Chapter 17

Avian Population Genomics Taking Off: Latest Findings and Future Prospects

Kira E. Delmore and Miriam Liedvogel

Abstract

Birds are one of the most recognizable and diverse groups of organisms on earth. This group has played an important role in many fields, including the development of methods in behavioral ecology and evolutionary theory. The use of population genomics took off following the advent of high-throughput sequencing in various taxa. Several features of avian genomes make them particularly amenable for work in this field, including their nucleated red blood cells permitting easy DNA extraction and small, compact genomes. We review the latest findings in the population genomics of birds here, emphasizing questions related to behavior, ecology, evolution, and conservation. Additionally, we include insights in trait mapping and the ability to obtain accurate estimates of important summary statistics for conservation (e.g., genetic diversity and inbreeding). We highlight roadblocks that will need to be overcome in order to advance work on the population genomics of birds and prospects for future work. Roadblocks include the assembly of more contiguous reference genomes using long-reads and optical mapping. Prospects include the integration of population genomics with additional fields (e.g., landscape genetics, phylogeography, and genomic mapping) along with studies beyond genetic variants (e.g., epigenetics).

Key words Population genomics, Bird, Avian, Demography, Conservation, Speciation, Ecology

1 Introduction

Birds have played a central role in our understanding of many research fields. Notable examples include (1) the development of methods essential for behavior and ecology by Margaret Nice [1] using populations of song sparrows (*Melospiza melodia*), and (2) the definition of species as groups of populations reproductively isolated from one another by Ernst Mayr [2], inspired by the geographic distribution of birds and galvanizing the field of evolution. Current declines in natural populations of birds worldwide may cause work by researchers like Nice and Mayr to be overshadowed by efforts in conservation. Concern about the loss of birds to the millinery trade triggered Harriet Hemenway and Mina Hall to establish the National Audubon Society in 1886. This

society is one of the first nonprofit environmental organizations and generates datasets essential for population monitoring and forecasting to this day (e.g., annual Christmas and Great Backyard Bird Counts to census birds worldwide and eBird, an online database of bird observations).

The influential role of birds across varied research fields continued with the development of population genetics. This field emerged in the 1980s following the advent of sequencing technologies to quantify marker based genetic variation, including sequence variation in mitochondrial DNA (mtDNA) and length polymorphisms in microsatellites. Tools from population genetics can be used to evaluate the role mutation, genetic drift, selection and gene flow play in generating variation within and between populations [3, 4], with relevance to behavior, ecology, evolution, and conservation. Avian blood has nucleated red blood cells, making it ideally suited for DNA extraction and subsequent population genetic analyses, early examples include the use of mtDNA to identify taxonomic units for conservation (e.g., dusky seaside sparrows [*Ammodramus maritimus nigrescens*], [5, 6]).

The last 10 years has seen a change in both the scale and depth of genetic analyses, with the transition from the use of one or a few genetic markers to tens of thousands of markers genome-wide marking the development of population genomics. This transition was stimulated by de novo assembly of reference genomes along with the advent of *high-throughput sequencing* (HTS). HTS is discussed in detail elsewhere (e.g., [7, 8]), but briefly, is a set of platforms that sequence DNA from multiple genomic regions and individuals in parallel. HTS has increased the proportion of the genome that can be sampled and decreased the time and cost of sequencing, allowing its use on most organisms of interest. Again, birds are well-suited for this extension. Not only can good amounts of DNA be generated easily, but they also have small (mean ~1.45 billion base pairs; [9]), compact (e.g., fewer transposable elements [TEs] and repetitive regions; [10]) genomes, allowing for relatively easy genome assembly and mapping at high coverages. The first avian reference genomes were the chicken (*Gallus gallus*, [11]) and zebra finch (*Taeniopygia guttata*, [12]), assembled using Sanger sequencing and followed by an explosion of bird genomes assembled using HTS [13].

We review the population genomics in birds here, emphasizing the role this group has played in both the development of this field and its application to questions in behavior, ecology, evolution, and conservation. Population genomics can be used to answer questions at both the genome and locus levels. Work at the genome level informs our understanding of population processes (e.g., demography and population structure), while work at the locus level helps identify genomic regions affected by mutation, drift, selection and/or gene flow [3]. We follow this division in this chapter,

introducing some of the latest findings from birds, highlighting the benefits of applying population genomics tools to these questions (vs. traditional population genetic techniques with fewer markers), and finishing by outlining future prospects along this trajectory.

2   Latest Findings

2.1   Relevance of Genomic Insight for Evolution

Demography is the study of changes in effective population size (N_e) through time, gene flow and divergence. Information on these dynamics is essential for understanding the evolution of species, populations and traits and important for setting baselines beyond which evolutionary processes can be examined. This is especially true in the current literature, where genome scans are being used to identify loci associated with phenotypic traits and/or involved in adaptation and speciation (*see* below; e.g., [14–17]). Early demographic analyses were based on the coalescent (or divergence), assuming that pairwise sequence divergence is proportional to the time of the coalescent, and relying on nonrecombining loci (e.g., mtDNA, Fig. 1a left, [18]). These analyses were expanded to include nuclear loci and permit estimates of changes in population size and gene flow, but remained limited to a small number of demographic scenarios and markers and were computationally expensive [19]. The availability of genome-wide data and new tools for their analysis has revolutionized this field and its progression is demonstrated well by a series of studies on collared and pied flycatchers (*Ficedula albicollis* and *F. hypoleuca*).

Early demographic work with collared and pied flycatchers suggested these species diverged during the Pleistocene, expanded their ranges following the last glacial maximum, and came into secondary contact in Central Europe where introgression from pied into collared flycatchers is greater [20, 21]. Nadachowska-Brzyska et al. [22] used *whole genome resequencing* (WGS) data to expand on these findings, comparing 15 demographic models using an Approximate Bayesian Computation (ABC) approach. A model with recent divergence time (230,000–240,000 years ago [ya]), unidirectional gene flow (0.16–0.36 pied individuals per generation) and ancient reductions in population size (e.g., 600,000–80,000 in collared) provided the best fit to the data (Fig. 1 right). This work represents one of the first times an ABC approach with HTS was applied to study demography in a non-model system, providing a level of detail unattainable with earlier methods [22]. Flycatchers exhibit strong reproductive isolation, with evidence for both pre- and postzygotic barriers to gene flow (e.g., nearly complete female sterility, [23]). Accordingly, these results suggest speciation in birds can occur quite rapidly.

