 2007.
Dunning LT, et al. Introgression and repeated co-option facilitated the recurrent emergence of C4 photosynthesis among close relatives. Evolution. 2017;71:1541–55.
Edmands S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol. 2007;16:463–75. https://doi.org/10.1111/j.1365-294X.2006.03148.x.
Edmonds H. Taxon plan for northern and southern Fiordland tokoeka (*Apteryx australis australis*): strategic plan for the recovery of northern and southern Fiordland tokoeka, for the period 2015–2025 and beyond. Te Anau: Department of Conservation; 2015.
Feigin CY, et al. Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. Nat Ecol Evol. 2018;2:182.
Fowler ME. Comparative clinical anatomy of ratites. J Zoo Wildl Med. 1991;22:204–27.
Frankham R, Ballou JD, Eldridge MDB, Lacy RC, Ralls K, Dudash MR, Fenster CB. Predicting the probability of outbreeding depression. Conserv Biol. 2011;25:465–75.
Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489–96.
Funk WC, Zamudio KR, Crawford AJ. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_61.
Germano J, et al. Kiwi Recovery Plan 2018–2028 Mahere Whakaora Kiwi 2018–2028. Wellington: New Zealand Department of Conservation; 2018.
Gould SJ. Of kiwi eggs and the Liberty Bell. Nat Hist. 1986;95:20–9.
Grealy A, et al. Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. Mol Phylogenet Evol. 2017;109:151–63.
Harshman J, et al. Phylogenomic evidence for multiple losses of flight in ratite birds. Proc Natl Acad Sci U S A. 2008;105:13462–7.
Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 2017;11:1189–204.
Heather BD, Robertson HA. The field guide to the birds of New Zealand. Auckland: Viking; 2005.
Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, Glenn TC. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour. 2016;16:1264–78.
Hoffman JI. Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome. Mol Ecol Resour. 2011;11:703–10. https://doi.org/10.1111/j.1755-0998.2011.02999.x.
Hoffman J, Thorne M, Trathan P, Forcada J. Transcriptome of the dead: characterisation of immune genes and marker development from necropsy samples in a free-ranging marine mammal. BMC Genomics. 2013;14:52. https://doi.org/10.1186/1471-2164-14-52.
Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010a;6:e1000862.

Hohenlohe PA, Phillips PC, Cresko WA. Using population genomics to detect selection in natural populations: key concepts and methodological considerations. Int J Plant Sci. 2010b;171:1059–71.
Holzapfel SA, Robertson HA, McLennan JA, Sporle W, Hackwell K, Impey M. Kiwi (*Apteryx* spp) recovery plan 2008–2018. Wellington: Department of Conservation; 2008.
Hurov KE, Cotta-Ramusino C, Elledge SJ. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. Genes Dev. 2010;24:1939–50.
Huxley TH. On the classification of birds and on the taxonomic value of the modifications of certain of the cranial bones observable in that class. Proc Zool Soc London. 1867;415–72.
IUCN. The IUCN Red List of Threatened Species. Version 2019-2. 2019. http://www.iucnredlist.org. Accessed 18 July 2019.
Jamieson IG, Allendorf FW. How does the 50/500 rule apply to MVPs? Trends Ecol Evol. 2012;27:578–84.
Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics. 2010;11:724.
Johnson RN, et al. Adaptation and conservation insights from the koala genome. Nat Genet. 2018;50:1102.
Jolly JN, Daugherty CH. Comparison of little spotted kiwi (*Apteryx owenii*) from Kapiti and D'Urville Islands. Wellington: Department of Conservation; 2002.
Jordan G, Goldman N. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol Biol Evol. 2011;29:1125–39.
Kanno Y, Vokoun JC, Letcher BH. Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. Mol Ecol. 2011;20:3711–29.
Kardos M, Luikart G, Allendorf F. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. 2015;115:63.
Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. Evol Appl. 2016;9:1205–18.
Kinsky F. The consistent presence of paired ovaries in the kiwi (*Apteryx*) with some discussion of this condition in other birds. J Ornithol. 1971;112:334–57.
Kober KM, Pogson GH. Genome-wide signals of positive selection in strongylocentrotid sea urchins. BMC Genomics. 2017;18:555.
Krishna KH, Kumar MS. Molecular evolution and functional divergence of eukaryotic translation initiation factor 2-alpha kinases. PLoS One. 2018;13:e0194335.
Kummrow MS. Ratites or Struthioniformes: Struthiones, Rheae, Cassuarii, Apteryges (ostriches, rheas, emus, cassowaries, and kiwis), and Tinamiformes (tinamous). In: Miller RE, Fowler ME, editors. Fowler's zoo and wild animal medicine, vol. 8. St. Louis: Elsevier/Saunders; 2014. p. 75–82.
Le Duc D, et al. Kiwi genome provides insights into evolution of a nocturnal lifestyle. Genome Biol. 2015;16:1–15.
Li R, et al. The sequence and *de novo* assembly of the giant panda genome. Nature. 2010;463:311.
Lim C-S, Mian IS, Dernburg AF, Campisi J. *C. elegans* clk-2, a gene that limits life span, encodes a telomere length regulator similar to yeast telomere binding protein Tel2p. Curr Biol. 2001;11:1706–10.
Liu J, et al. Genetic variants in the microRNA machinery gene GEMIN4 are associated with risk of prostate cancer: a case-control study of the Chinese Han population. DNA Cell Biol. 2012;31:1296–302.
Liu J, Ding Q-X, Gao L-Z. The complete mitochondrial genome of North Island brown kiwi (*Apteryx mantelli*). Mitochondrial DNA Part B. 2017;2:1–2.
Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered. 1998;89:238–47.

Luikart G, Kardos M, Hand B, Rajora OP, Aitken S, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 3–79.
Maderspacher F. Evolution: flight of the ratites. Curr Biol. 2017;27:R110–3.
Margres MJ, Jones ME, Epstein B, Kerlin DH, Comte S, Fox S, Fraik AK, et al. Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer. Mol Ecol. 2018;27:4189–99.
Martchenko D, Prewer E, Latch EK, Kyle CJ, Shafer ABA. Population genomics of ungulates. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG, part of Springer Nature; 2018. https://doi.org/10.1007/13836_2018_30.
Martin GR. Sensory capacities and the nocturnal habit of owls (Strigiformes). Ibis. 1986;128:266–77.
Mayr G. Paleogene fossil birds. Berlin: Springer; 2009.
McCall R, Nee S, Harvey P. The role of wing length in the evolution of avian flightlessness. Evol Ecol. 1998;12:569–80. https://doi.org/10.1023/A:1006508826501.
McLennan JA, McCann T. Genetic variability, distribution, and abundance of great spotted kiwi (*Apteryx haastii*). Wellington: Department of Conservation; 2002.
McLennan J, et al. Role of predation in the decline of kiwi, *Apteryx* spp., in New Zealand. N Z J Ecol. 1996;20:27–35.
McLennan J, Dew L, Miles J, Gillingham N, Waiwai R. Size matters: predation risk and juvenile growth in North Island brown kiwi (*Apteryx mantelli*). N Z J Ecol. 2004;28:241–50.
Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res. 2009;38:D204–10.
Miles J, Potter M, Fordham R. Northern brown kiwi (*Apteryx australis mantelli*) in Tongariro National Park and Tongariro Forest – ecology and threats, vol. 51. Wellington: Department of Conservation; 1997.
Miller HC, Bowker-Wright G, Kharkrang M, Ramstad KM. Characterisation of class II B MHC genes from a ratite bird, the little spotted kiwi (*Apteryx owenii*). Immunogenetics. 2011;63:223–33. https://doi.org/10.1007/s00251-010-0503-7.
Miller H, Biggs P, Voelckel C, Nelson N. De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (*Sphenodon punctatus*). BMC Genomics. 2012;13:439.
Mitchell KJ, et al. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. Science. 2014;344:898–900.
Neel MC, et al. Estimation of effective population size in continuously distributed populations: there goes the neighborhood. Heredity. 2013;111:189–99.
Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. Mol Ecol. 2017;26:1060–74.
Ogawa A, Murata K, Mizuno S. The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. Proc Natl Acad Sci. 1998;95:4415–8.
Oyler-McCance SJ, Oh KP, Zimmerman SJ, Aldridge CL. The transformative impact of genomics on sage-grouse conservation and management. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer Nature Switzerland AG; 2019. https://doi.org/10.1007/13836_2019_65.
Palstra FP, Ruzzante DE. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Mol Ecol. 2008;17:3428–47.
Parkes KC, Clark GA. An additional character linking ratites and tinamous, and an interpretation of their monophyly. Condor. 1966;68:459–71.

Parwanto MLE. The genetic aspect and morphological appearance of achondrogenesis. Int J Reprod Contracept Obstet Gynecol. 2017;6:3203–12.
Paton T, Haddrath O, Baker AJ. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc R Soc Lond B Biol Sci. 2002;269:839–46.
Pereira SL, Baker AJ. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol. 2006;23:1731–40.
Phillips MJ, Gibb GC, Crimp EA, Penny D. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. Syst Biol. 2010;59:90–107. https://doi.org/10.1093/sysbio/syp079.
Pimm SL, Dollar L, Bass OL. The genetic rescue of the Florida panther. Anim Conserv. 2006;9:115–22.
Prince DJ, et al. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Sci Adv. 2017;3:e1603198.
Prinzinger R, Dietz V. Pre-and postnatal energetics of the North Island brown kiwi (*Apteryx mantelli*). Comp Biochem Physiol A Mol Integr Physiol. 2002;131:725–32.
Pross J, et al. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. Nature. 2012;488:73.
Ramstad KM, Pfunder M, Robertson HA, Colbourne RM, Allendorf FW, Daugherty CH. Fourteen microsatellite loci cross-amplify in all five kiwi species (*Apteryx* spp) and reveal extremely low genetic variation in little spotted kiwi (*A. owenii*). Conserv Genet Resour. 2010;2:333–6. https://doi.org/10.1007/s12686-010-9233-2.
Ramstad KM, Colbourne RM, Robertson HA, Allendorf FW, Daugherty CH. Genetic consequences of a century of protection: serial founder events and survival of the little spotted kiwi (*Apteryx owenii*). Proc R Soc B Biol Sci. 2013;280:20130576. https://doi.org/10.1098/rspb.2013.0576.
Ramstad KM, Miller HC, Kolle G. Sixteen kiwi (*Apteryx* spp) transcriptomes provide a wealth of genetic markers and insight into sex chromosome evolution in birds. BMC Genomics. 2016;17:410.
Robertson HA, Colbourne RM. Survival of little spotted kiwi (*Apteryx owenii*) on Kapiti Island. Notornis. 2004;51:161–3.
Roux C, Fraisse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 2016;14:e2000234.
Ryder O. Conservation genomics: applying whole genome studies to species conservation efforts. Cytogenet Genome Res. 2005;108:6–15.
Sackton TB, et al. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. bioRxiv. 2018; https://doi.org/10.1101/262584.
Sæther B-E, Bakke Ø. Avian life history variation and contribution of demographic traits to the population growth rate. Ecology. 2000;81:642–53.
Sakaguchi M, Cai W, Wang C-H, Batista TM, Araki E, Kahn CR. FoxK1/K2 are new, important components of IR and IGF1R signaling and control of cell proliferation and metabolism. Am Diabetes Assoc. 2018;67:356.
Sales J. The endangered kiwi: a review. Folia Zool. 2005;54:1–20.
Salmona J, Heller R, Lascoux M, Shafer A. Inferring demographic history using genomic data. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 511–37.
Shafer AB, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol. 2015;30:78–87.
Shafer A, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JB. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol. 2016;8:907–17.
Shen Y, Song R, Pe'er I. Coverage tradeoffs and power estimation in the design of whole-genome sequencing experiments for detecting association. Bioinformatics. 2011;27:1995–7.

