Pierre Pontarotti *Editor*

Evolutionary Biology: Biodiversification from Genotype to Phenotype

Springer

Evolutionary Biology: Biodiversification from Genotype to Phenotype

Pierre Pontarotti
Editor

Evolutionary Biology: Biodiversification from Genotype to Phenotype

Springer

Editor
Pierre Pontarotti
Evolution Biologique et Modélisation
I2M UMR 7373
Aix Marseille Université/CNRS
Marseille
France

ISBN 978-3-319-19931-3 ISBN 978-3-319-19932-0 (eBook)
DOI 10.1007/978-3-319-19932-0

Library of Congress Control Number: 2015942234

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

For the 18th time, the Evolutionary Biology Meeting in Marseille took place from 16 to 19 September 2014. The goal of this annual meeting is to allow scientists of different disciplines, who share a deep interest in evolutionary biology concepts, knowledge and applications, to meet and exchange and enhance interdisciplinary collaboration.

The Evolutionary Biology Meeting in Marseille is now recognised internationally as an important exchange platform and a booster for the use of evolutionary-based approaches in biology and also in other scientific areas.

This year, more than 100 presentations were selected by the evolutionary biology meeting scientific committee. We have further selected 20 of the most representative ones for the book.

The book will give the reader an overview of the state-of-the-art research in the evolutionary biology field. The book is the eighth one that we have published further to the meeting. We would like to underline that the seven books are complementary one to another and should be considered as volumes.

The readers of the evolutionary biology books as well as the meeting participants will maybe like us witness years after years during the different meetings and book editions a shift in the evolutionary biology concepts and knowledge. The fact that the chapters of the book are selected from one meeting enables the quick diffusion of the novelties.

Concerning the book, the chapters are organised in the following categories

Genotype to Phenotype (Chaps. 1–3)
Genetic Mechanisms of Diversification (Chaps. 4–8)
Evolutionary Mechanisms (Chaps. 9–14)
Speciation and Biodiversity (Chaps. 15–20)

We would like to thank all the authors, the meeting participants, and the sponsors: Aix Marseille Université (AMU), CNRS, ITMO GGB and BCDE, ECCOREV Federation, Conseil Général 13, Ville de Marseille.

We also wish to thank the Springer editing staff and in particular Andrea Schlitzberger for her competence and help.

We wish also to thank members of the Association pour l'Etude de l'Evolution Biologique (AEEB) for their help in the meeting organisation, as well as AMU for its logistic support.

Marseille, France
April 2015

Pierre Pontarotti

Contents

Part IV Speciation and Biodiversity

Part I
Genotype to Phenotype

Chapter 1
Functional Mapping: How to Map Genes for Phenotypic Plasticity of Development

Lidan Sun, Libo Jiang, Meixia Ye, Xuli Zhu, Jing Wang, Kirk Gosik and Rongling Wu

Abstract Functional mapping is a statistical tool derived to map genes or quantitative trait loci (QTLs) that control the dynamic process of complex traits. In this chapter, we describe an innovative modification of functional mapping to characterize the genetic basis of phenotypic plasticity for the developmental pattern of phenotypic traits. Phenotypic plasticity is a phenomenon by which multiple phenotypes are produced by a single genotype in response to changing environment. Although phenotypic plasticity has been extensively studied in the past decades, new insights into its formation mechanisms can be gained by integrating developmental principles because environmentally induced phenotypes require time to form and build. The new framework for functional mapping enables geneticists to illustrate the genetic architecture of how QTLs cope with environment to regulate the developmental pattern and timing of phenotypic formation. Because of their role in guiding the evolution of complex phenotypes through environmental adaptation, the discoveries of these QTLs facilitate the synthesis of evo-devo and eco-devo.

Keywords Functional mapping · QTL · Phenotypic plasticity · Development · Genetic mapping

L. Sun
Beijing Key Laboratory of Ornamental Plants Germplasm Innovation and Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment and College of Landscape Architecture, Beijing Forestry University, Beijing 100083, China

L. Jiang · M. Ye · X. Zhu · J. Wang
Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

K. Gosik · R. Wu (✉)
Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA
e-mail: rwu@phs.psu.edu

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_1

1.1 Introduction

Phenotypic plasticity, the ability of one genotype to produce multiple phenotypes when exposed to different environments, has been thought to be a key regulator of an organism's adaptation to environmental perturbations (Schlichting 1989; Scheiner 1993). Because of its adaptive significance, much research has been devoted to reveal the causes, mechanistic underpinnings, and consequences of phenotypic plasticity (Sultan 2000, 2003; Agrawal 2001; Schlichting and Smith 2002; Wu et al. 2004). In particular, an increasing body of studies has focused on how phenotypic plasticity assists the organism in maintaining biodiversity and evolution in response to environmental change (Pfennig et al. 2010; Wennersten and Forsman 2012). It has been widely recognized that substantial genetic variation exists in the capacity of phenotypic plasticity, leading to a so-called genotype–environment interaction phenomenon, through change in the relative magnitude or order of phenotypic values among different genotypes over environment (Via and Lande 1985). Genetic mapping has been used to identify specific genes or quantitative trait loci (QTLs) that mediate phenotypic plasticity (Wu 1998; Ma et al. 2008; Lacaze et al. 2009; Tetard-Jones et al. 2011; Fabbrini et al. 2012; Zhai et al. 2014; Zhou et al. 2015). Several theoretical hypotheses have been established to explain the genetic basis of phenotypic plasticity, including overdominance, allelic sensitivity, gene regulation, and epigenesist (El-Soda et al. 2014).

The genetic studies of phenotypic plasticity depend on how a phenotypic trait is defined and described. The formation of any environmentally induced phenotypes require time to take effect, thus a trait can be better viewed as a developmentally, functionally, and phenotypically integrated complex unit. Several studies have begun to explore how the organism is plastic for the developmental pattern and timing of complex traits (Sultan 2000, 2010; Moczek et al. 2011). As shown by a hypothetical example (Fig. 1.1), Amphibian species may display different pathways of larval full length growth when exposed to different stressful environments. Such environment-induced differences in growth trajectories are an excellent representation of developmental plasticity. Wang et al. (2013c) integrated growth equations to characterize four patterns of developmental plasticity in terms of developmental timing, rate of development, length of development, and developmental interactions (Fig. 1.2). By estimating and testing the mathematical parameters for growth equations, these patterns can be quantified, which can gain new insights into the developmental mechanisms that cause phenotypic plasticity.

Genetic mapping has been implemented to map specific QTLs responsible for the phenotypic plasticity of development. This implementation was based on a dynamic version of genetic mapping—functional mapping, a model derived from the integration of biological principles underlying developmental processes into a setting of QTL mapping (Ma et al. 2002; Wu and Lin 2006; He et al. 2010). Functional mapping is a versatile framework within which many biological

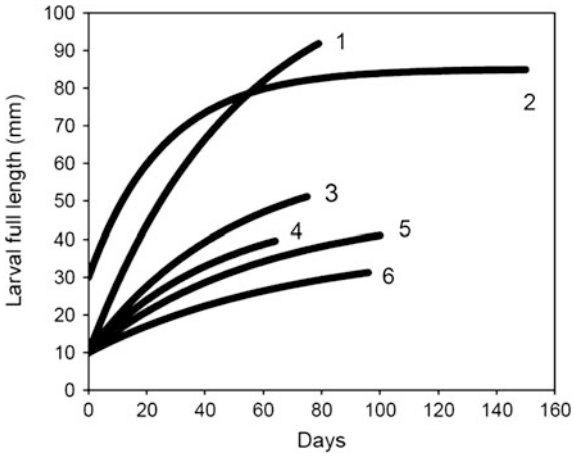


Fig. 1.1 Diagram showing developmental plasticity of larvae full length for an Amphibian species reared in six different stressful environments. In environment 2, the organism has more than two times longer duration of growth, compared to its growth duration in environment 4. In environment 1 and 2, the organism has the largest growth rate, followed by that in environment 3, 4, 5, and 6

phenomena can be incorporated with mathematical equations. Zhao et al. (2004a, b) integrated functional mapping and developmental plasticity to study the genetic architecture of the pattern of development. The renovated model was used to identify environment-induced QTLs for the phenotypic plasticity of body mass growth in mice (Zhao et al. 2004a) and plant height growth in rice (Zhao et al. 2004b) and soybean (Li et al. 2010b). Sun et al. (2014) implemented the concept of heterochrony into functional mapping to detect QTLs that regulate developmental timing and events. By analyzing leaf size and leaf mass growth trajectories for the common bean grown in two contrasting environments, Sun et al.'s model has successfully identified a few heterochrony QTLs (*h*QTLs) that are expressed, depending on the environment (Jiang et al. 2015). The integration of functional mapping, phenotypic plasticity, and heterochrony is leading to new insights into the adaptive nature of developmental plasticity, its underlying genetic mechanisms and its role in the evolutionary diversification of plants.

In this chapter, we provide a brief review of functional mapping in terms of its modeling structure and organization. We focus on the description of how functional mapping sheds light on the genetic control mechanisms of phenotypic plasticity of developmental process and pattern. A few examples of the application of functional mapping to identifying genotype–environment interactions (GEI) expressed by plasticity QTLs are illustrated.

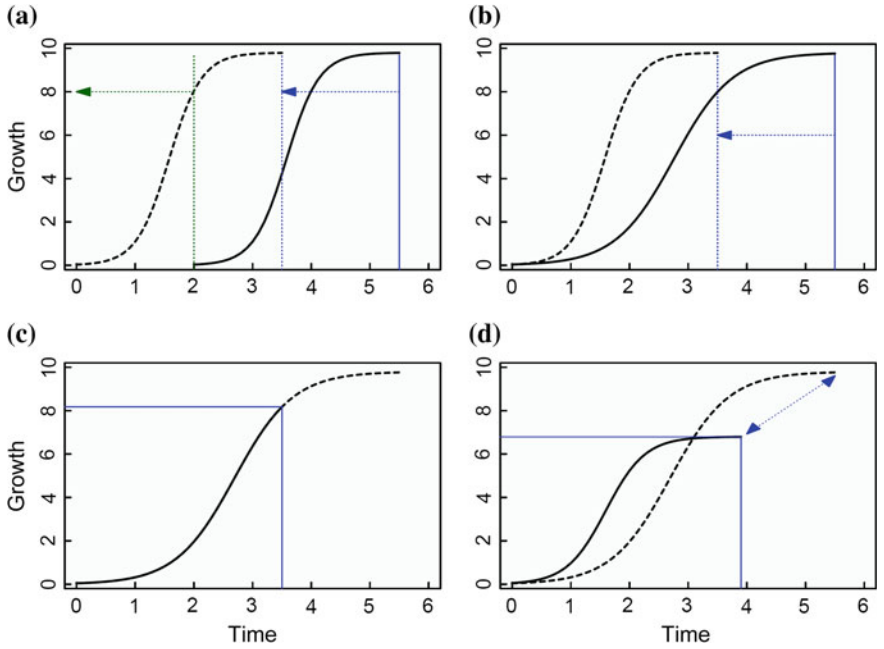


Fig. 1.2 Four patterns of developmental plasticity which arises from the transplantation of the same genotype from its original environment (*solid*) to a new environment (*dash*). **a** Early–late plasticity, in which development starts and stops earlier (shown by *arrows*) in response to environmental change. **b** Slow–fast plasticity, in which the rate of growth is accelerated in the new environment, thus using a shorter time to achieve the maximum growth (shown by an *arrow*). **c** Short–long plasticity, in which growth is prolonged in the new environment (shown by a *dashed curve*). **d** Sequential plasticity, in which the environment leads to decreasing growth in the early stage of development but increasing growth in the late development. Adapted from Wang et al. (2013c)

1.2 Model for Functional Mapping

1.2.1 Basic Principle

Because of tremendous efforts by geneticists worldwide, a vast amount of mapping materials used to map complex traits for a wide range of species have been established. Here, we assume that such materials, such as recombinant inbred lines (RILs), are available to map QTLs. Consider the mapping population composed of n RILs, each genotyped for a panel of molecular markers from which to construct a genetic linkage map that covers the entire genome. Since each RIL presenting a genotype can generate multiple copies, we rear the mapping population repeatedly in multiple different environments (say K), allowing GEI to be characterized. In each environment, the same trait of each RIL, such as plant height or diameter, is measured at a series of T time points spanning its lifecycle.

Let $\mathbf{y}_{ki} = (y_{ki}(1), \ldots, y_{ki}(T))$ denote a vector of time-dependent trait values for RIL i at environment k ($k = 1, \ldots, K$). By hypothesizing that growth curves are controlled by a set of QTLs located on the linkage map, functional mapping formulates a mixture likelihood model expressed as

$$\mathrm{L}(\mathbf{y}) = \sum_{i=1}^{n} \left[\omega_{1|i} f_1(\mathbf{y}_{1i}; \ldots; \mathbf{y}_{Ki}) + \cdots + \omega_{J|i} f_J(\mathbf{y}_{1i}; \ldots; \mathbf{y}_{Ki})\right] \tag{1.1}$$

where $\omega_{j|i}$ is the proportion of the jth mixture proportion ($j = 1, \ldots, J$), corresponding to the jth QTL genotype, described by the conditional probability of this QTL genotype given the adjacent marker genotype of RIL I (Wu et al. 2007); $f_j(\mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki})$ is the multivariate normal distribution density function with mean vectors expressed

$$\boldsymbol{\mu}_j = (\mu_{1j}(1), \ldots, \mu_{1j}(T); \ldots; \mu_{Kj}(1), \ldots, \mu_{Kj}(T)) \tag{1.2}$$

and ($KT \times KT$) covariance matrix $\boldsymbol{\Sigma}$, expressed as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \cdots & \boldsymbol{\Sigma}_{1K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{K1} & \cdots & \boldsymbol{\Sigma}_K \end{pmatrix} \tag{1.3}$$

Functional mapping can model time-dependent genotypic means for a QTL genotype j in an environment k (2) by a growth equation, expressed as

$$\mu_{kj}(t) = a_{kj}\left[1 + b_{kj}\mathrm{e}^{-r_{kj}t}\right]^{\frac{1}{1-k_{kj}}} \tag{1.4}$$

where a is the asymptotic value of the trait, b is a parameter to position the curve on the time axis, r is the growth rate that determines the spread of the curve along the time axis, and k is the shape parameter of the curve. By estimating a set of growth parameters (a_{jk}, b_{jk}, r_{jk}, k_{jk}), we can compare genotype-dependent differences over environment and further determine whether and how a QTL affects the phenotypic plasticity of developmental curves.

Functional mapping also chooses a parsimonious statistical model, such as autoregressive, antedependence, or autoregressive moving average (Li et al. 2010a), to structure the covariance $\boldsymbol{\Sigma}$ (3) by a particular set of parameters, thereby increasing the statistical power for QTL detection. Several nonparametric or semi-parametric approaches have also been proposed to model the longitudinal structure of covariance (Yap et al. 2009).

From functional mapping, we can formulate a procedure for testing how the QTL regulates the timing of development and environment-induced differences in growth form. Sun et al.'s (2014) model allows the test of the genetic control pattern of the timing of inflection point, timing of maximum acceleration, timing of maximum deceleration, and the duration of linear growth. Furthermore, by

identifying *h*QTLs, their number, genomic locations and actions and interactions, this model can characterize the genetic mechanisms of developmental plasticity and illustrate the role of these mechanisms in driving the evolutionary change of growth form.

1.2.2 Applications

Cheverud et al. (1996) used two strains, the large (LG/J) and small (SM/J), to generate an F_2 population of 535 progeny from which to construct one of the first genetic linkage maps for the mouse. The same cross experiment was repeated by Vaughn et al. (1999), in which 502 F_2 mice were generated and a linkage map of 1780 cM long was constructed. In both experiments, each F_2 progeny was measured for body mass at 10 weekly intervals starting at age seven days. The raw weights were corrected for the effects of each covariate due to dam, litter size at birth and parity. Zhao et al. (2004a) reported the identification of QTLs causing the phenotypic plasticity of growth curves due to different sexes from the second cross.

Figure 1.3 shows growth trajectories of body mass, separately for different sexes, in the F_2 population of mice. The two sexes exhibit different patterns of growth. Male mice are generally heavier than female mice, due to greater growth rate in the early stage from week 1 to 5 (Fig. 1.3a, b). Such sex-specific differences may be due to specific growth QTLs that display genotype–sex interactions, which can be detected by functional mapping (Zhao et al. 2004a). One QTL on chromosome 11 was identified to affect growth trajectories, but its expression depends on the sex background considerably (Fig. 1.3c). The pattern of age-increasing differences in male mice among three genotypes at this QTL implies its genetic effect on the body mass of male mice to increase with age. It is also interesting to note that this QTL operates in a highly overdominant manner because the two homozygotes have a similar form of growth curves remarkably different from that of the heterozygote. Given its subtle effect on female mice's growth, this QTL is thought to contribute the genotype–sex interaction of growth processes. Although this QTL affects the amount of growth more strikingly in male than in female mice, its effect on the timing of inflection point displays an inverse pattern of sex-specific difference (Fig. 1.3c).

In a doubled-haploid (DH) population of rice derived from semi-dwarf IR64 and tall Azucena, a genetic linkage map that covers 12 rice chromosomes was constructed using 175 markers (Huang et al. 1997). The DH lines were planted each with two replicates at two climatically contrasting locations, Hangzhou (subtropical, 30° 16′N) and Hainan (tropical, 19° 57′N), China. After 10 days of transplanting into the field trial, plant heights were measured every 10 days until all lines had headed. Figure 1.4 illustrates growth curves of 90 DH plants separately for Hangzhou and Hainan. On average, plants in the two environments have different forms of growth trajectories, with the Hangzhou plants growing slightly better than the Hainan plants. Substantial variation in growth curve between the two locations

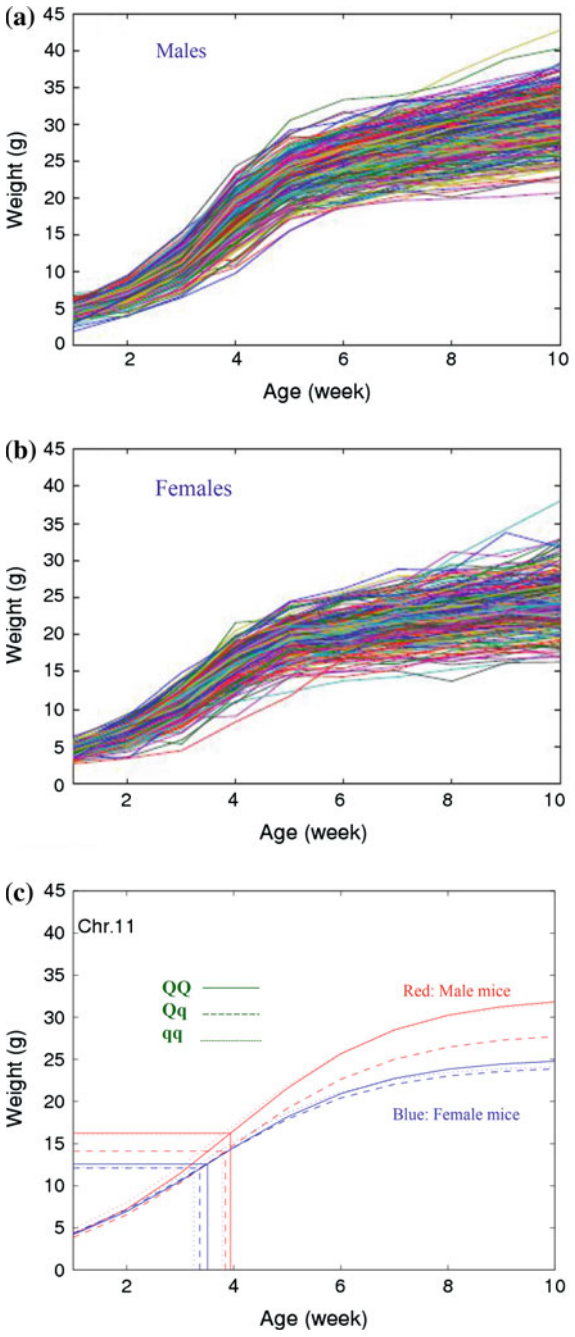


Fig. 1.3 Growth trajectories of body mass for an F_2 mapping population of male (**a**) and female mice (**b**). In **c**, three growth curves each present a genotype, *QQ* (*solid curves*), *Qq* (*dot curves*), or *qq* (*broken curves*), at a QTL detected on mouse chromosome 11, segregating in the male (*red*) and female (*blue*) mice. The times at the inflection point are indicated by the *vertical lines* each corresponding to a QTL genotype in each sex

suggests that specific QTL may be involved in shaping the phenotypic plasticity of developmental trajectories.

Zhao et al. (2004b) used functional mapping to identify several significant QTLs for plant growth curves, but none of these QTLs display significant genotype × environment interactions, suggesting that they are more general in regulating growth. However, these QTLs detected to affect overall growth trajectories also control the timing of maximum height growth rate. It is interesting to find that most QTLs displayed significant genotype × environment interaction effects on this timing. For example, the QTL on chromosome 1 exerts a significant effect on the timing of maximum growth rate for rice grown in Hangzhou, whereas no such an effect was detected in Hainan (Fig. 1.4c). Thus, this QTL can be regarded as an *h*QTL, because it contributes to GEI of growth through affecting the plastic response of developmental timing to environmental change.

An RIL population of soybean was produced by crossing Kefeng No. 1 and Nannong 1138-2 successively for 7 generations. A total of 184 RILs were genotyped for 488 molecular markers to construct a linkage map with 25 linkage groups covering 4,151.2 cM of the soybean genome (Zhang et al. 2004). The RILs, each

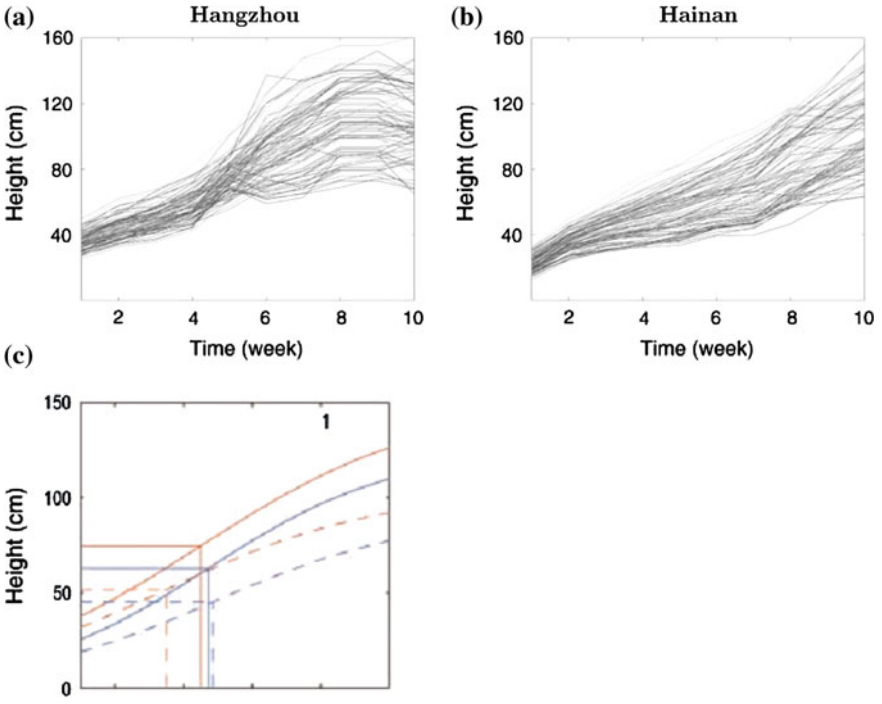


Fig. 1.4 Plant height growth trajectories of rice DH lines plants grown in Hangzhou (**a**) and Hainan (**b**). In **c**, two genotypes of a QTL detected on chromosome 1 are each expressed as a growth curve in Hangzhou (*red curves*) and Hainan (*blue curves*). The times at the inflection point are indicated by the *vertical lines* each corresponding to a QTL genotype in each environment

planted with multiple replicates in the field, were measured for plant height growth for six to eight times with a 10-day interval starting at the 28th day after emergence. The same study was repeated for year 2005 and 2006. This allows genotype–year interactions to be characterized.

Figure 1.5 plots the plant height growth of soybean RILs over time in each year. At the early stage of growth, the population showed a similar amount of growth in both years, but the plants grew better at the late stage in 2005 than 2006. In conjunction with much more variation among RILs observed in 2005 than 2006, different growth forms in two years implicate the possible involvement of year-varying QTLs in plant growth processes. Li et al. (2010b) used functional mapping to identify three significant QTLs for genotype–year interactions. For example, the QTL detected on linkage group 6 affects growth trajectories in both years, but displays tremendous year-dependent difference in the temporal pattern of genetic effect (Fig. 1.5). This QTL can be thought of being contributing to genetic variation in the Phenotypic plasticity of growth curves.

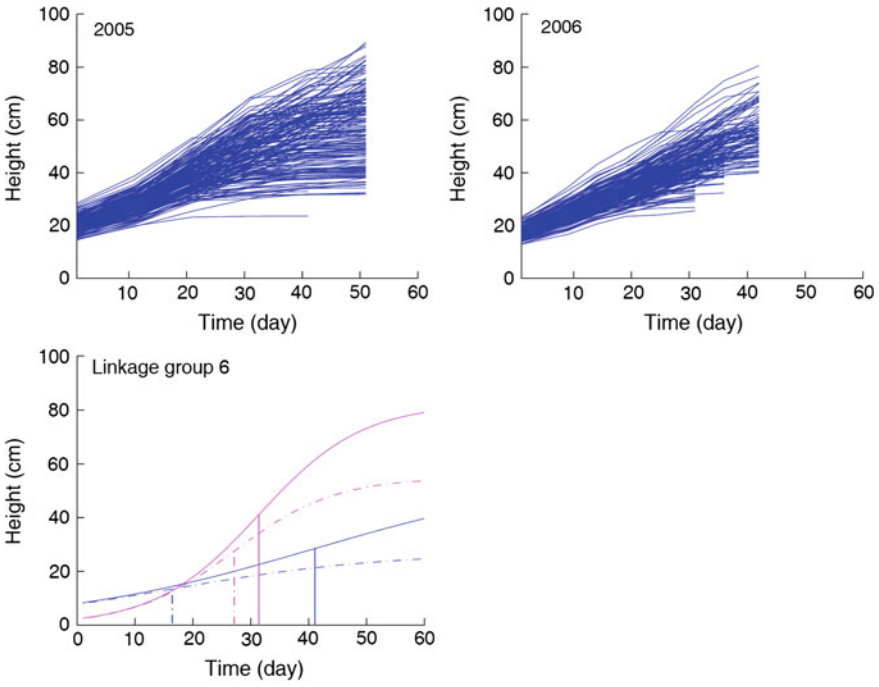


Fig. 1.5 Growth curves for plant heights in recombinant inbred lines of soybeans planted in year 2005 and 2006. Two genotypes at a QTL detected on linkage group 6 are each expressed as a growth curve in year 2005 (*purple*) and 2006 (*blue*). The times at the inflection point of a growth curve are shown by *vertical lines*

1.3 Mapping Reaction Norms as a Dynamic Process

When an organism is grown over a range of environments, it will produce an array of phenotypic values, i.e., reaction norms. If the environments exemplify a gradient of states, such as continuous temperature, irradiation or nutritional level, these reaction norms, expressed as response curves, can be described as a mathematical function of environmental factor (Wu and O'Malley 1998). The idea of functional mapping that incorporates environment-varying mathematical equations can be directly used to map and quantify specific QTLs that control variation in response curves. In developmental ecology, a universal law states that the metabolic rate of an organism increases with increasing temperature but decreases after temperature reaches a certain point (Kingsolver and Gomulkiewicz 2003). This law can be described by a mathematical equation, such as a quadratic function, whose implementation into functional mapping can foster the biological relevance and interpretations of results.