Nadachowska-Brzyska et al. [24] validated findings from the ABC analysis using the Pairwise Sequentially Markovian Coalescent
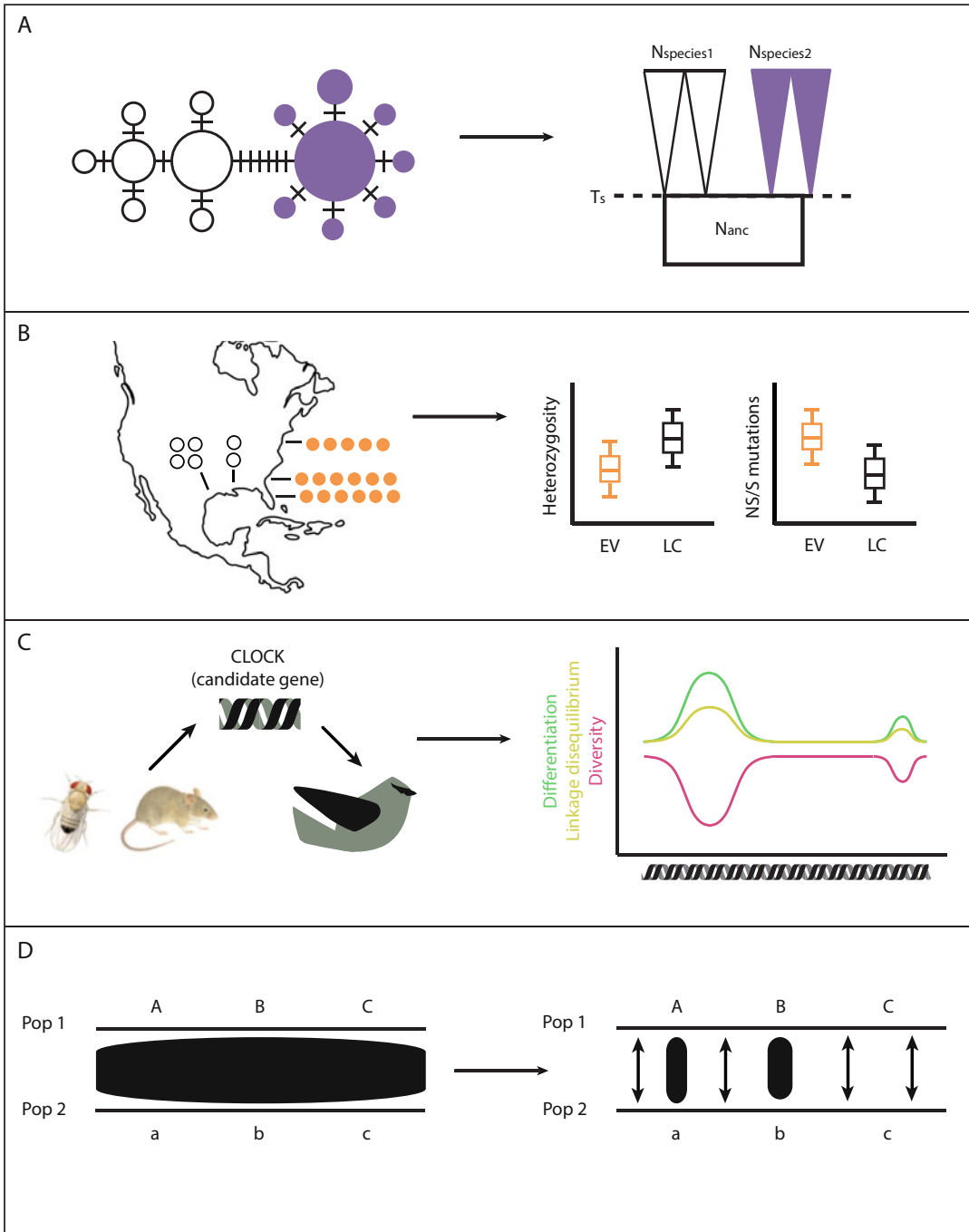
Fig. 1 Conceptualized figure showing examples of questions that were answered using genetic data and how they have been expanded with the transition from population genetics (left) to genomics (right). Panel **a** is relevant to *demography* and show a minimum spanning network where circles represent haplotypes, sizes of circles reflect abundance, and bars across branches represent single nucleotide changes that can be used to estimate divergence time (yellow), and more detailed analyses using genomic data where a demographic scenario with recent migration and population size changes derived from recent ABC analyses (N—effective

(PSMC, Li and Durbin [25]; similar to the model described in Chapter 7) and WGS data. PSMC estimated a slightly later divergence time between the species (386,000–888,000 ya) and provided additional information on changes in population size over time, with collared flycatchers undergoing an expansion 100,000–200,000 ya and subsequent decline, likely corresponding with glaciation period marine isotope stage (MIS) 6 (191,000–130,000 ya). Pied flycatchers did not exhibit the same decline, perhaps because they can tolerate glacial climates more effectively than collared flycatchers that occur at lower latitudes and altitudes. Instead, pied flycatchers appear to be increasing in population size, suggesting they are outcompeting collared flycatchers in the present day, which falls in line with behavioral observations in the system [23]. Results from this study on flycatchers support earlier work using data from 38 bird species [26], documenting similar variability in estimates of population size over time. Together these results provide important inferences about the population dynamics of temperate species that have experienced glacial cycles throughout their history and caution against assuming simple demographic history (e.g., constant population sizes, a single expansion, and/or similar trends for populations).

Of perhaps less evolutionary interest, but equal importance, Nadachowska-Brzyska et al. [24] conducted a thorough analysis of the impact of sequence coverage and missing data on the accuracy of demographic inference with PSMC. They conclude at least $18\times$ mean coverage is needed and no more than 25% missing data can be permitted. These kinds of considerations are important when implementing analyses with genomic data as biases resulting from poor filtering can have a significant impact on results. In the case of PSMC, low filtering cutoffs lead to homozygous sites being considered heterozygous, affecting the size of recombination blocks and estimates of N_e (both magnitude and the shape of curves). It is important to note that newer versions of PSMC exist and provide many benefits but are not easily applied to birds. For example,

Fig. 1 (continued) population size [anc—ancestral population], T_s—divergence time). Panel **b** is relevant to *conservation* and shows the geographic distribution of individuals differentiated at a small number of markers and potentially requiring separate management status, and more recent comparisons of heterozygosity and the proportion of nonsynonymous to synonymous mutations between an endangered/vulnerable species and one of least concern providing more detail when establishing management plans. Panel **c** is relevant to the *genetic basis of phenotypes* and shows a candidate gene identified in *Drosophila* and *Mus musculus* and surveyed in populations of birds as a potential regulator of migration and more recent uses of genomic data to identify selective sweeps related to a phenotype in an unbiased way (based on [30]). Panel **d** is relevant to the *genetics of speciation* and shows how speciation was originally conceived as complete loss of gene flow between population, and more recent ideas where speciation can still occur if regions of the genome continue to exchange genes, with black blocks representing areas where gene flow is no longer occurring and fixed alleles between population arise (based on [40])

multiple sequentially Markovian coalescent (MSMC, *see* Chapter 7) can provide information on more recent population dynamics but requires accurate *phasing*, which can be difficult to obtain for birds as reference panels of known haplotypes are often used and derive from trios (adult and both parents) or pedigreed populations which are also difficult to obtain for birds.