Steidl C, et al. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature. 2011;471:377.
Steimle V, Otten LA, Zufferey M, Mach B. Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). Cell. 1993;75:135–46.
Steiner CC, Putnam AS, Hoeck PE, Ryder OA. Conservation genomics of threatened animal species. Annu Rev Anim Biosci. 2013;1:261–81.
Subramanian S, Huynen L, Millar C, Lambert D. Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. BMC Evol Biol. 2010;10:387.
Taylor HR. Detecting inbreeding depression in a severely bottlenecked, recovering species: the little spotted kiwi (*Apteryx owenii*). Wellington: Victoria University of Wellington; 2014.
Taylor HR. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. Ecol Evol. 2015;5:3140–50.
Taylor HR, Kardos MD, Ramstad KM, Allendorf FW. Valid estimates of individual inbreeding coefficients from marker-based pedigrees are not feasible in wild populations with low allelic diversity. Conserv Genet. 2015;16:901–13. https://doi.org/10.1007/s10592-015-0709-1.
Taylor HR, Colbourne RM, Robertson HA, Nelson NJ, Allendorf FW, Ramstad KM. Cryptic inbreeding depression in a growing population of a long-lived species. Mol Ecol. 2017;26:799–813.
Tennyson AJD, Palma RL, Robertson HA, Worthy TH, GILL BJ. A new species of kiwi (Aves, Apterygiformes) from Okarito, New Zealand. Records Auckland Institute Museum. 2003;40:55–64.
Traka M, Millen KJ, Collins D, Elbaz B, Kidd GJ, Gomez CM, Popko B. WDR81 is necessary for purkinje and photoreceptor cell survival. J Neurosci. 2013;33:6834–44.
Van Wyngaarden M, et al. Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. Evol Appl. 2017;10:102–17.
Vega R, Vázquez-Domínguez E, White TA, Valenzuela-Galván D, Searle JB. Population genomics applications for conservation: the case of the tropical dry forest dweller *Peromyscus melanophrys*. Conserv Genet. 2017;18:313–26.
vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY. Redefining the role of admixture and genomics in species conservation. Conserv Lett. 2018;11:e12371.
Wan Q-H, et al. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. Cell Res. 2013;23:1091.
Waples RS, Luikart G, Faulkner JR, Tallmon DA. Simple life-history traits explain key effective population size ratios across diverse taxa. Proc R Soc B Biol Sci. 2013;280:20131339. https://doi.org/10.1098/rspb.2013.1339.
Waples RS, Antao T, Luikart G. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics. 2014;197:769–80.
Wayne RK, Shaffer HB. Hybridization and endangered species protection in the molecular era. Mol Ecol. 2016;25:2680–9.
Weir JT, Haddrath O, Robertson HA, Colbourne RM, Baker AJ. Explosive ice age diversification of kiwi. Proc Natl Acad Sci. 2016;113:E5580–7.
White DJ, et al. Discovery and complete genome sequence of a novel circovirus-like virus in the endangered rowi kiwi, *Apteryx rowi*. Virus Genes. 2016;52:727–31.
White DJ, Ramón-Laca A, Amey J, Robertson HA. Novel genetic variation in an isolated population of the nationally critical Haast tokoeka (*Apteryx australis* 'Haast') reveals extreme short-range structure within this cryptic and flightless bird. Conserv Genet. 2018;19:1401–10.
Wink M, Dyrcz A. Mating systems in birds: a review of molecular studies. Acta Ornithol. 1999;34:91–109.
Worthy T, Worthy JP, Tennyson AJ, Salisbury SW, Hand SJ, Scofield RP. Miocene fossils show that kiwi (*Apteryx*, Apterygidae) are probably not phyletic dwarves. In: Göhlich UB, Kroh A, editors. Paleornithological research 2013 (Proceedings of the 8th International Meeting of the Society of Avian Palaeontology and Evolution). Vienna: Verlag Naturhistorisches Museum; 2013. p. 63–80.

Xie T, et al. Crystal structure of the γ-secretase component nicastrin. Proc Natl Acad Sci. 2014;111:13349–54. https://doi.org/10.1073/pnas.1414837111.

Xu H, Tian W, Fu Y, Oyama TT, Anderson S, Cohen DM. Functional effects of nonsynonymous polymorphisms in the human TRPV1 gene. Am J Physiol Renal Physiol. 2007;293:F1865–76.

Xu L, Sin SYW, Grayson P, Janes DE, Edwards SV, Sackton TB. Evolutionary dynamics of sex chromosomes of palaeognathous birds. bioRxiv. 2018; https://doi.org/10.1101/295089.

Yamada M, et al. Cholinergic dilation of cerebral blood vessels is abolished in M5 muscarinic acetylcholine receptor knockout mice. Proc Natl Acad Sci. 2001;98:14096–101.

Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13:555–6.

Yonezawa T, et al. Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of the ratites. Curr Biol. 2017;27:68–77.

Zhang G, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014a;346:1311–20.

Zhang Y, Lv Y, Zhang Y, Gao H. Regulation of p53 level by UBE4B in breast cancer. PLoS One. 2014b;9:e90154.

Zhou Q, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 2014;346:1246338.

Ziesemann B. The social organisation and mating system for the brown kiwi (*Apteryx mantelli*): a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Ecology at Massey University, Albany, New Zealand. Palmerston North: Massey University; 2011.

# The Transformative Impact of Genomics on Sage-Grouse Conservation and Management


**Sara J. Oyler-McCance, Kevin P. Oh, Shawna J. Zimmerman, and Cameron L. Aldridge**

**Abstract** For over two decades, genetic studies have been used to assist in the conservation and management of both Greater Sage-grouse (*Centrocercus urophasianus*) and Gunnison Sage-grouse (*C. minimus*), addressing a wide variety of topics including taxonomy, parentage, population connectivity, and demography. The field of conservation genetics has been transformed by dramatic improvements in sequencing technology, facilitating genomic studies in many wildlife species. The quality and amount of data generated by genomic methods vastly exceed that of traditional genetic studies, allowing for increased precision in estimating genetic parameters of interest. Perhaps more importantly, genomic methods can provide insight into non-neutral evolution such as adaptive divergence. Here we recount the shift from genetic to genomic methods using two wildlife species of substantial conservation interest, focusing on the improved capabilities and advantages of genomic methods. For instance, reassessment of divergence in sage-grouse using genomic methods confirmed strong differentiation between the two species and revealed that a small population in the state of Washington was more genetically distinct than previously recognized. Further, new genomic resources and approaches have been used to identify a family of genes linked to local dietary adaptation suggesting that sage-grouse may possess digestive and metabolic adaptations that mitigate the effects of consuming plant secondary metabolites like those found in sagebrush. Genetic variation among populations in these gene regions is thought to be involved with local dietary adaptations, and therefore maintaining the tie between sage-grouse and the chemistry of local sagebrush may be an important management consideration. We posit that the integration of newly developed genomic resources combined with the vast wealth of ecological and behavioral data for sage-grouse has

S. J. Oyler-McCance (✉) · K. P. Oh
U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO, USA
e-mail: soyler@usgs.gov

S. J. Zimmerman · C. L. Aldridge
Natural Resource Ecology Laboratory and Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA

Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)], https://doi.org/10.1007/13836_2019_65, 

the potential to shed light on mechanistic relationships that ultimately are vital to the conservation and management of these species.

**Keywords** Adaptive genetic variation · *Centrocercus* · Conservation genetics · Landscape genetics · Whole-genome sequencing

# 1 Introduction

## *1.1 Background*

Molecular genetic methods (i.e., those methods involving a small number of anonymous and presumed neutral markers) have been used to address a wide variety of conservation and management issues for both Greater Sage-grouse (*Centrocercus urophasianus*) and Gunnison Sage-grouse (*C. minimus*). Both species have been well-studied from a genetic perspective with over 25 peer-reviewed publications in the past 20 years. The earliest research began in the mid-1990s examining taxonomy and distinct populations (Kahn et al. 1999; Oyler-McCance et al. 1999; Young et al. 2000; Benedict et al. 2003). Since those early papers, the range in topics tackled by genetic studies has been broad, examining questions ranging from lek formation and mating system (Gibson et al. 2005; Semple et al. 2001; Bird et al. 2012) to detecting gene flow and identifying landscape features impacting population connectivity (Bush et al. 2011; Oyler-McCance et al. 2005a, b; Cross et al. 2016; Row et al. 2015). Dramatic improvements in DNA sequencing technology (Mardis 2008; Shendure and Ji 2008; Metzker 2010) have facilitated the ability to collect genomic data for virtually any organism and the ability to parse anonymous versus putative adaptive genetic variation (Allendorf et al. 2010; Luikart et al. 2019; Hohenlohe et al. 2019). Such genomic approaches have recently been applied to sage-grouse, broadening our understanding about evolutionary history, current genomic structure, and potential adaptation – all of which are important for successful management and conservation. The aim of this chapter is to discuss the 20-year progression of molecular studies on Greater and Gunnison Sage-grouse, highlighting the expanded capabilities and advantages of genomic approaches and considering future research directions.

## *1.2 Conservation Status of Sage-Grouse*

The distributions of both Greater and Gunnison Sage-grouse have contracted considerably across North America since the time of European settlement (Fig. 1). As of 2004, Greater Sage-grouse were thought to occupy roughly 56% of their historical distribution, while Gunnison Sage-grouse occupy only 10% (Schroeder et al. 2004).