For discrete environments, such as different sexes, races or host species for polyphagous insects, there are no obvious mathematical equations that can describe the response curve of reaction norms. In spite of this, discrete values of phenotypes in multiple environments can be still viewed as a graded response by using the environmental index (EI) as the independent variable (Finlay and Wilkinson 1963; Lacaze et al. 2009). Wang et al. (2013a) embedded the mathematical form of grading response to discrete environments within the framework of functional mapping, enabling geneticists to map QTLs that affect reaction norm trajectories. This modification has been shown to be powerful for studying the genetic architecture of phenotypic plasticity.

In rice, a genotyped DH population of 123 lines derived from shorter IR64 and taller Azucena were grown in seven distinct locations of the Philippines, China, India, and Thailand spanning from 13.5° to 31.5°N and from 76° to 121.5°E (Li et al. 2003). These locations represent 11 different environments, in each of which plant heights were measured at the same age. By calculating the difference of the mean of all DH lines in a location from the mean of all DH lines over all locations as the EI for individual locations, we can see phenotypic trajectories of plant heights across 11 ordered environments as a function of normalized EIs (Fig. 1.6).

Wang et al. (2013b) used functional mapping to detect several significant QTLs and epistatic interactions for the environment-dependent plasticity of plant height. Figure 1.6 illustrates how four different genotypes at each of two pairs of QTLs respond to the EI. The first pair comes from two QTLs located on chromosome 1 and 2 (Fig. 1.6a), whereas the second pair from two QTLs located on chromosome 1 and 3 (Fig. 1.6b). First, the genotypes perform differently over a continuum of EI, suggesting the involvement of significant QTLs in mediating reaction norms. Second, these four distinct genotypes are not in parallel over EI, which implies that genetic effects of QTLs change with environment, leading to significant GEI. Third, pronounced difference between genotype AABB versus AABb and genotype AaBB versus AaBb or between genotype AABB versus AaBB and genotype AABb versus

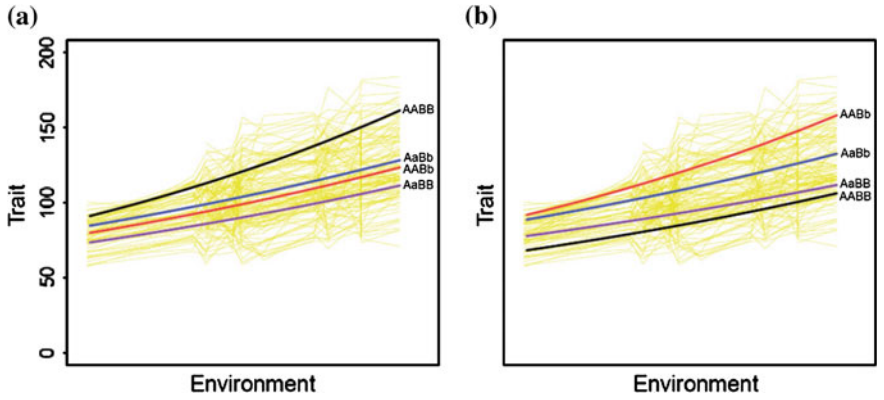


Fig. 1.6 Environment-dependent trajectory curves of plant heights for four genotypes at a pair of QTLs located on chromosome 1 and 2 (**a**) and on chromosome 1 and 3 (**b**). *Capital* and *lowercase letters* stand for alleles derived from two different parents. Each *lighter line* (in *yellow*) represents a DH progeny line

AaBb indicates that epistatic interactions between the two QTLs are responsible for variation in reaction norm curves of plant height. Thus, Wang et al.'s (2013b) model can clearly dissect a general trend of phenotypic plasticity over a large number of discrete environments into the underlying genetic components.

1.4 Concluding Remarks

Synthesizing development into evolution and ecology has been a major focus in modern biology, as an incentive to explore the mechanistic influence of developmental processes on phenotypic novelties created in a response to changing environments (Muller 2007). This synthesis can be accelerated by our understanding of how genes act and interact with each other and with environmental factors to determine the process of development. Functional mapping invented by R. Wu and group provides a state-of-the-art tool for genetic analysis of environment- or age-dependent phenotypes. This tool has proven to gain several key biological insights into the genetic mechanisms of development which cannot be afforded by traditional genetic mapping methods.

Functional mapping was invented to accommodate the developmental feature of a phenotypic trait in the context of genetic mapping. With the rapid advent of genotyping technologies, phenotypic traits can be partitioned into their underlying genes or QTLs distributed through the genome (Morrison et al. 2013), revolutionizing classic quantitative genetic analysis purely based on phenotypic variation. By integrating developmental principles for trait formation and progression through mathematical equations, functional mapping pushes conventional QTL mapping of

phenotypes measured at a static time point up to a dynamic level at which the interplay between gene and development can be characterized in a quantitative way. The characteristic of functional mapping lies in its mechanistic and parsimonious modeling of longitudinal covariance structure in which phenotypes between similar ages are more strongly correlated with each other than with those from widely separated ages. Because of this, functional mapping is more powerful from both biological and statistical perspectives than univariate or multivariate analysis of variance.

Functional mapping fits repeated measurements at discrete time points using a curve; tests differences in the form and pattern of curves at species, population, or genotype levels; and enables biological and evolutionary interpretation of phenotypic variation. Functional mapping shows its remarkable potential for all these tasks; more importantly, it is particularly equipped with a capacity to dissolve longitudinal phenotypic variation into genotype-specific curve components through a parametric or nonparametric function and test phenotypic variation at the molecular level. Some pioneering uses of functional mapping have led to the identification of specific QTLs that govern the process of dynamic traits in plants (Yang et al. 2011; Sillanpää et al. 2012), animals (Xiong et al. 2011), and humans (Li et al. 2009). Functional mapping has been used to study the genetic architecture of molecular and cellular interactions toward biological phenomena of significant importance, such as circadian rhythm, drug response, allometric scaling, phenotypic plasticity, and morphological shape (Wang et al. 2013a; Fu et al. 2013). The central idea of functional mapping has been extended to simultaneously model multiple variables that are interconnected through design principles to function as a whole. Examples include the integration of functional mapping and differential equations (Fu et al. 2011; Wang et al. 2013d which serve as a powerful means for describing and quantifying the emergent properties of dynamic systems.

Functional mapping offers several advantages and opportunities to enable the genetic dissection of developmental processes and construct the precise genotype–phenotype map through developmental pathways. First, next-generation sequencing techniques have made it possible to produce a vast amount of genetic, epigenetic, genomic, and proteomic data that are related to the formation of end-point phenotypes (Morrison et al. 2013). By reconstructing genetic networks using these data, functional mapping can be readily renovated to unravel those key intermediate steps and mechanisms that cause phenotypic variation. Recent statistical development in variable selection allows functional mapping to build a powerful model for the choice of most significant predictors. Second, in an emerging field of evolutionary and ecological developmental biology (evo-devo and eco-devo), we are not only interested in identifying the genetic variation of final phenotypes in heterogeneous environments, but also genetic actions and interactions that affect the timing and rate of development. Functional mapping provides a quantitative framework by which to test the genetic effects of specific QTLs on any developmental events during ontogeny. Third, functional mapping can be applied to a wide array of evolution and ecology in medicine, such as tumor growth, host–pathogen

interactions, pharmacodynamics reactions and neural systems, and software tools developed for the analysis of dynamic phenotypes in plants and animals could easily be adapted to the study of these medical problems.

Acknowledgments This research was funded by grants NIH/NHLBI-1U10HL098115, U01 HL119178, and UL1 TR000127.

References

Agrawal AA (2001) Phenotypic plasticity in the interactions and evolution of species. Science 294:321–326

Cheverud JM, Routman EJ, Duarte FAM, Swinderen BV, Cothran K, Perel C (1996) Quantitative trait loci for murine growth. Genetics 142:1305–1319

El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MG (2014) Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends Plant Sci 19:390–398

Fabbrini F, Gaudet M, Bastien C, Zaina G, Harfouche A, Beritognolo I, Marron N, Morgante M, Scarascia-Mugnozza G, Sabatti M (2012) Phenotypic plasticity, QTL mapping and genomic characterization of bud set in black poplar. BMC Plant Biol 12:47

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant breeding program. Aust J Agr Res 14:742–754

Fu GF, Wang Z, Li JH, Wu RL (2011) A mathematical framework for functional mapping of complex systems using delay differential equations. J Theor Biol 289:206–216

Fu GF, Bo WB, Pang XM, Wang Z, Chen L, Song YP, Zhang ZY, Li J, Wu RL (2013) Mapping shape QTLs using a radius-entroid-contour model. Heredity 110:511–519

He QL, Berg A, Li Y, Vallejos CE, Wu RL (2010) Modeling genes for plant structure, development and evolution: functional mapping meets plant ontology. Trends Genet 26:39–46

Huang N, Parco A, Mew T, Magpantay G, McCough S et al (1997) RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. Mol Breed 3:105–113

Jiang LB, Clavijo JA, Sun LD, Zhu XL, Bhakta MS, Gezan SA, Carvalho M, Vallejos CE, Wu R (2015) Plastic expression of heterochrony quantitative trait loci (*h*QTL) for leaf growth in the common bean (*Phaseolus vulgaris* L.). New Phytol. DOI:10.111/mph.13386

Kingsolver JG, Gomulkiewicz R (2003) Environmental variation and selection on performance curves. Integr Comp Biol 43:470–477

Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analysis in barley, *Hordeum vulgare*. Heredity 102:163–173

Li ZK, Yu SB, Lafitte HR et al (2003) QTL × environment interactions in rice. I. Heading date and plant height. Theor Appl Genet 108:141–153

Li N, Das K, Wu RL (2009) Functional mapping of human growth trajectories. J Theor Biol 261:33–42

Li N, McMurry T, Berg A, Wang Z, Berceli SA, Wu RL (2010a) Functional clustering of periodic transcriptional profiles through ARMA (p, q). PLoS ONE 5(4):e9894

Li Q, Huang ZW, Xu M, Wang CG, Gai JY, Huang YJ, Pang XM, Wu RL (2010b) Functional mapping of genotype-environment interactions for soybean growth by a semiparametric approach. Plant Methods 6:13

Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751–1762

Ma C-X, Yu QB, Berg A, Fu GF, Yap JS, Tan AX, Kirst M, Cui YH, Wu RL (2008) A statistical model for testing the pleiotropic control of phenotypic plasticity of a count trait. Genetics 179:627–636

Moczek AP, Sultan S, Foster S, Ledon-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW (2011) The role of developmental plasticity in evolutionary innovation. Proc R Soc B 278:2705–2713

Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, Muzny D, Yu F, Rice K, Zhu C, Bis J, Heiss G, O'Donnell CJ, Psaty BM, Cupples LA, Gibbs R, Boerwinkle E (2013) Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. Nat Genet 45:899–901

Muller GB (2007) Evo-devo: extending the evolutionary synthesis. Nat Rev Genet 8:943–949

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP (2010) Phenotypic plasticity's impact on diversification and speciation. Trends Ecol Evol 25:459–467

Scheiner SM (1993) Genetics and evolution of phenotypic plasticity. Annu Rev Ecol Syst 24:35–68

Schlichting CD (1989) Phenotypic integration and environmental change. Bioscience 39:460–464

Schlichting CD, Smith H (2002) Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. Evol Ecol 16:189–201

Sillanpää MJ, Pikkuhookana P, Abrahamsson S, Knürr T, Fries A, Lerceteau E, Waldmann P, García-Gil MR (2012) Simultaneous estimation of multiple quantitative trait loci and growth curve parameters through hierarchical Bayesian modeling. Heredity 108:134–146

Sultan SE (2000) Phenotypic plasticity for plant development, function and life history. Trends Plant Sci 5:537–542

Sultan SE (2003) Phenotypic plasticity in plants: a case study in ecological development. Evol Dev 5:25–33

Sultan SE (2010) Plant developmental responses to the environment: eco-devo insights. Curr Opin Plant Biol 13:96–101

Sun LD, Ye Hao H, Wang NT, Wang YQ, Cheng TR, Zhang QX, Wu RL (2014) A model framework for identifying genes that guide the evolution of heterochrony. Mol Biol Evol 31:2238–2247

Tetard-Jones C, Kertesz MA, Preziosi RF (2011) Quantitative trait loci mapping of phenotypic plasticity and genotype–environment interactions in plant and insect performance. Phil Trans R Soc B 366:1368–1379

Vaughn TT, Pletscher LS, Peripato A, King-Ellison K, Adams E, Erikson C, Cheverud JM (1999) Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. Genet Res 74:313–322

Via S, Lande R (1985) Genotype-environment interactions and the evolution of phenotypic plasticity. Evolution 39:505–522

Wang YQ, Wang NT, Wang JX, Wang Z, Wu RL (2013a) Delivering systems pharmacogenomics towards precision medicine through mathematics. Adv Drug Deliv Rev 65:905–911

Wang Z, Pang XM, Lv YF, Xu F, Zhou T, Li X, Feng SS, Li JH, Li ZK, Wu RL (2013b) A dynamic framework for quantifying the genetic architecture of phenotypic plasticity. Brief Bioinform 14:82–95

Wang Z, Pang XM, Wu WM, Wang JX, Wang ZH, Wu RL (2013c) Modeling phenotypic plasticity in growth trajectories: a statistical framework. Evolution 68:81–91

Wang ZH, Wang Z, Fu GF, Luo JT, Wu RL (2013d) Stochastic modeling of systems mapping for drug response. Adv Drug Deliv Rev 65:912–917

Wennersten L, Forsman A (2012) Population-level consequences of polymorphism, plasticity and randomized phenotype switching: a review of predictions. Biol Rev 87:756–767

Wu RL (1998) The detection of plasticity genes in heterogeneous environments. Evolution 52:967–977

Wu R, Lin M (2006) Functional mapping—how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7:229–237

Wu RL, O'Malley DM (1998) Nonlinear genotypic response to macro- and microenvironments. Theor Appl Genet 96:669–675

Wu RL, Grisson JE, McKe SE, O'Malley DM (2004) Phenotypic plasticity of fine roots and increased plant growth in pines. BMC Ecol 4:14

Wu RL, Ma C-X, Casella G (2007) Statistical genetics of quantitative traits: linkage, maps, and QTL. Springer, New York

Xiong H, Goulding EH, Carlson EJ, Tecott LH, McCulloch CE, Sen S (2011) A flexible estimating equations approach for mapping function-valued traits. Genetics 189:305–316

Yang R, Li J, Wang X, Zhou X (2011) Bayesian functional mapping of dynamic quantitative traits. Theor Appl Genet 123:483–492

Yap J, Fan J, Wu RL (2009) Nonparametric modeling of covariance structure in functional mapping of quantitative trait loci. Biometrics 65:1068–1077

Zhai Y, Lv YF, Li X, Wu WM, Bo WH, Sheng DF, Xu F, Pang XM, Zheng BZ, Wu RL (2014) A synthetic framework for modeling the genetic basis of phenotypic plasticity and its costs. New Phytol 20:357–365

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine* max L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139

Zhao W, Ma CX, Cheverud JM, Wu RL (2004a) A unifying statistical model for QTL mapping of genotype-sex interaction for developmental trajectories. Physiol Genom 19:218–227

Zhao W, Zhu J, Gallo-Meagher M, Wu RL (2004b) A unified statistical model for functional mapping of genotype × environment interactions for ontogenetic development. Genetics 168:1751–1762

Zhou T, Lyu YF, Xu F, Wu WM, Zhang J, Pang XM, Zheng BS, Wu RL (2015) A QTL model to map the common genetic basis for correlative phenotypic plasticity. Brief Bioinform 16:24–31

Chapter 2
The Genomics of Sexual Ornaments, Gene Identification and Pleiotropy

Martin Johnsson

Abstract Sexual ornaments, which are traits that make an individual attractive to potential mates, have a long history in evolutionary biology. These adaptations to mate choice have been the subject of research from the perspective of genetics, ecology and theoretical biology. The rapid development of genomic methods has equipped modern genetics with new tools to answer old questions and open up new areas of analysis. For research into sexual ornamentation, this has meant the application of genetic mapping, in particular quantitative trait locus (QTL) methods, and transcriptomics to search for genes and biological pathways affecting ornamental traits and investigate pleiotropy between ornamental traits, ornament preference and fitness-related traits. Examples come from QTL studies of beak colour in the zebra finch, colocalisation between loci for ornaments and other traits in crickets and moths, QTL mapping and population genomics of colour in guppies and cichlids, genetical genomics and pleiotropy mapping of comb size in the chicken, transcriptomic studies of handicap mechanisms in the grouse, and genetics and molecular evolution of several sexual traits in *Drosophila*. Genomic methods help reveal the variety of mechanisms involved in sexual ornamentation and are complementary to quantitative genetics, population genetics and organismal studies.

2.1 Introduction

Sexual ornaments are traits that help make an individual attractive to potential mates of the opposite sex. Here, ornaments are taken to include secondary sexual traits favoured by mate choice, but not sexual weaponry used in same-sex competition. Traits may serve a dual purpose, and in actual cases, the function of a trait is an empirical question that may be difficult to answer definitively. Ornamental

M. Johnsson (✉)
Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden
e-mail: martin.johnsson@liu.se

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_2

traits may be morphological or behavioural, such as exaggerated structures, colourful displays, acoustic signals and sexual behaviours.

Genomic methods allow the simultaneous investigation of genetic variants, transcripts and many other molecular features on a genome-wide scale. Such methods successively become available to evolutionary biologists and geneticists studying evolutionary phenomena, such as sexual ornamentation. With the exception of *Drosophila* and possibly the chicken, the study organisms for sexual ornaments are not model organisms for genetics. Hence, the genomic resources are comparatively limited and often have to be constructed by the researchers themselves. Improvements in sequencing technology have made this kind of work easier and will most likely continue to do so. Genomic research on sexual ornaments largely consists of genetic mapping studies in the linkage mapping framework. There is great promise in combining this kind of mapping with genome-scale molecular work and population genetics, made possible by recent advances in genomics.

There are different hypotheses about the basis of sexual ornamentation. Moreover, there is empirical support for several of them in different organisms. In the Fisher process, mate preference, once established, builds up genetic correlations between preference alleles and ornament alleles and runaway selection for ornament exaggeration (Mead and Arnold 2004). Except the Fisher process, most mechanistic hypotheses about sexual ornaments presuppose some form of pleiotropy between variants affecting the ornament and variants affecting other traits. In all cases, sufficiently tight linkage disequilibrium can substitute for genuine pleiotropy. The difference between hypotheses is what other traits that are coupled with the ornament and how they relate to fitness. In "good genes"-related hypotheses, variants affecting ornaments are expected to also pleiotropically affect fitness-related traits. More specifically, the "genic capture" variety of good gene hypotheses predicts many small-effect loci with pleiotropic effects on overall condition (Rowe and Houle 1996). Finally, "handicap hypotheses" add a layer of mechanistic detail, predicting these effects to be mediated by effects on physiological traits such as immunocompetence (Hamilton and Zuk 1982). In contrast, direct benefit hypotheses would predict no effect link between ornamentation and offspring genetic quality, but instead links between ornamentation and direct provisioning of resources by ornamented individuals to their mates.

Pleiotropy is a commonly used term with several slightly different meanings. Paaby and Rockman (2013) make a useful distinction between three different perspectives. The first is molecular gene pleiotropy, which means that a molecular gene, gene product or sequence element has several biological roles. While molecular gene pleiotropy is a prerequisite for any genetic variant to affect multiple traits, what is most interesting for the study of the evolution of any trait is when genetic variants have multiple phenotypic effects. This is what Paaby and Rockman call developmental pleiotropy. Finally, selectional pleiotropy is when a variant affects multiple fitness components. This perspective is also crucial for

understanding of sexual selection. Ornaments that act by means of handicap signalling or antagonistic pleiotropy would be clear examples of selectional pleiotropy. However, even when handicap signalling is not the central mechanism of an ornament, exaggerated displays are expected to be associated with trade-offs.

The establishment of a trait as a sexual ornament requires studies of mate preference and knowledge of the ecology and behaviour of the study organism. When it comes to the genetic basis of ornamentation, genomic approaches are complementary to quantitative and population genetic methods. Statistical quantitative genetic methods are based on the measurements of traits and pedigrees, and population genetic methods consider variants in populations. For example, a quantitative genetic technique to investigate pleiotropy is to estimate the genetic correlation between two traits. One genomic approach is to map quantitative trait loci associated with both traits and see whether they overlap. In one sense, these experiments study the same phenomenon but from different angles. However, the genomic analysis also raises new questions. Are the co-localising loci genuinely pleiotropic or are they built up of closely linked variants that effectively act as pleiotropic? What is the molecular basis of pleiotropic effects at this locus? Does the variant affect several biological pathways? Genomic analysis can help test old hypotheses about the mechanisms behind sexual ornaments, but they also raise new levels of analysis that are largely independent of the classical questions.

2.2 Genetic Mapping and Transcriptomic Techniques

Genetic mapping is localisation of genetic variants affecting a trait by means of statistical models that use genetic markers as predictors and phenotypic traits as response variables. When the trait in question is measured on some quantitative scale, rather than presence/absence, this is called quantitative trait locus (QTL) mapping (Soller et al. 1976; Lander and Botstein 1989). All kinds of genetic mapping rely on genetic markers that are linked to the causative variants. In linkage mapping, one uses related individuals with a known pedigree with markers that are informative of the segregation of chromosomes in these particular families. Linkage mapping can be undertaken either with natural pedigrees and using variance component methods or with experimental line crosses. QTL mapping in experimental crosses finds variants that differ between the founder individuals and are usually constructed from divergent populations, ideally from inbred founder individuals, to maximise the genetic variation that can be detected. Line crosses can be applied to within-population variation by, for example, repeated QTL mapping from inbred individuals or crosses of selection lines established from the same population.

As genotyping and sequencing technology have progressed and a high marker density has become easier to achieve, genome-wide association studies (GWASs) have risen to prominence (Risch and Merikangas 1996). GWAS relies on linkage

disequilibrium on the population level, rather than linkage within a family, and hence has much higher mapping resolution. Linkage disequilibrium blocks can be short enough for GWAS to even isolate a single candidate gene for an association. Also, GWAS works on unrelated individuals, so natural populations can be investigated without the need to construct a pedigree. However, a GWAS is a major undertaking in terms of both sample size and construction of marker maps, if such genomic resources are not already available for the species in question.

A complementary approach to genetic mapping is transcriptome-wide gene expression measurement. Expressed sequence tag (EST) sequencing and microarrays have been available for some time, and with the rise of massively parallel sequencing, RNA sequencing will likely be more common in the future. Gene expression studies can be applied both to the environmental and to the genetic mechanisms of ornamental traits, by either comparing individuals of different genotype or exposed to different environments. The application of genetic mapping to gene expression values is called expression QTL or eQTL mapping (Jansen and Nap 2001). eQTL mapping allows screening for genes that are affected by segregating variants in the population. The co-localisation of eQTL and QTL can be used as a way to filter the genes under the QTL to find those that could be a causative gene affected by a regulatory variant. When the genomic location of the gene being measured is known, eQTL can be divided into local eQTL that map back to the location of the gene and distal eQTL that map to other regions of the genome. Local eQTL are putative cis-eQTL, that is QTL affected by variants in some regulatory sequence such as a promoter, enhancer or insulator. Distal eQTL are likely trans-acting, meaning that they affect some upstream gene in a regulatory pathway. While trans-eQTL often have relatively smaller effect sizes and are more difficult to detect, this distinction in principle can allow one to find downstream consequences of causative genes. Gene expression comparisons between genetically distinct stocks such as different populations or selection lines are similar in spirit to eQTL mapping. These experimental designs also find genes affected by genetic variants differing between populations or lines. However, they do not give any genomic localisation of the regulatory variant and do not distinguish local and distal regulatory effects.

2.3 Beak Colour in the Zebra Finch

Beak colour in the zebra finch is one ornamental trait that has been investigated in quantitative genetic and QTL mapping studies. The redness of the beak is sexually dimorphic, and females of some populations prefer males with darker red beaks. Beak redness is one example of a carotenoid-based colour, which is suggested to be involved in trade-offs between colouration and immune function (Blount et al. 2003). Schielzeth et al. (2012) mapped QTL for beak colour as measured with spectrometry. They used variance component mapping in a within-population design, which allows mapping naturally segregating variants. In this case, the

population was a pedigree of captive birds. They report four QTL at the suggestive threshold level that still appear to explain most of the genetic variation within the population: 29 %, when the heritability was estimated to 34 %. They take this to mean that beak colour is oligogenic, with these four loci explaining the bulk of the additive genetic variance. This is at odds with genic capture, because genic capture implies a polygenic architecture built up of many small-effect loci. They hypothesise that antagonistic pleiotropy could be at work. However, there is an alternative interpretation, which has to do with the statistical limitations of QTL mapping, even in large studies such as this one. If a study is underpowered, that is it has a small sample size in relation to the expected size of the effect to be estimated, it is likely to result in overestimates. The problem with genetic mapping is that under a polygenic model, genetic effects are expected to be very small. Furthermore, variance component-based genetic mapping using natural pedigrees requires much larger sample sizes than line-cross QTL mapping to achieve good power. In a simulation study, Slate (2013) suggested that the results of the beak colour study are consistent with a polygenic architecture.

2.4 Eye Span and Meiotic Drive in the Stalk-Eyed Fly

The stalk-eyed fly *Cyrtodiopsis dalmanni* provides an example of how an ornamental trait can reflect genetic quality of a mate in the form of the absence of or protection against a meiotic driver. In this species, some males carry an X-linked variant that makes them produce offspring with very skewed sex ratios. The X-linked driver causes them to pass on itself, rather than the Y chromosome, producing female offspring that also carry the driver. This effect causes the driver chromosome to spread in the population because of this advantage. Females who avoid mating with driver-carrying males, however, can gain a reproductive advantage in that they produce sons. In these flies, eye span is associated with brood sex ratio and hence seems to work as a signal of meiotic drive (Wilkinson et al. 1998). Johns et al. (2005) used QTL mapping to investigate the genomic overlap between drive and eye span, investigating whether the traits are linked and in particular if there is an X-linked eye locus linked to the driver. They did find five eye stalk QTL, the largest of which was X-linked. It appears that the driver is indeed linked with a variant affecting eye span. They also investigated the offspring sex ratio of males in relation to their X driver genotype. They found an association between one autosomal QTL and sex ratio of the offspring. This suggests that the eye stalks not only signal the presence or absence of the meiotic driver, but also at least one autosomal modifier locus that counteracts the effect of the driver. The length of the eye stalk allows mate choice before mating, but there is also a potential for post-copulatory sexual selection. In particular, promiscuity in these flies opens the door to sperm competition, where variation in sperm traits might be important for the mating success of the males. In a later QTL study, Johns and

Wilkinson (2007) found that the X-linked driver and one autosomal eye stalk QTL also co-localised with QTL for sperm tail length.

There is also a study (Baker et al. 2009) that applies transcriptomics to the study of eye span in the stalk-eyed fly. The authors measured gene expression in eye stalk tissue at the larval, pupal and adult stage in individuals from lines selected for long or short eye span by means of EST cloning and sequencing. They measured differential expression between selection lines and also investigated gene duplication and protein coding gene evolution with alignment against *Drosophila melanogaster* genes. If a reference genome for the stalk-eyed fly were to be constructed, comparisons of differentially expressed and quickly evolving genes against the eye stalk QTL identified in previous studies could aid in identifying quantitative trait genes for these QTL.

2.5 Acoustic Signalling in Crickets and Moths

Even though Fisherian runaway selection does not presuppose physical linkage between variants for the ornament and the preference, mate preference and ornament expression QTL sometimes co-localise, revealing direct functional or physical links between preference and ornament. In the case of behavioural ornamental traits, such as certain patterns of sound production, one might hypothesise that the faculties required to produce the signal and receive it might be the same and affected by the same genetic variants. In the Hawaiian cricket genus *Laupala*, there is indeed direct linkage or pleiotropy of male song and female song preference loci. This was investigated in a series of QTL mapping studies (Shaw et al. 2007; Shaw and Lesnick 2009; Wiley and Shaw 2010) with crosses of closely related species *L. paranigra* and *L. kohalensis*. When performing genome-wide scan for female preference (Shaw and Lesnick 2009), they detected one preference QTL. Investigating preference, however, is in general more difficult and time-consuming than investigating the signalling trait. A strength in this case is that the preference was measured using synthesised songs and hence avoided the variation introduced by interactions with actual stimulus males. In follow-up studies, the preference effects were tested in directed crosses (Wiley and Shaw 2010) and introgression lines (Wiley et al. 2012) using the five male song QTL. Four out of five pulse rate QTL also affected female song preference. The fact that these QTL were readily detected in between-species crosses suggests that they have indeed been fixed by selection. Possibly, the fixation of preference for and production of songs of different pulse rate are part of the rapid speciation of these crickets. Direct genetic linkage between preference and signal may make Fisherian runaway selection even more effective, since it ensures high genetic correlation from the outset.