2.2   Relevance of Genomic Insight for Conservation

Demographic analyses provide us with information on population dynamics mostly in the past. Population genomics can also be used to understand dynamics in present day populations and are especially relevant for conservation (e.g., helping inform the management of threated species). Of prime importance is the application of these tools to study genetic factors that can compound reductions in population size already experienced by threatened species and transfer of findings from common species to those that are under threat. Here the availability of HTS is allowing researchers to obtain far more accurate and precise estimates not only of population structure and dynamics, but also loss of genetic diversity and inbreeding (Fig. 1b). Using a comparative framework and genome sequences spanning nearly the full phylogenetic spectrum of birds Li et al. [27] highlighted the potential of these data for conservation. We will discuss this work below but want to emphasize the importance of the application of population genomics to these questions for birds where 1375/10,000 species (13%) are threatened with extinction (IUCN https://www.iucn.org/). Habitat loss is among the main threats and the ramifications of extinction in birds will be far reaching, as they are essential for ecosystem functioning (e.g., as seed dispersers) and serve as important indicators of ecosystem health (e.g., tracking changes in habitat, water and climate).

Li et al. [27] is one of 27 papers that were released by the Avian Phylogenomics Project in December 2014 based on 48 genomes assembled using HTS. These authors classified each species as endangered/vulnerable (EV) or of no conservation concern and observed that EV species exhibited lower levels of heterozygosity and more nonsynonymous (and potentially deleterious) mutations than species of no conservation concern. Nonsynonymous mutations were associated with increased *linkage disequilibrium* (*see* also Chapter 1) across the genome. Combined, these findings suggest EV species may be at risk of *inbreeding depression*. Estimates of inbreeding can also be obtained using pedigrees in the form of inbreeding coefficients (the probability of a locus being identical by decent). Nevertheless, as noted already pedigrees are rare in wild bird populations so the application of HTS in the framework used by Li et al. [27] will be imperative for understanding risks to natural populations of birds [28].

Li et al. [27] took their work one step further, obtaining WGS data from eight individuals of the critically endangered crested ibis

(*Nipponia nippon*). These data were used to document changes in population size similar to work described by Nadachowska-Brzyska et al. [24] with flycatchers. However, of potentially greater importance for conservation, Li et al. [27] also used these data to develop a set of genetic loci to track individuals of the species. 166,000 degenerate STR (short tandem repeats) loci were identified. Among these loci, 23 were informative and will have many applications, including the estimation of sex and paternity along with the reconstruction of pedigrees to identify birds for breeding programs. Similar applications are being promoted by many authors as a way of implementing rapid biodiversity screening for risk assessment, including the quantification of genetic diversity and population structure in natural populations. This work will allow researchers to identify taxonomic units quickly, permitting the delineation of geographic areas for conservation and management.

2.3 Locus-Level Work to Examine the Genetic Basis of Phenotypic Traits

The application of population genomics to questions of demography and conservation are examples of genome-level analyses. Population genomics can also be used at the locus-level and include bottom-up approaches to examining the genetic basis of phenotypic traits, including morphological and behavioral traits. These analyses involve the identification of populations that differ in a trait of interest with a strong genetic component and the estimation of summary statistics along the genomes of these populations to detect *selective sweeps* (*see* also Chapter 1). Selective sweeps can derive from positive or divergent selection and evidence for these events include reductions in nucleotide diversity and increased linkage disequilibrium within populations. Elevated differentiation between populations also provides evidence for selective sweeps [29, 30]. The number of studies examining the genetic basis of phenotypic traits in birds is increasing and one pattern that is emerging is the importance of inversions for the control of phenotypic traits. Inversions are a form of rearrangement where portions of a chromosome are flipped, disrupting chromosome pairing during meiosis and preventing recombination from occurring. These regions should be inherited largely as single units, explaining how loci involved in the expression of phenotypic traits can work in concert.

One well-known example of an inversion associated with a phenotypic trait in birds comes from the ruff (*Philomachus pugnax*) where an inversion controls the expressive of different reproductive morphs. Additional examples are beginning to accumulate and include the willow warbler (*Phylloscopus trochilus*) and white-crowned sparrow (*Zonotrichia leucophrys gambelii*). Two subspecies of willow warblers form a ring around the Baltic Sea and differ in migratory orientation, forming a migratory divide in central Scandinavia. Using genome-wide SNP data, Lundberg et al. [31] identified three regions of the genome that exhibit extremely high

differentiation compared to the rest of the genome and form distinct haplotype clusters. These clusters suggest recombination is rare in these regions and could be explained by inversion polymorphisms. Haplotypes at one region correlated with environmental features (altitude and latitude) while the other two correlated with migratory orientation. Tuttle et al. [32] used WGS to identify a series of potential inversions on avian chromosome 2 between two color morphs of white-crowned sparrows that also differ in reproductive behavior (promiscuous vs. monogamous). This region spans ~100 Mb and includes 1100 genes. F_{ST} between morphs is elevated in this region and linkage disequilibrium is high. One morph is homozygous for alleles at this inversion and the other heterozygous. These authors used a phylogenetic analysis to show the inversion evolved before sparrows diverged from their most closely related relative.

There are several benefits of using the bottom-up approach of population genomics to study the genetic basis of phenotypic traits. Early work on this topic was limited to a set of candidate genes that were often identified in model organisms that were distantly related from the focal species (e.g., [33]). Work with HTS allows researchers to study all genes in the genome, permitting unbiased assays of genomic variation and allowing for the de novo identification of candidate genes and new biological processes underlying traits (Fig. 1c). This genome-wide perspective also provides a broader understanding of how phenotypic traits are controlled, including the number, size and distribution of loci that underlie these traits. The bottom-up approach also has considerably more power than other methods. For example, genome-wide association studies (GWAS) can be used, but often require data from hundreds of individuals as they are conducted in single populations that vary in a trait of interest and have low levels of linkage disequilibrium. Bottom-up approaches can use of as few as 10 individuals/population which can be important for nonmodel organisms like birds where large numbers of individuals may be hard to sample. Nevertheless, there are still some problems associated with the bottom-up approach, including the fact that processes other than positive or divergent selection can generate signals similar to selective sweeps (e.g., *background selection*, *see* also Chapter 1). We discuss these problems and potential solutions below.