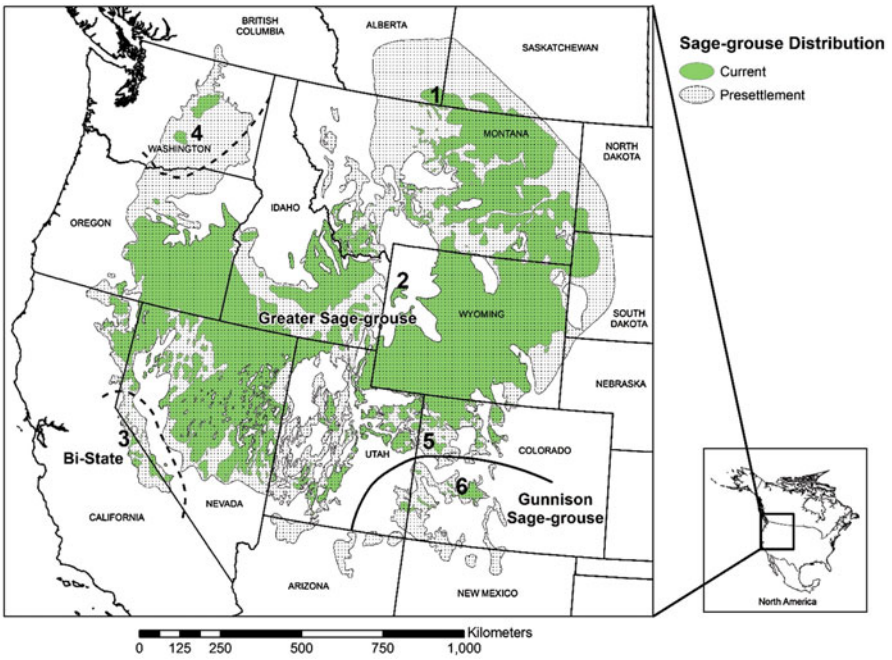


**Fig. 1** Current and presettlement distribution of sage-grouse (modified from Schroeder et al. 2004). The boundary for the Bi-State population of Greater Sage-grouse as well as the Washington populations is delineated by dotted lines, whereas the boundary for the Gunnison Sage-grouse distribution is delineated by a solid line. The numbers represent sampling locations for the whole-genome resequencing study (1, Alberta; 2, Jackson Hole; 3, Bi-State; 4, Washington; 5, Piceance Basin; 6, Gunnison Sage-grouse) of Oh et al. (2019)

The causes of range contraction vary in different parts of the ranges yet likely involve habitat fragmentation, degradation, and loss associated with agriculture, resource extraction, livestock grazing, fencing, powerlines, invasive plants, and changes in the fire cycle (Connelly and Braun 1997; Braun 1998; Oyler-McCance et al. 2001; Knick et al. 2003; Connelly et al. 2004; Green et al. 2017; Monroe et al. 2017). Despite significant range contraction, the Greater Sage-grouse persists across much of the remaining western North American landscape that is dominated by sagebrush (*Artemisia* spp.). Some populations (e.g., in the states of Washington and Utah and the Jackson Hole population in the state of Wyoming) are small and isolated, while others persist in relatively continuous habitat (Fig. 1). Conservation and management efforts often cross state and federal boundaries and, by necessity, focus on large-scale processes. Both species have been petitioned to be listed under the US Endangered Species Act, with Gunnison Sage-grouse currently listed as threatened under US law (USFWS 2014) and Greater Sage-grouse listed as endangered under the Canadian Federal Species at Risk Act (Environment Canada 2014) for the northernmost populations in Alberta and Saskatchewan. The threat of potential listing has facilitated a plethora of studies on habitat requirements, population

trends, impacts of management actions, and causes of decline, particularly for Greater Sage-grouse.

## *1.3 Sagebrush Specialist*

Sage-grouse are considered sagebrush obligate species (Beever and Aldridge 2011), depending on sagebrush throughout their entire life cycle (Patterson 1952). They require sagebrush for cover and nesting, and while they forage on sagebrush throughout the year, they rely on it exclusively for food in the winter months (Patterson 1952; Dalke et al. 1963; Wallestad et al. 1975; Braun et al. 1976; Connelly et al. 2000; Young et al. 2000). There are six main species and subspecies of sagebrush that are important for sage-grouse (described in Connelly et al. 2000), and their occurrence varies widely across the landscape due to differences in climate, soil type, topography, and disturbance (West 1983; Miller et al. 2011). Sagebrush leaves contain high levels of plant secondary metabolites (PSMs) such as monoterpenes (Kelsey et al. 1982) that act as a defense against herbivory by inhibiting digestive enzymes in herbivores (Kohl et al. 2015). Different varieties of sagebrush have distinct combinations and concentrations of toxins that vary across the landscape (Frye et al. 2013). Sage-grouse are dietary specialists and consume sagebrush leaves selectively, targeting leaves with higher nutrient content and lower concentration of PSMs (Remington and Braun 1985; Welch et al. 1988; Frye et al. 2013) and have coevolved mechanisms to deal with the inhibition of digestive enzymes associated with PSMs (Kohl et al. 2015).

## *1.4 Mating System*

Both species of sage-grouse have a polygynous mating system that has been the focus of numerous studies over many decades (Wiley 1974; Wittenberger 1978; Gibson and Bradbury 1986; Bergerud 1988; Gibson et al. 1991; Young et al. 2000). In the spring, males congregate on leks, where they engage in an elaborate strutting display to attract females. Males establish territories on leks and defend them throughout the breeding season (Gibson and Bradbury 1986). Behavioral observations suggest that females arrive at leks later in the breeding season and typically mate with one of the dominant males on the lek (Wiley 1974; Gibson and Bradbury 1986; Gibson et al. 1991). Thus, reproductive success is highly variable among males, with a small proportion of males monopolizing all matings, which has important implications for management and conservation. Such highly skewed mating success among males implies strong sexual selection which can lead to rapid changes in morphology and behavior that can facilitate divergence and speciation (Ellsworth et al. 1994; Uy and Borgia 2000; Panhuis et al. 2001; Spaulding 2007; Oyler-McCance et al. 2010), processes that favor the formation of

evolutionarily significant units that are important to identify and protect. Further, this mating skew decreases the overall genetic diversity and effective population size of sage-grouse (Stiver et al. 2008), which may be important for surviving and adapting to future stressors such as novel diseases or environmental change. Importantly, lekking behavior and the leks themselves provide a predictable time and location for biologists to count, monitor, and sample sage-grouse for research.

## *1.5 Why Sage-Grouse Are Good Candidates for Genetic and Genomic Research*

Unlike many species of conservation concern, sage-grouse have been closely monitored for decades as they are game birds that have been sought by hunters and are therefore actively managed by governmental wildlife agencies. Research and monitoring efforts by these agencies have produced a wealth of data on population trends and habitat needs (see volume edited by Knick and Connely 2011), and concern over listing has prompted further research to better understand threats to sage-grouse and their habitat. This resulted in an abundance of information regarding sage-grouse populations, habitats, and threats, providing an extensive baseline into which molecular data can be integrated. The collection of samples for genetic analyses has been relatively straightforward as wings from hunter-killed grouse are collected each fall by most state agencies to determine demographic information, and DNA can be extracted from the muscle tissue of those wings (Oyler-McCance et al. 1999, 2005a; Benedict et al. 2003). In addition, blood collected from birds in radio telemetry studies has served as a good source of DNA (Oyler-McCance et al. 2005b, 2014; Bush et al. 2011). More recently, DNA extracted from feathers and fecal pellets collected noninvasively on leks during the breeding season has successfully been used in genetic analyses (Bush et al. 2005, 2010; Baumgardt et al. 2013; Row et al. 2015; Cross et al. 2016; Shyvers et al. 2019). While many genomic methods (e.g., whole-genome or reduced representation (re)sequencing) require relatively large quantities of high molecular weight DNA from tissue or blood (see Oyler-McCance et al. 2016 for a discussion of DNA quantity and quality in genomic applications), a few (e.g., targeted sequence capture; for review, see Jones and Good (2016)) have successfully generated genome-wide SNP markers using low-quality DNA samples such as from preserved museum specimens (Bi et al. 2013), suggesting that feather or fecal-derived DNA may be suitable for some genomic applications. Finally, sage-grouse are closely related to two agriculturally important galliform species (domestic turkey [*Meleagris gallopavo*; Dalloul et al. 2010] and chicken [*Gallus gallus*; International Chicken Genome Sequencing Consortium 2004]), thereby providing extensive genomic resources, including functional and structural genomic datasets and experimental validation that can serve as invaluable resources for assigning putative gene functions to sage-grouse orthologs.

## 2 How Traditional Genetics and the Shift to Genomics Help Conserve Sage-Grouse

### *2.1 Delineating Units for Conservation and Management*

Historically, sage-grouse were considered to be one species. Research in the 1990s revealed dramatic morphological (Hupp and Braun 1991) and behavioral (Young et al. 1994) differences between sage-grouse in southwestern Colorado and southeastern Utah compared to the rest of the range, raising the possibility that this group of sage-grouse may be a new species. Genetic analyses using mitochondrial DNA and microsatellite loci were then employed to explore genetic differentiation between sage-grouse in northern and southern Colorado (Kahn et al. 1999; Oyler-McCance et al. 1999), comparing across the purported species boundary located within Colorado. These studies revealed a lack of gene flow between the two morphologically and behaviorally distinct groups of sage-grouse in Colorado, consistent with the idea that sage-grouse in southwestern Colorado were a distinct species. This new species was subsequently named, described (Young et al. 2000), and recognized by the American Ornithologists' Union (2000). Further, these studies revealed that the newly described Gunnison Sage-grouse had much less genetic diversity than was found within Greater Sage-grouse in northern Colorado (Kahn et al. 1999; Oyler-McCance et al. 1999).

The recognition of the Gunnison Sage-grouse as a separate species led to the renaming of all other sage-grouse as Greater Sage-grouse and a correction to its range distribution (Fig. 1). Within the revised large range of the Greater Sage-grouse, the species had historically been divided into two taxonomic groups; an eastern (*C. u. urophasianus*) and a western (*C. u. phaios*) subspecies (Aldrich 1946). This delineation was based on plumage and coloration differences in 11 individual Greater Sage-grouse collected from Washington, Oregon, and California (Aldrich 1946). The western subspecies presumably occurred in southern British Columbia (Aldridge and Brigham 2003), central Washington, east-central Oregon, and northeastern California (Aldrich 1946), although Aldrich and Duvall (1955) considered the birds in California to be intermediate. Populations in other areas of the range were considered to be the eastern subspecies. The validity of the subspecies distinction was later questioned (Johnsgard 1983). Using genetic techniques similar to those used to evaluate the validity of the Gunnison Sage-grouse, Benedict et al. (2003) and Oyler-McCance et al. (2005a) examined the subspecific boundary and found no genetic evidence to support the original subspecies distinction (Fig. 1). Instead, they found several populations that were notable for other reasons. Along the border between Nevada and California, the "Bi-State" population (alternately referred to as "Lyon/Mono") was found to be unusual, with mitochondrial DNA sequences largely unique compared to the rest of the range (Benedict et al. 2003). Further examination of the Bi-State population revealed that, unlike Gunnison Sage-grouse, they are neither morphologically nor behaviorally distinct from other Greater Sage-grouse (Taylor and Young 2006; Schroeder 2008). Interestingly, Benedict

et al. (2003) also found that the two populations in Washington contained the lowest level of haplotype diversity observed (with one of the few haplotypes a common, widespread haplotype), perhaps resulting from a recent genetic bottleneck given that these populations now occupy just 8–10% of their original range and have shown significant declines in population size (Schroeder et al. 2000).

Managing populations with unique genetic diversity could be extremely important if those populations are to be conserved. Benedict et al. (2003) suggested that the unique allelic composition of the Bi-State population might be of particular importance for conservation. Since the likelihood that distinctiveness of anonymous genetic markers extends to genes under adaptive selection, they suggested this population should be managed independently, avoiding translocation of other Greater Sage-grouse into this area. They also surmised that the probable loss of genetic variation in Washington should be addressed, recommending that translocation of birds from neighboring populations may be justified to ensure continued persistence of the populations in Washington (Benedict et al. 2003). A subsequent study spanning the species' entire range using both mitochondrial DNA and nuclear microsatellites came to similar conclusions (Oyler-McCance et al. 2005a).