In the acoustic moth *Achroia grisella*, on the contrary, QTL mapping with inbred lines based on two populations from Florida and Kansas, USA, did not find any evidence of shared architecture between female preference and male signalling with ultrasound pulses (Limousin et al. 2012). A following study (Alem et al. 2013)

investigated whether song preference and response to bat echolocation signals share overlapping loci. They found no overlap. However, there are complications with trying to learn the evolutionary history of a trait from current genetic variation. Even if male acoustic signalling did evolve by co-opting mechanisms from predator avoidance, which does seem likely despite the lack of overlap, the pleiotropic variants involved in the early evolution of this signal must still be different between populations for QTL mapping to detect them. The authors interpret the lack of overlap to mean the independent architecture might reflect later modifications of the female responses.

2.6 Comb Size in the Chicken

The chicken comb is an example of a mutual sexual ornament that is preferred by both sexes. It is used by females to guide largely post-copulatory mate choice and by males to allocate sperm (Pizzari and Birkhead 2000; Pizzari et al. 2003). Comb size reflects status and reproductive investment (Cornwallis and Birkhead 2007). There is also evidence of handicap effects on males (Von Schantz et al. 1995). Because relative comb size has increased under chicken domestication, the genetic and molecular basis of comb size can be studied in intercrosses of wild and domestic chickens. QTL mapping in two F_2 crosses and one eighth-generation advanced intercross (Wright et al. 2008; Johnsson et al. 2012, 2014) found six replicable QTL regions with additive moderate to large effects. The advanced intercross is a way of increasing the mapping density of experimental intercrosses. Starting from the F_2 intercross, the intercross line is interbred for a number of generations, accumulating recombinations. The QTL regions of the advanced intercross are considerably tighter than the F_2 study (Johnsson et al. 2012), but most QTL still have many positional candidate genes. The exception in this case is one major comb QTL that covers the non-coding region between the genes *bone morphogenetic protein 2* and *hydroxyacid oxidase 1*. A local eQTL study of these two genes in comb base tissue suggests that both of them are regulated by genetic variants at the QTL. A microarray-based eQTL study (Johnsson et al. 2014) targeted to five of the replicated comb QTL found an additional three putative quantitative genes that display local eQTL effects and an association between comb gene expression and relative comb mass.

These QTL studies also provide evidence that comb size reflects different mate quality traits. Egg production in the female chicken is physiologically coupled with bone metabolism, since bone mineralisation and eggshell production share a common calcium reserve. Female chicken bone is continually being remodelled with the laying cycle (Kerschnitzki et al. 2014). In line with this phenomenon, comb mass QTL overlapped with both egg production and bone mineral density QTL (Rubin et al. 2007). Additionally, QTL for female onset of sexual maturity was found to overlap with comb QTL at all three loci that were detected (Wright et al. 2012). Moreover, multivariate QTL models were used to help disentangle

pleiotropy from close linkage and found that in several cases, the data were better explained by multiple linked QTL. This analysis is subject to the statistical uncertainties of model selection and limited by the recombinations available in the cross. However, it does suggest that tight linkage as well as genuine pleiotropy plays a role in the ornamental function of the chicken comb. The presence of clusters of linked QTL seems to be a general feature in the genetic architecture of chicken domestication (Wright et al. 2010). In the case of sexual ornaments, linkage between fitness-influencing variants and ornament variants could be a way for ornaments to capture variation in different traits. This mechanism of coupling requires no functional connection but the contingencies of genome organisation.

2.7 Comb Size in the Red Grouse

Studies in the field suggest that the size of the comb of the red grouse *Lagopus lagopus scoticus* is a condition-dependent sexual ornament dependent on testosterone signalling (Mougeot et al. 2004). Males that were implanted with testosterone displayed larger combs, more parasites and a reduction in T-cell-mediated immunity. This is consistent with immunocompetence handicap, where the cost of a large comb is a steroid-induced impairment of the immune system. Subsequent transcriptomic studies (Webster et al. 2011; Wenzel et al. 2013) investigated the mechanistic basis of testosterone effects on red grouse males. The authors manipulated the testosterone as well as the caecal parasite load of the birds in a factorial design under natural conditions (Webster et al. 2011). They measured gene expression in caecal tissue with microarrays and found genes affected by parasite load and testosterone. The treatment of high testosterone and high parasite load is particularly relevant to males with a large comb, who should suffer the handicap effect of testosterone and parasite load. Fifty-two transcripts were differentially expressed under these conditions, and 51 of them were down-regulated. These putative handicap-related transcripts were not necessarily from immune-related genes. A later analysis of the same field experiment (Wenzel et al. 2013) explicitly compared two mechanistic hypotheses for the handicap effect: immunocompetence handicap and oxidative stress handicap, where the testosterone increase is thought to cause either immunosuppression or increased oxidative stress. The authors compiled one set of immune system genes and one of oxidative stress genes based on Gene Ontology annotation. More genes from the immune set than the oxidative stress set were differentially expressed, but most of the detected genes fell in neither category. In short, the transcriptome evidence was not clear in favour of either hypothesis. This could be due to the limits of available gene annotation, but also raises the possibility that handicap in this species is mediated by a third unknown mechanism or a mix of physiological mechanisms.

2.8 Pigmentation in Guppies and Cichlids

The guppy, *Poecilia reticulata*, displays heritable genetic variation in male colour pattern, which is polymorphic within populations. Experiments with mate choice in the laboratory and with constructed populations in the field have found negative frequency-dependent selection for male colour (Hughes et al. 2013; Zajitschek et al. 2006). The number of offspring produced favoured males with rare colour patterns. This indicates that male colour polymorphisms are maintained by balancing selection, something that is theoretically known to be able to maintain variation. The mechanism favouring the preference for rare males is not known. Colour polymorphism in guppies has been the subject of QTL mapping in a cross of two distinct populations (Tripathi et al. 2009). The authors mapped 12 aspects of male colouration and found between two and eleven QTL per trait. This suggests a polygenic architecture with several QTL on the linkage group corresponding to the sex chromosome. A population genetic study of guppies from different populations (Willing et al. 2010) found evidence for selection at two of these loci. While variation can be maintained within a population, these between-population studies are evidence of directional selection as well.

Cichlid fishes are another example of diverse colour variation both within and between species. Cichlid colour has also been studied with QTL mapping. Albertson et al. (2014) crossed two related species from Lake Malawi that have different colour patterns and measured black and red yellow pixel counts with digital photography on 12 regions of the body. Mapping resulted in 41 QTL for various aspects of pigmentation. They genotyped the fishes by massively parallel sequencing of restriction site associated regions. In addition to the F_2 intercross, genotyping included wild-caught fishes from each population that were used for population genomics of the detected QTL regions. They found scaffolds of the reference cichlid genome matching their QTL regions and estimated genetic differentiation between the populations. Seventeen out of 366 of the single nucleotide variants found in one large QTL for red yellow colour had a high degree of differentiation. One of them was fixed between species and located within the 5' untranslated region of the *pax3a* gene. Levels of *pax3a* expression, as measured with qPCR, correlated with number of pigment-bearing xanthophores, and *pax3a* displayed allele-specific expression in the F_1 generation. Taken together, this suggests that this QTL is explained by a cis-acting regulatory variant acting on *pax3a* expression. It also demonstrates the power of combining QTL mapping with population genomics and gene expression for gene identification.

2.9 Sexual Ornaments in *Drosophila*

The fruit flies *Drosophila* spp. are genetic model organisms par excellence but also have sophisticated systems of mate choice. Studies on *Drosophila* highlight the importance of integrating multiple ornamental traits for an understanding of how

preference and ornamentation work in a natural setting. Cuticular hydrocarbons, one class of sexually selected trait in *Drosophila*, are inherently multivariate, since the signal is made up of a complex chemical mixture. However, *Drosophila* also utilise courtship dance and song and pigment patterns to guide mate choice.

Related species of *Drosophila* provide a model for how sexually selected traits contribute to isolation between species. Interspecific crosses of *D. simulans* and *D. sechellia* make up such a system. Gleason and Richie (2004) measured the interpulse distance in male courtship song in backcrosses and found six QTL. None of them overlapped the within-species QTL detected for the same trait detected in *D. melanogaster* (Gleason et al. 2002). However, the latter study was based on two laboratory strains and found three QTL, which can only cover a limited sampling of the variation within *D. melanogaster*. Similarly, Gleason et al. (2005, 2009) mapped QTL for cuticular hydrocarbons between *D. simulans* and *D. sechellia*. These studies estimate rather large effect sizes, explaining between 30 and 80 % of the variance in the crosses with few QTL per trait. Since very small-effect sizes are quite possible for polygenic traits, there is always a risk of QTL effect overestimation. One also needs to keep in mind that the proportion of variance explained in an intercross does not necessarily reflect the variance explained by that QTL in any natural setting. However, there is at least the potential for large-effect genes contributing to species isolation. As a counterpoint, consider a QTL study of within-population genetic variation in cuticular hydrocarbons in *D. melanogaster* (Foley et al. 2007). The authors used recombinant inbred lines based on a single pair of files derived from field-caught females. They identified 25 QTL in females and 15 in males, indicative of polygenic within-population variation.

Several, but by no means all, of the QTL for cuticular hydrocarbon production that has been found in the above studies may be explained by variants in genes of the desaturase gene family. *desatF* (Shirangi et al. 2009) is potentially one of the genes responsible for the differences between *D. simulans* and *D. sechellia*. It is also sexually dimorphic in some *Drosophila* species and monomorphic in some. The authors use in situ hybridisation and reporter assays to find regulatory sequences that drive *desatF* expression. They find that sexually dimorphic expression is caused by the presence of a binding site for the transcription factor *doublesex* and find a single nucleotide difference in this motif that abolishes its function in *D. takahasii*. Arguably, these differences have more to do with mate recognition than sexual selection, but they are still examples of the things that can be achieved when causative genes have been isolated.

2.10 The Value of Genome-Wide Studies Compared to Candidate Gene Studies

The QTL mapping of beak colouration (Schielzeth et al. 2012) also exemplifies the difficulties of candidate applying gene approaches to sexual ornaments. The authors compile a list of zebra finch orthologs of known carotenoid-related genes and

searched for these genes under their QTL. There was no evidence for any association with the candidate genes in this population. Similarly, enrichment testing of Gene Ontology annotation categories is a popular method to try to extract biological knowledge from gene sets derived from genomic experiments. Gene Ontology annotation enrichment on the genes located in QTL regions in this study found only an enrichment of the "metabolic process" category, which is extremely general. QTL regions are usually broad and contain many more genes than the expected number of causative genes. Hence, Gene Ontology enrichment on QTL regions is unlikely to yield anything but noise, except possibly in the most high-resolution studies such as GWAS. The gene expression study on stalk-eyed flies (Baker et al. 2009) also illustrates the difficulty of Gene Ontology enrichment in gene expression studies. The comparison between flies selected for long and short eye span resulted in 367 differentially expressed genes, but no overrepresented categories. The study found a handful of particularly promising candidates based on both gene expression and sequence evolution; however, at least two of the five candidates highlighted by the authors have little to no functional literature. This emphasises the value of genome-wide approaches such as these, even if genome-wide top-down studies require more work than targeted candidate gene studies.

The functions of genes are often unknown, particularly since molecular pleiotropy means that a given molecular gene may participate in processes it is not currently known for. Candidate gene approaches may be applicable in certain cases when the functional biology of the trait is well known or when high-quality candidate gene data from mutation studies in the same species are available. For example, Kottler et al. (2013) performed a successful candidate gene study of pigment patterns in the guppy based on genes from the zebrafish, and some of the QTL found in the abovementioned *Drosophila* QTL studies overlaps mutational candidate genes for those traits. This level of functional information is usually not the available when investigating complex traits in non-model organisms.

2.11 Maintenance of Genetic Variation

A recurring question in sexual selection is how, if at all, genetic variance in sexually selected traits is maintained. If a sexual ornament is under directional selection, the genetic variance should be depleted eventually (Taylor and Williams 1982). Several mechanisms of maintenance are possible and have empirical support in different cases. For instance, frequency-dependent selection that favours rare variants can maintain variation, and this is the case for colour morphs in the guppy (Hughes et al. 2013). Another mechanism is heterozygote advantage, where selection favours heterozygotes with intermediary trait levels, possibly because of life history trade-offs between ornamentation and survival. Sexual weaponry in the form of horn shape in the Soay sheep of St Kilda is an example of heterozygote advantage in the wild (Johnston et al. 2013).

However, quantitative trait variation is not necessarily maintained. The genic capture model of good genes suggests that variation is continually replenished, because ornaments reflect polygenic variation in fitness. When between-population QTL mapping works and reveals loci, it works best for differences that are fixed between populations. This could reflect local adaptation, but also that different populations go through and fix different subsets of the possible mutations that affect an ornament. Moreover, even if there is detectable genetic variation in a trait, selection could still have exhausted all available genetic variation. Whether apparent genetic variation in a single trait is available to selection depends on the interactions with other genetically correlated traits. This curse of dimensionality potentially affects all traits subject to multivariate selection, which in a naturalistic setting may be most ornamental traits. Quantitative genetic studies of cuticular hydrocarbons in *D. serrata* suggest that sexual selection has actually depleted the variance in the direction of selection and that univariate estimates of genetic variance that suggest maintenance can be misleading in this case (Hine et al. 2004; Blows et al. 2004).

2.12 Promises and Limitations of Genomic Methods

Sexual ornaments are a diverse set of morphological and behavioural traits that provide case studies for functional and evolutionary genomics. Researchers have used genomics, principally QTL mapping and microarrays, to investigate the genetic architecture of ornaments, to search for the underlying causative genes and to study mechanisms of ornament function. All three perspectives will likely benefit from improvements in genomic sequencing, transcriptomics and other genome-wide molecular assays. Also, laboratory and field studies of phenotypes and fitness as well as pedigree-based quantitative genetic studies will remain crucial.

The main limiting factors to untangling the genetic architecture of a trait through mapping are power and resolution. Low power prevents one from finding loci and leads to overestimation of the ones identified. Low resolution prevents one from distinguishing linkage from pleiotropy. Both effects are related to sample size and bias mapping towards making genetic architecture look simpler than it is. Mapping resolution in a genome-wide association study is superior to linkage mapping and can often cut associated regions down to a handful of genes. Therefore, GWAS is an attractive alternative for within-population mapping studies. GWAS requires much higher marker density than linkage mapping. With modern sequencing and genotyping methods, marker discovery is still a major undertaking, but achievable. However, the main limitation will be collecting and phenotyping sufficient samples of individuals.

Genomics opens up for the top-down identification of genes and genetic variants affecting ornamental traits without relying on previous knowledge of molecular mechanisms and candidate genes, which is often not available. However, such gene identification studies require high mapping resolution and likely the combination of

multiple techniques. A promising approach that several groups have taken is to integrate mapping with gene expression or population genomics. Other functional genomic data, such as chromatin immunoprecipitation of transcription factors or chromatin marks, could be also be overlaid on top of QTL and genetic variant data to aid the search for causative variants. Genetic variants and molecular genes that have been found through such methods can then be the starting point for molecular evolution studies that trace the evolution of ornaments through extant populations and phylogenetic lineages. Evolutionary studies of *desat* genes (Shirangi et al. 2009; Keays et al. 2011; Fang et al. 2009) in *Drosophila* provide tantalising examples.

A lot of the value of genomics comes from harnessing functional information about genes and pathways. While databases such as RefSeq, Ensembl, UniProt, the Gene Ontology Annotation project and many others are vast troves of information, much is still unknown about any given biological process. The path from molecular function to organismal trait is often uncharted. This complicates any inference that relies heavily on annotation derived from other species and contexts. Additionally, the final demonstration of any causative variant is experimental manipulation in transgenic organisms. Such molecular genetic tools take time and effort to develop and are not readily available for most species. However, these limitations are not specific to this field. Even human genomics has to rely on study systems such as laboratory mice for molecular testing of causative genes.

2.13 Conclusions

Sexual ornaments have long stimulated evolutionary research from genetical, ecological and theoretical perspectives. It appears that a multitude of mechanisms is at work in sexual ornamentation in different species and possibly within the same populations. Some ornamental traits have been revisited several times with different approaches over decades of work. Lately, genetic mapping and gene expression methods have been brought to bear on ornamentation and many quantitative trait loci have been mapped for different ornamental traits. However, going from quantitative trait loci to genes and molecular mechanisms is not an easy feat. There is good reason to think that future developments in genomics will improve the chances of isolating causative genes and help illuminate the basis of pleiotropy between ornaments and other traits.

References

Albertson RC, Powder KE, Hu Y, Coyle KP, Roberts RB, Parsons KJ (2014) Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. Mol Ecol 23(21):5135–5150

Alem S, Streiff R, Courtois B, Zenboudji S, Limousin D, Greenfield M (2013) Genetic architecture of sensory exploitation: QTL mapping of female and male receiver traits in an acoustic moth. J Evol Biol 26(12):2581–2596

Baker RH, Morgan J, Wang X, Boore JL, Wilkinson GS (2009) Genomic analysis of a sexually-selected character: EST sequencing and microarray analysis of eye-antennal imaginal discs in the stalk-eyed fly *Teleopsis dalmanni* (Diopsidae). BMC Genom 10(1):361

Blount JD, Metcalfe NB, Birkhead TR, Surai PF (2003) Carotenoid modulation of immune function and sexual attractiveness in zebra finches. Science 300(5616):125–127

Blows MW, Chenoweth SF, Hine E (2004) Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. Am Nat 163(3):329–340

Cornwallis CK, Birkhead TR (2007) Experimental evidence that female ornamentation increases the acquisition of sperm and signals fecundity. Proc R Soc B Biol Sci 274(1609):583–590. doi:10.1098/rspb.2006.3757

Fang S, Ting C-T, Lee C-R, Chu K-H, Wang C-C, Tsaur S-C (2009) Molecular evolution and functional diversification of fatty acid desaturases after recurrent gene duplication in *Drosophila*. Mol Biol Evol 26(7):1447–1456

Foley B, Chenoweth S, Nuzhdin S, Blows M (2007) Natural genetic variation in cuticular hydrocarbon expression in male and female *Drosophila melanogaster*. Genetics 175:1465–1477

Gleason JM, Ritchie MG (2004) Do quantitative trait loci (QTL) for a courtship song difference between *Drosophila simulans* and *D. sechellia* coincide with candidate genes and intraspecific QTL? Genetics 166(3):1303–1311

Gleason J, Nuzhdin S, Ritchie M (2002) Quantitative trait loci affecting a courtship signal in *Drosophila melanogaster*. Heredity 89(1):1–6

Gleason JM, Jallon J-M, Rouault J-D, Ritchie MG (2005) Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between *Drosophila simulans* and *D. sechellia*. Genetics 171(4):1789–1798

Gleason J, James R, Wicker-Thomas C, Ritchie M (2009) Identification of quantitative trait loci function through analysis of multiple cuticular hydrocarbons differing between *Drosophila simulans* and *Drosophila sechellia* females. Heredity 103(5):416–424

Hamilton WD, Zuk M (1982) Heritable true fitness and bright birds: a role for parasites? Science 218(4570):384–387

Hine E, Chenoweth SF, Blows MW (2004) Multivariate quantitative genetics and the lek paradox: genetic variance in male sexually selected traits of *Drosophila serrata* under field conditions. Evolution 58(12):2754–2762

Hughes KA, Houde AE, Price AC, Rodd FH (2013) Mating advantage for rare males in wild guppy populations. Nature 503(7474):108–110

Jansen R, Nap J (2001) Genetical genomics: the added value from segregation. Trends Genet 17 (7):388–391

Johns P, Wilkinson G (2007) X chromosome influences sperm length in the stalk-eyed fly *Cyrtodiopsis dalmanni*. Heredity 99(1):56–61

Johns PM, Wolfenbarger LL, Wilkinson GS (2005) Genetic linkage between a sexually selected trait and X chromosome meiotic drive. Proc R Soc B Biol Sci 272(1576):2097–2103

Johnsson M, Gustafson I, Rubin C-J, Sahlqvist A-S, Jonsson KB, Kerje S, Ekwall O, Kämpe O, Andersson L, Jensen P, Wright D (2012) A sexual ornament in chickens is affected by pleiotropic alleles at HAO1 and BMP2, selected during domestication. PLoS Genet 8(8): e1002914. doi:10.1371/journal.pgen.1002914

Johnsson M, Rubin CJ, Höglund A, Sahlqvist AS, Jonsson K, Kerje S, Ekwall O, Kämpe O, Andersson L, Jensen P (2014) The role of pleiotropy and linkage in genes affecting a sexual ornament and bone allocation in the chicken. Mol Ecol 23(9):2275–2286

Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, Slate J (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature

Keays MC, Barker D, Wicker-Thomas C, Ritchie MG (2011) Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. Mol Ecol 20 (17):3617–3630

Kerschnitzki M, Zander T, Zaslansky P, Fratzl P, Shahar R, Wagermaier W (2014) Rapid alterations of avian medullary bone material during the daily egg-laying cycle. Bone

Kottler VA, Fadeev A, Weigel D, Dreyer C (2013) Pigment pattern formation in the guppy, *poecilia reticulata*, involves the kita and csf1ra receptor tyrosine kinases. Genetics 194(3):631–646

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Limousin D, Streiff R, Courtois B, Dupuy V, Alem S, Greenfield MD (2012) Genetic architecture of sexual selection: QTL mapping of male song and female receiver traits in an acoustic moth. PLoS ONE 7(9):e44554

Mead LS, Arnold SJ (2004) Quantitative genetic models of sexual selection. Trends Ecol Evol 19 (5):264–271

Mougeot F, Irvine JR, Seivwright L, Redpath SM, Piertney S (2004) Testosterone, immunocompetence, and honest sexual signaling in male red grouse. Behav Ecol 15(6):930–937

Paaby AB, Rockman MV (2013) The many faces of pleiotropy. Trends Genet 29(2):66–73

Pizzari T, Birkhead TR (2000) Female feral fowl eject sperm of subordinate males. Nature 405:787

Pizzari T, Cornwallis CK, Lovlie H, Jakobsson S, Birkhead TR (2003) Sophisticated sperm allocation in male fowl. Nature 426:70–73

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273(5281):1516–1517

Rowe L, Houle D (1996) The Lek paradox and the capture of genetic variance by condition dependent traits. Proc R Soc Lond B Biol Sci 263(1375):1415–1421. doi:10.1098/rspb.1996.0207

Rubin C-J, Brändström H, Wright D, Kerje S, Gunnarsson U, Schutz K, Fredriksson R, Jensen P, Andersson L, Ohlsson C, Mallmin H, Larsson S, Kindmark A (2007) Quantitative trait loci for BMD and bone strength in an intercross between domestic and wildtype chickens. JBMR 22 (3):375–384

Schielzeth H, Kempenaers B, Ellegren H, Forstmeier W (2012) QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. Evolution 66(1):18–30

Shaw KL, Lesnick SC (2009) Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. Proc Natl Acad Sci 106(24):9737–9742

Shaw KL, Parsons YM, Lesnick SC (2007) QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket Laupala. Mol Ecol 16(14):2879–2892

Shirangi TR, Dufour HD, Williams TM, Carroll SB (2009) Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. PLoS Biol 7(8):e1000168

Slate J (2013) From Beavis to beak color: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. Evolution 67(5):1251–1262

Soller M, Brody T, Genizi A (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47(1):35–39

Taylor PD, Williams GC (1982) The lek paradox is not resolved. Theor Popul Biol 22(3):392–409

Tripathi N, Hoffmann M, Willing E-M, Lanz C, Weigel D, Dreyer C (2009) Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. Proc R Soc B Biol Sci 276(1665):2195–2208

Von Schantz T, Tufvesson M, Göransson G, Grahn M, Wilhelmson M, Wittzell H (1995) Artificial selection for increased comb size and its effects on other sexual characters and viability in *Gallus domesticus* (the domestic chicken). HEREDITY-LONDON 75:518–518

Webster L, Paterson S, Mougeot F, MARTINEZ-PADILLA J, Piertney J (2011) Transcriptomic response of red grouse to gastro-intestinal nematode parasites and testosterone: implications for population dynamics. Mol Ecol 20(5):920–931

Wenzel MA, Webster LM, Paterson S, Mougeot F, Martínez-Padilla J, Piertney SB (2013) A transcriptomic investigation of handicap models in sexual selection. Behav Ecol Sociobiol 67 (2):221–234

Wiley C, Shaw KL (2010) Multiple genetic linkages between female preference and male signal in rapidly speciating Hawaiian crickets. Evolution 64(8):2238–2245
Wiley C, Ellison CK, Shaw KL (2012) Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket Laupala. Proc R Soc B Biol Sci 279(1731):1203–1209
Wilkinson GS, Presgraves DC, Crymes L (1998) Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. Nature 391(6664):276–279
Willing E, Bentzen P, Van Oosterhout C, Hoffmann M, Cable J, Breden F, Weigel D, Dreyer C (2010) Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. Mol Ecol 19(5):968–984
Wright D, Kerje S, Brändström H, Schütz K, Kindmark A, Andersson L, Jensen P, Pizzari T (2008) The genetic architecture of a female sexual ornament. Evolution 62:86–98
Wright D, Rubin CJ, Martinez Barrio A, Schütz K, Kerje S, Brändström H, Kindmark A, Jensen P, Andersson L (2010) The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. Mol Ecol 19(23):5140–5156
Wright D, Rubin C, Schutz K, Kerje S, Kindmark A, Brandström H, Andersson L, Pizzari T, Jensen P (2012) Onset of sexual maturity in female chickens is genetically linked to loci associated with fecundity and a sexual ornament. Reprod Domest Anim 47:31–36
Zajitschek SR, Evans JP, Brooks R (2006) Independent effects of familiarity and mating preferences for ornamental traits on mating decisions in guppies. Behav Ecol 17(6):911–916

Chapter 3
Life's Dual Nature: A Way Out of the Impasse of the Gene-Centred 'Versus' Complex Systems Controversy on Life

Alexis De Tiège, Koen B. Tanghe, Johan Braeckman and Yves Van de Peer

Abstract Living cells and organisms are complex physical systems. Does their organization or complexity primarily rely on the intra-molecular crystalline structure of genetic nucleic acid sequences? Or is it, as critics of the 'gene-centred' perspective claim, predominantly a result of the inter- and supra-molecular—thus 'holistic'—network dynamics of genetic and various extra-genetic factors? The twentieth-century successes in several branches of genetics caused intensive focus on the causal role of genes in the biochemistry, development and evolution of living organisms, resulting in a relative abstraction or even neglect of life's complex systems dynamics. Today, however, partly due to the success of systems biology, a number of authors defend life's systems complexity while criticizing the gene-centred approach. Here, we offer a way out of the impasse of the gene-centred 'versus' complex systems perspective to arrive at a more balanced and complete understanding of life's multifaceted nature. After sketching the conceptual and historical background of the controversy, we show how the present state of knowledge in biology vindicates both the holistically complex and gene-centred nature of life on Earth, but decisively falsifies extreme genetic 'determinism' and 'reductionism' as well as extreme 'gene-de-centrism'. Contrary to what is often claimed, the fact that genes are one among many extra-genetic causal factors contributing to the biochemistry and development of cells and organisms, only undermines or falsifies genetic determinism and reductionism but not necessarily gene-centrism. Some implications for evolutionary theory, i.e. for the controversy between the Modern Synthesis and an 'Extended Synthesis', are outlined.