2.4 Locus-Level Work to Understand the Genetics of Adaptation and Speciation

Similar to work focused on identifying the genetic basis of phenotypic traits, the estimation of summary statistics along the genome can be used to study the processes of adaptation and speciation more broadly. In this case, studies normally compare closely related populations from the same or related species and focus on estimates of genomic differentiation like F_{ST}. One of the chief findings from this work is that differentiation can be highly variable across the genome, with areas of elevated F_{ST} interspersed with areas of

reduced F_{ST} (e.g., [34–39]). An important inference drawn from these findings is that speciation can proceed through a few focal changes and does not require divergence across the entire genome (Fig. 1d, [40]). While this conclusion does not seem to be controversial, the processes that generate variable patterns of differentiation have received considerable interest and include two main models, divergence-with-gene-flow and selection-in-allopatry.

The divergence-with-gene-flow model posits that divergent selection at loci involved in speciation must be protecting some regions of the genome from gene flow, elevating an otherwise homogenized (or low) landscape of differentiation [41, 42]. This model received considerable enthusiasm when it was first developed as it suggests researchers can identify loci involved with speciation relatively easily, by looking for areas of elevated differentiation between closely related populations. Recent work has encouraged caution with this approach and work with birds has been at the forefront of this wave, promoting a second model to explain variation in F_{ST}, selection-in-allopatry. The selection-in-allopatry model posits that variation in the strength of selection can explain variation in differentiation on its own [43]. This model derives from the observation that F_{ST} is a relative measure of differentiation that includes a term for within population variation. As a result, it can be elevated by reductions in within population variation alone. These reductions can derive from any form of linked selection, including *genetic hitchhiking* and background selection that is unrelated to speciation ([43, 44]; also increases in neutral variance generated by population structure [45]) and mean that gene flow is not necessarily needed to explain variation in genomic differentiation.

Contrasts between windowed-estimates of F_{ST} and d_{XY} have been used to support the selection-in-allopatry model. d_{XY} is an absolute measure of differentiation that does not include a term for within population variation. Accordingly, it should be unaffected by reductions in within population variation and show limited associations with estimates of F_{ST} across the genome. Work with birds supports this prediction. Burri et al. [46] estimated F_{ST} and d_{XY} between collared and pied flycatchers and noted that d_{XY} was not elevated where F_{ST} was elevated. In fact, d_{XY} seemed to show the opposite pattern of F_{ST}, being reduced where F_{ST} was elevated. Burri et al. [46] argued that recurrent linked selection in regions of reduced recombination in ancestral populations were responsible for this pattern. This form of selection would reduce d_{XY} to zero prior to population splitting. Similar findings have been documented between Swainson's thrushes [35], greenish warblers (*Phylloscopus trochiloides*, [36]), stonechats (*Saxicola rubicola*, [37]), and Darwin's finches (*Geospiza fortis*, [38]).

Burri et al. [46] added an additional dimension to the selection-in-allopatry model. These authors documented an association between F_{ST} and recombination rates in flycatchers,

suggesting genomic features like reduced recombination that extend the effects of linked selection by preventing linked neutral sites from recombining off their shared background could also play a role in generating variation in F_{ST}. Delmore et al. [47] evaluated this idea further using a comparative analysis, estimating genomic differentiation (F_{ST} and d_{XY}) between eight population pairs of birds that span a broad taxonomic scale (sharing a common ancestor ~52 million ya). Features of the local genomic landscape are highly conserved across birds, including chromosome number, recombination rate and synteny [48–54]. Accordingly, Delmore et al. [47] predicted that if genomic features are generating variation in differentiation across genomes they should generate correlated patterns of differentiation across population pairs of birds. In support of this prediction, a significant proportion of variation in windowed-estimates of F_{ST} and d_{XY} could be explained by correlations across pairs (up to 3% for F_{ST} and 26% for d_{XY}). In addition, genomic regions showing high repeatability across pairs were correlated with several genomic features (e.g., reduced recombination rates [approximated using GC content], elevated gene densities and chromosome size [higher on micro- vs. macrochromosomes]).

As a final note on the genetics of adaptation and speciation, support for the divergence-with-gene-flow model versus the selection-in-allopatry model will likely depend on the geographic context in which speciation occurs. Much of the work on speciation genomics in birds focuses on species in the temperate region that have experienced periods of allopatry and sympatry [55]. Accordingly, a model including allopatric periods (selection-in-allopatry) will likely be more relevant. In addition, there are variants to the selection-in-allopatry model. For example, Delmore et al. [35] and Irwin et al. [36] describe a scenario where selective sweeps upon secondary contact could also reduce d_{XY} in regions of elevated F_{ST}, with globally adaptive alleles evolving during allopatric periods and sweeping across both populations when they come into secondary contact. These alternatives are not mutually exclusive.

3   Roadblock: Genome Assemblies, Novel Genes and Structural Variants

Much of the work in population genomics makes use of reference genomes to place (or map) resequencing reads from individuals or populations. Reference genomes for birds are of variable quality [56]. As noted earlier, the first references were assembled using Sanger sequencing. They represent the most complete reference genomes for birds, but are not perfect. For example, annotations are still lacking for some of the microchromosomes in these references [57] and most chromosomes still include random sequences that cannot be placed. Each reference also includes unassigned scaffolds with unknown chromosome location. More recent

genomes assembled using HTS often integrate data from short insert libraries and mate pairs. Data from small insert libraries are used to construct *contigs* and contigs are combined into *scaffolds* using mate pair libraries. Mate pair libraries help bridge complex regions of the genome that cannot be assembled (e.g., highly repetitive or heterozygous regions) but resulting genomes remain quite incontiguous, consisting of thousands of scaffolds for genome that consist of only 33 chromosomes (on average). In addition, these scaffolds are often assigned to chromosomes using synteny with the chicken or zebra finch genome. While karyotype and synteny are conserved across birds, intrachromosomal rearrangements are quite common [58] and it is possible that regions of the genome containing genes relevant to your focal species will not be present in these highly domesticated species. It is also highly likely that structural variations (e.g., inversions and duplications) controlling phenotypic traits and involved in adaptation and speciation are missing or misassembled.