Compared to other Greater Sage-grouse populations, the Bi-State population has a somewhat similar amount of genetic divergence (based on anonymous neutral markers) as the Gunnison Sage-grouse, yet it lacks the morphological and behavioral differences present between the two species (Taylor and Young 2006; Schroeder 2008; Oyler-McCance et al. 2014). This has led to lingering confusion over the taxonomic status of the Bi-State population. A shift to genomic markers has helped to resolve this taxonomic uncertainty. Using a reduced representation approach (RAD-Seq), Oyler-McCance et al. (2015a) identified over 11,000 single-nucleotide polymorphisms (SNPs) among three groups: Gunnison Sage-grouse, Bi-State, and the southern portion of the range of Greater Sage-grouse. Contrary to previous findings with microsatellites and mitochondrial DNA, they found much higher differentiation between Gunnison and Greater Sage-grouse than within Greater Sage-grouse (e.g., Bi-State population versus populations in the remainder of the species' range). When each SNP site was mapped onto the chicken genome, the most highly divergent SNPs between Greater and Gunnison Sage-grouse were located on the Z chromosome (sex-determining macrochromosome in birds), and for both species, genetic diversity on the Z chromosome was reduced compared to autosomes (Oyler-McCance et al. 2015b). Greater divergence on the Z chromosome could be the result of selection (including sexual selection) or genetic drift associated with a genetic bottleneck related to the speciation event. These recent findings highlight the added value of genomic approaches, which help to better characterize patterns of genetic variation in sage-grouse and add insights into the mechanisms underlying speciation in these birds.

In light of these studies, there is ongoing interest in better understanding the genetic distinctiveness of sage-grouse populations, particularly those with small populations that exist on the margins of the species range, often constrained to relatively isolated patches of suitable sagebrush habitat that may limit gene flow from neighboring populations. From a conservation perspective, an important

question is whether the patterns of genetic differentiation observed with anonymous markers in such populations have any functional genetic significance that might suggest local adaptation. While homology-based approaches provide a convenient option, species-specific genomic resources are necessary to fully characterize genetic variation observed. Recent efforts to enhance sage-grouse genomic studies have been bolstered by the de novo assembly of a high-coverage (ca. 170×) reference genome for Gunnison Sage-grouse (Oh et al. 2019). Comparative genomic analysis shows that 98% of scaffolds mapped with high confidence (e-value <1e-50) to the chicken genome, with all chicken chromosomes covered by three or more scaffolds. Preliminary genome annotation was also performed, utilizing both ab initio gene prediction and homology-based methods, resulting in a draft annotation containing 18,565 protein-coding genes. Taken together, this reference genome represents arguably the most comprehensive set of genomic resources available for a non-domesticated galliform species to date and was used to facilitate a whole-genome resequencing study (Oh et al. 2019) aimed at investigating anonymous and adaptive differentiation for several small, isolated, and potentially unique Greater Sage-grouse populations (Fig. 1): (1) at the northern extent of the contemporary species range in southeastern Alberta (Bush et al. 2011); (2) an isolated population near Jackson Hole, Wyoming (Schulwitz et al. 2014); (3) the Bi-State population (Benedict et al. 2003; Oyler-McCance et al. 2005a); and (4) a population in southern Washington (Benedict et al. 2003; Oyler-McCance et al. 2005a). Fifteen individual genomes were resequenced from each population, in addition to 15 samples from Greater Sage-grouse from the Piceance Basin in northwestern Colorado, which were expected to be more genetically representative of the largest populations occurring across relatively contiguous habitat of the Wyoming Basin (Oyler-McCance et al. 2005a). Fifteen Gunnison Sage-grouse samples were also included to evaluate previous interspecific genetic comparisons at a finer resolution. Analyses of approximately 1.5 million SNPs in the resulting dataset suggested distinct clustering by population (Fig. 2), with a largely hierarchical population structure, consistent with a pattern of postglacial recolonization from multiple refugia (Taberlet et al. 1998; Oyler-McCance et al. 2005a; Meirmans 2012). Evaluation of divergence at the whole-genome level (Oh et al. 2019) suggested greatest levels of differentiation at the interspecific level (mean pairwise $F_{ST}$ at autosomal SNPs for *C. urophasianus* × *C. minimus* populations = 0.460), largely corroborating previous results. Interestingly, comparisons among Greater Sage-grouse populations indicated relatively high levels of divergence in pairwise contrasts involving Washington (mean pairwise $F_{ST}$ at autosomal SNPs = 0.231) compared to the mean values among Greater Sage-grouse populations in the northeastern core of the species range (Alberta, Jackson Hole, and Piceance Basin: mean pairwise $F_{ST}$ at autosomal SNPs = 0.103). The Bi-State population also showed comparatively elevated levels of genome-wide differentiation (mean pairwise $F_{ST}$ at autosomal SNPs = 0.137). Importantly, because tests of population structure over relatively large geographic ranges can be biased by limited dispersal (i.e., isolation-by-distance), a partial Mantel test confirmed the evidence of genetic clustering, while controlling for interpopulation distance (Meirmans 2012). While evidence from

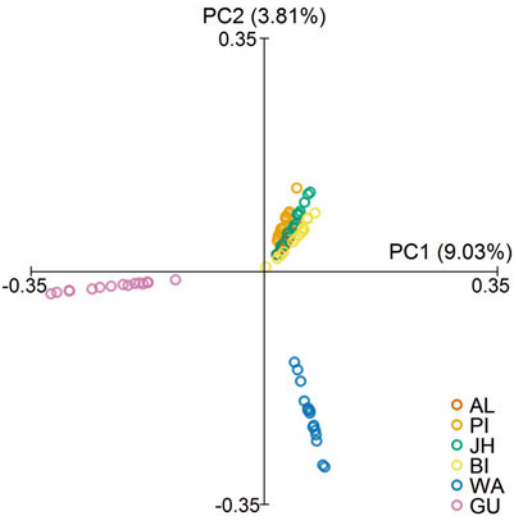


**Fig. 2** Principal component analysis of complete dataset representing five populations of Greater Sage-grouse (AL, Alberta, Canada; PI, Piceance Basin, Colorado; JH, Jackson Hole, Wyoming; BI, Bi-State population spanning the border between California and Nevada; WA, Washington) and the Gunnison Sage-grouse (GU in southwestern Colorado), based on 1,500,781 nuclear SNPs. Axes represent first (PC1) and second (PC2) principal components, with percentage of total genetic variance explained by each component shown in parentheses (Oh et al. 2019, reprinted with permission from Oxford University Press, Genome Biology and Evolution)

previous genetic analyses of Washington sage-grouse has been consistent with a history of isolation and dramatic reductions in population size (Oyler-McCance et al. 2005a), these results at the whole-genome level provide new quantitative evidence for greater genetic distinctiveness of Washington birds than previously appreciated, which likely has implications for management priorities (Oh et al. 2019).

Whole-genome sequences were also used to investigate historical demographic trends in both sage-grouse species. Utilizing the Gunnison Sage-grouse genome along with a reference genome for Greater Sage-grouse (sequenced to a moderate depth of ~27× and then aligned to the *C. minimus* reference), Oh et al. (2019) used the pairwise sequentially Markovian coalescent model (Li and Durbin 2011) to infer changes in effective population size ($N_e$) over the past ~4 million years. Both species showed evidence of declines in $N_e$ (Fig. 3), part of which coincides with the last glacial period in North America (c. 110,000–11,700 BP). However, while the Greater Sage-grouse genome revealed some evidence of population size stabilization (consistent with postglacial range expansion), the inferred $N_e$ for Gunnison Sage-grouse exhibited consistent decline, suggesting that the ancestral population to this species may have been demographically isolated from other sage-grouse populations and undergone a more severe bottleneck, perhaps contributing to initial population divergence and the speciation process (Oh et al. 2019).

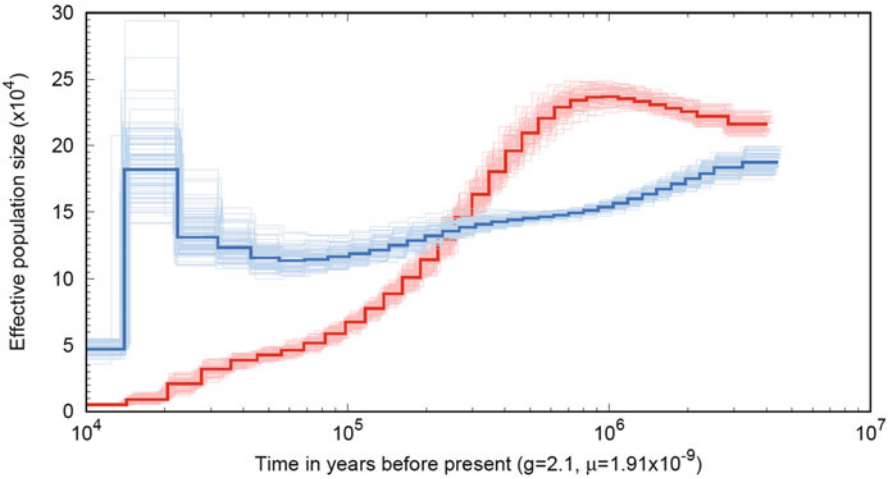


**Fig. 3** Inferred changes in ancestral effective population size for Greater Sage-grouse (blue) and Gunnison Sage-grouse (red) from Oh et al. (2019) (Reprinted with permission from Oxford University Press, Genome Biology and Evolution). Thick lines are median estimate from PSMC analysis of autosomes from a single individual of each species; lighter lines are from 100 bootstrap replicates. Values were scaled by generation time ($g$) (Stiver et al. 2008) and lineage-specific estimated mutation rate ($\mu$) (Nam et al. 2010). The median divergence time for the two species is estimated to be roughly 1.24 million years ago (range 0.58–1.64 million years ago, Kumar et al. 2017; Jetz et al. 2012)

## 2.2 *Population Connectivity and the Factors Influencing Gene Flow*

Dispersal of individuals has important effects on population dynamics and persistence, as well as patterns of diversity and population structure (e.g., Garant et al. 2005; Row et al. 2010, 2016; Fedy et al. 2017). Thus, documenting effective dispersal (i.e., dispersal that results in gene flow) across landscapes can inform management actions on how to improve or maintain population connectivity. Genetic studies have examined gene flow at both large (Oyler-McCance et al. 2005a, b; Cross et al. 2018; Row et al. 2018) and small scales (Bush et al. 2011; Oyler-McCance et al. 2014; Schulwitz et al. 2014; Davis et al. 2015; Cross et al. 2016; Row et al. 2016). A common theme that emerges from these studies is that sage-grouse follow an isolation-by-distance pattern where neighboring populations tend to be more closely related than those that are separated by larger geographic distances (i.e., gene flow occurs more readily among neighboring populations). Further, populations in discrete patches of habitat isolated from other populations in more continuous sagebrush (e.g., satellite populations of Gunnison Sage-grouse, Washington, Jackson Hole, Bi-State populations within Greater Sage-grouse) are less connected than populations in more contiguous habitat (Oyler-McCance et al. 2005a, b; Schulwitz et al. 2014).