A. De Tiège (✉) · K.B. Tanghe · J. Braeckman
Department of Philosophy and Moral Science, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
e-mail: alexis.detiege@ugent.be

Y. Van de Peer
Department of Plant Systems Biology (VIB), Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_3

3.1 Introduction and Chapter Outline

During the first half of the twentieth century, some biological disciplines, such as classical and population genetics, were inevitably gene-centred, whereas others, such as developmental biology and/or embryology, were more focussed on life's holistic complexity. The spectacular advances in molecular biology and genetics during the second half of the twentieth century, however, caused intensive focus on the causal role of genes in the biochemistry, development and evolution of living organisms. This strong emphasis on genetic causation was at the cost of a clear understanding of life's holistic network dynamics (Gilbert and Sarkar 2000). As 'reductionism' flourished, there even was an unwillingness to recognize the latter aspect of life, with Conrad Waddington being a notable exception of this trend (Waddington 1957). Today, we see an opposite tendency: due to the success of systems biology, the focus is on dynamic *network complexity*. Some authors (e.g. Goodwin 1984, 1994; Oyama 1985; Oyama et al. 2001; Keller 2000; Moss 2003; Callebaut et al. 2007; Noble 2008, 2010, 2012; Noble et al. 2014) have been defending life's holistic systems complexity while denying gene-centeredness as a property of life. The present chapter's objective is to end the swinging of the pendulum from one extreme to the other, as recognizing just one of life's characteristics *at the cost* of another one seriously stands in the way of a more balanced and complete understanding of life's *multifaceted nature*. Life is constituted by *inter- and supra-molecular* or holistic network dynamics which is, however, *permeated* by functional gene products (i.e. functional RNAs and proteins) and hence by genetic sequence information derived from the *intra-molecular crystalline* structure of nucleic acid sequences. Both characteristics, holistic complexity and gene-centeredness, are *essential* to a clear understanding of the nature of life on Earth.

In Sect. 3.2, we further sketch the conceptual and historical background of the controversy. In Sect. 3.3, we demonstrate how the present state of knowledge in biology vindicates both life's holistically complex and gene-centred nature, but at the same time falsifies extreme genetic 'determinism' and 'reductionism' (according to which biochemical, cellular and organismal organization are quasi-exhaustively determined by and reducible to genetic sequence information)[1] as well as extreme 'gene-de-centrism' (according to which genes are quasi-entirely 'de-centralized' and/or 'relativized' within the complex biochemical and developmental dynamics of cells and organisms). We show, among others, that the fact that genes are one among many extra-genetic causal factors contributing to the biochemistry and development of cells and organisms (cf. Oyama et al. 2001; Jablonka and Lamb 2005; Noble 2008, 2010, 2012; De Tiège et al. 2014; Noble et al. 2014), only undermines or falsifies genetic determinism and reductionism but not gene-centrism. In Sect. 3.4, we conclude with some implications for evolutionary

[1]The idea that gene-centrism should be distinguished from genetic determinism and reductionism has already been put forward by one of us (Tanghe 2013) and even goes back to Dawkins (1982).

theory, i.e. for the controversy between the Modern Synthesis and a so-called Extended Synthesis (e.g. Pigliucci and Müller 2010; Noble et al. 2014; Laland et al. 2014).

3.2 Conceptual and Historical Background of the Controversy

A number of biological research disciplines, such as classical, population, molecular and developmental genetics, are gene-centred in the sense that their conceptual and methodological research framework is centred on the concept of the gene. Their 'methodological gene-centrism', however, can be contrasted with the 'meta-biological' perspective that life is also in essence a gene-centred phenomenon or process, a position that could be termed 'ontological gene-centrism'.[2] Both methodological and ontological gene-centrism reinforced each other during the course of twentieth-century biology. Research successes in methodologically gene-centred disciplines such as classical, population and molecular genetics reinforced the belief in life's gene-centred nature (i.e. ontological gene-centrism) and, vice versa, theoretical accounts on life's ontologically gene-centred foundation (e.g. Muller 1922, 1966; Schrödinger 1944; Dawkins 1976) catalysed the development of methodologically gene-centred disciplines such as population genetics, molecular genetics and gene-selectionism.

The idea of the gene/genome as the 'central (re)source' of order in cells and organisms (ontological gene-centrism) can be traced back to August Weismann's thesis—later dubbed the 'Weismann barrier'—which stated that structural specifications or instructions can be transferred from germ-plasm to soma-plasm but not the other way around (Weismann 1889, 1893, 1904) (for more details, see Mayr 1982; Gould 2002; Haig 2007; Tanghe 2013; De Tiège et al. 2014). A few decades later, the geneticist Hermann J. Muller wrote in his 'Variation Due to Change in the Individual Gene' (Muller 1922, p. 32) that 'genes exist as ultramicroscopic particles', that 'their influences nevertheless permeate the entire cell', that 'they play a fundamental role in determining the nature of all cell substances, cell structures and cell activities', and that 'through these cell effects, in turn, the genes affect the entire organism'. Muller's ideas on the *centrality of the gene in life* were backed up by his research on X-ray-induced genetic mutations causing changes in the biochemistry and development of cells and organisms (i.e. gene-centrism at the proximate physical–biochemical–physiological–developmental level) which can, thereafter, be transgenerationally inherited and subjected to natural selection (i.e. gene-centrism at the ultimate-evolutionary level). His views were highly influential on the rise of both population genetics and molecular biology and genetics (Witkin 2001). They

[2]A distinction between methodological and ontological gene-centrism was already made by Thange (2013, pp. 294–295).

also indirectly—via biophysicist Max Delbrück—inspired quantum physicist Erwin Schrödinger to identify the gene as an 'aperiodic crystal', i.e. as an 'equilibrium structure' characterized by low statistical entropy and held together by the strong chemical (covalent) bond (Schrödinger 1944). Schrödinger argued that the aperiodically crystalline genetic material is the main negentropic and/or 'informational' contributor to cellular and organismal organization, allowing the latter to metabolize and develop resources into ordered biomass—a process he termed 'order from disorder' (cf. gene-centrism at the proximate physical–biochemical–physiological–developmental level). This intra-generational process was contrasted with the inter-generational process of 'order from order', meaning cellular and organismal reproduction underpinned by the replication of the genetic material (cf. gene-centrism at the ultimate-evolutionary level). Schrödinger's book further catalysed the gene-centred direction taken by post-war molecular biology and biology in general (Olby 1974; Morange 1998; Keller 2000; Moss 2003). It was, among others, influential on Crick's (1958, 1970) 'central dogma' of molecular biology, a somewhat 'molecularized' version of the Weismann barrier, which states that genetic sequence-specificity or information cannot pass from protein to protein nor from protein 'backwards' to nucleic acid (DNA/RNA).[3] The development of gene-centrism through the work of Weismann, Muller, Schrödinger, Crick and others resulted in the understanding of the gene/genome as a kind of 'central source' (Griesemer 2002, 2005) or 'ROM-device' (Shapiro 2011, 2013) from which there is a quasi-unidirectional flow of order, negentropy and/or information into cellular biochemistry and development.

Gene-centrism, which is centred on the *intra-molecular crystalline* structure of genetic nucleic acid sequences, can be contrasted with a more holistic complex systems perspective on life, emphasizing the *inter- and supra-molecular*—thus holistic—network dynamics of many molecular species of which nucleic acids are one of them. Here too, the distinction can be made between a 'methodological holism/complexity', exemplified by disciplines such as pre-war embryology and present-day systems biology, and the 'ontological holism/complexity' that is argued for by more theoretically and meta-biologically inclined biologists interested in the very nature of life, such as Goodwin (1984, 1994) and Kauffman (1993, 1995, 2000). From the 1980s onwards, these and other theorists (e.g. Oyama 1985; Oyama et al. 2001; Gilbert and Sarkar 2000; Keller 2000; Van Speybroeck 2000; Griesemer 2002; Moss 2003; Jablonka and Lamb 2005; Stotz 2006a, b; Callebaut et al. 2007; Shapiro 2009, 2011, 2013; Noble 2008, 2010, 2011, 2012; Griffiths and Stotz 2013; De Tiège et al. 2014; Noble et al. 2014) have been criticizing the 'reductionist' gene-centred orthodoxy of reducing the complex inter- and supra-molecular organization of cells and organisms to the intra-molecular order of

[3]The central dogma is derived from some mechanically implausible if not impossible transfers of linear sequence information. Sequence information can be transferred between DNA- and RNA-sequences (transcription and reverse transcription) and from DNA-/RNA-sequences to amino acid sequences (translation), but not from amino acid sequences to nucleic acid sequences (no reverse translation), nor from 3D protein to either protein or nucleic acid.

just one class of molecules—nucleic acid sequences. The gene-centred orthodoxy sees the structure of genetic nucleic acid sequences as the primary or main (re) source of order in cellular and organismal organization, while the complex systems perspective casts genes as 'one among many resources' within the collective inter- and supra-molecular dynamics or self-organization responsible for cellular and organismal organization. That is, the latter perspective pleads for a profound 'contextualization', and perhaps even a 'de-centralization', of the gene within the inter- and supra-molecular dynamics of cells and organisms as complex systems. As Kauffman (1995, p. 83; also cited in Moss 2003, p. 75) aptly summarized: 'At its heart, the debate centres on the extent to which the sources of order in biology lie predominantly in the stable bond structures of molecules, Schrödinger's main claim, or in the collective dynamics of a system of such molecules'. That is, the debate revolves around whether genes/genomes present the 'crystalline core' of cells and organisms *versus* just 'one among many components' making up cells and organisms.

3.3 How the Current State of Knowledge in Biology Vindicates Both Life's Holistic Systems Complexity and Gene-Centeredness (And Falsifies Both Genetic Determinism–Reductionism and Gene-de-Centrism)

We will now demonstrate how the present state of knowledge in biology vindicates both life's holistically complex and gene-centred nature, while at the same time falsifies extreme genetic 'determinism' and 'reductionism' (according to which biochemical, cellular and organismal organization are quasi-exhaustively determined by and reducible to genetic sequence information) as well as extreme 'gene-de-centrism' (according to which genes are quasi-entirely 'de-centralized' and/or 'relativized' within the complex biochemical and developmental dynamics of cells and organisms).

3.3.1 The Vindication of Holistic Systems Complexity and the Falsification of Genetic Determinism–Reductionism

The understanding of the gene/genome as a kind of 'central source' (Griesemer 2002, 2005) or 'ROM-device' (Shapiro 2011, 2013) from which there is a quasi-unidirectional flow of order, negentropy and/or information into cellular biochemistry and development, has been challenged by data from molecular, developmental and systems biology. One of the first inroads on the orthodox 'DNA-centric' perspective prevalent during the 1950s and 1960s was due to the

discovery of *reverse transcription* (Crick 1970; Temin and Mizutani 1970). If reverse transcription would be non-existent or uncommon, the negentropic/informational constraints from the genome on the transcriptome would be stronger than vice versa. However, due to the high frequency rates of reverse transcription (Temin 1985; Brosius 1999, 2003; Shapiro 2009, 2011), the flow of negentropy/information among the two is far from unidirectional. In humans, for instance, over one-third of the genetic DNA-variation in the genome stems from reverse-transcribed RNA (International Human Genome Sequencing Consortium 2001). Reverse-transcribed RNA contributes considerably to the origin of new genetic DNA-variation in eukaryotes as well as in archaea and bacteria. Brosius (2003) has therefore argued that it may be a relic of the evolutionary transition from the proto- or early-biotic RNA world towards the current DNA/RNA world, when RNA-genes would have been gradually replaced by and, hence, transcribed into DNA-genes. The pervasive bidirectionality of transcription among DNA and RNA seriously corrodes DNA-centrism and delivers a picture of a kind of global 'NA-genome' containing both the DNA- and RNA-sequences of the cell, as such enforcing an extension from DNA-centrism to 'NA-centrism' (De Tiège et al. 2014).

Moreover, as several critics of gene-centrism (e.g. Thieffry and Sarkar 1998; Griesemer 2002; Moss 2003; Stotz 2006a, b; Noble 2008, 2011, 2012; Shapiro 2009, 2011, 2013; Griffiths and Stotz 2013; De Tiège et al. 2014; Noble et al. 2014) pointed out, the causal flow of negentropy and information from NA-sequences all the way to cellular and organismal (phenotypic) organization, too, is far from purely unidirectional and is 'corroded', 'diluted' and 'contextualized' due to substantial *extra-genetic input*, i.e. due to *causal co-determination* by factors not reducible to the NA-sequence-specificity in the organism's genome, such as follows:

- enzyme-, cell- and environmentally mediated regulations of gene activity;
- enzyme-, cell- and environmentally mediated modifications of the genome architecture (e.g. mobile genetic elements, lateral gene transfer);
- enzyme-, cell- and environmentally mediated *pre-translational* modifications of DNA- and RNA-sequences (e.g. directed mutagenesis, changes in DNA-sequence due to proofreading and repair, DNA-methylation causing the mutation of methylated cytosine into thymine, RNA-editing, RNA-splicing);
- enzyme-, cell- and environmentally mediated *translational* recoding such as frameshifting, programmed bypassing and codon redefinition (Baranov et al. 2003; Stotz 2006b);
- enzyme-, cell- and environmentally mediated *post-translational* protein modifications due to covalent alterations on the ribosomes (Shapiro 2009, 2011);
- the fact that gene products (functional RNAs and proteins) are not the sole components making up cells and organisms and the fact that genes are not the sole factors being inherited during cellular and organismal reproduction (e.g. Jablonka and Lamb 2005; Rando and Verstrepen 2007; Jablonka and Raz 2009; Lamm 2014).

These facts show how genetic sequence information is undeniably 'diluted' and 'contextualized' within the collective biochemical and developmental dynamics of

cells and organisms as complex systems (cf. Callebaut et al. 2007; Noble et al. 2014). Perspectives such as genetic 'determinism' and 'reductionism' become seriously flawed if not falsified: cells and organisms are not exhaustively determined or specified by their genetic sequences—they are not reducible to the sequence information in their genomes alone. Therefore, the current state of knowledge in biology decisively falsifies genetic determinism and reductionism while at the same time vindicating life's holistic complexity.

3.3.2 *The Vindication of Gene-Centrism and the Falsification of Gene-de-Centrism*

Does the falsification of genetic determinism and reductionism also imply a falsification of gene-centrism? According to most critics, it does (e.g. Oyama et al. 2001; Jablonka and Lamb 2005; Stotz 2006a, b; Callebaut et al. 2007; Shapiro 2009, 2011; Noble 2008, 2010, 2011, 2012; Noble et al. 2014). They do not really distinguish gene-centrism from genetic determinism and reductionism. Gene-centrism, however, does *not* demand:

(i) that every aspect of biochemical, cellular and organismal (phenotypic) organization is exhaustively determined by and reducible to genetic sequence information (i.e. genetic determinism and reductionism), *nor*
(ii) that the gene/genome is an absolutely 'sealed' central source, i.e. that there are no 'Lamarckian' violations of the Weismann barrier and the central dogma at all—thus that 'natural genetic engineering' (term Shapiro 2011, 2013) and 'downward causation' (Noble 2008, 2012) are non-existent.

Further relying on Tanghe's (2013, pp. 371–372) comparison between gene-centrism and heliocentrism, one might compare the situation with the heliocentred nature of our planetary system. Analogous to gene-centrism, heliocentrism does *not* imply:

(i) that all planetary processes and behaviours are exhaustively determined by and reducible to solar processes (i.e. no 'helio-determinism' and 'helio-reductionism'), *nor*
(ii) that there are no causal influences from the planets on the Sun at all.

Rather, heliocentrism refers to the fact that the Sun is the (gravitational and electromagnetic) 'power centre' of the planetary system, i.e. to the fact that the causal (gravitational and electromagnetic) power or constraints from the Sun on the planets are significantly *stronger* than the constraints from the latter on the former. Analogously, gene-centrism would hold if genes are at the (negentropic and informational) 'power centre' of biochemical, cellular and organismal organization, i.e. if the causal (negentropic and informational) constraints from genetic sequence-specificity on biochemical, cellular and organismal organization would be

significantly *stronger* than the constraints from the latter on the former (in the 'Lamarckian', 'natural genetic engineering' and/or 'downward causation' direction). Just as the Sun cannot be fully 'de-centralized' within the planetary system—the Sun is not simply a heavenly body 'among' the planets—genes too could then not be fully 'de-centralized' within cells and organisms, i.e. genes would then not simply be molecules 'among' the other molecules that make up cells and organisms.

It is important to realize this, because gene-centrism is often attacked with the help of correct but inappropriate arguments, arguments that, in reality, falsify genetic determinism and reductionism but not gene-centrism in the proper sense of the word. A constantly recurring example is the (correct) statement that genes are one among many extra-genetic causal factors contributing to the biochemistry and development of cells and organisms (Oyama et al. 2001; Jablonka and Lamb 2005; Noble 2008, 2010, 2012; Noble et al. 2014).[4] Biochemistry and development are indeed not exclusively determined by and reducible to genetic information. However, although this argument works successfully against genetic determinism and reductionism, it is not an appropriate strategy to counter or falsify gene-centrism. An analogous argument, i.e. that the Sun is one among many non-solar causal factors contributing to planetary processes and behaviour—thus that not everything occurring on planets is exclusively determined by and reducible to solar processes, could be used to counter or falsify 'helio-determinism' and 'helio-reductionism' but *not* heliocentrism. Indeed, the latter nevertheless holds, as the causal (gravitational and electromagnetic) constraints from the Sun on the planets are significantly *stronger* than the constraints from the latter on the former, thereby allowing the Sun to be the (gravitational and electromagnetic) 'power centre' of the planetary system. Hence, the key question for gene-centrism is as follows:

> Are genes effectively at the (negentropic and informational) 'power centre' of biochemical, cellular and organismal organization? That is, are the causal (negentropic and informational) constraints from genetic sequence-specificity on biochemical, cellular and organismal organization effectively *stronger* than the constraints from the latter on the former, thus than the constraints in the 'Lamarckian', 'natural genetic engineering' and/or 'downward causation' direction from phenotypic (cellular and organismal) organization and gene products (such as enzymes) on genetic NA-sequences?[5]

The answer is almost incontrovertibly 'yes'. The causal (negentropic and informational) constraints from genetic NA-sequences via gene products on cellular and

[4]See also Stegmann (2012) for a philosophical analysis of the argument.

[5]In De Tiège et al. (2014), we demonstrated that the causal-informational constraints from genetic NA-sequences on both the conformational structure and functioning of gene products (functional RNAs and proteins) are significantly stronger and more pervasive than the other way around (i.e. than in the 'natural genetic engineering' direction), thereby justifying a modest kind of gene-centrism or 'NA-sequence-centrism' confined to the subcellular level of NA/protein-based biochemistry. However, the question on any extension or generalization of gene-centrism beyond this limited domain, i.e. to more 'peripheral' domains such as higher levels of biological organization, remained unanswered, as we did not give a reason why the (inevitably weaker) causal-informational constraints from genetic sequences on such more peripheral zones would still be stronger than those involved in the 'natural genetic engineering' of DNA- and RNA-sequences.

organismal (phenotypic) organization in the 'upward', 'ontogenetic', 'Weismannian' and/or 'central dogma' direction is—in spite of all the 'corrosive' processes mentioned in Sect. 3.3.1—much more *robust, canalized and/or statistically reliable* than in the 'downward', 'Lamarckian' and/or 'natural genetic engineering' direction. If we consider the most *radical and intrusive* instances of 'Lamarckian' and/or 'downward' causation, i.e. those involving effective enzyme-, cell- and environmentally mediated *modifications of NA-sequence-specificity* (e.g. mobile genetic elements, directed mutagenesis, changes in DNA-sequence due to proofreading and repair, DNA-methylation causing the mutation of methylated cytosine into thymine, RNA-editing and RNA-splicing), then we see that, even here, there is no such thing as a robust, canalized and/or statistically reliable 'reverse transformation' from specific functional states in gene products 'feed-backwards' into linear genetic NA-sequences, and certainly not a robust, canalized and/or statistically reliable 'reverse ontogeny' from specific functional states in cellular and organismal phenotypic organization 'feed-backwards' into linear genetic NA-sequences.

The heliocentrism analogy is particularly clarifying in this regard. Heliocentrism is based on the fact that the Sun is at the (gravitational and electromagnetic) 'power centre' of the planetary system—thus on the fact that the causal (gravitational and electromagnetic) constraints from the Sun on the planets are significantly stronger than the other way around. Therefore, a physical (e.g. gravitational or electromagnetic) change in the state of the Sun may cause a change in the state of the planets, but a change in the state of a planet is *less likely* to cause a change in the state of the Sun. And indeed, similarly, a change or mutation in genetic NA-sequence may cause a change in gene products and in phenotypic (cellular and organismal) organization, but a change in phenotypic organization or in a gene product is *less likely* to cause a change or mutation in genetic NA-sequence. That is, the negentropic-informational power from genetic sequence-specificity on gene products and phenotypic organization is indeed stronger than the power from the latter on the former. *This simple fact puts genetic sequence-specificity at the (negentropic and informational) 'power centre' of biochemical, cellular and organismal organization, in an analogous way as the Sun is at the (gravitational and electromagnetic) 'power centre' of the planetary system, thus rendering invalid nearly all criticism on gene-centrism—although not on genetic determinism and reductionism—formulated during the past three decades.* While obviously causally 'integrated' and 'contextualized' within the collective biochemical and developmental dynamics of cells and organisms as complex systems, genes nevertheless *resist a full 'de-centralization' or 'relativization'*—which is in line with the position of many biologists (e.g. Gilbert et al. 1996; Hall 2001, 2003; Gilbert 2003; Haig 2007; Wagner 2014).

It is important to realize that, due to an increasing amount of intervening co-determining extra-genetic factors on the ontogenetic 'upward' causal pathway from genetic sequences via gene products all the way to biochemical, cellular and organismal organization, there is—the 'higher-up' we go—an increasing 'independency' and/or 'plasticity' of these higher levels of organization towards the genes/genome. Analogous with the planetary system: the further away from the Sun, the weaker the causal (gravitational and electromagnetic) constraints from the

Sun on that planet or satellite. However, as long as the causal (gravitational and electromagnetic) constraints from the Sun on a satellite are strong enough, the satellite is still 'captured' in the causal (gravitational and electromagnetic) 'field' of the Sun. Similarly, the further away from the genes/genome, the weaker the causal (negentropic and informational) constraints from the genes/genome on that level of biological organization. However, as long as the constraints from the genes/genome on a particular level of organization are strong enough, that level would still be 'captured' in the causal (negentropic and informational) 'field' of the genes/genome. Even levels of biological organization outside the cellular boundaries of an organism, i.e. so-called extended phenotypic organization (cf. Dawkins 1982, 2004), could still be 'captured' in the causal (negentropic and informational) 'field' of the organism's genes/genome. To use the planetary system analogy again, such extended phenotypes would 'circle' in a (admittedly loose) 'trajectory' around the organism's crystalline 'power centre'—its genes/genome.

3.3.3 *Preliminary Conclusion*

The preceding considerations can be summarized as follows: the present state of knowledge in biology strongly suggests that life on Earth should be regarded both (i) holistically and dynamically complex and (ii) gene-centred. (i) Holism/complexity refers to the causal 'integratedness' or 'contextualizedness', and *not* to a radical 'de-centeredness', of genes within the complex biochemical and developmental dynamics of cells and organisms. And (ii) gene-centrism refers to the causal (negentropic and informational) constraints from genetic sequence-specificity on biochemical, cellular and organismal organization being significantly stronger than the other way around, and *not* to a radical genetic 'determinism' and 'reductionism' according to which biochemical, cellular and organismal organization are quasi-exhaustively determined by and reducible to genetic information. The extreme and radical perspectives are falsified, while the moderate perspectives are vindicated by the current state of knowledge of life on Earth.

3.3.4 *The Vindication of Both Life's Holistic Systems Complexity and Gene-Centeredness (and the Falsification of Both Genetic Determinism–Reductionism and Gene-de-Centrism) by Cross-Species Genome Transplantations*

In this final subsection, we want to indicate that the results of bacterial genome transplantation and cross-species cloning, in one time, empirically vindicate both life's holistic systems complexity and gene-centeredness while falsifying both

radical genetic determinism–reductionism and gene-de-centrism.[6] In 2007, the J. Craig Venter Institute transplanted the genome of one bacterial species (*Mycoplasma mycoides*) into another, closely related (genome-deprived) bacterial species (*Mycoplasma capricolum*), thereby turning the recipient species into the donor species (Lartigue et al. 2007). As Pennisi (2007) remarked, the experiment has only been carried out between two closely related microbial species lacking cell walls. Indeed, donor genetic sequences cannot do anything without recipient biochemical and cellular organization, not even without recipient organization that is closely related and somehow 'compatible' with the donor genetic material, thereby unambiguously displaying life's holistically complex nature. But contrary to what one might think at first sight, this does not undermine gene-centrism. Relevant for gene-centrism is the question whether the negentropic-informational (transformational) constraints from (donor) genetic sequences on (recipient) biochemical and cellular organization are *stronger* than those from the latter on the former (cf. Sect. 3.3.2). This is undeniably the case here: the donor genes transform the recipient cell into a donor species cell, whereas the recipient biochemical and cellular organization does *not* transform the donor genes into recipient species genes. More precisely, when we *balance* the negentropic-informational (transformational) power of (donor) genetic sequences *versus* (recipient) biochemical and cellular organization, then—after fusing both powers—there are three theoretical possibilities:

- The resulting species is more like the recipient species (which is not the case): the negentropic-informational constraints of the (recipient) biochemical and cellular dynamics would be stronger than those of the (donor) genetic sequences.
- The resulting species is a hybrid (which is also not the case): the negentropic-informational constraints of the (recipient) biochemical and cellular dynamics and the (donor) genetic sequences would be about equally strong.
- The resulting species is more like the donor species (which it is): the negentropic-informational constraints of the (donor) genetic sequences are stronger than those of the (recipient) biochemical and cellular dynamics.

That is, the negentropic-informational constraints imposed by (donor) genetic sequence-specificity on (recipient) biochemical and cellular organization appear to be definitively *stronger* than the constraints from the latter on the former, resulting in a bacterial cell that definitively belongs to the donor genetic species, thus displaying genes at the 'power centre' of (microbial) life.

A similar rationale applies to *cross-species cloning* experiments in eukaryotes: the nuclear genome of one (endangered or even extinct) species is inserted into the enucleated egg cell of a related (more common) species, thereby converting the recipient species into the donor species (e.g. Lanza et al. 2000; Loi et al. 2001, 2007; Gomez et al. 2004; Williams et al. 2006; Folch et al. 2009; Hajian et al. 2011). Life's holistically complex nature is apparent from the indispensable role of the

[6]Some of the reasoning in this section can already be found in Tanghe (2013, pp 386–389).

biochemical and cellular organization of the recipient egg cell, among others, in the process of 'nuclear reprogramming', which refers to the structural and functional chromatin modifications that are imposed by the enucleated host oocyte on the inserted nuclear genome to restore the totipotency of the zygotic nucleus. Until now, in only a minority of cross-species cloning experiments, the totipotency of the zygotic nucleus is successfully restored; in the majority of cross-species clones, epigenetic drift leads to abnormal, non-viable epiphenotypes, explaining the large number of early deaths both before and after birth. These difficulties notwithstanding, in all of the cases the (non-viable and viable) embryos and born individuals definitively belong to the nuclear donor species and not to the enucleated host egg cell species (neither are they hybrids) (cf. Tanghe 2013, pp 386–389). It is therefore beyond any doubt that, as in the bacterial genome transplantation experiment, the negentropic-informational (transformational) constraints from (donor) genetic sequence-specificity on (recipient) biochemical and cellular organization are definitively *stronger* than the constraints from the latter on the former, thereby putting genes at the 'power centre' and displaying life's gene-centred aspect.