One of the next steps in the population genomics of birds will be to improve these genome assemblies. Linkage maps would help join scaffolds but are hard to generate in birds as for several crosses cannot easily be made in the lab and a limited number of pedigreed populations in the wild exist. Nevertheless, alternative methods to join scaffolds are being developed and include both long-read technologies (e.g., Nanopore and $10\times$ genomics) and optical mapping with BioNano technology (Fig. 2a). Optical maps are generated by shearing DNA into large molecules (>250 kb), linearizing them in nanochannels, and barcoding them with restriction enzymes. These maps are visualized with fluorescence microscopy and generated for each fragment before being combined into a consensus map. Nick sites from restriction enzymes are used to order and orient HTS scaffolds generated using traditional techniques and estimate gap size between them, ultimately producing hybrid assemblies with an increased N50 (defined as the minimum threshold of sequence length above which all scaffolds cover at least 50% of the total assembly size) and less, longer scaffolds for each genome. One example of a hybrid genome assembly generated using optical mapping comes from the ostrich (*Struthio camelus*). The original Illumina-based assembly had an N50 of 3.59 Mb and 414 scaffolds. Using optical mapping the N50 was increased to 17.71 and the number of scaffolds was reduced to 75 [59].

The desired quality of reference genomes will depend on the objectives of each study. If for example, researchers are interested in specific genes associated with a trait of interest, they may be missing from annotation. If structural variants are of interest (and results from birds thus far suggest they may be), more contiguous genomes will be needed to identify them. On the other hand, many studies, especially those interested in genome-level processes like demography and population structure, may not need a high quality
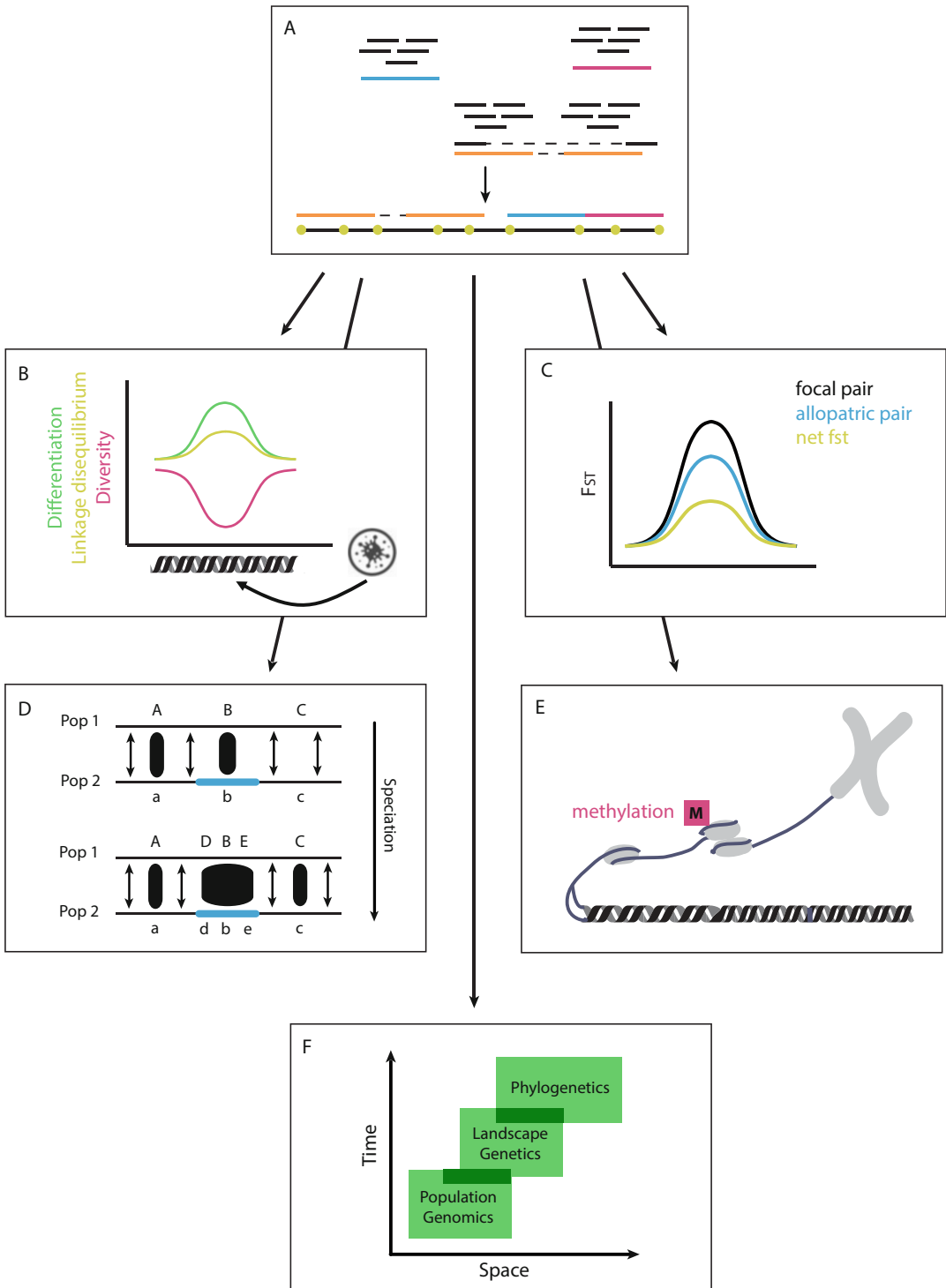
Fig. 2 Conceptualized figure showing roadblocks and prospects for future work for population genomics in birds. Panel **a** shows a roadblock that will need to be overcome (incomplete *genome assemblies*), using optical mapping (yellow) to extend traditional genomes based on contigs (blue and pink) and scaffolds (orange)

reference genome as specific loci or variants are not of interest. The same considerations apply to choosing sequence technologies for resequencing individuals or populations. If locus-level questions are of interest, whole genome resequencing data will likely be needed as linkage disequilibrium can drop off quite quickly in birds [35, 60]. If specific loci are not important (e.g., for analyses of population structure or preliminary work characterizing genomic regions), restriction site associated DNA sequencing (RAD) or Genotyping-by-sequencing (GBS), where DNA is cut with restriction enzymes and sequencing limited to those cut sites may be sufficient [61]. Target capture approaches can also be used if specific loci are not important, including enrichment of ultraconserved elements and their flanking regions [62].