While documenting levels of gene flow among populations is an important first step, understanding how different landscapes actually influence gene flow provides a logical progression and can be critical for management and prioritization of areas for protection. The field of landscape genetics combines landscape modelling and genetic data to better comprehend how landscape features influence gene flow across a given region (Manel et al. 2003; Storfer et al. 2007; Forester et al. 2018). Several studies have used microsatellite data to examine such relationships in both Greater and Gunnison Sage-grouse at vastly different extents (Shirk et al. 2015; Row et al. 2015, 2018; Zimmerman 2019). Landscape features that impact gene flow in Greater Sage-grouse are scale-dependent and vary across the range (Row et al. 2018). Similarly, the habitat composition also had a scale-dependent facilitation of gene flow for Gunnison Sage-grouse, with the presence of sagebrush habitat facilitating gene flow among populations and high-quality nesting habitat and a tall shrub component facilitating gene flow among leks within the largest population (Zimmerman 2019). In general, sage-grouse gene flow tends to be greatest in areas of high-quality breeding habitat, yet conifers, rugged terrain, and agriculture impeded gene flow in many areas (Shirk et al. 2015; Row et al. 2015, 2018; Zimmerman 2019). Thresholds can be identified for the amount of breeding habitat or other important variables (positive or negative) that might influence gene flow (see Row et al. 2018), providing guidance on how to best manage landscapes to promote connectivity and gene flow. Genomic methods have the potential to add precision to landscape genetic studies due to the large number of markers. For instance, Jahner et al. (2016) analyzed variation at 27,866 SNPs in 140 male Greater Sage-grouse in a small region in central Nevada and found that geographic distance and suitable habitat best predicted genetic differentiation. Landscape genetic studies have produced maps that depict the strength and redundancy of connectivity that can help inform conservation actions that maintain and restore functional connectivity for sage-grouse. The added precision from genomic studies could further refine such efforts (Forester et al. 2018). Moreover, genomic methods could greatly expand the types of landscape genetic research questions being asked for sage-grouse by including adaptive loci. For instance, ties between adaptive genetic loci and environmental gradients could be examined (Waits and Storfer 2016) and used to predict potential responses to changing habitats under differing climate change scenarios.

Genetic data have frequently been used to estimate diversity within and differentiation between populations. Although one key feature of genomic data is being able to evaluate functional genetic regions, using thousands of anonymous loci can increase the precision of population parameter estimates (Allendorf et al. 2010). For instance, Gunnison Sage-grouse samples have been used to compare population parameter estimates from two datasets, one composed of 22 microsatellite loci from 254 individuals across populations and another composed of 14,072 SNP loci from 60 individuals (a subset of the 254) across populations (Zimmerman et al. 2019b). Both datasets generally showed the same pattern of differentiation, diversity, and clustering, although the SNP data had some increased precision of estimates and identification of distinct groups, as expected (Fig. 4). However, this increased precision was not always realized with differentiation metrics (see $F_{ST}$; Fig. 4). As

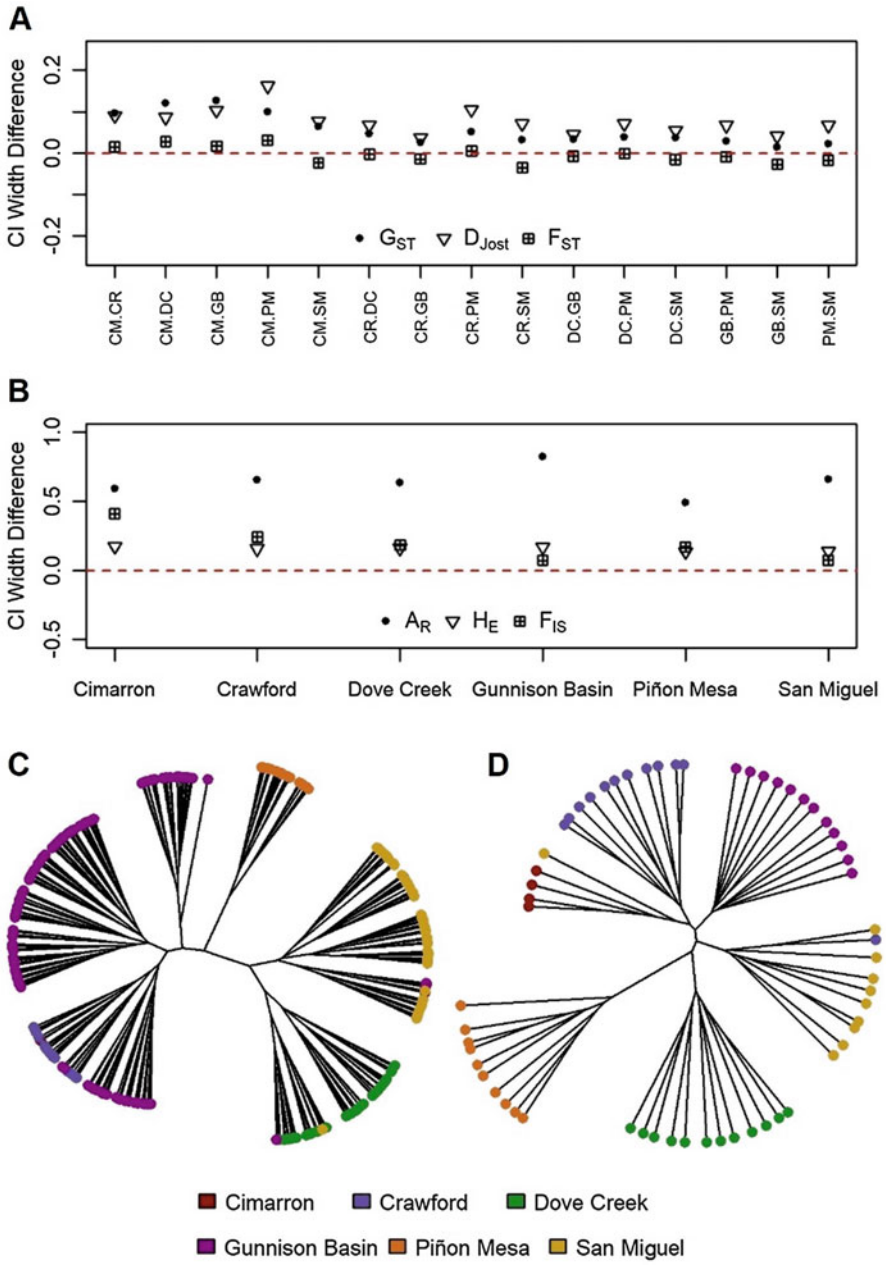


**Fig. 4** Increased precision in Gunnison Sage-grouse population genetic metrics for 14,072 putatively neutral SNP loci from 60 individuals versus 22 microsatellite loci from 254 individuals for multiple measures (Figure created from data presented in Zimmerman et al. 2019b). For (**a**) differentiation metrics ($F_{ST}$ (Weir and Cockerham 1984); $D_{Jost}$ (Jost 2008); $G_{ST}$ (Hedrick 2005); calculated in diveRsity R package (Keenan et al. 2013) with 1,000 bootstraps) and (**b**) diversity

other studies have demonstrated (Willing et al. 2012; Defaveri et al. 2013), precision in bootstrapped confidence intervals for pairwise differentiation is impacted by how many SNPs are used in combination with the number of individuals sampled for each population (more of both results in greater precision).

## 2.3 *Managing Genetic Diversity*

Small and isolated populations often face a risk of severe inbreeding and the resulting expression of deleterious recessive alleles. Additionally, for species with ranges that span diverse habitats, natural selection may promote highly localized adaptations that could warrant consideration when setting conservation priorities or contemplating certain management practices such as translocation or captive breeding. Thus, an important challenge for conservation genetics is to balance the maintenance of genetic diversity with the retention of potentially locally adapted genetic variants. The transition to genomic studies in sage-grouse conservation has provided important advances toward this goal.

Previous genetic analyses revealed low genetic diversity in Gunnison Sage-grouse satellite populations in comparison to the larger Gunnison Basin population (Oyler-McCance et al. 2005b). One of the management actions taken to mitigate both population size and genetic diversity was to translocate individuals from Gunnison Basin to satellite populations (Fig. 5). Recently, genetic data from 22 microsatellite loci were used to estimate change in diversity, differentiation, and population admixture among samples collected before, and 9 years after translocation efforts began (Zimmerman et al. 2019a). Satellite populations that received translocated birds all had increased genetic diversity, decreased genetic differentiation from the larger Gunnison Basin population, and showed signals of population admixture within individuals, indicating reproduction between Gunnison Basin transplants and resident satellite population birds. Though this work was completed using microsatellite loci, large numbers of anonymous loci from genomic techniques would likely identify finer signatures of change as a result of translocation. For one of the datasets, Zimmerman et al. (2019a) used a large number of noninvasively collected genetic samples, which were low in quality and unsuitable for many

**Fig. 4** (continued) metrics ($A_R$ = allelic richness, $H_E$ = expected heterozygosity, $F_{IS}$ = inbreeding coefficient; calculated in diveRsity R package), increased precision is illustrated as the difference in 95% confidence interval width for estimates calculated from microsatellites and SNPs (>0 indicates SNPs have a smaller width). Populations in pairwise comparisons for differentiation metrics (**a**) are abbreviated along the *x*-axis: *CM* Cimarron, *CR* Crawford, *DC* Dove Creek, *GB* Gunnison Basin, *PM* Piñon Mesa, *SM* San Miguel; *CM.CR* $F_{ST}$ between Cimarron and Crawford. Clustering approaches (hclust function in R, the complete method and with dissimilarity matrix (Nei and Kumar 2000) and 1,000 bootstraps) show a clear increase in precision of identifiable groups when using SNPs (**d**) as opposed to microsatellites (**c**)

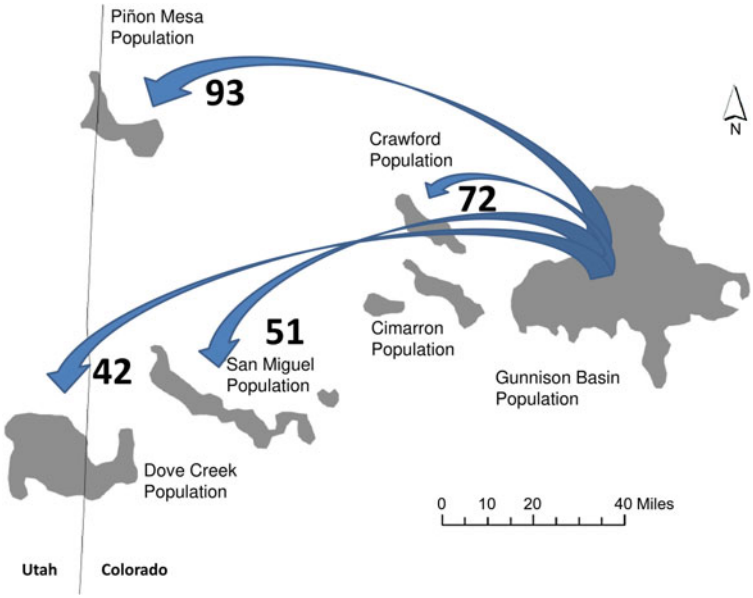


**Fig. 5** Range of Gunnison Sage-grouse in southwestern Colorado and southeastern Utah (modified from Zimmerman et al. 2019a). The largest (core) population is the Gunnison Basin. All other populations are considered to be satellite populations. The blue arrows represent translocation of Gunnison Sage-grouse from the Gunnison Basin to satellite populations, and the numbers represent the number of birds that were translocated between 2005 and spring of 2014 (USFWS 2014). Although birds were translocated into the Cimarron population, none of those birds survived. Analysis of pre- and post-translocation genetic data generally revealed increased genetic variation in the satellite populations and a decrease in differentiation between satellites and the Gunnison Basin population (Zimmerman et al. 2019a)

genomic techniques. Once anonymous loci are identified from high-quality samples, additional samples collected noninvasively could be used to continue tracking change as a result of the management action.