Although in *cross-genus* cloning the causal constraints from (recipient) cytoplasmic egg cell factors on (donor) gene regulation and expression are somewhat more striking (Sun and Zhu 2014), there is no mention of substantial modifications of (donor) genetic *sequence-specificity*. Therefore, here too, the negentropic-informational (transformational) constraints from (donor) genetic sequence-specificity on (recipient) biochemical and cellular organization appear to be definitively *stronger* than the constraints from the latter on the former, thus displaying life's gene-centeredness. For example, although the implementation of the nuclear genome of a common carp (*Cyprinus carpio*) into the enucleated egg cell of a goldfish (*Carassius auratus*) resulted in offspring with a goldfish number of vertebrae (due to gene regulation mediated by recipient cytoplasmic egg cell factors), the offspring definitively belongs to the nuclear donor carp species. As Sun et al. (2005, p. 513) report, 'the morphological data are solid evidence that the common carp nuclei directed the development of the cross-genus cloned fish'.

The indispensability of extra-genetic such as epigenetic, cytoplasmic and ecological elements of the recipient egg cell species (Noble 2008, 2011) does not falsify the gene-centred aspect of life (cf. Tanghe 2013, pp 386–389). Surely, donor genetic sequences cannot do anything without recipient extra-genetic organization, not even without recipient organization that is relatively closely related to and compatible with the donor genetic material, thus displaying life's holistically complex nature. However, this does not undermine life's gene-centeredness, since the latter relies on the negentropic-informational (transformational) constraints from (donor) genetic sequence-specificity on (recipient) biochemical and cellular organization being significantly *stronger* than the constraints from the latter on the former, which is also displayed in these experiments. Therefore, the data on bacterial genome transplantation and cross-species/genus cloning unequivocally (i) support or vindicate *both* life's holistic complexity and gene-centeredness and (ii) disprove or falsify *both* extreme or radical genetic determinism–reductionism and gene-de-centrism.

3.4 From the Proximate to the Evolutionary Level: Some Implications for Evolutionary Theory

Without entering into extensive details about the theoretical and bio-philosophical discussions that take place within modern evolutionary theory, we will nevertheless point to some important implications for the controversy between the Modern Synthesis and a so-called Extended Synthesis (e.g. Pigliucci and Müller 2010; Noble et al. 2014; Laland et al. 2014). The Modern Synthesis originally grew out of the synthesis between Mendelian genetics and Darwinian evolutionism (Huxley 1942; Mayr and Provine 1980). From the beginning, the focus was on the role of genes in evolution: gene and genotype selection, genetic drift, genetic mutation, genetic recombination and gene flow. Advocates of an Extended Synthesis plead for the incorporation of extra-genetic higher level causation in evolution. Some important, partially overlapping examples are directed mutagenesis (Jablonka and Lamb 2005, Chap. 3; Rando and Verstrepen 2007) and natural genetic engineering (Shapiro 2011, 2013), epigenetic causation and inheritance (Rando and Verstrepen 2007; Jablonka and Raz 2009), developmental, phenotypic and behavioural plasticity and inheritance (West-Eberhard 2003; Moczek et al. 2011), niche construction and inheritance (Scott-Phillips et al. 2014) and multilevel selection processes (Okaska 2006; Godfrey-Smith 2009).

The empirical data on extra-genetic causation in evolution are solid, although relatively scarce compared to the overwhelming data on genetic causation (gene and genotype selection, genetic drift, genetic mutation, genetic recombination and gene flow). On the one hand, as already underscored, gene products and phenotypic (cellular and organismal) organization are not exhaustively determined by and reducible to genetic sequence-specificity (i.e. no genetic determinism and reductionism). This not-to-genes-reducible complexity and specificity causally (biochemically, physiologically, developmentally, behaviourally) interacts with the environment, and is thus expected to play an equally not-to-genes-reducible causal role in evolutionary processes. As such, life's holistically complex aspect would naturally extend from the proximate to the evolutionary level. That is, biological evolution would not be exhaustively reducible to genetic evolution and, for example, evolution by natural selection would not be exhaustively reducible to gene and genotype selection (cf. Okaska 2006; Godfrey-Smith 2009).

On the other hand, since the causal (negentropic-informational) constraints from genetic sequence-specificity on gene products and phenotypic (cellular and organismal) organization are significantly *stronger* than those from the latter on the former (see supra), extra-genetic higher level processes—which are virtually *permeated* by functional gene products (functional RNAs and proteins) and, hence, by genetic sequence-specificity—could never be fully divorced from their genetic base or 'centre'. As Wray et al. (in Laland et al. 2014, p. 164), for instance, write on epigenetic causation in evolution: 'we know of no case in which a new trait has been shown to have a strictly epigenetic base divorced from gene sequence'. Although epigenetic specificity is not exhaustively determined by and reducible to

genetic sequence-specificity (i.e. no genetic determinism and reductionism), the negentropic-informational constraints from genetic sequence-specificity on epigenetic specificity are significantly *stronger* than the constraints from epigenetic causation (through 'natural genetic engineering'—see Sect. 3.3.1) on genetic sequence-specificity. That is, the flow of statistical negentropy and information from the genetic sequence level to the epigenetic level is significantly *stronger* than in the opposite direction. More generally, since genes resist a full 'de-centralization' in the biochemistry and development of cells and organisms (see supra), genes could also not be fully 'de-centralized' in the causal (biochemical, physiological, developmental, behavioural) interaction processes of cells and organisms with their environments and, thus, in the evolutionary process. Hence, life's gene-centred aspect, too, would naturally extend from the proximate to the evolutionary level.

Thus, while the Modern Synthesis was arguably too narrowly focussed on genetic causation, the latter cannot simply be de-centralized either. Therefore, the term 'Postmodern Synthesis' (Whitfield 2008) seems inappropriate. But an 'Extended Synthesis' that takes into account irreducible extra-genetic causation in evolution (Pigliucci and Müller 2010; Laland et al. 2014), as well as an 'Evolutionary Systems Biology' that takes into account the different levels of network complexity in which life is organized (Medina 2005; Koonin and Wolf 2006; Soyer 2012), seems preferable, however, without losing sight or touch with the genetic or 'crystalline' centre or baseline around or upon which all those levels are organized. Indeed, biological evolution should be viewed as *both* a holistically complex and gene-centred process; *both* aspects of life's nature most likely extend from the proximate to the evolutionary level.

3.5 Conclusion and Further Prospects

Both intra-molecular crystalline genetic sequence-specificity and inter- and supra-molecular holistic network complexity are constitutive properties of life on Earth. On the one hand, life is holistically and dynamically complex: it is constituted by complex *inter- and supra-molecular* network dynamics and interconnectivity. On the other hand, this inter- and supra-molecular network organization is virtually *permeated and constrained* by functional gene products (i.e. functional RNAs and proteins) and, hence, by genetic sequence information derived from the *intra-molecular crystalline* structure of nucleic acid sequences. Recognizing only one of life's characteristics, whether this be holistic network complexity or gene-centeredness, prevents a more balanced and complete understanding of life's multifaceted nature. Due to the ground-breaking advances in molecular biology and genetics during the second half of the twentieth century, intensive focus was laid on life's gene-centred aspect, while holistic network dynamics was under-investigated or even neglected. At present, however, due to current successes in systems biology, some authors underestimate—or are even unwilling to recognize—life's

gene-centred aspect. As such, they 'over-jump' or 'cover up' the very 'subtle' and 'fine-grained' intra-molecular crystalline aspect of life.

Here, we showed that the main argument raised against gene-centrism, viz. that genes are one among many extra-genetic causal factors contributing to the biochemistry and development of cells and organisms (e.g. Oyama et al. 2001; Jablonka and Lamb 2005; Noble 2008, 2010, 2012; Noble et al. 2014), only undermines or falsifies genetic determinism and reductionism but not gene-centrism. More broadly, we showed how the current state of knowledge in biology strongly suggests that life on Earth should be regarded as both (i) holistically and dynamically complex and (ii) gene-centred. (i) Holism/complexity refers to the causal 'integratedness' or 'contextualizedness', and *not* to a radical 'de-centeredness', of genes within the complex biochemical and developmental dynamics of cells and organisms. And (ii) gene-centrism refers to the causal (negentropic and informational) constraints from genetic sequence-specificity on biochemical, cellular and organismal organization being significantly stronger than the other way around, and *not* to a radical genetic 'determinism' and 'reductionism' according to which biochemical, cellular and organismal organization are quasi-exhaustively determined by and reducible to genetic information. We also underscored how bacterial genome transplantation and cross-species cloning experiments, in one time, empirically falsify the extreme or radical perspectives while vindicating the moderate perspectives. Finally, we indicated that life's dual nature most likely extends from the proximate to the evolutionary level, a conclusion that bears some relevance to the controversy between the Modern Synthesis and an 'Extended Synthesis'.

Concerning future prospects, much work remains to be done, especially on the conceptual integration between twentieth-century gene-centred disciplines (e.g. population and evolutionary genetics, molecular biology and genetics) and other, more recent branches such as systems biology and evolutionary developmental biology (cf. De Backer et al. 2010; Pigliucci and Müller 2010; Noble et al. 2014). Of crucial importance in this endeavour is not to undermine or neglect one of life's constitutive aspects. It is obvious that a strategy that tries to 'reduce' all inter- and supra-molecular dynamics and extra-genetic complexity to genetic variation is not an option. However, neither is a strategy that tries to radically 'de-centre' or 'relativize' genetic sequence-specificity within the overall dynamics of cells and organisms. A more subtle approach is required: integration/contextualization *through* centralization, or centralization *through* integration/contextualization.

Acknowledgements Preparation of this manuscript was made possible by the Fund for Scientific Research Flanders (FWO), Belgium (Project number: G001013N).

References

Baranov PV, Gurvich OL, Hammer AW, Gesteland RF, Atkins JF (2003) RECODE 2003. Nucleic Acids Res 31:87–89

Brosius J (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238:115–134

Brosius J (2003) The contribution of RNAs and retroposition to evolutionary novelties. Genetica 118:99–116

Callebaut W, Müller GB, Newman SA (2007) The organismic systems approach: EvoDevo and the streamlining of the naturalistic agenda. In: Sansom R, Brandon R (eds) Integrating evolution and development: from theory to practice. MIT Press, Cambridge, pp 25–92

Crick FHC (1958) On protein synthesis. Symp Soc Exp Biol 12:138–163

Crick FHC (1970) Central dogma of molecular biology. Nature 227:561–563

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dawkins R (1982) The extended phenotype. Oxford University Press, Oxford

Dawkins R (2004) Extended phenotype—but not too extended. A reply to Laland, Turner and Jablonka. Biol Philos 19:377–396

De Backer P, De Waele D, Van Speybroeck L (2010) Ins and outs of systems biology vis-à-vis molecular biology: continuation or clear cut? Acta Biotheor 58:15–49

De Tiège A, Tanghe K, Braeckman J, Van de Peer Y (2014) From DNA- to NA-centrism and the conditions for gene-centrism revisited. Biol Philos 29:55–69

Folch J, Cocero MJ et al (2009) First birth of an animal from an extinct subspecies (*Capra pyrenaica pyrenaica*) by cloning. Theriogenology 71:1026–1034

Gilbert SF (2003) Evo-devo, devo-evo, and devgen-popgen. Biol Philos 18:347–352

Gilbert SF, Sarkar S (2000) Embracing complexity: organicism for the 21st century. Dev Dyn 219:1–9

Gilbert SF, Opitz JM, Raff RA (1996) Resynthesizing evolutionary and developmental biology. Dev Biol 173:357–372

Godfrey-Smith P (2009) Darwinian populations and natural selection. Oxford University Press, NY

Gomez MC, Pope CE et al (2004) Birth of African wildcat cloned kittens born from domestic cats. Cloning Stem Cells 6:247–258

Goodwin BC (1984) A relational or field theory of reproduction and its evolutionary implications. In: Ho M-W, Saunders PT (eds) Beyond Neo-Darwinism: an introduction to the new evolutionary paradigm. Academic Press, London, pp 219–241

Goodwin BC (1994) How the Leopard changed its spots: the evolution of complexity. Weidenfeld and Nicolson, London

Gould SJ (2002) The structure of evolutionary theory. Harvard University Press, Cambridge (MA)

Griesemer J (2002) What is "epi" about epigenetics? Ann NY Acad Sci 981:97–110

Griesemer J (2005) The informational gene and the substantial body: on the generalization of evolutionary theory by abstraction. In: Jones MR, Cartwright N (eds) Idealization XII: correcting the model. idealization and abstraction in the sciences (Poznan studies in the philosophy of sciences and the humanities, vol 86). Rodopi, Amsterdam/NY, pp 59–115

Griffiths P, Stotz K (2013) Genetics and philosophy: an introduction. Cambridge University Press, Cambridge

Haig D (2007) Weismann rules! OK? Epigenetics and the Lamarckian temptation. Biol Philos 22:415–428

Hajian M, Hosseini SM et al (2011) "Conservation cloning" of vulnerable Esfahan mouflon (*Ovis orientalis isphahanica*): in vitro and in vivo studies. Eur J Wildl Res 57:959–969

Hall BK (2001) The gene is not dead, merely orphaned and seeking a home. Evol Devel 3:225–228

Hall BK (2003) Unlocking the black box between genotype and phenotype: cell condensations as morphogenetic (modular) units. Biol Philos 18:219–247

Huxley JS (1942) Evolution: the modern synthesis. Allen and Unwin, London
International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
Jablonka E, Lamb M (2005) Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life. MIT Press, Cambridge (MA)
Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q Rev Biol 84:131–176
Kauffman SA (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, Oxford
Kauffman SA (1995) What is life? Was Schrödinger right? In: Murphy MP, O'Neill AJ (eds) What is Life? The next fifty years. Cambridge University Press, Cambridge, pp 83–114
Kauffman SA (2000) Investigations. Oxford University Press, Oxford
Keller EF (2000) The century of the gene. Harvard University Press, Cambridge
Koonin EV, Wolf YI (2006) Evolutionary systems biology: links between gene evolution and function. Curr Opin Biotechnol 17:481–487
Laland K, Uller T, Feldman M et al (2014) Does evolutionary theory need a rethink? Nature 514:161–164
Lamm E (2014) Inheritance systems. In: Zalta EN (ed) The stanford encyclopedia of philosophy (Spring 2014 Edition) URL:http://plato.stanford.edu/archives/spr2014/entries/inheritance-systems/
Lanza RP, Cibell JB, Diaz F, Moraes CT, Farin PW, Farin CE, Hammer CJ, West MD, Damiani P (2000) Cloning of an endangered species (*Bos gaurus*) using interspecies nuclear transfer. Cloning 2:79–90
Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA III, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. Science 317:632–638
Loi P, Ptak G, Barboni B, Fulka J, Cappai P, Clinton M (2001) Genetic rescue of an endangered mammal by cross-species nuclear transfer using post-mortem somatic cells. Nat Biotechnol 19:962–964
Loi P, Galli C, Ptak G (2007) Cloning of endangered mammalian species: any progress? Trends Biotechnol 25:195–200
Mayr E (1982) The growth of biological thought: diversity, evolution, and inheritance. Harvard University Press, Massachusetts London
Mayr E, Provine WB (eds) (1980) The evolutionary synthesis: perspectives on the unification of biology. Harvard University Press, Massachusetts London
Medina M (2005) Genomes, phylogeny, and evolutionary systems biology. PNAS 102:6630–6635
Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, Abouheif A, Pfennig DW (2011) The role of developmental plasticity in evolutionary innovation. Proc R Soc B 278:2705–2713
Morange M (1998) A history of molecular biology. Harvard University Press, Cambridge
Moss L (2003) What genes can't do. MIT Press, Cambridge
Muller HJ (1922) Variation due to change in the individual gene. Am Nat 56:32–50
Muller HJ (1966) The gene material as the initiator and the organizing basis of life. Am Nat 100:493–517
Noble D (2008) Genes and causation. Phil Trans R Soc A 366:3001–3015
Noble D (2010) Biophysics and systems biology. Phil Trans R Soc A 368:1125–1139
Noble D (2011) Neo-Darwinism, the modern synthesis, and selfish genes: are they of use in physiology? J Physiol 589:1007–1015
Noble D (2012) A biological theory of relativity: no privileged level of causation. J R Soc Interface Focus 2:55–64
Noble D (2013) Physiology is rocking the foundations of evolutionary biology. Exp Physiol 98:1235–1243
Noble D, Jablonka E, Joyner MJ, Müller GB, Omholt SW (2014) Evolution evolves: physiology returns to centre stage. J Physiol 592:2237–2244
Okasha S (2006) Evolution and the levels of selection. Oxford University Press, Oxford

Olby R (1974) The path to the double helix: the discovery of DNA. Dover Publications Inc., New York
Oyama S (1985) The ontogeny of information: developmental systems and evolution. Cambridge University Press, Cambridge
Oyama S, Griffiths PE, Gray RD (eds) (2001) Cycles of contingency: developmental systems and evolution. MIT Press, Cambridge
Pennisi E (2007) Replacement genome gives microbe new identity. Science 316:1827
Pigliucci M, Müller GB (2010) Evolution—the extended synthesis. The MIT Press, Cambridge
Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell 128:655–668
Schrödinger E (1944) What is Life? The physical aspect of the living cell. Cambridge University Press, Cambridge
Scott-Phillips TC, Laland KN, Shuker DM, Dickins TE, West SA (2014) The niche construction perspective: a critical appraisal. Evolution 68:1231–1243
Shapiro JA (2009) Revisiting the central dogma in the 21st century. Natural genetic engineering and natural genome editing. Ann NY Acad Sci 1178:6–28
Shapiro JA (2011) Evolution: a view from the 21st century. FT Press Science, Upper Saddle River
Shapiro JA (2013) How life changes itself: the read-write (RW) genome. Phys Life Rev 10:287–323
Soyer OS (ed) (2012) Evolutionary systems biology. Advances in experimental medicine and biology 751. Springer, New York
Stegmann U (2012) Varieties of parity. Biol Philos 27:903–918
Stotz K (2006a) Molecular epigenesis: distributed specificity as a break in the central dogma. Hist Phil Life Sci 28:527–544
Stotz K (2006b) With 'genes' like that, who needs an environment? Postgenomics' argument for the 'ontogeny of information'. Philos Sci 73:905–917
Sun Y-H, Zhu Z-Y (2014) Cross-species cloning: influence of cytoplasmic factors on development. J Physiol 592:2375–2379
Sun Y-H, Chen S-P, Wang Y-P, Hu W, Zhu Z-Y (2005) Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. Biol Reprod 72:510–515
Tanghe KB (2013) The non-mendelian revolution: a conceptual reinterpretation of the genetic revolution. Unpublished Doctoral Dissertation, Ghent University, Ghent
Temin HM (1985) Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol Biol Evol 2:455–468
Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226:1211–1213
Thieffry D, Sarkar S (1998) Forty years under the central dogma. Trends Biochem Sci 23:312–316
Van Speybroeck L (2000) The organism: a crucial genomic context in molecular epigenetics? Theory Biosci 119:187–208
Waddington CH (1957) The strategy of the genes. Macmillan, New York
Wagner GP (2014) Homology, genes, and evolutionary innovation. Princeton University Press, Princeton
Weismann A (1889) Essays upon heredity and kindred biological problems, vol 1. Clarendon Press, Oxford
Weismann A (1893) The all-sufficiency of natural selection: a reply to Herbert Spencer. Contemp Rev 64:309–338
Weismann A (1904) The evolution theory (trans: Thomson JA, Thomson M). Edward Arnold, London
West-Eberhard MJ (2003) Developmental plasticity and evolution. Oxford University Press, Oxford
Whitfield J (2008) Postmodern evolution? Nature 455:281–284
Williams JB, Shin T et al (2006) Cloning of exotic/endangered species: desert bighorn sheep. Methods Mol Biol 348:169–181
Witkin EM (2001) H. J. Muller and the nature of the gene. Genetics 157:461–463

Part II
Genetic Mechanisms of Diversification

Chapter 4
Retrotransposons: Genomic and Trans-Genomic Agents of Change

David L. Adelson, Reuben M. Buckley, Atma M. Ivancevic, Zhipeng Qu and Lu Zeng

Abstract Genome structure in higher eukaryotes is highly dependent on the type and abundance of transposable elements, particularly retrotransposons, in their non-coding DNA. Retrotransposons are generally viewed as genomic parasites that must be suppressed in order to ensure genome integrity. This perception is based on the instances of retrotransposons having caused deleterious structural variation in genomes. Recent data are beginning to provide a more positive view of the impact of retrotransposons, particularly in mammals, where the evolution of the placenta has depended on the exaptation of a type of retrotransposon, endogenous retroviruses. Finally, exosome trafficking of retrotransposons between cells has been shown to induce the innate immune system gene expression, possibly indicative of a role for retrotransposons in the regulation of the innate immune system. It may be time for us to review the status of retrotransposons and reclassify them as symbionts rather than parasites.

4.1 Evolutionary Origin and Structure of Retrotransposons

Genome structure and function are two sides of the same coin, and retrotransposons (AKA retrotransposable elements, retroelements and retroposons), self-replicating DNA sequences that are found in all eukaryotic taxa, have the capacity to make larger changes to genome structure than other sources of variation—such as DNA polymerase errors that lead to single nucleotide variation (SNV). Because retrotransposons can account for the majority of the genome sequence in eukaryotes, their accumulation and clade specificity have been implicated in speciation, regulation of gene expression, exaptation and structural variation. Understanding the

D.L. Adelson (✉) · R.M. Buckley · A.M. Ivancevic · Z. Qu · L. Zeng
School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
e-mail: david.adelson@adelaide.edu.au

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_4

mechanisms that govern retrotransposon distribution and replication is thus of fundamental importance.

The evolutionary origin of retrotransposons is a matter of debate, but sequence similarity of their reverse transcriptases with the catalytic subunit of telomerase (Eickbush 1997; Lingner et al. 1997) and phylogenetic studies of reverse transcriptase sequences can be interpreted to indicate that reverse transcriptase may have evolved from telomerase, or telomerase is the result of co-opting reverse transcriptase. However, there are also good arguments for the ancient, prokaryotic origin of reverse transcriptase as a descendant of group II introns, which are mobile, self-splicing introns (Boeke 2003).

Retrotransposons can be divided into four major classes (Eickbush and Jamburuthugoda 2008). This classification is based on the reverse transcriptase enzyme required for replication and encoded by these elements. In vertebrates, retrotransposons can account for half of the genome sequence, and in plants, up to 70 % of the genome. This chapter is focused on the mammalian/vertebrate retrotransposons and these are commonly described as falling into two broad categories: those containing long terminal repeats (LTR) and those not containing LTR (non-LTR) (Jurka et al. 2007).

Non-LTR retrotransposons encode their own internal promoter and one or two open reading frames (ORFs) with reverse transcriptase and endonuclease activities that are used for replication (Fig. 4.1). LTR containing retrotransposons resemble (endogenous) retroviruses (ERVs) in that they can contain additional ORFs similar to those found in retroviruses, and these are referred to as endogenous retrovirus-like elements (ERVL). ERVL LTR retrotransposons are believed to have evolved from DNA transposons (Bao et al. 2010) and then acquired additional genes from viruses such as *env*, allowing them to become retrovirus-like and to produce infectious particles.

4.2 The Retrotransposon Life cycle

Retrotransposons replicate via an RNA intermediate that is reverse transcribed and reinserted into the genome (Fig. 4.1) at short target motifs (Fig. 4.2) (Cost and Boeke 1998). For non-LTR retrotransposons, also called long interspersed elements (LINE), transcription is initiated by an internal Pol II promoter and the resulting transcript is then translated to produce two proteins, one of which, ORF2p has both reverse transcriptase and endonuclease activities (Feng et al. 1996; Moran et al. 1996). ORF2p has the ability to recognise short target sequences and initiate nicks at those locations which subsequently serve to prime the reverse transcription of the retrotransposon RNA directly into the genome (Eickbush and Jamburuthugoda 2008; Morrish et al. 2002).

Some retrotransposons do not contain ORFs (non-autonomous) and are dependent on retrotransposons that do (autonomous) (Jurka et al. 2007). Autonomous retrotransposons are longer (LINEs), whereas the shorter, non-autonomous

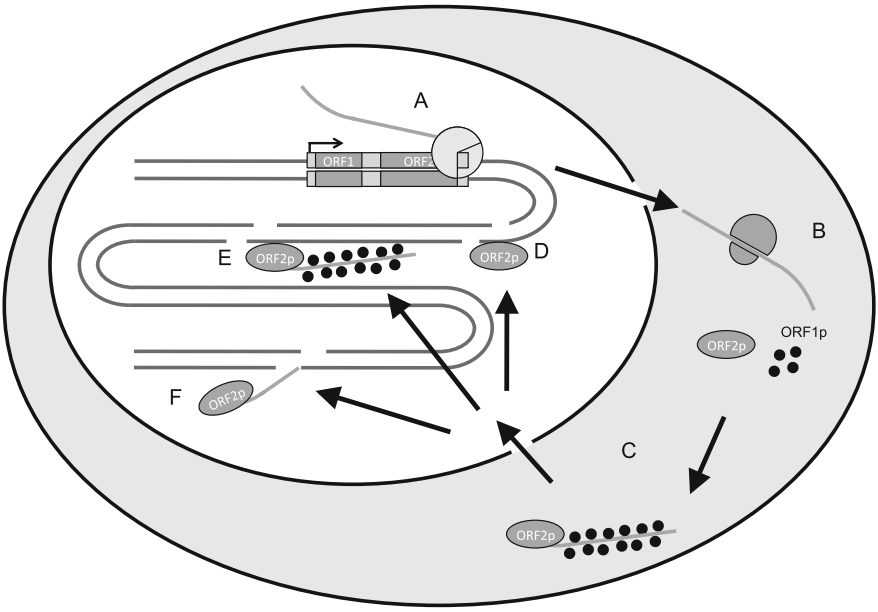


Fig. 4.1 Retrotransposon life cycle: *A* TEs are transcribed by RNA Pol II and exported to the cytoplasm (Swergold 1990). *B* In the cytoplasm, ORF1 and ORF2 are both translated. The ORF1 protein (*ORF1p*) is an RNA-binding protein believed to aid the entry of LINE L1 RNA into the nucleus (Martin 2006). The ORF2 protein (*ORF2p*) has both endonuclease and reverse transcriptase activities (Feng et al. 1996; Moran et al. 1996). *C* To enter the nucleus, ORF1p and ORF2p form a complex with the L1 RNA known as a ribonuclear protein (*RNP*) (Martin 2006). *D* The endonuclease activity of ORF2p creates double-stranded breaks without insertion of TEs (Gasior et al. 2006). *E* The endonuclease activity is essential for the process of target-primed reverse transcription (*TPRT*). TPRT requires that ORF2p creates a nick in each strand at the integration site. The LINE L1 RNA is then used as a template for the reverse transcriptase activity of ORF2p (Cost et al. 2002). *F* L1 RNA is able to insert into and aid in repairing double-stranded breaks independent of the endonuclease activity of ORF2p (Morrish et al. 2002)

elements are called short interspersed elements (SINEs). While LINEs are usually ubiquitously distributed across taxa, SINEs are usually clade specific, as they result from the fusion of an internal promoter containing transcript with the 3' end of a LINE.