4   Prospects

4.1   Continued Application of Population Genomics for Conservation

The use of HTS data can provide more accurate estimates of genetic diversity in threatened populations and allow for the development of reference platforms to quickly survey populations. Another major goal of conservation genetics is to identify population structure and delineate species boundaries. A few studies have begun doing this (e.g., Bell's vireo [*Vireo bellii*, [62]]; Mottled duck [*Anas fulvigula*, [63]]), but caution will be needed as the increased resolution provided by HTS may permit the identification of fine scale population structure that does not warrant separate taxonomic status. Work to identify adaptive variants important for population survival and maintenance will also be important (Fig. 2b). To the best of our knowledge this has not yet been done in birds, but there are examples from fisheries; for example, two ecotypes of kokanee salmon (*Oncorhynchus nerka*) that differ in their reproductive strategies (stream vs. shore-spawning) exist that are panmictic outside the breeding season and morphologically indistinguishable. These ecotypes are currently managed as separate

Fig. 2 (continued) from short read HTS assemblies. The remaining panels show prospects, including (**b**) the potential to identify *selective sweeps in populations of conservation* concern associated with traits of importance (here a region protecting the population from pathogen infection), (**c**) controlling for the effects *processes unrelated to adaptation and speciation* have on differentiation—estimating F_{ST} between a focal pair and allopatric populations that do not differ in the trait of interest, subtracting allopatric F_{ST} from focal F_{ST} to obtain net F_{ST}, (**d**) extension of Fig. 1d (right) with genomic features (shown in blue, e.g., reduced recombination) that extend the effects of linked selection causing alleles nearby and those under selection to fix at later *stages of speciation*, (**e**) complete reference genomes with high quality annotation allow for analyses of general *regulatory and epigenetic mechanisms*, such as characterization of open chromatin regions, histones, and mapping of chemical marks (e.g., methylation), and (**f**) *integrating* three fields together that differ in their temporal and geographic scales to gain a broader understanding of behavior, ecology, evolution, and conservation (based on [80])

stocks and marker based genetic characterization supports this separation [64, 65]. These authors used RAD sequencing to show these ecotypes are genetically differentiated from one another and identified 12 contigs using transcriptome sequencing that matched pathogens known to reduce the fitness of salmonids. These contigs were limited to stream-spawners, suggesting this ecotype has evolved a way to reduce pathogen load. Combined these results suggest different management strategies for both ecotypes. *See* Chapter 14 for more information on the genomics of fishes.

4.2 Control for Alternative Processes in Genome Scans and Expand Studies to Focus on the *Process* of Speciation

The estimation of genomic differentiation between closely related populations had led to the observation that levels of differentiation are highly variable across the genome. There is interest in using these differentiation patterns to identify loci involved in adaptation and speciation but processes other than positive or divergent selection can elevate differentiation and include background selection that is not related to adaptation or speciation. Accordingly, future work on this topic must control for regions affected by these processes and the framework outlined by Vijay et al. [66] using the crow species complex (genus *Corvus*) provides one approach. These authors obtained WGS data from populations at different stages of differentiation and from different color morphs (all-black or pied). One of the objectives of their study was to identify genomic regions associated with this color difference using populations in three hybrid zones between the morphs. They estimated F_{ST} between allopatric populations that did not differ in color and used these values as a null beyond which F_{ST} in hybrid zones had to exceed to provide evidence of positive selection relevant to color (Fig. 2c). Interestingly, these hybrid zones did not share the same peaks of differentiation, suggesting different genes were involved in each color transition. Instead, each contact zone had at least one gene in a differentiated region that was associated with the Wnt signalling component of the melanogenesis pathway suggesting the pathways but not necessarily the same genes to be important.

Scans of genomic differentiation across genomes can also be expanded to study the *process* of speciation, rather than a snapshot in what is a continuous process from initial population divergence to complete reproductive isolation (Fig. 2d). Speciation is also not a deterministic process; populations can go extinct of fuse back into a single unit before reaching speciation. Results from Delmore et al. [47] provide one example. Recall, these authors documented consistent patterns of differentiation across eight species pairs of birds implicating linked selection at genomic features in generating variation in genomic differentiation. Linked selection at genomic features is expected to have a greater effect at later stages of speciation when enough mutations have accumulated in populations to reflect processes occurring at these features [67]. Under this prediction, patterns of species pairs at later stages of speciation should show

increasing levels of consistency. The species pairs included in Del-more et al. [47] span the full speciation continuum and an analysis comparing correlation coefficients between population pairs in windowed-estimates of F_{ST} support this prediction, with pairs later in the continuum (e.g., with more narrow hybrid zones) exhibiting more consistent patterns. This pattern was not observed using d_{XY} and likely relates to the fact that d_{XY} reflects processes that have accumulated over multiple speciation events. Accordingly, it will not matter when d_{XY} is estimated in the process of speciation, it will always reflect linked selection at genomic features.

4.3   Expand Beyond Studies of Genetic Variation Alone

Thus far we have focused mainly on genetic variants, SNPs or structural features like inversions. As noted, more direct evaluations of structural variants are needed but additional expansions would also profit research in birds. For example, it is possible that in addition to genetic sequence variation, expression dynamics and/or epigenetic mechanisms often play an important role in the expression of complex traits (Fig. 2e). These traits are typically controlled by many loci of small effect [68, 69] that may depend on the regulation of gene expression to produce the downstream phenotype ([70] but *see* [71–73]). Changes in gene expression can derive from chemical and molecular modifications, including methylation, phosphorylation, acetylation, accessibility of DNA in chromatin, and occupancy of regulatory sequences by transcription factors.

Epigenetic studies are beginning to accumulate in birds and include investigations of vocalization (zebra finch [74]); learning and cognition (great tit [*Parus major*, [75]]), and beak size and shape (Darwin's finches, [76, 77]). It is likely that these studies will increase in prevalence as the field moves from studying relatively simple traits (e.g., color) to more complex traits (e.g., behavior, such as migration). Along with this expansion, greater precision will be needed, and study designs require a carefully controlled experimental framework, making sure hindering noise (diet has a huge effect on expression profiles, profiles vary significantly across sexes, developmental stages) is kept to a minimum. Even within tissues, especially within brain, epigenetic makeup depends on the exact location of brain area. In birds this has been studied for bird song: behaviorally regulated gene expression profiles vary depending on time and between anatomical structures. It was hypothesized that drivers of this variability are signalling cascades modulate by transcription factors, cis and trans regulatory regions and epigenetic chromatin states [74]. This study focused on identifying transcriptionally active chromatin regions for song nuclei involved in their focal trait (singing). Their results suggest the presence of epigenetic modifications that prime gene regulation differ between brain regions, so that specific target regions are in a chromatin state

that allows to immediately modulate transcription of behavior specific genes once the behavior kicks in (upon neuronal firing).