Despite relatively close geographic proximity, Gunnison Sage-grouse satellite populations have relatively high levels of genetic differentiation (Oyler-McCance et al. 2005b) with conspicuous differences in habitat quantity, quality, and ecological composition (GSRCC 2005) leading to the possibility of local adaptation across populations. Understanding the underlying genetic basis of such adaptations could be important for implementing conservation measures (Savolainen et al. 2013) such as translocations. The small effective population size of satellite populations could present a risk of translocations overwhelming any locally adapted alleles with variation from the larger Gunnison Basin. Zimmerman et al. (2019b) used 15,033 SNP genotypes in genomic outlier analyses, genotype-environment associations, and gene ontology (GO) enrichment analyses to examine patterns of putatively adaptive genetic differentiation in six of the Gunnison Sage-grouse populations.

A total of 411 loci linked to 289 putative genes associated with biological functions that were overrepresented in the assemblage of outliers were identified. Of particular interest was the identification of candidate SNPs linked to four genes which are members of the cytochrome P450 gene family (CYP4V2, CYP2R1, CYP2C23B, CYP4B1) which could indicate adaptive divergence for genes involved in sagebrush PSM metabolism and candidate loci that were linked to genes potentially involved in antiviral response (DEAD box helicase gene family and SETX) (Zimmerman et al. 2019b). Additionally, seven of the candidate SNPs corresponded to predicted non-synonymous amino acid substitutions in putative genes; this included two putative genes associated with antiviral response (DDX60 and SETX), as well as one (CYB5R4) that was previously associated with heat stress response (Zimmerman et al. 2019b; Zimmerman 2019).

Genomic methods have also been applied to investigate potentially adaptive genetic variation in small and isolated Greater Sage-grouse populations (Oh et al. 2019). Utilizing the whole-genome resequencing datasets (see above), a population genomics study was carried out to identify SNPs that bear the signature of selection (Oh et al. 2019). Briefly, the analysis utilized a Bayesian method that first estimates overall genetic covariance among populations and then identifies outlier loci that deviate from the expectations of this background population structure (Gautier 2015). This analysis identified 8,630 outlier SNPs that exhibited extreme differentiation among populations (i.e., exceeded 1% probability threshold). Of these, 147 SNPs occurred within exons of predicted protein-coding genes, with 50 identified as causing non-synonymous changes. Another 2,099 SNPs occurred within 5 kb up- or downstream regions of genes, thus suggesting potential regulatory effects on nearby genes. Gene ontology analysis of predicted genes associated with outlier SNPs revealed participation in diverse organismal functions, including spermatogenesis (HOOK1, MYCBP-associated protein) and immune function (CFI, GAB3), suggesting a potential role of sexual and pathogen-mediated natural selection, respectively, in shaping patterns of protein variation. In a parallel approach, the same study tested for evidence of positive natural selection on cytochrome P450 genes, along with a panel of candidate genes that are likely related to metabolism of PSMs, identified from pharmacological literature. Multiple genomic regions containing outlier SNPs that were associated with candidate genes related to the metabolism of xenobiotic compounds were identified, suggesting that inter-population variation could underlie consequential local dietary adaptations (Oh et al. 2019). These potential links between sage-grouse and the chemistry of the local sagebrush plants within which they reside are highly relevant to consider for conservation and management strategies. For instance, sagebrush restoration efforts could consider using local sagebrush material to avoid mismatches in PSMs with the local sage-grouse population. Additionally, it may be prudent for translocation efforts to carefully consider the sagebrush communities associated with source and recipient sage-grouse populations.

## 3 Future Directions in Sage-Grouse Genomics

### 3.1 Identifying and Conserving Adaptive Genetic Variation

The genetic and genomic research described above highlights the many ways that molecular data have contributed to the management of sage-grouse. The new genomic resources available for both species of sage-grouse greatly expand the types of questions that can now be answered, with a particular focus on understanding and conserving adaptation. Given the recently discovered relationship between specific sagebrush varieties and the sage-grouse that coevolved with them, genomic methods could further explore this relationship. A comprehensive analysis of diet, for example, can be obtained using metabarcoding approaches (Jarman et al. 2004; Deagle et al. 2009; Pompanon et al. 2012) and is particularly compelling as it can be completed noninvasively through analysis of fecal pellets. Moreover, the role of sage-grouse gut microbiome in metabolizing PSMs in sagebrush may be important (Kohl et al. 2015) and could be further investigated using genomic techniques. Adaptive genetic variation can also be identified by testing for associations between genomic variation and environmental variables important for sage-grouse. As the range of Greater Sage-grouse remains large, encompassing a wide variety of habitat and environmental conditions, this type of analysis could be particularly useful. Finally, genomic approaches can provide insights into the susceptibility of sage-grouse to disease. Rudimentary exploration of genetic diversity at immune genes has shown that both species of sage-grouse have lower levels of diversity compared to other prairie grouse and that Gunnison Sage-grouse is particularly low (Minias et al. 2016, 2018). Newly developed genomic resources for both species (Oh et al. 2019) should facilitate the expansion of this line of research.

### 3.2 Exploring the Impacts of Low Genetic Diversity

Both species of sage-grouse have experienced significant population declines that can result in loss of genetic diversity, which may decrease evolutionary adaptive potential and increase the likelihood of inbreeding depression (Allendorf et al. 2010; Steiner et al. 2013). Issues with low genetic diversity and inbreeding depression have been documented in a close relative, the greater prairie-chicken (*Tympanuchus cupido*; Westemeier et al. 1998), and was suspected in at least one population of Gunnison Sage-grouse (Stiver et al. 2008). Although the range and overall number of Greater Sage-grouse (>100,000 individuals) are relatively large, some populations (e.g., two populations in Washington and one in Strawberry Valley, Utah) show low levels of genetic diversity (Oyler-McCance et al. 2005a; Oh et al. 2019). Gunnison Sage-grouse, on the other hand, have a small and fragmented range, small number of individuals (<5,000), and much lower genetic diversity (Oyler-McCance et al. 2015a; Oh et al. 2019), which may make impacts of low

genetic diversity more problematic. Genomic methods can elucidate the underlying genetic basis of inbreeding depression and provide a mechanistic link between phenotypes and the molecular processes behind them (Steiner et al. 2013). Identifying genes that contribute to inbreeding depression can be achieved through genome-wide association studies comparing genomic variation of individuals with different fitness levels. This approach is currently being investigated in captive Attwater's prairie-chicken (*T. cupido attwateri*, J Johnson, pers. comm) and could potentially be useful in sage-grouse. Captive breeding of both species of sage-grouse has been attempted (Pyrah 1964; Johnson and Boyce 1991; Thompson et al. 2015; Apa and Wiechman 2015) and potentially could be used as a tool to augment wild sage-grouse populations as is taking place in Canada (D McKinnon, pers. comm), increasing both the size and genetic diversity of populations. Genomic analyses can provide more precise estimates of relatedness and inbreeding coefficients both in captive and wild populations that may assist in release, translocation, and genetic rescue efforts (Kardos et al. 2015).

### 3.3 *Document and Better Understand Physiological Response to Stress*

To date, an understanding of how sage-grouse respond to stress has relied on measuring corticosteroid metabolites in fecal samples (Jankowski et al. 2009, 2014; Blickley et al. 2012). Genomic methods have the potential to expand such investigations through gene expression experiments involving transcriptome sequencing. Such investigations could examine the physiological response of sage-grouse to biotic and abiotic stressors that occur both naturally (e.g., seasonal changes in temperature) and due to anthropogenic origin (e.g., noise) (Kleist et al. 2018). By providing measures of relative changes in gene expression in response to exposure to stressors, these analyses can yield insight not only into the molecular basis of these responses but may also serve as biological indicators for monitoring ecosystem health (Isaksson 2015). The main limitation for gene expression studies is that they require systems that are amenable to experimental manipulation, something that has proven to be difficult for sage-grouse. Improvements in our ability to maintain sage-grouse in captivity from captive breeding programs may provide an avenue to move forward with such experimental studies, such as testing adaptability to different food resources and thus variation in sagebrush PSMs.

### 3.4 *Incorporate Genomic Data into Comprehensive Monitoring Programs*

Baseline microsatellite data across the range of both species have been collected (Zimmerman et al. 2019a; Cross et al. 2018; Row et al. 2018), providing current

information on connectivity among populations and levels of anonymous genetic diversity within them. While these microsatellite datasets provide useful information, reliance on these markers for future monitoring may not be ideal for several reasons. First, microsatellite allelic variation is based on fragment size, typically estimated from electrophoretic methods, which often vary among different techniques and conditions, thereby requiring constant standardization across instruments and laboratories, and may sometimes be influenced by subjectivity in scoring. Second, while microsatellite markers may be useful for documenting connectivity and levels of anonymous genetic diversity, they are typically not useful alone for identifying and monitoring adaptive genetic variation that may be important for conservation efforts. Finally, genomic genotyping methods are likely now less expensive and more repeatable than traditional microsatellite genotyping. Thus, a standardized set of genome-wide SNPs representing both anonymous and adaptive processes could be developed from existing genomic resources and archived samples and used as a baseline for future monitoring programs.

In the past, monitoring of sage-grouse populations relied on yearly lek counts in the field and analyses of trends that were tracked through time by individual states. Recent more comprehensive monitoring efforts now coordinate monitoring at different hierarchical scales (e.g., lek, lek cluster, region, or management zone) across the range of Greater Sage-grouse (Coates et al. 2017; Edmunds et al. 2017). These efforts identify when a lek or lek cluster is declining, identifying when trends deviate from a broader-scale pattern(s), and ultimately will link causal mechanisms to those declines which will identify potential management actions. Genomic monitoring could be incorporated into such programs as feathers could be collected periodically from a subset of leks and analyzed to watch for changes in connectivity, isolation of populations, or loss of anonymous or adaptive genetic diversity, evaluating potential links to population trends.

## 4 Conclusion

Information from genetic studies has informed conservation and management of sage-grouse for nearly two decades, addressing a broad variety of questions from taxonomy and gene flow to investigations of mating systems and unique identification of individuals for demographic analyses. Genomic methods, however, can build significantly on these foundations, greatly expanding the types of questions that can now be addressed. Novel genomic techniques coupled with the recently developed genomic resources for sage-grouse facilitate more precise estimates of parameters of interest (e.g., gene flow, inbreeding coefficients) and provide a more comprehensive understanding of the genetic basis of adaptation in sage-grouse. The integration of these new genomic resources with existing ecological and behavioral data for sage-grouse promises to shed light on mechanistic relationships that ultimately are vital for the conservation and management of these species.

**Acknowledgments** We thank M. Schroeder, J. Fike, and the editors of this volume for their helpful comments on this chapter and J. Saher for her help with figures. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

## References

Aldrich JW. New subspecies of birds from western North America. Proc Biol Soc Wash. 1946;59:129–36.

Aldrich JW, Duvall AJ. Distribution of American gallinaceous game birds. USFWS Circular 34, Washington, DC; 1955.

Aldridge CL, Brigham RM. Distribution, status and abundance of greater sage-grouse, *Centrocercus urophasianus*, in Canada. Can Field Nat. 2003;117:25–34.

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

American Ornithologist's Union. Forty-second supplement to the American Ornithologists Union check-list for North American birds. Auk. 2000;117(3):847–58.