The mechanism of SINE creation is still an open question, but most likely is a function of aspects of the LINE life cycle. SINEs have a composite structure: a 5' end similar to 5' tRNA, 7SL RNA or 5S rRNA promoters, a unique region and a 3' end similar to the 3' tail of LINEs (Piskurek and Jackson 2012). The most accepted hypothesis on SINE origins is based on the proposed template-switching mechanism of Buzdin et al. (Buzdin et al. 2002; Gilbert and Labuda 2000; Gogvadze and Buzdin 2009, Kramerov and Vassetzky 2005; Ohshima and Okada 2005). This template-switching mechanism is based on the study of pseudogenes, where the LINE (L1) reverse transcriptase switches from its own L1 mRNA to other nearby

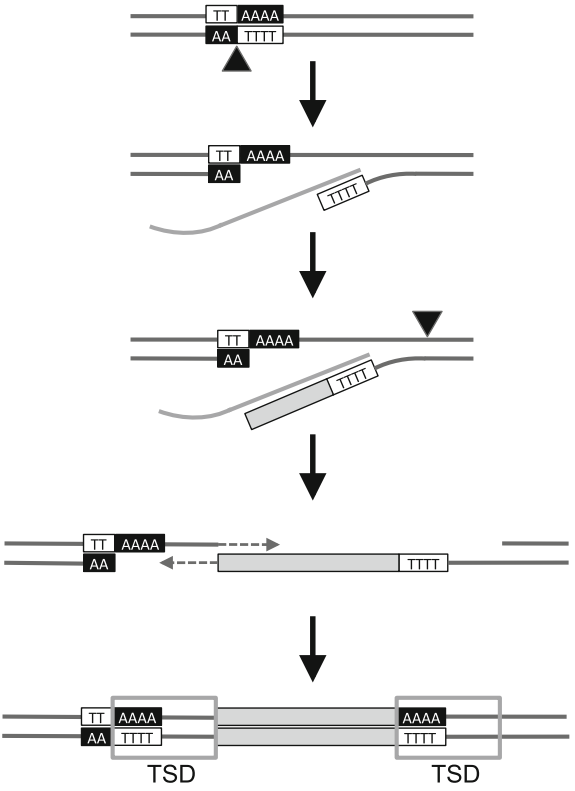


Fig. 4.2 Target-primed Reverse Transcription (*TPRT*) is how retrotransposons are inserted into the genome. ORF2p endonuclease activity creates a nick in the DNA at the AA/TTTT target site (Cost and Boeke, 1998). ORF2p reverse transcriptase activity then uses the cDNA copy as a template for DNA synthesis. Next ORF2p endonuclease activity creates a second nick in the DNA. The second DNA strand is then synthesised via double-strand break (*DSB*) repair and results in the formation of short target site duplications (*TSD*)

mRNA sequences through an RNA–RNA recombination process, thus creating new recombinant pseudogenes (and possibly SINEs) during L1 insertion (Buzdin et al. 2002; Gogvadze et al. 2007; Ichiyanagi et al. 2007; Piskurek and Jackson 2012). However, other investigators have suggested direct transposon into transposon (TnT) insertion as an alternative mechanism for the creation of novel transposable elements (Giordano et al. 2007; Ichiyanagi et al. 2007; Kriegs et al. 2007). The TnT mode of retrotransposon generation is what has led to the formation of SVA (SINE/VNTR/Alu) elements in humans, which are chimeric elements that can be mobilised by L1 elements and contain Alu-like sequence, Variable Number of Tandem Repeats (VNTR) sequence and SINE-R sequence resulting from a series of TnT events (Ostertag et al. 2003). The template-switching and TnT mechanisms are not mutually exclusive, and it is clear that both operate to create new SINEs, but at present we do not know which mechanism dominates.

Because retrotransposons can control their own expression through internal promoters [Pol II for LINEs and Pol III for SINEs and ERVs (Belancio et al. 2010a; Dieci et al. 2013)], expression is inextricably linked to the retrotransposon replication and to the evolution of new SINEs. As a result of this ability to autonomously insert new copies from expressed sequences into the genome, eukaryotes

have evolved mechanisms to keep retrotransposon expression in check in order to avoid large-scale deleterious structural variation.

4.2.1 Retrotransposon Suppression

There appear to be two main mechanisms for retrotransposon suppression: transcriptional repression and post-transcriptional degradation (Fig. 4.3). Transcriptional repression can be caused by methylation of retrotransposon promoters or alteration of chromatin state to make retrotransposons transcriptionally inaccessible. Proof for the importance of methylation is evident from the phenotype of *dnmt3l* (DNA (cytosine-5)-methyltransferase 3-like) knockout mice (Bourc'his and Bestor 2004; Webster et al. 2005), which undergo meiotic catastrophe associated with the rampant expression of retrotransposons in male germ cells. The *dnmt3l* locus encodes a protein that regulates methyl transferase activity required to methylate and suppress the activity of CpG islands in retrotransposon promoters (Vlachogiannis et al. 2015). In addition to CpG island methylation, transcription can be repressed by the alteration of chromatin status (Fadloun et al. 2013), and this may be mediated by piRNA transported to the nucleus (Kuramochi-Miyagawa et al. 2008).

Post-transcriptional degradation of retrotransposon RNA in the male germ line is mediated by piRNAs derived from retrotransposon sequences and amplified by the ping-pong reaction (Aravin et al. 2008). In the female germ line, the situation appears to be different, with siRNAs shown to mediate retrotransposon transcript destruction via the RNA-induced silencing complex (RISC) pathway (Ciaudo et al. 2013; Watanabe et al. 2008).

There may also be additional mechanisms that can suppress retrotransposons at the translational level (Grivna et al. 2006; Tanaka et al. 2011) or even at the post-translational level to interfere with ORF proteins binding to retrotransposon transcripts (Fig. 4.3) (Goodier et al. 2012). In spite of all of these mechanisms to suppress retrotransposons at various steps in their life cycle, they are still transcribed at some developmental stages and in many somatic tissues (Belancio et al. 2010b). Perhaps suppression is a loaded term in this context and perhaps what we are observing is actually the regulation of retrotransposon expression.

4.2.2 Retrotransposon Expression

At certain phases of the mammalian life cycle, retrotransposons are negatively regulated to a lesser degree and are therefore transcribed and able to retrotranspose. Because methylation of cytosine to 5-methyl-cytosine (5mC) is critical to retrotransposon silencing, retrotransposons are potentially most active at times of low genomic 5mC content, which occurs in mouse embryos at around 3.5 days of embryonic development and also in primordial germ cells (Hackett and Surani 2013).

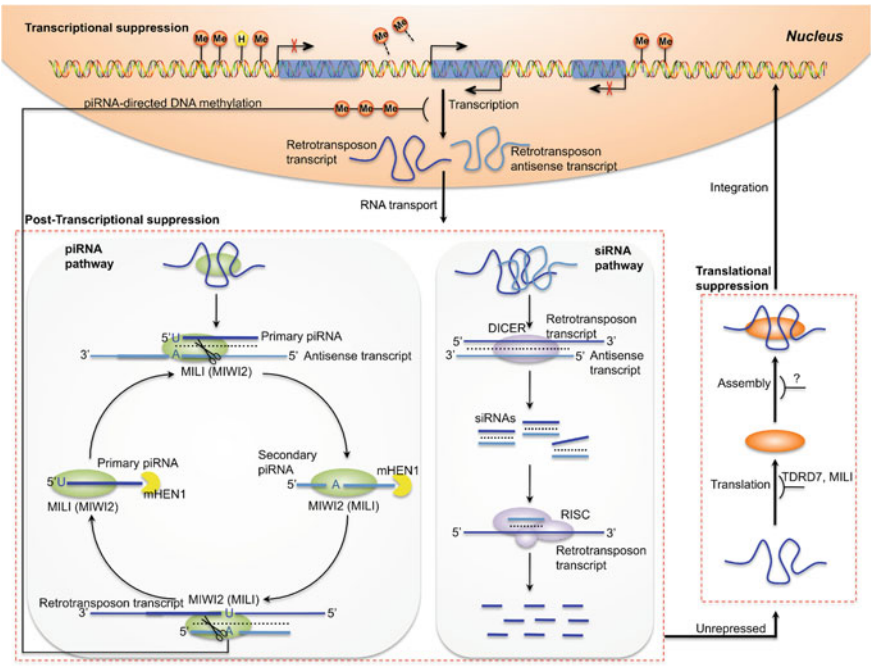


Fig. 4.3 A schematic overview of retrotransposon suppression. Retrotransposons can be suppressed by different mechanisms throughout their life cycle (Crichton et al. 2014). Transcriptional suppression: In most cell types, retrotransposons are in a repressed state due to high levels of DNA methylation or histone modifications (Fadloun et al. 2013; Meissner et al. 2008). In some specific developmental stages and cell types, some retrotransposon RNAs can be transcribed bidirectionally and transported from the nucleus to the cytoplasm (Fadloun et al. 2013). Post-transcriptional suppression: Retrotransposon RNAs can be silenced through the piRNA pathway (mostly in the male germ line) or siRNA pathway (mostly in the female germ line). The ping-pong cycle is a well-characterised model for piRNA synthesis. In the mouse, sense retrotransposon RNAs are processed into primary piRNAs. MILI (or MIWI2) is recruited to cleave antisense retrotransposon RNAs into secondary piRNAs with the guidance of primary piRNAs, and mHEN1 is used to subsequently methylate their 3' termini. Secondary piRNAs then bind with MIWI2 (or MILI) to cleave sense retrotransposon RNAs into primary piRNAs and close the loop of the ping-pong cycle (Aravin et al. 2008). piRNAs can also be transported to the nucleus to repress the transcription of retrotransposon by directing DNA methylation (Kuramochi-Miyagawa et al. 2008). For the siRNA pathway, sense and antisense retrotransposon transcripts can form double-strand RNAs, which are cleaved into double-strand siRNAs by DICER. Then, double-stranded siRNAs are unwound and loaded into the RISC to guide the degradation of retrotransposons (Ciaudo et al. 2013; Watanabe et al. 2008). Translational suppression: The Tudor domain-containing protein TDRD7 and MILI might be involved in the suppression of retrotransposon activity during translation (Grivna et al. 2006; Tanaka et al. 2011). Other repression mechanisms may also exist at later stages, such as the assembly stage of retrotransposon RNA and retrotransposon-encoded proteins (Goodier et al. 2012)

However, it is primarily in early embryos that L1 retrotransposons are transcribed and retrotranspose (Kano et al. 2009). Presumably, other suppression mechanisms keep retrotransposons in check in primary germ cells. In spite of significant levels of global 5mC in the genome at other stages of development, retrotransposons are also activated in specific somatic tissues, indicating that retrotransposon suppression is more complex than just ensuring high levels of 5mC, and it may be less stringent in some tissues/cell types. Faulkner et al. (2009) showed that up to 30 % of mouse or human transcripts from all tissues are of retrotransposon origin and that retrotransposons were transcribed in all tissues surveyed. Retrotransposon expression per se does not always mean that retrotransposition is occurring, as some retrotransposons have inserted into UTRs and are therefore transcribed as part of a mRNA. However, it has been shown in both neural progenitor cells and in the human brain that retrotransposition does occur at a detectable level, altering the genomic landscape of that tissue (Baillie et al. 2011; Coufal et al. 2009).

Retrotransposon expression and subsequent retrotransposition have significant impacts on the genomes of both germ line (via germ line insertions and early embryonic insertions) and soma. Germ line insertions can then be transmitted through vertical inheritance, while somatic insertions are not currently believed to contribute to the vertical inheritance of novel insertions. However, there is another mode of retrotransposon transmission: horizontal transfer, where retrotransposon sequences jump to another cell or species, and this type of transfer may be the result of a more general mechanism of intercellular retrotransposon transfer.

4.3 Horizontal Transfer

Horizontal transfer of transposons has been demonstrated in plants, insects and vertebrates. In the context of retroviruses (including ERVs that have maintained ORFs to support an infectious life cycle), horizontal transfer is a relatively commonplace event. For example, in plants, horizontal transfer of transposable elements is both widespread and frequent (El Baidouri et al. 2014). In animals, horizontal transfer of DNA transposons is also widespread (Ivancevic et al. 2013). A good example is in *Drosophila melanogaster* where P-elements swept through the population starting in the 1950s via horizontal transfer (Daniels et al. 1990). *Mariner* elements are also horizontally transmitted between species, including both insects and mammals (Lampe et al. 2003; Lohe et al. 1995; Maruyama and Hartl 1991). Furthermore, Space Invader (*SPIN*) elements have been horizontally transferred in mammals and other tetrapods, as have OC1 elements (Gilbert et al. 2010; Pace et al. 2008). It was not until the 1990s that the first evidence for horizontal transfer of retrotransposons was published, when the patchy phylogenetic distribution and likely horizontal transfer of BovB retrotransposons was first reported (Kordis and Gubensek 1998, 1999a).

4.3.1 BovB: An Example of Widespread Horizontal Transfer

The BovB retrotransposon (also known as LINE-RTE) is a 3.2 kb LINE with at least one large ORF encoding a reverse transcriptase and a possible small ORF1 overlapping with the large ORF (Malik and Eickbush 1998). In cattle and sheep, over a thousand full length BovB, hundreds of thousands of 5' truncated BovB fragments and derived SINEs (Bov-tA and Bov-tA2 (Lenstra et al. 1993; Okada and Hamada 1997) account for ~25 % of the genome sequence (Adelson et al. 2009; Jiang et al. 2014). The high degree of sequence conservation of BovB with sequences detected from the venom gland of *Vipera ammodytes* gave the first support to the idea of horizontal transfer of this retrotransposon (Kordis and Gubensek 1998, 1999b). BovB is now known to have a widespread, but patchy phylogenetic distribution, coupled to a high degree of sequence conservation, two of the hallmarks of horizontally transferred DNA (Fig. 4.4).

Even though BovB has horizontally transferred across a wide range of species, it has not always colonised the genome to the same extent in different species. Some

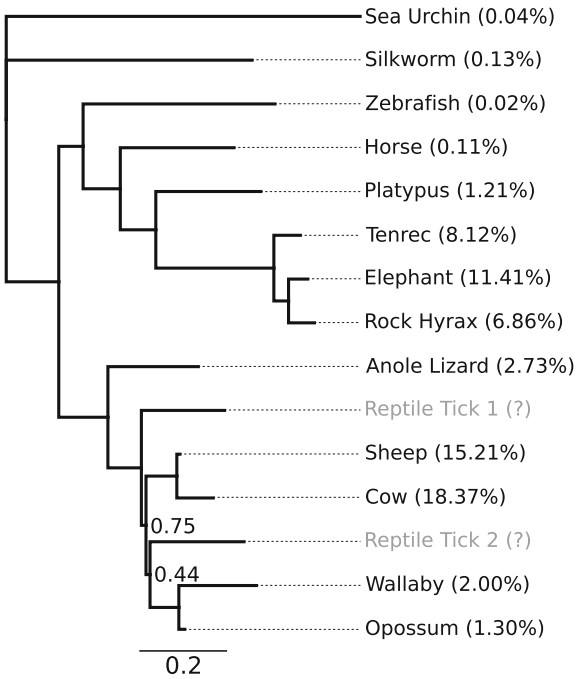


Fig. 4.4 BovB phylogeny Maximum likelihood tree of aligned BovB sequences based on Walsh et al. (2013), showing the sporadic distribution, sequence similarity and abundance of BovB elements across taxa. Local support values are only shown if <0.9. The labels at each branch tip give the species common name and (*in brackets*) the percentage of genome sequence identified as BovB elements for that species. Reptile Tick 1 is *Bothriocroton hydrosauri*, Reptile Tick 2 is *Amblyomma limbatum*; and the BovB genome coverage for these ticks is unknown

lineages such as ruminants and afrotheria have a high percentage of their genomes derived from BovB, whereas in other species BovB has not retrotransposed as prolifically (Fig. 4.4). This difference may be indicative of either variability in how different species suppress retrotransposons or it may simply reflect stochasticity in the population dynamics of retrotransposon expansion in different genomes. Presumably, the initial horizontal transfer event that results in retrotransposition and replication needs only a single germ line incorporation which can either replicate exponentially or "fizzle out" within the "genomic ecosystem" (Brookfield 2005; Le Rouzic et al. 2007). It is clear based on the currently available small and biased (towards mammals) sample of available genome sequences that retrotransposons as exemplified by BovB are capable of widespread and near ubiquitous horizontal transfer, and that this transfer might be enabled by parasites, such as ticks, that feed on blood. However, what is currently lacking is/are the molecular mechanism(s) for these transfers.

4.3.2 Possible Mechanisms/Modes of Transfer

A number of vectors, including arthropods, viruses, snails and DNA transposons, have been proposed for horizontal transfer, and the current state of knowledge was recently summarised by Ivancevic et al. (2013). It is relatively easy to see how a virus or transposon might act as a vector to package or transpose retrotransposons, but at the molecular level, it is not as obvious how eukaryotic vectors might effect the transfer of retrotransposon sequences between species, let alone into the germ line of another species.

4.3.2.1 Viruses as Vectors

For retrotransposons, the only example at present of a molecular virus vector is the taterapox virus (a dsDNA virus) which may have mediated transfer of Sauria SINE between reptiles and West African rodents (Piskurek and Okada 2007). This can be viewed as a highly unusual transfer, as a non-autonomous retrotransposon should not be as likely to colonise a new genome after transfer as an autonomous retrotransposon, such as a LINE. However, if cognate autonomous LINEs are present in both source and recipient species, a non-autonomous SINE could replicate effectively in the recipient species. RNA viruses have also been proposed as vectors of horizontal transfer for retrotransposons as they might package non-LTR retrotransposon transcripts inside infectious virus particles, but a tangible example for this type of transfer has yet to be demonstrated. Interestingly, *Mariner*-like DNA transposons are the plausible vectors for transfer of the CR1 retrotransposon in butterflies and moths (Sormacheva et al. 2012).

4.3.2.2 Endogenous Retroviruses/LTR Retrotransposons

As mentioned in Sect. 4.1, LTR retrotransposons are believed to have arisen from retrotransposons that acquired viral genes allowing them to become infectious, possibly leading to the evolution of retroviruses (Shimotohno and Temin 1981). In addition, waves of retroviral invasions into eukaryotic genomes have resulted in the formation of ERVs. While some ERVs have remained endogenous, occasionally they are able to become infectious and transfer to other genomes, where they can cause disease and eventually become domesticated. This is currently the case for a rodent ERV that has infected Koalas and is causing leukaemia in its new host while colonising the germ line as a new ERV (Tarlinton et al. 2006). Over time, domesticated retroviruses (ERVs) have contributed significantly to the genomic landscape of eukaryotes and have been co-opted into various aspects of eukaryotic biology (Feschotte and Gilbert 2012). In addition to this evolution of the capacity for horizontal transfer via infection, it is possible that retroviruses could package non-infectious non-LTR retrotransposons as a part of their viral payload. While there is no solid evidence for such transfer, exosomes/microvesicles are able to incorporate virus particles and transfer them to adjacent cells. This raises the question of whether exosomes can also transfer retrotransposon sequences directly.

4.3.2.3 Exosomes/Vesicles as Vectors

Exosomes are a class of membrane vesicle that has recently been shown to contain protein and RNA including miRNAs, piRNAs and retrotransposon sequences that they can transport from cell to cell (Batagov and Kurochkin 2013, Li et al. 2013; Skog et al. 2008; Valadi et al. 2007; Villarroya-Beltri et al. 2013; Yuan et al. 2009). Furthermore, exosome transport of Pol III-produced retrotransposon sequences has been specifically shown to regulate cancer therapy resistance pathways, including interferon-stimulated genes by direct activation of retinoid acid-inducible gene 1 (RIG-I) (Boelens et al. 2014). One of the hallmarks of Pol III transcripts is their 5' triphosphate group, which is recognised specifically by RIG-I as a trigger for activation. Pol III is responsible for the transcription of primarily housekeeping-type genes such as tRNAs and rRNAs, but it also transcribes many other loci, including SINEs that have originated from a fusion of Pol III promoter containing transcripts with LINE 3' sequences (Belancio et al. 2010b; Dieci et al. 2013). Because retrotransposons are known to be somatically expressed (see Sect. 4.2.2) in many tissues and cell types, they are likely to be present in exosomes exported by those cell types.

In the context of horizontal transfer, one can envision a number of potential scenarios for intercellular transport of retrotransposon sequences by exosomes (Fig. 4.5). Exosome-mediated transfer could allow transfer of retrotransposon sequences from a mammal or reptile to somatic cells of a parasite such as a tick through blood-borne exosomes. Within the tick, exosome-mediated transfer could then allow transmission to the germ line from the soma and eventual transmission back to other species used as food sources by that species of tick.

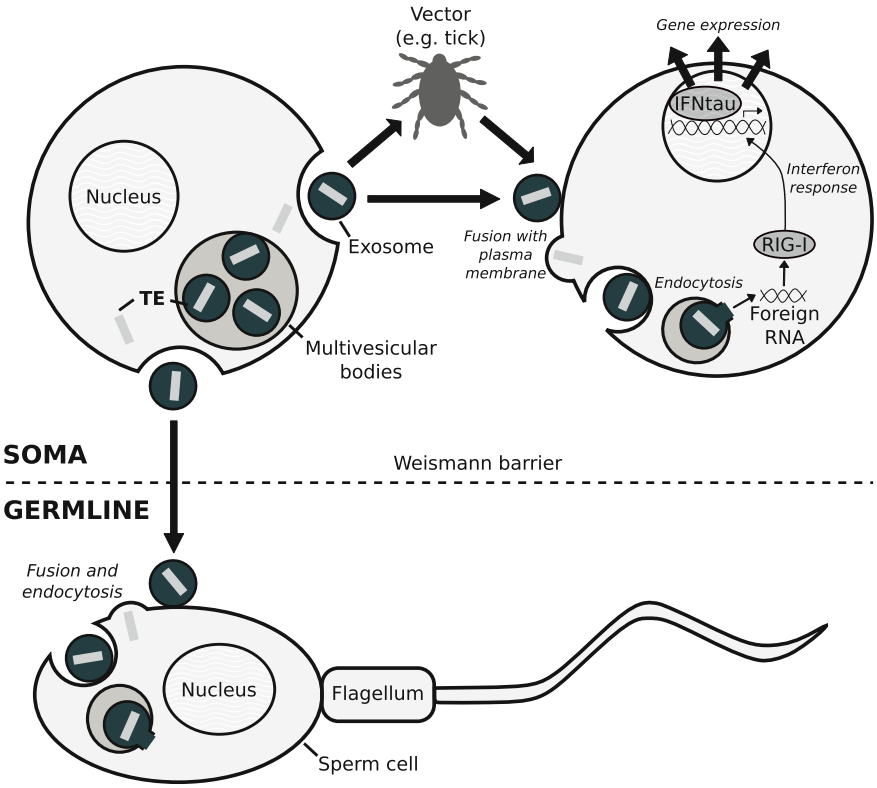


Fig. 4.5 Possible scenarios of intercellular transfer of transposable elements via exosomes. TEs packaged in exosomes can be transferred between both somatic and germline cells. Within an organism, a TE can travel from a somatic, exosome-generating cell directly (e.g. through the blood) into a somatic, exosome-target cell by fusing with the plasma membrane and undergoing endocytosis. Similarly, TEs can be horizontally transferred between the somatic cells of different organisms or species, via some kind of vector (e.g. a parasite). Exosomes can also carry TEs from the soma to the germ line, making them a permanent change in the genome that is eventually passed down to the offspring. Note that for simplicity only entry to the male germ line is shown above. In addition to the transfer of TEs, once inside the target cell, this "foreign RNA" from the TE can trigger an interferon pathway response by inducing the interferon signal transduction pathway via RIG-I. For example, in ruminants, exosomes loaded with ERV/TE RNAs trigger pattern recognition receptors, stimulating the innate immune system and production of interferon-tau, which plays a role in pregnancy recognition and placentation (see Sect. 4.4.4)

While one might envision that the existing piRNA-based suppression system might degrade these retrotransposon sequences rapidly, it also appears that retrotransposon sequences (as exosome cargo) have been co-opted into a signalling role for the innate immune system in vertebrates and used to activate interferon-stimulated genes in the absence of interferon (Dreux et al. 2012; Li et al. 2013). This would not be the first time that retrotransposon sequences have been co-opted for gene regulation (Feschotte 2008; Feschotte and Gilbert 2012), but it introduces a

new dimension of intercellular regulation of gene expression in the context of the evolutionary impact of retrotransposons.

4.4 Evolutionary Impacts

Retrotransposons are known to affect genome structure and hence function. The specific types of structural changes they introduce upon retrotransposition can have a wide-ranging set of subsequent effects in terms of genome structure, gene expression and gene function. More recently, it has become clear that retrotransposons have had a profound impact on the evolution of placentation in mammals.

4.4.1 Genome Structure

Retrotransposon insertion can directly perturb gene structure, but it can also have significant effects on a larger scale (Fig. 4.6). In particular, if retrotransposons form an array of elements with the same orientation on a chromosome, they can serve as

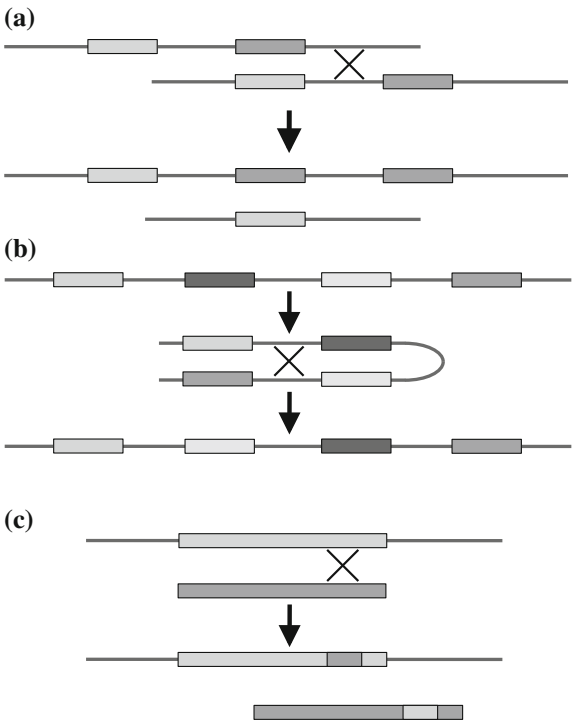


Fig. 4.6 Retrotransposons can lead to changes in genome structure. **a** Changes in CNVs result from non-allelic homologous recombination (*NAHR*) caused by the insertion of many TEs from the same family (Stankiewicz and Lupski 2002; Startek et al. 2015). **b** Chromosomal inversion is also the result of NAHR (Stankiewicz and Lupski 2002). **c** SINE elements have potential to drive change through gene conversion (Roy et al. 2000)

a substrate for non-allelic homologous recombination (NAHR) leading to segmental duplication (Fig. 4.6a) (Stankiewicz and Lupski 2002; Startek et al. 2015). However, statistical analysis of repeats in flanking regions of segmental duplications found that only ~10 % of segmental duplications could be attributed to flanking repetitive elements (Zhou and Mishra 2005). Other types of rearrangements have been shown to result from arrays of repeats such as inversions (Fig. 4.6b) and gene conversion (Fig. 4.6c).

While it is clear that retrotransposons can have indirect effects on genome structure as mentioned above, given the limitations inherent in identifying small segmental duplications and copy number variants the precise magnitude of these effects is unknown.

4.4.2 Gene Expression

As shown in Fig. 4.7, transposable elements can insert into and next to genes, affecting gene expression through multiple mechanisms, including epigenetic silencing of transcription, shortening a transcript via premature poly-Adenylation,

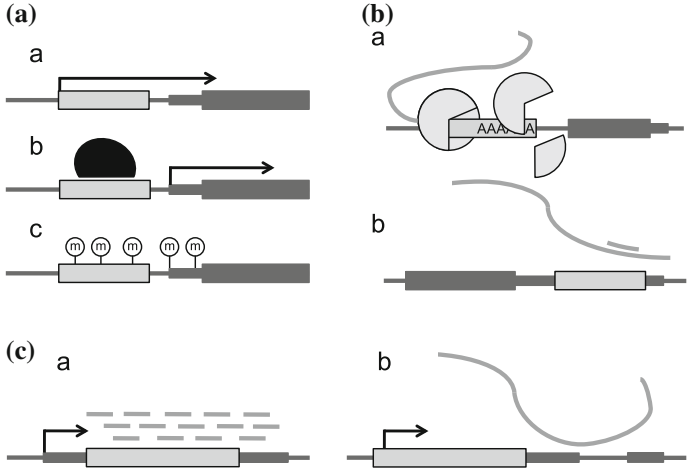


Fig. 4.7 Retrotransposons can alter gene expression. **a** 5' insertion of a retrotransposon with respect to a gene. *a* TEs are able to act as alternative promoters to adjacent genes (Faulkner et al. 2009; Speek 2001). *b* TEs are able to act as transcription factor binding sites (*TFBS*) and are thereby able to modulate gene expression (Bourque et al. 2008). *c* In plants, epigenetic silencing of TEs silences nearby genes; this is also likely to occur in animals (Buckley and Adelson 2014; Hollister and Gaut 2009). **b** 3' insertion of a retrotransposon *a* polyA signal/tail of the retrotransposon can result in shortened transcripts (Lee et al. 2008; Perepelitsa-Belancio and Deininger 2003). *b* Retrotransposon insertion in the 3' UTR of a gene can provide a target site for piRNAs which down-regulate gene expression (Watanabe et al. 2014). **c** Intergenic insertion of TEs. *a* Insertion of TEs into a piRNA cluster results in piRNAs that can target genes carrying TE-derived sequences (Yamamoto et al. 2013). *b* TEs involved in the origin and evolution of lncRNA (Kapusta et al. 2013)

driving piRNA expression or altering 3' UTR structure to affect mRNA stability. Analysis of retrotransposon insertions into or near genes has shown that many genes have been altered in ways that are likely to alter expression (Jjingo et al. 2011; Jordan et al. 2003) and analysis of enhancers has shown that retrotransposons drive the evolution of eukaryotic enhancers (McDonald et al. 1997). All of these effects on gene expression are subject to selection and are therefore part of the evolutionary process. Not all insertions into genes will affect regulation of gene expression, some can directly affect the coding sequence or coding potential of genes through exaptation.