4.4 Integrate Population Genomics with Additional Fields

We will finish our review by highlighting the importance of integrating population genetics with additional fields to answer broader questions relevant to ecology, behavior, conservation, and evolution. Specifically, the fields of landscape genetics and phylogeography were originally developed as bridges to population genetics, to study the interaction between landscape features and micro- or macroevolutionary processes, respectively. Both involve geographic sampling beyond what is done for population genetics. The advent of HTS has blurred the boundaries between these three fields, with principles from landscape genetics and phylogeography being used to (1) identify loci under selection, (2) identify correlations between genomic data and the environment, and (3) reconstruct their history of divergence within species [60, 80]. One example comes from [78] who focused on populations of white-breasted nuthatches (*Sitta carolinensis*) that occupy the sky islands of Arizona. RAD sequencing showed that genetic differentiation between these islands was mediated largely by ecological distance rather than geographic distance, identifying eight loci associated with ecological distance (e.g., strong associations with minimum precipitation of driest month and maximum temperature of the hottest month). This integration provides a more complete picture of adaptation and speciation (Fig. 2f) and could be especially important for conservation, allowing researchers to understand how organisms relate to their environment and will handle future changes to their landscape.

Similar integrations with genetic mapping have the potential of being quite powerful as well, using the top-down approach of genetic mapping to identify loci associated with a trait of interest and the bottom-up approach of population genomics to examine the selective context underlying these traits. Work by Delmore et al. [79] demonstrates the utility of this integration. These researchers focused on a hybrid zone between two subspecies groups of Swainson's thrushes (*Catharus ustulatus*) in western North America that differ in migratory orientation. Hybrids between these groups take intermediate and inferior routes, likely helping reduce gene flow between subspecies (i.e., maintain subspecies boundaries). Delmore et al. [79] focused on hybrids, tracking these birds on migration with light-level geolocators and genotyping them using GBS. *Admixture mapping* using these data identified a region on chromosome 4 associated with migratory orientation. This region includes 60 genes, many involved in cell signalling, the nervous system and the circadian clock and had been identified in smaller-scale studies of migration in other animal groups supporting the idea there may be a common genetic basis to migration in animals. Delmore et al. [79] took traditional genetic mapping one step

further by obtaining resequencing data form pure populations and showing this region is strongly differentiated between the subspecies, suggesting it is under divergent selection, supporting behavioral observations of inferior hybrid behavior showing differences in migration could be helping maintain subspecies boundaries.

5   Conclusion

We look forward to both participating and watching how the field of population genomics evolves for birds in the coming years. As sketched in this chapter, birds have played an important role in the development and application of population genomics, with work on demography and genetic variation at the genome level informing evolution and conservation and work at the locus level providing information on the genetic basis of phenotypic traits, adaptation and speciation. The improvement of reference genomes, use of long-read mapping for studying structural variation and expansion beyond genetic variation to epigenetics will take us a long way to expanding population genomics in birds. Integrations with other fields and beyond landscape genetics and phylogenetics (e.g., inclusion of functional genomics and genome editing) are also on the horizon for some species and will undoubtedly help to unravel many mysteries concerning the behavior, ecology, evolution, and conservation of birds.

Glossary

Admixture mapping: A genetic mapping approach that makes use of recombination and backcrossing in natural hybrid zones instead of producing crosses (Quantitative trait locus [QTL] mapping) and/or examining variation within a single population (Genome wide association studies [GWAS]).

Background selection: Reduction in diversity at sites linked to a variant under purifying selection, removing deleterious mutations from the populations.

Contigs: Continuous sequences derived from overlapping reads.

Genetic hitchhiking: Reduction in diversity at sites linked to variant under positive selection.

High-throughput sequencing: Set of platforms that sequence DNA from multiple genomic regions and individuals in parallel.

Inbreeding depression: Reduced fitness of a population resulting from inbreeding, the breeding of related individuals.

Linkage disequilibrium: Nonrandom association of loci in the genome, causing specific loci to be inherited together more often than expected by chance.

Phasing: Estimation of haplotypes from genotype data, of alleles on a particular chromosome that are inherited together.

Scaffolds: Series of contigs linked together. Noncontinuous—contigs separated by gaps of known length.

Selective sweeps: The reduction or loss of variation at loci linked to one experiencing positive or divergent selection.

Whole genome resequencing (WGS): Form of HTS that theoretically obtains data from the entire genome. When Illumina technology is used, this results in short reads (no more than 250 bp) with insert size of ~300 kb that are aligned/mapped to a reference genome and used to call variants, often SNPs.

References

1. Nice M (1939) The watcher at the nest. Macmillan, New York

2. Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, Michigan

3. Luikart G, England PR, Tallmon D et al (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994

4. Begun DJ, Holloway AK, Stevens K et al (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol 5:e310

5. Avise JC, Nelson WS (1989) Molecular genetic relationships of the extinct dusky seaside sparrow. Science 243:646–648

6. Zink RM, Kale HW (1995) Conservation genetics of the extinct dusky seaside sparrow *Ammodramus maritimus nigrescens*. Biol Conserv 74:69–71

7. Stapley J, Reger J, Feulner PGD et al (2010) Adaptation genomics: the next generation. Trends Ecol Evol 25:705–712

8. Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

9. Gregory TR (2005) Synergy between sequence and size in large-scale genomics. Nat Rev Genet 6:699–708

10. Organ CL, Shedlock AM, Meade A et al (2007) Origin of avian genome size and structure in non-avian dinosaurs. Nature 446:180–184

11. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

12. Warren WC, Clayton DF, Ellegren H et al (2010) The genome of a songbird. Nature 464:757–762

13. Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol 29:51–63

14. Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol 26:2097–2108

15. Zeng K (2013) A coalescent model of background selection with recombination, demography and variation in selection coefficients. Heredity 110:363

16. Bank C, Ewing GB, Ferrer-Admettla A et al (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. Trends Genet 30:540–546

17. Beissinger TM, Wang L, Crosby K et al (2016) Recent demography drives changes in linked selection across the maize genome. Nat Plants 2:16084

18. Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. Proc R Soc Lond B Biol Sci 271:1881–1887

19. Pinho C, Hey J (2010) Divergence with gene flow: models and data. Annu Rev Ecol Evol Syst 41:215–230

20. Borge T, Lindroos K, Nádvorník P et al (2005) Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. J Evol Biol 18:1416–1424

21. Sætre G-P, Borge T, Lindell J et al (2001) Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. Mol Ecol 10:737–749

22. Nadachowska-Brzyska K, Burri R, Olason PI et al (2013) Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. PLoS Genet 9:e1003942

23. Saetre G-P, Saether SA (2010) Ecology and genetics of speciation in *Ficedula* flycatchers. Mol Ecol 19:1091–1106

24. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H (2016) PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol 25:1058–1072

25. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. Nature 475:493–496. https://doi.org/10.1038/nature10231

26. Nadachowska-Brzyska K, Li C, Smeds L et al (2015) Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr Biol 25:1375–1380

27. Li S, Li B, Cheng C et al (2014) Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. Genome Biol 15:557

28. Kardos M, Taylor HR, Ellegren H et al (2016) Genomics advances the study of inbreeding depression in the wild. Evol Appl 9:1205–1218