Apa AD, Wiechman LA. Captive-rearing of Gunnison sage-grouse from egg collection to adulthood to foster proactive conservation and recovery of a conservation-reliant species. Zoo Biol. 2015;34:438–52.

Baumgardt JA, Goldberg CS, Reese KP, Connelly JW, Musil DD, Garton EO, Waits LP. A method for estimating sex ratio for sage-grouse using noninvasive genetic samples. Mol Ecol Resour. 2013;13:393–402.

Beever EA, Aldridge CL. Influences of free-roaming equids on sagebrush ecosystems, with a focus on greater sage-grouse. In: Knick ST, Connelly JW, editors. Greater sage-grouse: ecology and conservation of a landscape species and its habitats, Studies in avian biology, vol. 38. Berkley: University of California Press; 2011. p. 273–90.

Benedict NG, Oyler-McCance SJ, Taylor SE, Braun CE, Quinn TW. Evaluation of the eastern (*Centrocercus urophasianus urphasianus*) and western (*Centrocercus urophasianus phaios*) subspecies of Sage-grouse using mitochondrial control-region sequence data. Conserv Genet. 2003;4:201–310.

Bergerud AT. Mating systems in grouse. In: Bergerud AT, Gratson MW, editors. Adaptive strategies and population ecology of northern grouse. Minneapolis: University of Minnesota Press; 1988. p. 439–72.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. Unlocking the vault: next-generation museum population genomics. Mol Ecol. 2013;22:6018–32.

Bird KL, Aldridge CL, Carpenter JE, Paszkowski CA, Boyce MS, Coltman DW. The secret sex lives of sage-grouse: multiple paternity and intraspecific nest parasitism revealed through genetic analysis. Behav Ecol. 2012;24:29–38.

Blickley JL, Word KR, Krakauer AH, Phillips JL, Sells SN, Taff CC, Wingfield JC, Patricelli GL. Experimental chronic noise is related to elevated fecal corticosteroid metabolites in lekking male greater sage-grouse (*Centrocercus urophasianus*). PLOS One. 2012;7(11):e50462. https://doi.org/10.1371/journal.pone.0050462.

Braun CE. Sage grouse declines in western North America: what are the problems? Proc West Assoc of State Fish Wildl Agencies. 1998;78:139–56.

Braun CE, Baker MF, Eng RL, Gashwiler JS, Schroeder MH. Conservation committee report on effects of alteration of sagebrush communities on the associated avifauna. Wilson Bull. 1976;88:165–71.

Bush KL, Vinsky MD, Aldridge CL, Paszkowski CA. A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater sage-grouse (*Centrocercus urophasianus*). Conserv Genet. 2005;6:867–70.

Bush KL, Aldridge CL, Carpenter JE, Paszkowski CA, Boyce MS, Coltman DW. Birds of a feather do not always lek together: genetic diversity and kinship structure of greater sage-grouse (*Centrocercus urophasianus*) in Alberta. Auk. 2010;127:343–53.

Bush KL, Dyte CK, Moynahan BJ, Aldridge CL, Sauls HS, Battazzo AM, Walker BL, Doherty KE, Tack J, Carlson J, Eslinger D, Nicholson J, Boyce MS, Naugle DE, Paszkowski CA, Coltman DW. Population structure and genetic diversity of greater sage-grouse (*Centrocercus urophasianus*) in fragmented landscapes at the northern edge of their range. Conserv Genet. 2011;12:527–42.

Coates PS, Prochazka BG, Ricca MA, Wann GT, Aldridge CL, Hanser SE, Doherty KE, O'Donnell MS, Edmunds DE, Espinosa SP. Hierarchical population monitoring of greater sage-grouse (*Centrocercus urophasianus*) in Nevada and California – identifying populations for management at the appropriate spatial scale. U.S. Geological Survey Open-File Report 2017-1089.2017.

Connelly JW, Braun CE. Long-term changes in sage grouse *Centrocercus urophasianus* populations in western North America. Wildl Biol. 1997;3:229–34.

Connelly JW, Schroeder MA, Sands AR, Braun CE. Guidelines to manage sage grouse populations and their habitats. Wildl Soc Bull. 2000;28:967–85.

Connelly JW, Knick ST, Schroeder MA, Stiver SJ. Conservation assessment of greater sage-grouse and sagebrush habitats. Western Association of Fish and Wildlife Agencies: Cheyenne; 2004.

Cross TB, Naugle DE, Carlson JC, Schwartz MK. Hierarchical population structure in greater sage-grouse provides insight into management boundary delineation. Conserv Genet. 2016;17:1417–33.

Cross TB, Naugle DE, Oyler-McCance SJ, Row JR, Fedy BC, Schwartz MK. The genetic network of greater sage-grouse: range-wide identification of keystone hubs of connectivity. Ecol Evol. 2018;8:5394–412.

Dalke PD, Pyrah DB, Stanton DC, Crawford JE, Schlatterer EF. Ecology, productivity, and management of sage grouse in Idaho. J Wildl Manag. 1963;27:811–41.

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, et al. Multi-platform next-generation sequencing of the domestic Turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol. 2010;8:e1000475.

Davis DM, Reese KP, Gardner SC, Bird KL. Genetic structure of greater sage-grouse (*Centrocercus urophasianus*) in a declining, peripheral population. Condor Ornithol Appl. 2015;117:530–44.

Deagle BE, Kirkwood R, Jarman SN. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Mol Ecol. 2009;18:2022–38.

Defaveri J, Viitaniemi H, Leder E, Merilä J. Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Mol Ecol Resour. 2013;13:377–92.

Edmunds DR, Aldridge CL, O'Donnell MS, Monroe AP. Greater sage-grouse population trends across Wyoming. J Wildl Manag. 2017;82:397–412.

Ellsworth DL, Honeycutt RL, Silvy NJ, Rittenhouse KD, Smith MH. Mitochondrial-DNA and nuclear-gene differentiation in North American prairie grouse (genus *Tympanuchus*). Auk. 1994;111:661–71.

Environment Canada. Amended recovery strategy for the greater sage-grouse (*Centrocercus urophasianus uophasianus*) in Canada, Species at risk act recovery strategy series. Ottawa: Environment Canada; 2014.

Fedy BC, Row JR, Oyler-McCance SJ. Integration of genetic and demographic data to assess population risk in a continuously distributed species. Conserv Genet. 2017;18:89–104.

Forester BR, Landguth EL, Hand BK, Balkenhol N. Landscape genomics for wildlife research. In: Hohenlohe PA, Rajora OP, editors. Population genomics: wildlife. Cham: Springer International Publishing AG; 2018. https://doi.org/10.1007/13836_2018_56.

Frye GG, Connelly JW, Musil DD, Forbey JS. Phytochemistry predicts habitat selection by an avian herbivore at multiple spatial scales. Ecology. 2013;94:308–14.

Garant D, Kruuk LEB, Wilkin TA, McCleery RH, Sheldon BC. Evolution driven by differential dispersal within a wild bird population. Nature. 2005;433:60–5.

Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics. 2015;201:1555–79.
Gibson RM, Bradbury JW. Male and female mating strategies on sage grouse leks. In: Rubenstein DI, Wrangham RW, editors. Ecological aspects of social evolution. Princeton: Princeton University Press; 1986. p. 379–98.
Gibson RM, Bradbury JW, Vehrencamp SL. Mate choice in lekking sage grouse revisited: the roles of vocal display, female site fidelity and copying. Behav Ecol. 1991;2:165–80.
Gibson RM, Pires D, Delaney KS, Wayne RK. Microsatellite DNA analysis shows that greater sage-grouse leks are not kin groups. Mol Ecol. 2005;14:4453–9.
Green AW, Aldridge CL, O'Donnell MS. Investigating impacts of oil and gas development on greater sage-grouse. J Wildl Manag. 2017;81:46–57.
Gunnison Sage-grouse Rangewide Steering Committee (GSRSC). Gunnison sage-grouse rangewide conservation plan. Denver: Colorado Division of Wildlife; 2005.
Hedrick PW. A standardized genetic differentiation measure. Evolution. 2005;59:1633–8.
Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population genomics provides key insights in ecology and evolution. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 483–510.
Hupp JW, Braun CE. Geographic variation among sage grouse in Colorado. Wilson Bull. 1991;103:255–61.
International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432:695–716.
Isaksson C. Urbanization, oxidative stress and inflammation: a question of evolving, acclimatizing or coping with urban environmental stress. Funct Ecol. 2015;29:913–23.
Jahner JP, Gibson D, Weitzman CL, Bloomberg EJ, Sedinger JS, Parchman TL. Fine scale genetic structure among greater sage-grouse leks in Central Nevada. BMC Evol Biol. 2016;16:127.
Jankowski MD, Wittwer DJ, Heisey DM, Franson JC, Hofmeister EK. The adrenocortical response of greater sage-grouse (*Centrocercus urophasianus*) to capture ACTH injection, and confinement, as measured in fecal samples. Physiol Biochem Zool. 2009;82:190–201.
Jankowski MD, Russell RE, Franson JC, Dusek RJ, Hines MK, Gree M, Hofmeister EK. Corticosterone metaboloite concentrations in greater sage-grouse are positively associated with the presence of cattle grazing. Rangeland Ecol Manag. 2014;67:237–46.
Jarman SN, Deagle BE, Gales NJ. Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. Mol Ecol. 2004;13:1313–22.
Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. The global diversity of birds in space and time. Nature. 2012;491:444–8.
Johnsgard PA. The grouse of the world. Lincoln: University of Nebraska Press; 1983.
Johnson GD, Boyce MS. Survival, growth, and reproduction of captive reared sage grouse. Wildl Soc Bull. 1991;18:88–93.
Jones MR, Good JM. Targeted capture in evolutionary and ecological genomics. Mol Ecol. 2016;25:185–202.
Jost L. $G_{ST}$ and its relatives do not measure differentiation. Mol Ecol. 2008;17:4015–26.
Kahn NW, Braun CE, Young JR, Wood S, Mata DR, Quinn TW. Molecular analysis of genetic variation among large- and small-bodied Sage Grouse using mitochondrial control-region sequences. Auk. 1999;116:819–24.
Kardos M, Luikart G, Allendorf FW. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. 2015;115:63–72.
Keenan K, McGinnity P, Cross TF, Crozier WW, Prodohl PA. diveRsity: an R package for the estimation of population genetics parameters and their associated errors. Methods Ecol Evol. 2013;4:782–8.
Kelsey RG, Stephens JR, Shafizadeh F. The chemical constituents of sagebrush foliage and their isolation. J Range Manag. 1982;35:617–22.
Kleist NJ, Guralnick RP, Cruz A, Lowry CA, Francis CD. Chronic anthropogenic noise disrupts glucocorticoid signaling and has multiple effects on fitness in an avian community. Proc Natl Acad Sci. 2018;115:E648–57.