4.4.3 Exaptation

When retrotransposons contribute to non-coding or protein coding exon sequences, they are referred to as exaptations. These exaptations may or may not be subject to immediate purifying selection, depending on the type of change they cause. Some exaptations that prove beneficial are selected for, but these are rare. Many examples of exaptation come from non-coding transcripts, where retrotransposon insertions have led to novel piRNA and miRNA transcripts (Jurka et al. 2007; Yamamoto et al. 2013). In fact, only ~50 instances of coding sequences derived from LTR retrotransposons syntenic between human and mouse have been identified (Jurka et al. 2007). One of these encodes the PEG10 (paternally expressed gene 10) locus, which is required for placentation. Occasionally, insertion of a retrotransposon sequence into an intron can lead to exonisation of part of the retrotransposon sequence as an alternative transcript through the presence of splice donor/acceptor sites in the sequence (Fig. 4.8). When this happens, sometimes the alternative transcripts are deleterious because of impaired function, and the regulation of alternative splicing may then become an additional regulatory mechanism for the affected gene (Lorenz et al. 2007).

4.4.4 Innate Immunity/Pregnancy Recognition

Some exaptations of retrotransposon sequences have been well-characterised, particularly in terms of the evolution of placentation. There is strong evidence for exaptation of ERV genes in both mouse and hominoid primates required for placental function (Chuong 2013; Haig 2012; Mallet et al. 2004). One of the most striking such exaptations is the role of endogenous jaagsiekte retrovirus (enJSRV) in ruminant pregnancy recognition and placentation. The domestic ruminant conceptus expresses interferon-tau (IFNT) from days 10 to 12, which dramatically alters gene expression in the uterine epithelium and stroma (Bazer et al. 2008; Dunlap et al. 2006; Gray et al. 2006; Spencer and Bazer 1995). At the same time, enJSRVs are released into the ruminant reproductive tract and they are known to

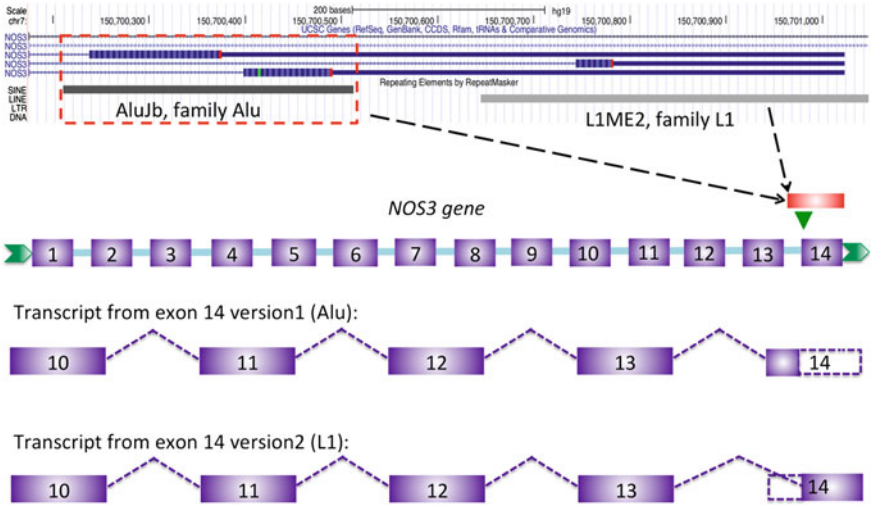


Fig. 4.8 Retrotransposon exaptation influences mRNA processing and can cause multiple splice variants. At the top, the UCSC browser (Kent et al. 2002) track for the human NOS3 gene is shown, including repeat element annotation. *Below*, a schematic of the 3' end of the human NOS3 gene illustrating an Alu element (*black bar*) inserted into intron 13. This retrotransposon provides exon 14 alternative splicing version 1. An adjacent L1 insertion can result in exon 14 alternative splicing version 2 (Lorenz et al. 2007). Dashed lines indicate a splicing event

regulate key peri-implantation development in the embryo and placenta (Dunlap et al. 2005, 2006). enJSRVs therefore have been exapted to regulate key aspects of development associated with implantation and placentation by virtue of their ability to trigger expression of IFNT expression in the conceptus. Recently, exosomes have been shown to be part of the specific mechanism used to trigger IFNT expression in this system, but without specifically testing for retrotransposon RNA content (Ruiz-Gonz ez et al. 2014, 2015). We speculate that exosomes loaded with retrotransposon sequences may also be involved in pregnancy recognition more generally in order to activate the STAT1 pathway in an interferon-free fashion.

SINE/ERV transcripts packaged into exosomes can trigger RIG-I in target cells leading to IFN independent activation of the IFN pathway, leading us to speculate that the role of retrotransposons is broader than previously thought, and that they may be involved in global regulation of the innate immune system.

4.5 Conclusion

Retrotransposons are abundant, found in a broad phylogenetic distribution and yet in spite of clade specific non-autonomous variants, exhibit a significant degree of commonality. Furthermore, their transcription is highly regulated, rather than

suppressed at all times. These facts, along with the evidence of pervasive and widespread horizontal transfer and an exosome-based mechanism for transfer that has likely co-evolved with the innate immune system and placentation, suggest to us that retrotransposons are not genomic parasites but rather genomic symbionts. We hypothesise that mammals and other vertebrates depend on these symbionts for cell-to-cell signalling in innate immunity and reproduction.

Acknowledgments The authors wish to thank R. Daniel Kortschak and Joy M. Raison for helpful discussions and advice.

References

Adelson DL, Raison JM, Edgar RC (2009) Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. Proc Natl Acad Sci USA 106:12855

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31:785

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA, Faulkner GJ (2011) Somatic retrotransposition alters the genetic landscape of the human brain. Nat Cell Biol 479:534

Bao W, Kapitonov VV, Jurka J (2010) Ginger dna transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons. Mob DNA 1(1):3. doi:10.1186/1759-8753-1-3

Batagov AO, Kurochkin IV (2013) Exosomes secreted by human cells transport largely mrna fragments that are enriched in the 3'-untranslated regions. Biol Direct 8:12. doi:10.1186/1745-6150-8-12

Bazer FW, Burghardt RC, Johnson GA, Spencer TE, Wu G (2008) Interferons and progesterone for establishment and maintenance of pregnancy: interactions among novel cell signaling pathways. Reprod Biol 8(3):179–211

Belancio VP, Roy-Engel AM, Deininger PL (2010a) All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol 20(4):200–210. doi:10.1016/j.semcancer.2010.06.001

Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P (2010 b) Somatic expression of LINE-1 elements in human tissues. Nucleic Acids Res 38:3909

Boeke JD (2003) The unusual phylogenetic distribution of retrotransposons: a hypothesis. Genome Res 13(9):1975–1983. doi:10.1101/gr.1392003

Boelens MC, Wu TJ, Nabet BY, Xu B, Qiu Y, Yoon T, Azzam DJ, Twyman-Saint Victor C, Wiemann BZ, Ishwaran H, Ter Brugge PJ, Jonkers J, Slingerland J, Minn AJ (2014) Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. Cell 159 (3):499–513. doi:10.1016/j.cell.2014.09.051

Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking dnmt3 l. Nature 431(7004):96–99. doi:10.1038/nature02886

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18(11):1752–1762. doi:10.1101/gr.080663.108

Brookfield JFY (2005) The ecology of the genome—mobile dna elements and their hosts. Nat Rev Genet 6(2):128–136. doi:10.1038/nrg1524

Buckley RM, Adelson DL (2014) Mammalian genome evolution as a result of epigenetic regulation of transposable elements. Biomol Concepts 5(3):183–194. doi:10.1515/bmc-2014-0013

Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E (2002) A new family of chimeric retrotranscripts formed by a full copy of u6 small nuclear rna fused to the 3' terminus of l1. Genomics 80(4):402–406

Chuong EB (2013) Retroviruses facilitate the rapid evolution of the mammalian placenta. Bioessays 35:853

Ciaudo C, Jay F, Okamoto I, Chen CJ, Sarazin A, Servant N, Barillot E, Heard E, Voinnet O (2013) Rnai-dependent and independent control of line1 accumulation and mobility in mouse embryonic stem cells. PLoS Genet 9(11):e1003,791. doi:10.1371/journal.pgen.1003791

Cost GJ, Boeke JD (1998) Targeting of human retrotransposon integration is directed by the specificity of the l1 endonuclease for regions of unusual dna structure. Biochemistry 37 (51):18081–18093

Cost GJ, Feng Q, Jacquier A, Boeke JD (2002) Human l1 element target-primed reverse transcription in vitro. EMBO J 21(21):5899–5910

Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH (2009) L1 retrotransposition in human neural progenitor cells. Nat Cell Biol 460:1127

Crichton JH, Dunican DS, Maclennan M, Meehan RR, Adams IR (2014) Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci 71(9):1581–1605. doi:10.1007/s00018-013-1468-0

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A (1990) Evidence for horizontal transmission of the p-transposable element between drosophila species. Genetics 124:339

Dieci G, Conti A, Pagano A, Carnevali D (2013) Identification of rna polymerase iii-transcribed genes in eukaryotic genomes. Biochim Biophys Acta 1829(3–4):296–305. doi:10.1016/j.bbagrm.2012.09.010

Dreux M, Garaigorta U, Boyd B, Décembre E, Chung J, Whitten-Bauer C, Wieland S, Chisari FV (2012) Short-range exosomal transfer of viral rna from infected cells to plasmacytoid dendritic cells triggers innate immunity. Cell Host Microbe 12(4):558–570. doi:10.1016/j.chom.2012.08.010

Dunlap KA, Palmarini M, Adelson DL, Spencer TE (2005) Sheep endogenous betaretroviruses (enjsrvs) and the hyaluronidase 2 (hyal2) receptor in the ovine uterus and conceptus. Biol Reprod 73(2):271–279. doi:10.1095/biolreprod.105.039776

Dunlap KA, Palmarini M, Varela M, Burghardt RC, Hayashi K, Farmer JL, Spencer TE (2006) Endogenous retroviruses regulate periimplantation placental growth and differentiation. Proc Natl Acad Sci USA 103(39):14390–14395. doi:10.1073/pnas.0603836103

Eickbush TH (1997) Telomerase and retrotransposons: which came first? Science (New York, NY) 277(5328):911–912

Eickbush TH, Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134:221

El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O (2014) Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24(5):831–838. doi:10.1101/gr.164400.113

Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla ME (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of line-1 by rna. Nat Struct Mol Biol 20(3):332–338. doi:10.1038/nsmb.2495

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest ARR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, Grimmond SM, Carninci P (2009) The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 41:563

Feng Q, Moran J, Kazazian H, Boeke J (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905

Feschotte C (2008) Opinion—transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397

Feschotte C, Gilbert C (2012) Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet 13(4):283–296. doi:10.1038/nrg3199

Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. Journal of Molecular Biology

Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1352

Gilbert N, Labuda D (2000) Evolutionary inventions and continuity of core-sines in mammals. J Mol Biol 298(3):365–377. doi:10.1006/jmbi.2000.3695

Giordano J, Ge Y, Gelfand Y, Abrusan G, Benson G, Warburton P (2007) Evolutionary history of mammalian transposons determined by genome-wide defragmentation. PLoS Comput Biol 3: e137

Gogvadze E, Buzdin A (2009) Retroelements and their impact on genome evolution and functioning. Cell Mol Life Sci 66(23):3727–3742. doi:10.1007/s00018-009-0107-2

Gogvadze E, Barbisan C, Lebrun MH, Buzdin A (2007) Tripartite chimeric pseudogene from the genome of rice blast fungus magnaporthe grisea suggests double template jumps during long interspersed nuclear element (line) reverse transcription. BMC Genomics 8:360. doi:10.1186/1471-2164-8-360

Goodier JL, Cheung LE, Kazazian HH Jr (2012) Mov10 rna helicase is a potent inhibitor of retrotransposition in cells. PLoS Genet 8(10):e1002941. doi:10.1371/journal.pgen.1002941

Gray CA, Abbey CA, Beremand PD, Choi Y, Farmer JL, Adelson DL, Thomas TL, Bazer FW, Spencer TE (2006) Identification of endometrial genes regulated by early pregnancy, progesterone, and interferon tau in the ovine uterus. Biol Reprod 74(2):383–394. doi:10.1095/biolreprod.105.046656

Grivna ST, Pyhtila B, Lin H (2006) Miwi associates with translational machinery and piwi-interacting rnas (pirnas) in regulating spermatogenesis. Proc Natl Acad Sci USA 103 (36):13415–13420. doi:10.1073/pnas.0605506103

Hackett JA, Surani MA (2013) Dna methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond B Biol Sci 368(1609):20110328. doi:10.1098/rstb.2011.0328

Haig D (2012) Retroviruses and the placenta. Current biology: CB

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19 (8):1419–1428. doi:10.1101/gr.091678.109

Ichiyanagi K, Nakajima R, Kajikawa M, Okada N (2007) Novel retrotransposon analysis reveals multiple mobility pathways dictated by hosts. Genome Res 17:33

Ivancevic AM, Walsh AM, Kortschak RD, Adelson DL (2013) Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. Bioessays 35:12

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Fuentes-Utrilla P, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas JW, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. Science 344(6188):1168–1173. doi:10.1126/science.1252806

Jjingo D, Huda A, Gundapuneni M, Mariño-Ramrez L, Jordan IK (2011) Effect of the transposable element environment of human genes on gene length and expression. Genome Biol Evol 3:259–271. doi:10.1093/gbe/evr015

Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68

Jurka J, Kapitonov VV, Kohany O, Jurka MV (2007) Repetitive sequences in complex genomes: structure and evolution. Ann Rev Genomics Hum Genet

Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev 23:1303

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 9:e1003470

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at ucsc. Genome Res 12(6):996–1006. doi:10.1101/gr.229102. Article published online before print in May

Kordis D, Gubensek F (1998) Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. Proc Natl Acad Sci USA 95(18):10704–10709

Kordis D, Gubensek F (1999a) Horizontal transfer of non-LTR retrotransposons in vertebrates. Genetica 107:121

Kordis D, Gubensek F (1999b) Molecular evolution of bov-b lines in vertebrates. Gene 238 (1):171–178

Kramerov DA, Vassetzky NS (2005) Short retroposons in eukaryotic genomes. Int Rev Cytol 247:165–221. doi:10.1016/S0074-7696(05)47004-7

Kriegs JO, Matzke A, Churakov G, Kuritzin A, Mayr G, Brosius J, Schmitz J (2007) Waves of genomic hitchhikers shed light on the evolution of gamebirds (aves: Galliformes). BMC Evol Biol 7:190. doi:10.1186/1471-2148-7-190

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T (2008) Dna methylation of retrotransposon genes is regulated by piwi family members mili and miwi2 in murine fetal testes. Genes Dev 22(7):908–917. doi:10.1101/gad.1640708

Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson HM (2003) Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. Mol Biol Evol 20(4):554–562. doi:10.1093/molbev/msg069

Le Rouzic A, Boutin TS, Capy P (2007) Long-term evolution of transposable elements. Proc Natl Acad Sci USA 104(49):19375–19380. doi:10.1073/pnas.0705238104

Lee JY, Ji Z, Tian B (2008) Phylogenetic analysis of mrna polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. Nucleic Acids Res 36(17):5581–5590. doi:10.1093/nar/gkn540

Lenstra JA, van Boxtel JA, Zwaagstra KA, Schwerin M (1993) Short interspersed nuclear element (sine) sequences of the bovidae. Anim Genet 24(1):33–39

Li CCY, Eaton SA, Young PE, Lee M, Shuttleworth R, Humphreys DT, Grau GE, Combes V, Bebawy M, Gong J, Brammah S, Buckland ME, Suter CM (2013) Glioma microvesicles carry selectively packaged coding and non-coding rnas which alter gene expression in recipient cells. RNA Biol 10(8):1333–1344. doi:10.4161/rna.25281

Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. Science (New York, NY) 276:561

Lohe AR, Moriyama EN, Lidholm DA, Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol Biol Evol 12 (1):62–72

Lorenz M, Hewing B, Hui J, Zepp A, Baumann G, Bindereif A, Stangl V, Stangl K (2007) Alternative splicing in intron 13 of the human enos gene: a potential mechanism for regulating enos activity. FASEB J 21(7):1556–1564. doi:10.1096/fj.06-7434com

Malik H, Eickbush T (1998) The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol Biol Evol 15:1123

Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B (2004) The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc Natl Acad Sci USA 101:1731

Martin SL (2006) The orf1 protein encoded by line-1: structure and function during l1 retrotransposition. J Biomed Biotechnol 2006(1):45621. doi:10.1155/JBB/2006/45621

Maruyama K, Hartl DL (1991) Evidence for interspecific transfer of the transposable element mariner between Drosophila and Zaprionus. J Mol Evol 33:514

McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ, Miller WJ (1997) Ltr retrotransposons and the evolution of eukaryotic enhancers. Genetica 100(1–3):3–13

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale dna methylation maps of pluripotent and differentiated cells. Nature 454(7205):766–770. doi:10.1038/nature07107

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr (1996) High frequency retrotransposition in cultured mammalian cells. Cell 87(5):917–927

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV (2002) Dna repair mediated by endonuclease-independent line-1 retrotransposition. Nat Genet 31(2):159–165. doi:10.1038/ng898

Ohshima K, Okada N (2005) Sines and lines: symbionts of eukaryotic genomes with a common tail. Cytogenet Genome Res 110(1–4):475–490. doi:10.1159/000084981

Okada N, Hamada M (1997) The 3' ends of trna-derived sines originated from the 3' ends of lines: a new example from the bovine genome. J Mol Evol 44(1):52–56

Ostertag E, Goodier J, Zhang Y, Kazazian H (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet 73:1444

Pace JK, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci USA 105:17023

Perepelitsa-Belancio V, Deininger P (2003) Rna truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 35(4):363–366. doi:10.1038/ng1269

Piskurek O, Jackson DJ (2012) Transposable elements: from dna parasites to architects of metazoan evolution. Genes (Basel) 3(3):409–422. doi:10.3390/genes3030409

Piskurek O, Okada N (2007) Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc Natl Acad Sci USA 104(29):12046–12051. doi:10.1073/pnas.0700531104

Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AO, Batzer MA, Deininger PL (2000) Potential gene conversion and source genes for recently integrated alu elements. Genome Res 10(10):1485–1495

Ruiz-González I, Xu J, Wang X, Burghardt RC, Dunlap K, Bazer FW (2014) Exosomes, endogenous retroviruses and toll-like receptors: pregnancy recognition in ewes. Reproduction

Ruiz-González I, Minten M, Wang X, Dunlap K, Bazer FW (2015) Involvement of TLR7 and TLR8 in conceptus development and establishment of pregnancy in Sheep. Reproduction

Shimotohno K, Temin HM (1981) Evolution of retroviruses from cellular movable genetic elements. Cold Spring Harb Symp Quant Biol 45(Pt 2):719–730

Skog J, Würdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, Curry WT Jr, Carter BS, Krichevsky AM, Breakefield XO (2008) Glioblastoma microvesicles transport rna and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol 10(12):1470–1476. doi:10.1038/ncb1800

Sormacheva I, Smyshlyaev G, Mayorov V, Blinov A, Novikov A, Novikova O (2012) Vertical evolution and horizontal transfer of cr1 non-ltr retrotransposons and tc1/mariner dna transposons in lepidoptera species. Mol Biol Evol 29(12):3685–3702. doi:10.1093/molbev/mss181

Speek M (2001) Antisense promoter of human l1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 21(6):1973–1985. doi:10.1128/MCB.21.6.1973-1985.2001

Spencer TE, Bazer FW (1995) Temporal and spatial alterations in uterine estrogen receptor and progesterone receptor gene expression during the estrous cycle and early pregnancy in the ewe. Biol Reprod 53(6):1527–1543

Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18(2):74–82

Startek M, Szafranski P, Gambin T, Campbell IM, Hixson P, Shaw CA, Stankiewicz P, Gambin A (2015) Genome-wide analyses of line-line-mediated nonallelic homologous recombination. Nucleic Acids Res. doi:10.1093/nar/gku1394

Swergold GD (1990) Identification, characterization, and cell specificity of a human line-1 promoter. Mol Cell Biol 10(12):6718–6729

Tanaka T, Hosokawa M, Vagin VV, Reuter M, Hayashi E, Mochizuki AL, Kitamura K, Yamanaka H, Kondoh G, Okawa K, Kuramochi-Miyagawa S, Nakano T, Sachidanandam R, Hannon GJ, Pillai RS, Nakatsuji N, Chuma S (2011) Tudor domain containing 7 (tdrd7) is essential for dynamic ribonucleoprotein (rnp) remodeling of chromatoid bodies during spermatogenesis. Proc Natl Acad Sci USA 108(26):10579–10584. doi:10.1073/pnas.1015447108

Tarlinton RE, Meers J, Young PR (2006) Retroviral invasion of the koala genome. Nature 442 (7098):79–81. doi:10.1038/nature04841

Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO (2007) Exosome-mediated transfer of mrnas and micrornas is a novel mechanism of genetic exchange between cells. Nat Cell Biol 9(6):654–659. doi:10.1038/ncb1596

Villarroya-Beltri C, Gutiérrez-Vázquez C, Sánchez-Cabo F, Pérez-Hernández D, Vázquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn M, Sánchez-Madrid F (2013) Sumoylated hnrnpa2b1 controls the sorting of mirnas into exosomes through binding to specific motifs. Nat Commun 4:2980. doi:10.1038/ncomms3980

Vlachogiannis G, Niederhuth CE, Tuna S, Stathopoulou A, Viiri K, de Rooij DG, Jenner RG, Schmitz RJ, Ooi SKT (2015) The dnmt3 l add domain controls cytosine methylation establishment during spermatogenesis. Cell Rep. doi:10.1016/j.celrep.2015.01.021

Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL (2013) Widespread horizontal transfer of retrotransposons. Proc Natl Acad Sci USA 110:1012

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H (2008) Endogenous sirnas from naturally formed dsrnas regulate transcripts in mouse oocytes. Nature 453(7194):539–543. doi:10.1038/nature06908

Watanabe T, Cheng EC, Zhong M, Lin H (2014) Retrotransposons and pseudogenes regulate mrnas and lncrnas via the pirna pathway in the germline. Genome Res. doi:10.1101/gr.180802.114

Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, Harrison DK, Aung H, Phutikanit N, Lyle R, Meachem SJ, Antonarakis SE, de Kretser DM, Hedger MP, Peterson P, Carroll BJ, Scott HS (2005) Meiotic and epigenetic defects in dnmt3l-knockout mouse spermatogenesis. Proc Natl Acad Sci USA 102(11):4068–4073. doi:10.1073/pnas.0500702102

Yamamoto Y, Watanabe T, Hoki Y, Shirane K, Li Y, Ichiiyanagi K, Kuramochi-Miyagawa S, Toyoda A, Fujiyama A, Oginuma M, Suzuki H, Sado T, Nakano T, Sasaki H (2013) Targeted gene silencing in mouse germ cells by insertion of a homologous dna into a pirna generating locus. Genome Res 23(2):292–299. doi:10.1101/gr.137224.112

Yuan A, Farber EL, Rapoport AL, Tejada D, Deniskin R, Akhmedov NB, Farber DB (2009) Transfer of micrornas by embryonic stem cell microvesicles. PLoS ONE 4(3):e4722. doi:10.1371/journal.pone.0004722

Zhou Y, Mishra B (2005) Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. Proc Natl Acad Sci USA 102(11):4051–4056. doi:10.1073/pnas.0407957102

Chapter 5
Horizontal Transfers and the New Model of TE-Driven Genome Evolution in Eukaryotes

Moaine El Baidouri and Olivier Panaud

Abstract In this chapter, we present a new model of TE-driven genome evolution in eukaryotes. This model is based on the recent discovery of the propensity of transposable elements to be transferred horizontally among plant and animal species. We propose that the horizontal transfer of transposable elements (HTTs) is a key mechanism of long-term survival of TEs in eukaryotic genomes, by allowing TEs to escape from the silencing machinery of their host genome. We provide a description of the most recent discoveries of HTTs among plants and animals, an up-to-date description of the TE silencing pathways in eukaryotes, and some characteristics of TE biology in terms of functional impact and of response to environmental stress.

5.1 Introduction

The availability of the genome sequence of over 2000 eukaryotic species, from plant, animal and fungi (www.gold.jgi-psf.org), has opened new paths towards understanding genome evolution with the endeavour of closing the gap between gene evolution at molecular level and biological evolution at large. This, however, necessitates to clearly distinguish genes from genomes at both functional and evolutionary levels, as tentatively illustrated in the present chapter. Genomes were first considered as the cellular entities containing the genes and *nothing else* apart from some DNA sequences with structural features, such as centromeres and telomeres. This view was challenged by the discovery of the extensive variation of

M. El Baidouri
Center for Applied Genetic Technologies, University of Georgia,
111 Riverbend Dr, Athens, GA 30602, USA

O. Panaud (✉)
Laboratoire Genome et Développement des Plantes UMR 5096 UPVD/CNRS,
Université de Perpignan via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan, France
e-mail: panaud@univ-perp.fr
URL: http://www.panaudlab.org

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_5

genomes size among plant and animals species with no correlation with biological complexity, which was stated as the *C-value paradox* by Thomas in the early 1970s (Thomas 1971). This paradox has now been resolved with the discovery that transposable elements (TEs, McClintock 1953) are the main components of eukaryotic genomes (Feschotte and Pritham 2007), thus outnumbering genes in many species (Hattori et al. 1986; SanMiguel et al. 1996). However, resolving this paradox leads to the new problematics of the origin and evolution of TEs, as well as (and far more importantly) of their putative biological impact on eukaryotic species, which was phrased as the *C*-value enigma (Gregory 2001).

The new findings on TE-driven genome evolution that we present here will be discussed in the frame of the increase/decrease model that was published earlier (Vitte and Panaud 2005). We may therefore briefly recall some of the definitions and concepts that we have adopted, helping the reader to critically review the models that we propose below: we consider a *GENE* as a DNA sequence containing some information necessary for the development and reproduction of a living organism. This broad definition concerns any DNA sequence with a function and thus under putative selection, including protein coding genes, regulatory sequences such as promoters and enhancers, miRNA precursor genes, etc. Any other DNA sequence with no function and therefore evolving neutrally is then defined as a *NON-GENE*. Genes and non-genes are conceptually mutually exclusive, and therefore, any genome could be divided into the genic and non-genic compartments, which are distinct and physically non-overlapping. Several remarks follow from this representation of eukaryotic genome:

5.1.1 Content and Physical Organization

Non-genic compartment is mostly composed of TEs in all eukaryotes (Kazazian 2004). These may be easily recognized as such in case of recent transpositional activity, while more ancient, extinct TE families are usually found as deleted and degenerated copies. Both genic and non-genic compartments are physically intermingled, as genes and therefore non-genes (composing the intergenic space) are scattered throughout the chromosomes (Choulet et al. 2014; International Human Genome Sequencing Consortium 2001). However, pericentromeric regions, particularly in plants, are more TE rich (and consequently gene-poor) than intersticial regions (Devos 2010).