29. Siol M, Wright SI, Barrett SC (2010) The population genomics of plant adaptation. New Phytol 188:313–332

30. Nielsen R, Williamson S, Kim Y et al (2005) Genomic scans for selective sweeps using SNP data. Genome Res 15:1566–1575

31. Lundberg M, Liedvogel M, Larson K et al (2017) Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evol Lett 1:155–168

32. Tuttle EM, Bergland AO, Korody ML et al (2016) Divergence and functional degradation of a sex chromosome-like supergene. Curr Biol 26:344–350

33. Delmore KE, Liedvogel M (2016) Investigating factors that generate and maintain variation in migratory orientation: a primer for recent and future work. Front Behav Neurosci 10:3

34. Ellegren H, Smeds L, Burri R et al (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760

35. Delmore KE, Hübner S, Kane NC et al (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol Ecol 24:1873–1888

36. Irwin DE, Alcaide M, Delmore KE et al (2016) Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Mol Ecol 25:4488–4507

37. Van Doren BA, Campagna L, Helm B et al (2017) Correlated patterns of genetic diversity and differentiation across an avian family. Mol Ecol 26:3982–3997

38. Han F, Lamichhaney S, Grant BR et al (2017) Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res 27:1004–1015

39. Poelstra JW, Vijay N, Bossu CM et al (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414

40. Wu C-I (2001) Genes and speciation. J Evol Biol 14:889–891

41. Nosil P (2012) Ecological speciation (Oxford series in ecology and evolution). Oxford University Press, Oxford

42. Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375–402

43. Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23:3133–3157

44. Noor MA, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444

45. Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it. . .? Why are FST outliers sometimes so frequent? Mol Ecol 22:2061–2064

46. Burri R, Nater A, Kawakami T et al (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res 25:1656–1665

47. Delmore KD, Lugo JS, Van Doren BM et al (2018) Comparative analysis examining patterns of genomic differentiation across multiple

episodes of population divergence in birds. Evol Lett 2:76–87

48. Dawson DA, Akesson M, Burke T et al (2007) Gene order and recombination rate in homologous chromosome regions of the chicken and a passerine bird. Mol Biol Evol 24:1537–1552

49. Griffin DK, Robertson LBW, Tempest HG, Skinner BM (2007) The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res 117:64–77

50. Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. Biol Lett 2:435–438

51. Stapley J, Birkhead TR, Burke T, Slate J (2008) A linkage map of the zebra finch Taeniopygia guttata provides new insights into avian genome evolution. Genetics 179:651–667

52. Kawakami T, Smeds L, Backström N et al (2014) 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 23:4035–4058

53. Zhang G, Li C, Gilbert M et al (2014) Comparative genomic data of the avian phylogenomics project. GigaSci 3:26

54. Kawakami T, Mugal CF, Suh A et al (2017) Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. Mol Ecol 26:4158–4172

55. Hewitt G (2000) The genetic legacy of the quaternary ice ages. Nature 405:907–913

56. Botero-Castro F, Figuet E, Tilak M-K et al (2017) Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. Mol Biol Evol 34:3123–3131

57. Miller MM, Robinson CM, Abernathy J et al (2013) Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. J Hered 105:203–215

58. Skinner BM, Griffin DK (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity 108:37–41

59. Zhang J, Li C, Zhou Q, Zhang G (2015) Improving the ostrich genome assembly using optical mapping data. GigaSci 4:24

60. Edwards SV, Shultz AJ, Campbell-Staton S (2015) Next-generation sequencing and the expanding domain of phylogeography. Folia Zool 64:187–206

61. Puritz JB, Addison JA, Toonen RJ (2012) Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. PLoS One 7:e34241

62. Faircloth BC, McCormack JE, Crawford NG et al (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol 61:717–726

63. Klicka LB, Kus BE, Burns KJ (2016) Conservation genomics reveals multiple evolutionary units within Bell's Vireo (*Vireo bellii*). Conserv Genet 17:455–471

64. Lemay MA, Donnelly DJ, Russello MA (2013) Transcriptome-wide comparison of sequence variation in divergent ecotypes of kokanee salmon. BMC Genomics 14:308

65. Lemay MA, Russello MA (2015) Genetic evidence for ecological divergence in kokanee salmon. Mol Ecol 24:798–811

66. Vijay N, Bossu CM, Poelstra JW et al (2016) Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun 7:13195

67. Burri R (2017) Interpreting differentiation landscapes in the light of long-term linked selection. Evol Lett 1:118–131

68. Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. Nature 452:169–175

69. Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577

70. Carroll SB (2000) Endless forms: the evolution of gene regulation and morphological diversity. Cell 101:577–580

71. Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995–1016

72. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206–216

73. Albert FW, Kruglyak L (2015) The role of regulatory variation in complex traits and disease. Nat Rev Genet 16:197–212

74. Whitney O, Pfenning AR, Howard JT et al (2014) Core and region-enriched networks of behaviorally regulated genes and the singing genome. Science 346:1256780

75. Laine VN, Gossmann TI, Schachtschneider KM et al (2016) Evolutionary signals of selection on cognition from the great tit genome and methylome. Nat Commun 7:10474

76. Skinner MK, Gurerrero-Bosagna C, Haque MM et al (2014) Epigenetics and the evolution

of Darwin's finches. Genome Biol Evol 6:1972–1989

77. McNew SM, Beck D, Sadler-Riggleman I et al (2017) Epigenetic variation between urban and rural populations of Darwin's finches. BMC Evol Biol 17:183

78. Manthey JD, Moyle RG (2015) Isolation by environment in white-breasted Nuthatches (Sitta carolinensis) of the Madrean Archipelago

sky islands: a landscape genomics approach. Mol Ecol 24:3628–3638

79. Delmore KE, Toews DPL, Germain RR et al (2016) The genetics of seasonal migration and plumage color. Curr Biol 26:2167–2173

80. Edwards SV, Potter S, Schmitt CJ et al (2016) Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proc Natl Acad Sci 113:8025–8032

Further Reading

Alkan C et al (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12:363–376

Carstens A et al (2013) How to fail at species delimitation. Mol Ecol 22:4369–4383

Edwards SV et al (2015) Next-generation sequencing and the expanding domain of phylogeography. Folia Zool 64:187–206

Feltus FA (2014) Systems genetics: a paradigm to improve discovery of candidate genes and

mechanisms underlying complex traits. Plant Sci 223:45–48

Manel N et al (2010) Perspectives on the use of landscape genetics to determine genetic adaptive variation in the field. Mol Ecol 19:3760–3772

Wolf JB, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet 18:87–100