Knick ST, Connely JW. Greater sage-grouse: ecology and conservation of a landscape species and its habitats, Studies in avian biology, vol. 38. Berkley: University of California Press; 2011.
Knick ST, Dobkin DS, Rotenberry JT, Schroeder MA, Vander Haegen WM, van Ripper C III. Teetering on the edge or too late? Conservation and research issues for avifauna of sagebrush habitats. Condor. 2003;105:611–34.
Kohl KD, Pitman E, Robb BC, Connelly JW, Dearing MD, Forbey JS. Monoterpenes as inhibitors of digestive enzymes and counter-adaptations in a specialist avian herbivore. J Comp Physiol B. 2015;185:425–34.
Kumar S, Stecher G, Suleski M, Hedges SB. Timetree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;4:1812–9.
Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475:493–6.
Luikart G, Kardos M, Hand BK, Rajora OP, Aitken SN, Hohenlohe PA. Population genomics: advancing understanding of nature. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Cham: Springer Nature Switzerland AG; 2019. p. 3–79.
Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 2003;18:189–97.
Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.
Meirmans PG. The trouble with isolation by distance. Mol Ecol. 2012;21:2839–46.
Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet. 2010;11:31–46.
Miller RF, Knick ST, Pyke DA, Meike CW, Hanser SE, Wisdom MJ, Hild AL. Characteristics of sagebrush habitats and limitations to long-term conservation. In: Knick SJ, Connelly JW, editors. Greater sage-grouse: ecology and conservation of a landscape species and its habitats, Studies in avian biology, vol. 38. Berkley: University of California Press; 2011. p. 145–84.
Minias P, Bateson ZW, Whittingham LA, Johnson JA, Oyler-McCance SJ, Dunn PO. Contrasting evolutionary histories of MHC class I and class II loci in grouse – effects of selection and gene conversion. Heredity. 2016;116:466–76.
Minias P, Bateson ZW, Whittingham LA, Johnson JA, Oyler-McCance SJ, Dunn PO. Extensive shared polymorphism at non-MHC immune genes in recently diverged North American prairie-grouse. Immunogenetics. 2018;70:195–204.
Monroe AP, Aldridge CL, Assal TJ, Veblen KE, Pyke DA, Casazza ML. Patterns in greater sage-grouse population dynamics correspond with public grazing records at broad scales. Ecol Appl. 2017;27:1096–107.
Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JBW, Backström N, Künstner A, Balakrishnan CN, Heger A, Ponting CP, Clayton DF, Ellegren H. Molecular evolution of genes in avian genomes. Genome Biol. 2010;11:R68.
Nei M, Kumar S. Molecular evolution and phylogenetics. New York: Oxford University Press; 2000.
Oh KP, Aldridge CL, Forbey JS, Dadabay CY, Oyler-McCance SJ. Genomic insights into neutral and adaptive variation in sage-grouse: implications for ecology and conservation. Genome Biol Evol. 2019;11:2023–2034.
Oyler-McCance SJ, Kahn NW, Burnham KP, Braun CE, Quinn TW. A population genetic comparison of large- and small-bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. Mol Ecol. 1999;8:1457–65.
Oyler-McCance SJ, Burnham KP, Braun CE. Influence of changes in sagebrush on Gunnison sage grouse in southwestern Colorado. Southwest Nat. 2001;46:323–31.
Oyler-McCance SJ, Taylor SE, Quinn TW. A multilocus population genetic survey of the greater sage-grouse across their range. Mol Ecol. 2005a;14:1293–310.
Oyler-McCance SJ, St. John J, Taylor SE, Apa AD, Quinn TQ. Population genetics of Gunnison sage-grouse: implications for management. J Wildl Manag. 2005b;69:630–7.
Oyler-McCance SJ, St. John J, Quinn TW. Rapid evolution in lekking grouse. Ornithol Monogr. 2010;67:114–22.

Oyler-McCance SJ, Casazza ML, Fike JA, Coates PS. Hierarchical spatial genetic structure in a distinct population segment of greater sage-grouse. Conserv Genet. 2014;15:1299–311.
Oyler-McCance SJ, Cornman RS, Jones KL, Fike JA. Genomic single-nucleotide polymorphisms confirm that Gunnison and greater sage-grouse are genetically well differentiated and that the bi-state population is distinct. Condor. 2015a;117:217–27.
Oyler-McCance SJ, Cornman RS, Jones KL, Fike JA. Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. Heredity. 2015b;115:452–9.
Oyler-McCance SJ, Oh KP, Langin KM, Aldridge CL. A field ornithologist's guide to genomics: practical considerations for ecology and conservation. Auk Ornithol Adv. 2016;4:626–48.
Panhuis TM, Butlin R, Zuk M, Tregenza T. Sexual selection and speciation. Trends Ecol Evol. 2001;16:364–71.
Patterson RL. The sage grouse in Wyoming. Denver: Sage Books; 1952.
Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, Taberlet P. Who is eating what: diet assessment using next generation sequencing. Mol Ecol. 2012;21:1931–50.
Pyrah DB. Sage chickens in captivity. Game Bird Breed Pheas Fanciers Avicultur Gaz. 1964;13:10–1.
Remington TE, Braun CE. Sage grouse food selection in winter, North Park, Colorado. J Wildl Manag. 1985;49:1055–61.
Row JR, Blouin-Demers G, Lougheed SC. Habitat distribution influences dispersal and fine-scale genetic population structure of eastern foxsnakes (*Mintonius gloydi*) across a fragmented landscape. Mol Ecol. 2010;19:5157–71.
Row JR, Oyler-McCance SJ, Fike JA, O'Donnell MS, Doherty KE, Aldridge CL, Bowen ZH, Fedy BC. Landscape characteristics influencing the genetic structure of greater sage-grouse within the stronghold of their range: a holistic modeling approach. Ecol Evol. 2015;5:1955–69.
Row JR, Oyler-McCance SJ, Fedy BC. Differential influences of local subpopulations on regional diversity and differentiation for greater sage-grouse (*Centrocercus urophasianus*). Mol Ecol. 2016;25:4424–37.
Row JR, Doherty K, Cross T, Schwartz M, Oyler-McCance SJ, Naugle D, Knick ST, Fedy B. Quantifying functional connectivity: the role of breeding habitat, abundance, and landscape features on range-wide gene flow in sage-grouse. Evol Appl. 2018;11:1305–21.
Savolainen O, Lascoux M, Merila J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20.
Schroeder MA. Variation in greater sage-grouse morphology by region and population. Spokane: US Fish and Wildlife Service Report; 2008.
Schroeder MA, Hays DW, Livingston MF, Stream LE, Jacobson JE, Pierce DJ. Changes in the distribution and abundance of sage grouse in Washington. Northwest Nat. 2000;81:104–12.
Schroeder MA, Aldridge CL, Apa AD, Bohne JR, Braun CE, Bunnel SD, Connelly JW, Deibert PA, Gardner SC, Hilliard MA, Kobringer GD, McAdam SM, McCarthy CW, McCarthy JJ, Mitchell DL, Rickerson EV, Stiver SJ. Distribution of sage-grouse in North America. Condor. 2004;106:363–76.
Schulwitz S, Bedrosian B, Johnson JA. Low anonymous genetic diversity in isolated greater sage-grouse (*Centrocercus urophasianus*) populations in northwest Wyoming. Condor Ornith Appl. 2014;116:560–73.
Semple K, Wayne RK, Gibson RM. Microsatellite analysis of female mating behavior in lek-breeding sage grouse. Mol Ecol. 2001;10:2043–8.
Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotech. 2008;26:1135–45.
Shirk AJ, Schroeder MA, Robb LA, Cushman SA. Empirical validation of landscape resistance models: insights from the greater sage-grouse (*Centrocercus urophasianus*). Landsc Ecol. 2015;30(10):1837–50. https://doi.org/10.1007/s10980-015-0214-4.
Shyvers JL, Walker BL, Oyler-McCance SJ, Fike JA, Noon BR. Genetic mark-recapture analysis of winter faecal pellets allows estimation of population size in Sage Grouse *Centrocercus urophasianus*. Ibis. 2019 (in press).

Spaulding A. Rapid courtship evolution in grouse (Tetraonidae): contrasting patterns of acceleration between the Eurasian and North American polygynous clades. Proc R Soc London B. 2007;274:1079–86.
Steiner CC, Putnam AS, Hoeck PEA, Ryder OA. Conservation genomics of threatened animal species. Annu Rev Anim Biosci. 2013;1:261–81.
Stiver JR, Apa AD, Remington TE, Gibson RM. Polygyny and female breeding failure reduce effective population size in the lekking Gunnison sage-grouse. Biol Conserv. 2008;141:472–81.
Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP. Putting the "landscape" in landscape genetics. Heredity. 2007;98:128–42.
Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol. 1998;7:453–64.
Taylor SE, Young JR. A comparative behavioral study of three greater sage-grouse populations. Wilson J Ornithol. 2006;118:36–41.
Thompson TR, Apa AD, Reese KP, Tadvick KM. Captive rearing sage-grouse for augmentation of surrogate wild broods: evidence for success. J Wildl Manag. 2015;79:998–1013.
USFWS. Endangered and threatened wildlife and plants; determination for the Gunnison Sage-grouse as a threatened or endangered species. Fed Regist. 2014;79:69192–310.
Uy JAC, Borgia G. Sexual selection drives rapid divergence in bowerbird display traits. Evolution. 2000;63:153–64.
Waits LP, Storfer A. Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetic studies. In: Balkenhol N, Cushman SA, editors. Landscape genetics: concepts, methods, applications. Hoboken: Wiley; 2016. p. 35–57.
Wallestad R, Peterson JG, Eng RL. Foods of adult sage grouse in central Montana. J Wildl Manag. 1975;39:628–30.
Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.
Welch BL, Pederson JC, Rodriquez RL. Selection of big sagebrush by sage grouse. Great Basin Nat. 1988;48:274–9.
West NE. Western intermountain sagebrush steppe. In: West NE, editor. Ecosystems of the world, Temperate deserts and semi-deserts, vol. 5. New York: Elsevier; 1983. p. 331–49.
Westemeier RL, Brawn JD, Simpson SA, Esker TL, Jansen RW, Walk JW, Kershner EL, Bouzat JL, Paige KN. Tracking the long-term decline and recovery of an isolated population. Science. 1998;282:1695–8.
Wiley RH. Evolution of social organization and life-history patterns among grouse. Q Rev Biol. 1974;49:201–27.
Willing E-M, Dreyer C, van Oosterhout C. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. PLOS One. 2012;7:1–7.
Wittenberger JF. The evolution of mating systems in grouse. Condor. 1978;80:126–37.
Young JR, Hupp JW, Bradbury JW, Braun CE. Phenotypic divergence of secondary sexual traits among sage grouse populations. Anim Behav. 1994;47:1353–62.
Young JR, Braun CE, Oyler-McCance SJ, Hupp JW, Quinn TW. A new species of sage-grouse (Pahsianidae: Centrocercus) from southwestern Colorado. Wilson Bull. 2000;112:445–53.
Zimmerman, SJ. Genetics, genomics, grouse, and conservation: use of genetic and genomic data to evaluate conservation actions and characterize popualtions of Gunnison Sage-grouse. PhD dissertation, Colorado State University, Fort Collins; 2019.
Zimmerman SJ, Aldridge CL, Apa AD, Oyler-McCance SJ. Evaluation of the genetic change from translocation among Gunnison Sage-grouse (*Centrocercus minimus*) populations. Condor. 2019a;121:1–14.
Zimmerman SJ, Aldridge CL, Oyler-McCance SJ. Signatures of adaptive divergence among populations of an avian species of conservation concern. Evol Appl. 2019b; https://doi.org/10.1111/eva.12825.

# Index




Paul A. Hohenlohe and Om P. Rajora (eds.), *Population Genomics: Wildlife*, Population Genomics [Om P. Rajora (Editor-in-Chief)], https://doi.org/10.1007/978-3-030-63489-6,