5.1.2 Permeability Between the Two Compartments

An important feature of the evolutionary dynamics of both genic and non-genic compartments is their permeability: gene duplication (either through retroprocessing, segmental duplication or polyploidization) often leads to pseudogenization, as an

alternative to neo- and sub-functionalization (Edger and Pires 2009). Pseudogenes are non-functional, degenerated form of the genes and are therefore part of the non-genic space, while their non-duplicated ancestor once belonged to the gene space. Moreover, TEs have been found to play a functional role, both as mutagenic agents or as regulatory sequences (Rebollo et al. 2012). The latter case referred to as TE domestication (Feschotte 2008) illustrates that a sequence originating from non-genic space may then belong to the gene space.

5.1.3 Evolutionary Constraints of Non-Genes

One consequence of the definition of the non-gene given above is that TEs should evolve neutrally, which is debatable if one considers that some TEs, like the retrotransposon, exhibit a higher sequence conservation in their functional domains than in the other regions of the element. This indeed suggests that these domains may be under some selective constraint. One would expect that a particular TE, in order to be active, should harbour all the functional domains that are necessary for its transposition cycle (Wicker et al. 2007). In this regard, recently active TEs should be the ones with intact domains, which could explain the difference in divergence rates observed within the sequence of a given element. However, once inserted, a newly transposed element will accumulate mutations in a neutral manner (as long as it belongs to the non-genic space), as predicted by the model. This has been verified experimentally (SanMiguel et al. 1998) and therefore shows that there is no ambiguity in the model.

5.2 The Increase/Decrease Model

5.2.1 The Model

The *increase/decrease* model (Vitte and Panaud 2005) posits that genome size in eukaryotes results from the action of two counteracting forces: retrotransposition (that adds DNA to the genome) and deletion. Polyploidization, an obvious mechanism of genome size increase, is not considered in this model (mainly focussing on the evolution of diploid genomes). Retrotransposons are Class I elements that transpose via a RNA intermediate through a copy and paste mechanism (Wicker et al. 2007). They are distinct from transposons, the Class II elements, that transpose through a "cut and paste" mechanism, i.e. excised and directly reintegrated elsewhere in the genome. Retrotransposons multiply their copy number while active. Genome-wide retrotransposition may reach such an extent that it leads to significant variation in genome size in short evolutionary time span (SanMiguel

et al. 1998). However, TE-related sequences are eliminated from their host genome either through deletions or recombinations, both being considered as an efficient evolutionary force counteracting retrotransposition (Petrov et al. 2000).

5.2.2 *Supporting Evidences for Retrotransposition-Based Genome Size Increase*

Over the past few years, many studies lead to conclude that retrotransposition is indeed the main factor of genome size increase: the prevalence of LTR-retrotransposons in plant genomes, first evidenced by the pioneering work on maize by SanMiguel et al. (1996) has been confirmed by subsequent genome sequencing projects, except for that of the unusual small genome species *Arabidopsis thaliana*. In cereals, TE annotation of the rice, sorghum, *Brachypodium* and maize genomes clearly demonstrated that there is a positive correlation between TE content and genome size (El Baidouri and Panaud 2013). The largest plant genome sequenced and assembled so far is that of the 20 Gbp gymnosperm *Picea abies* (Nystedt et al. 2013). Similarly to what has been observed in angiosperms, this gymnosperm genome is almost exclusively composed of LTR-retrotransposon-related sequences. The next question is wheather genomes increase in size through a slow accumulation of TE-related sequences or rather through discontinuous, catastrophic bursts of retrotransposition. This question was answered through comparative genomic studies in both plant and animals. Piegu et al. (2006) showed that the transpositional activity of three LTR-retrotransposons families has doubled the genome size of the wild rice species *O. australiensis* in a short time, posterior to speciation. A similar observation was made in another wild rice species *O. granulata* (Ammiraju et al. 2007), where other TE families contributed to a significant genome size increase over the last two million years. In cotton, Hawkins et al. (2006) showed that one LTR-retrotransposon underwent a burst, specifically in K-genome, reaching more than 80,000 copies in some species of the lineage. In animals, partial genomic data of several salamander species evidenced the role played by LTR-retrotrans posons in genomic gigantism observed in this taxonomic group (Sun et al. 2011). Finally, a comparative study of the TE dynamics among 8 angiosperms species showed that, regardless the size, all genomes exhibit traces of recent transpositional activity, i.e. within the last 5 million years (El Baidouri et al. 2013). Moreover, all species also exhibit an L-shaped distribution of the copy number of TE families, thus showing that only few families are highly repeated, while most families are single- or low-copy. This suggests that genome size change occurs rapidly through the activity of only few TE families in angiosperms. It should be mentioned that these results were not confirmed for the species *P. abies*, where it was shown that many TE-related sequences are of more ancient origin (tens of My; Sun et al. 2011).

5.2.3 Supporting Evidences for TE Elimination

In-depth analysis of angiosperms genomes shows that, like in the case of the gymnosperm genome of *P. abies*, they often harbour traces of more ancient transposition bursts, but these are in the form of short, deleted and degenerated copies of TE families. This suggests that, in this group, TE-related sequences are eliminated through deletions. There are two main mechanisms of LTR-retrotransposon elimination: small deletions (Ma and Bennetzen 2004) and illegitimate recombinations between LTRs, leading to solo-LTR sequences and the elimination of circular molecules harbouring one LTR and the internal region of the element (Shirasu et al. 2000; Vitte and Panaud 2003). Several studies tentatively estimated the rate at which TEs are eliminated from angiosperm genomes following amplification bursts. Estimations of LTR-retrotransposon << half-life >> range from 2 My (Vitte et al. 2007) to 6 My (Bennetzen et al. 2005), based on rice and Arabidopsis (for the latter) genome analyses. In animals, early studies in *Drosophila* showed a strong bias of mutations towards deletions in TE-related sequences (Petrov and Hartl 1998). In this paper, the TE half-life estimation in the genus was estimated at 12 My, which is significantly longer than the first estimates in plants.

Several authors have proposed that difference in genome size could be explained in some extent by a difference in TE-removal rate, i.e. that small genomes should exhibit a higher removal rate than large ones. This was indeed shown when compared elimination rate between Laupala cricket and *Drosophila* (Petrov et al. 2000) and mountain grasshopper and *Drosophila* (Bensasson et al. 2001). Similar observations were made in salamanders (Frahry et al. 2015) and Norway spruce, *P. abies* (Nystedt et al. 2013), two species with very large genomes. In their genome-wide comparative study of eight plant genomes, El Baidouri and Panaud (2013) did not find the correlation between genome size and TE-removal rate being always validated in pairwise comparisons: for instance, removal rate in sorghum genome (700 Mbp) is lower than in soybean (975 Mbp). Whether the difference in TE-removal rate could systematically explain difference in genome size only in the case of very large genomes remains to be demonstrated.

Except for the very large genomes, both transposition bursts and TE elimination seem to occur at a very fast rate, resulting in a quick turnover of intergenic space. Comparative genomic studies indeed show that while gene order along the chromosomes may be conserved at some extent over long evolutionary periods (tens of millions of years), there is a striking lack of homology between TE-related sequences in the intergenic space, even among closely related species within a genus (Sanyal et al. 2010).

5.3 Epigenetic Control of Transposition and Phylogenetic Discrepancies

Over the past ten years, significant advances have been made in our understanding of how transposition is suppressed through epigenetic mechanisms. Transposition in plants and animals is notably limited through TE silencing ensured by epigenetic mechanisms (Lisch 2009; Slotkin and Martienssen 2007; Holoch and Moazed 2015). These mechanisms include the deposition of epigenetic marks such as DNA methylation and heterochromatic histone post-translational modifications. DNA methylation is maintained during cell replication through the activity of specific DNA methyltransferases and chromatin remodellers (MET1, CMT3 and DDM1) and is reinforced specifically at TE loci through the action of the RNA-directed DNA methylation (RdDM) mechanism in plants (Law and Jacobsen 2010). In the RdDM pathway, the plant-specific RNA polymerase IV is responsible for the production of heterochromatic small RNAs (24 nucleotides in length) that direct the de novo methylation machinery to TE sequences in order to silence them. In case a particular family would escape the transcriptional silencing, the post-transcriptional gene silencing (PTGS) pathway is then activated. PTGS consists first in the production of short RNAs of 21 nucleotides long that target and form double-stranded RNA with the mRNA of the TE being silenced. This double-stranded RNA is recognized by a RISC complex that leads to the destruction of the mRNA or to inhibition of its translation. In addition microRNAs could play a role in TE recognition (Creasey et al. 2014). In animals, a specific class of small RNAs called piwi RNAs is involved in both in PTGS and TGS and silences TE in germ line and in the soma (Ross et al. 2014; Luteijn and Ketting 2013). Piwi RNAs can be amplified to form secondary piwi RNAs that reinforce the silencing of TEs.

In such context, one would expect that most TEs in a given genome are efficiently kept inactive. This strict control combined with the strong bias of mutations of TE-related sequences towards deletions should lead to the elimination of TEs from most eukaryotic genomes, which is exactly the opposite of what is observed: among the hundreds of eukaryotic genomes that have been sequenced so far, only two (i.e. that of the protozoan *Plasmodium falciparum*, Durand et al. 2006 and of the microplankton *Micromonas pusilii*, Worden et al. 2009) appear to be devoid of any trace of recent transpositional activity. This questions the maintenance of endogenous TE families in a given genome over long evolutionary periods. Furthermore, the phylogenetic relationships found among LTR-retrotransposons in eukaryotes (Fig. 5.1a, Llorens et al. 2010) often exhibit incongruences (Figs. 5.1b, c). If strict vertical transmission were to be the rule for TEs, then such tree topology could only be obtained if the eukaryotic ancestral genome harboured all the various TE families found today in plant and animal genomes (Fig. 5.1d). This implies that a TE family found today in distant evolutionary lineages survived and was vertically transmitted from their ancestral genome over up to hundreds of million of years despite silencing and deletion! In addition, the sequence conservation

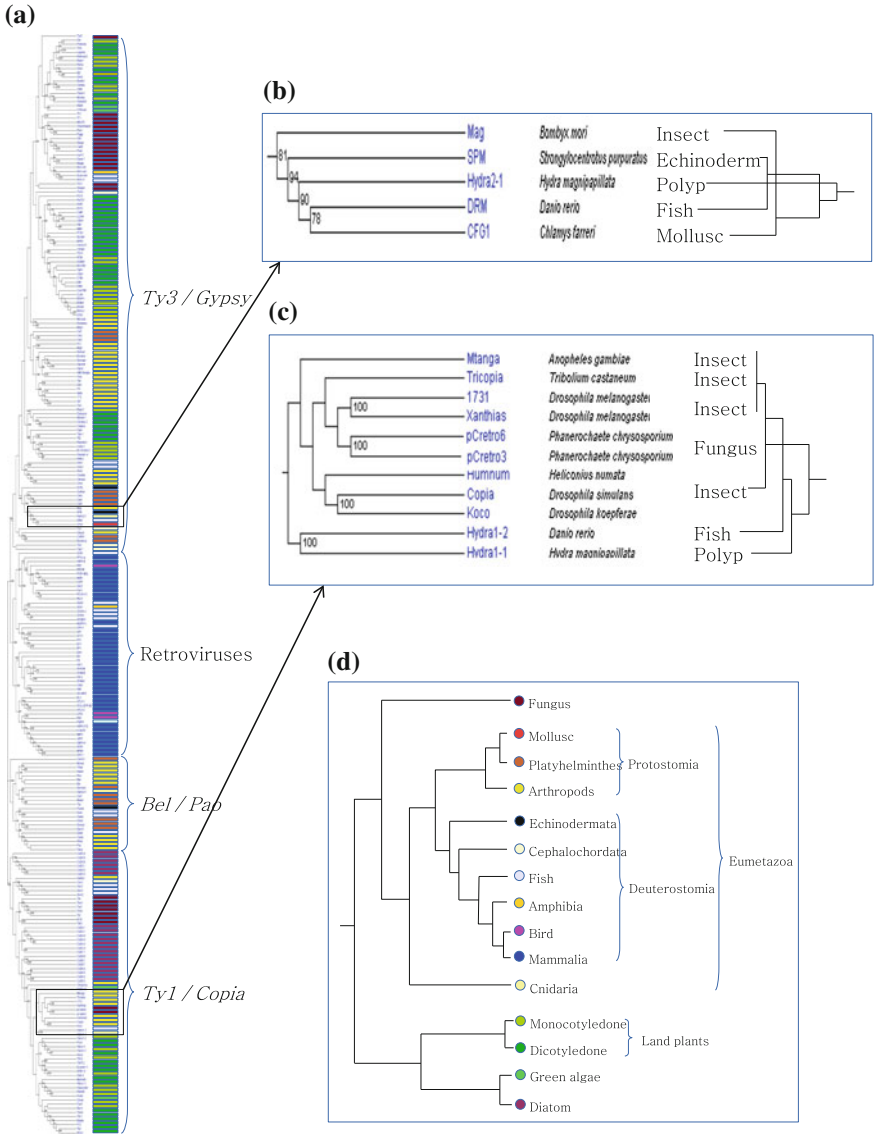


Fig. 5.1 Phylogenetic evidence of HTTs in eukaryotes **a** phylogenetic tree of eukaryotic LTR-retrotransposons and retroviruses. The tree was extrapolated from the Gy-db database (www.gydb.org; Llorens et al. 2011). Each leaf of the tree represents one family, whereas the *coloured boxes* represent the taxonomic group. Correspondence between colours and groups is given in Fig. 5.1d. **b** and **c** Examples of HTTs evidenced through phylogenetic incongruence. Close-up of a clade of Fig. 5.1a phylogenetic tree showing phylogenetic incongruence between a TE family and the species phylogeny for Ty3/gypsy and Ty1/copia LTR-retrotransposons, respectively. **d** Phylogenetic relationships among all the species where a LTR-retrotransposon has been found and used for building the phylogenetic tree in Fig. 5.1a

observed among related TE families in distant lineages could only be the result of very strong selective constraints. Even in the case of most essential house-keeping genes is not sequence conservation observed over such long evolutionary periods.

5.4 Horizontal Transfers of Transposable Elements and the New Birth and Death Model

Horizontal transfers (HTs) are defined as the transmission of genetic material among species that are sexually isolated. Horizontal TE transfers (HTTs) have been proposed first by Hartl et al. (1997) as a possible dissemination mechanism of TEs in eukaryotes. This model was then revived and further elaborated by Schaack et al. (2010) in the light of the latest discoveries on epigenetic-mediated TE silencing. This model, based on the birth and death process, posits that HTTs could be the mechanism ensuring the long-term survival of TEs among eukaryotic lineages. TEs could circumvent the silencing machinery of their host genome by being horizontally transferred to a new « naive » genome where it could transpose (the birth) before being in turn silenced and eliminated (the death), unless it is again transferred. If this model is right, then most of the active TE families found in plant and animal genomes should originate from other species through HTT.

Until recently, HTs had been mostly described in prokaryotes where they are considered to play a major role in adaptation and speciation. HTs occur so frequently among bacteria that they are put forward as the main cause of the reticulation of phylogenetic trees in this phylum. In eukaryotes, documented cases of horizontal gene transfers are much more scarce (Mower et al. 2004). However, because TEs (whether they belong to Class I or Class II) exist as a free molecule in the cell at least at one point in their transposition cycle, one could envisage that they may be more prone to HTs than any other genomic sequence. In addition, some elements contain envelope-like coding domains, which make them structurally similar to retroviruses. The first case of horizontal TE transfer (HTT) was discovered decades ago between and *Drosophila willistoni* and *Drosophila melanogaster* (Daniel et al. 1990). For many years, the scarcity of more documented cases of HTT was probably attributable to the difficulties raised by their detection in large genomes (see below for the description of the detection methods). With the availability of the genome sequence of many eukaryotic species, several cases of HTTs have been reported recently, mainly using comparative genomics approaches (Wallau et al. 2012).

5.5 Detection of HTTs from Full Genome Sequences

The detection of a HT relies upon three criteria (Wallau et al. 2012): High sequence identity between TEs found in the genomes of distantly related species (the HS criterion, for high similarity); the patchy distribution of TEs in phylogenies (PD) and the phylogenetic incongruence between the TEs and their host (PI). When the full genome sequence is available, the HS criterion has a good detection power because one could compare the sequence identity of the horizontally transferred TE with that of the complete gene repertoire of the two distantly related species (El Baidouri et al. 2014). In comparison, the other two criteria may not be as efficient, because the species for which full genome sequence is available obviously exhibit a patchy distribution in phylogenetic trees (until all taxa have been sequenced), which also impedes the use of the PI criterion since only incomplete phylogenetic trees may be drawn from these species. If one combines full comparative genomic approaches with wet laboratory experiments, such as PCR amplification and sequencing of TEs in representative sample of species, then both PI and PD criteria may efficiently complement the use of the HS criterion in the detection process.

Hundreds of cases of HTTs have been evidenced over the past years, thus providing support for the new birth and death model of TE-driven genome evolution. Below is a non-exhaustive description of the newest reports from the past year that illustrates the contribution of NGS (Next Generation Sequencing) technologies to the field of evolutionary genomics: Ortiz et al. (2015) looked for HTTs among several *Drosophila* species, together with some parasitic wasp species using a metagenomic approach based on full genome sequencing of several individuals from each species. They identified five HTT between *Drosophila* species that do not share the same ecological habitat. Zhang et al. (2014) found multiple HTTs of the *Chapaev* transposon among mammals, reptiles, jawed fishes, lampreys, insects as well as an insect bracovirus. Such wide spectrum of transfers suggests that some TE family may be more HTT prone than others. Interestingly, the presence of the transposon in an insect bracovirus also suggests that viruses may be the vectors of HTTs (see below on the mechanisms of HTTs). Yamada et al. (2014) evidenced the occurrence of HTTs of a *Mariner*-like element between a bee, a wasp and a butterfly in the south-west islands of Japan. Dupeyron et al. (2014) identified HTTs between crustacean (terrestrial isopods) and insects. They used a combination of in silico and wet laboratory approaches to evidence multiple transfers of two Mariner transposon families *Crmar2* and *Mariner-5-Dbi*. Fewer reports concern HTTs in plants. El Baidouri et al. (2014) performed a genome-wide search of all possible HTTs of LTR-retrotransposons among 46 sequenced plant genomes and found 32 clear cases of recent transfers, which leads to an estimation of millions of HTTs among all angiosperms in a recent evolutionary past. The authors also evidenced that TEs have been active in their new host genome after their transfer, which is in favour of the new birth and death model.

All these recent discoveries show that HTTs are far more frequent among eukaryotes than previously thought. Even if they do not provide a direct evidence for the necessity of this mechanism for the long-term survival of TEs in plants and animals, they shall change our views on the impact of TEs on genome evolution and justify to include HTTs in a new model.

5.6 Mechanisms of HTTs

The mechanisms through which TEs could be horizontally transferred among species remain not fully understood, even if several reports suggest that host–parasite interactions may favour HTTs in animals: Using a comparative genomics strategy, Gilbert et al. (2010) showed that the parasite triatomine bug *Rhodnius prolixus* harbours in its genome four TE families sharing a high sequence identity with several of its hosts (opossum and squirrel monkey), while it also shares another TE family with the pond snail *Lymnaea stagnalis* known to be the vector of trematodes parasites of several mammalian species. Ivancevic et al. (2013) showed that the LINE *BovB* has been horizontally transferred to several taxonomic groups as diverse as marsupials, ruminents, squamates or monotremes, through an arthropod vectors (ticks). Gilbert et al. (2014) recently showed that two transposons (*Mar1* and *IFP2*) initially found in the genome of the cabbage looper *Trichoplusia ni*, a moth feeding on a wide spectrum of plant species, could transpose in the genome of the baculovirus *AcMNPV* upon its infection of the insect. The low frequency of transposition events together with the polymorphisms of insertions sites in the 240-kbp viral genome strongly suggests that the transposition of these two families occurred very recently, which leads to hypothesize that the HTTs are concommittent with viral infection. Moreover, as mentioned above, Zhang et al. (2014) found the presence of one *Chapaev* transposon in a bracovirus, while several members of this super-family are also found in insects, fish and reptiles, thus evidencing one of the most widespread horizontally transferred TE. Because of their obligatory relationships with parasitic wasps (where they replicate in the ovaries) and their transmission to lepidopteran larvea, bracoviruses are one of the most obvious vectors of HTT between insects. However, no experimental evidence of this process has been reported yet. In plants, no evidence of host/parasite-driven HTTs has been provided yet. The only case of HT in a host/parasite system concerns the *atp1* mitochondrial gene between the parasitic genus *Cuscuta* and its host *Plantago* (Mower et al. 2004). With the accumulation of new genome sequences of both eukaryotes and prokaryotes in a near future, one should be able to further test the hypothesis of the involvement of parasitism in HTTs.

5.7 Functional Impact of TEs

Because of their propensity to invade and thus densely populate the genomes of most eukaryotes, the question of the functional impact of TEs arose as a key issue in the field of genomics and more broadly in biology at large. As mentioned above, TEs were first considered as mutagenic. As a consequence, only neutral TE insertions (i.e. not disrupting the function of the genes) were expected to be found in genomes, while those causing deleterious mutations should be quickly eliminated from populations. However, several lines of evidence accumulated over the past decades have changed this paradigm (Bennetzen 2005). The recruitment of TE-related sequences by their host genome, referred to as "TE domestication", has been shown in several lineages (Lin et al. 2007; Casola et al. 2007) and concerns some major biological features of some organisms, like VdJ recombination, one of the key molecular mechanisms of the vertebrate immune system (Kapitonov and Jurka 2005), or placental pregnancy in mammals (Lynch et al. 2015). As mentioned above, TEs are under the strict control of several silencing pathways, some acting on the methylation of TE-related sequences in the genome. The change of methylation status of a genomic region could in some cases affect the expression of the genes located nearby the targeted TE (Weil and Martienssen 2008). This was for instance shown as being the molecular basis of fruit colour in grape (Kobayashi et al. 2004). In this regard, TE can be considered as epigenetic mediators that could act indirectly on gene expression (Weil and Martienssen 2008). Finally, TEs can be involved in gene movements: genes can be transducted by TEs, as it has been shown in rice with *Mu* like elements (i.e. The packMULEs, Jiang et al. 2004). This mechanism is at the origin of some of the largest gene families in the species. Similarly, Morgante et al. (2005) showed that helitrons, a particular type of Class II elements, were associated with gene translocation in maize. Moreover, genes can be translocated through ectopic recombination mechanisms between paralogous repeated TEs, as was recently shown by Wicker et al. (2010) based on a comparative survey of the *Brachypodium*, rice and maize full genome sequences.

5.8 Stress and Transposition

As mentioned above, one of the characteristics of TEs, regardless their type or their host genome, is their propensity to be transcriptionally activated by environmental stress, for instance (and not exhaustively) wounding (*Corky* from *Quercus,* Rocheta et al. 2012; *CLCoy1* in lemon, Felice et al. 2008; *OARE1* in oat, Kimura et al. 2001; *Tnt1* in tobacco, Mhiri et al. 1997), salt (*AtCopeg1* in *Arabidopsis*, Duan et al. 2008), drought (*BARE1* in barley, Kalendar et al. 2000; *CCR* in rice, Neumann et al. 2007) heat (*ONSEN* in *Arabidopsis*, Ito et al. 2011) or cold (*mPing* in rice, Jiang et al. 2003). Some of these studies also demonstrated that this activation is caused by the presence of transcription factor binding sites in the promoter region of the

element (Mhiri et al. 1997). In addition, Ito et al. (2011) showed that this stress-triggered transcriptional activation leads to the transpositional activity of the *ONSEN* element in *Arabidopsis*. These last observations, if generalized to the plant kingdom could open new perspectives for the genome-wide functional impact of TEs: this process, particularly in the case of Class I element, could lead to the spreading of stress-response promoter regions in the genome that could in turn be domesticated by the host genome and form a network of stress-response genes.

5.9 Conclusion

The new model that we present in this chapter is schematized in Fig. 5.2. It synthesizes most recent data on TE dynamics and genome evolution. HTTs play a central role in this model, as being the main process ensuring long-term survival of TEs in eukaryotic lineages. The functional impact of TEs is also addressed, as well as stress-induced transposition. Even if several aspects of the model remain speculative, it provides a paradigm those hypotheses could be experimentally tested through systematic genomic approaches thanks to the development of the newest sequencing technologies. In this regard, next investigations may focus on the actual mechanisms of transfers, either to establish parasitism as the main cause of HTTs or to find alternative modes of transfers. On a longer term, one of the main endeavours of evolutionary genomicists involved in TE biology is to demonstrate the involvement of TEs in biological innovations as a general trend in eukaryotic evolution.

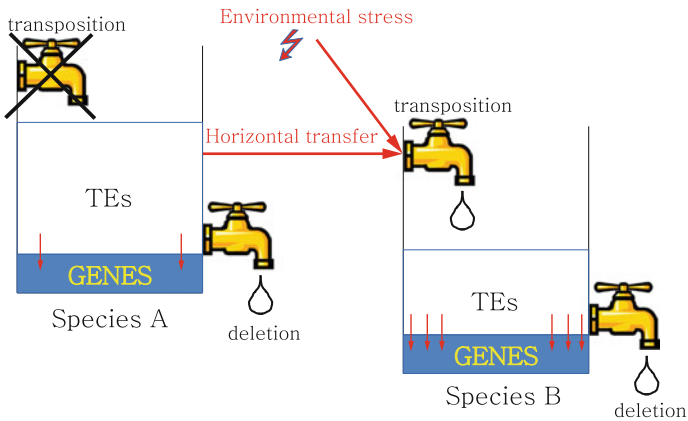


Fig. 5.2 A schematic view of the birth and death model

References

Ammiraju JS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar DS, SanMiguel PJ, Jiang N, Jackson SA, Panaud O, Wing RA (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J 52(2):342–351

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95(1):127–132

Bennetzen JL (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev Dec 15(6):621–627. Review

Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol Biol Evol 18(2):246-53

Casola C, Lawing AM, Betrán E, Feschotte C (2007) PIF-like transposons are common in *Drosophila* and have been repeatedly domesticated to generate new host genes. Mol Biol Evol 24(8):1872–1888

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345(6194):1249721

Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature 508(7496):411–415

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A (1990) Evidence for horizontal transmission of the P transposable element between *Drosophila* species. Genetics 124(2):339–355

Devos KM (2010) Grass genome organization and evolution. Curr Opin Plant Biol 13(2):139–145

Duan K, Ding X, Zhang Q, Zhu H, Pan A, Huang J (2008) AtCopeg1, the unique gene originated from AtCopia95 retrotransposon family, is sensitive to external hormones and abiotic stresses. Plant Cell Rep 27(6):1065–1073

Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C (2014) Horizontal transfer of transposons between and within crustaceans and insects. Mob DNA 5(1):4

Durand PM1, Oelofse AJ, Coetzer TL (2006). An analysis of mobile genetic elements in three Plasmodium species and their potential impact on the nucleotide composition of the P. falciparum genome. BMC Genomics 4 7:282

Edger PP, Pires JC (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 17(5):699–717. Review

El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5(5):954–965

El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O (2014) Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24(5):831–838

Felice B, Wilson RR, Argenziano C, Kafantaris I, Conicella C (2008) A transcriptionally active copia-like retroelement in Citrus limon. Cell Mol Biol Lett 14:289–304

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368. Review

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9(5):397–405

Frahry MB, Sun C, Chong RA, Mueller RL (2015) Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. J Mol Evol 80(2):120–129

Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ, Feschotte C (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464(7293):1347–1350

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou EA, Cordaux R (2014) Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun 5:3348

Gregory TR (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc 76(1):65–101. Review

Hartl DL et al (1997) Modern thoughts on an ancient marinere: function, evolution, regulation. Annu Rev Genet 31:337–358

Hattori M, Kuhara S, Takenaka O, Sakaki Y (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature 321(6070):625–628

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res 16(10):1252–1261

Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet 16(2):71–84

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472 (7341):115–119

Ivancevic AM, Walsh AM, Kortschak RD, Adelson DL (2013) Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution. BioEssays 35(12):1071–1082

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421(6919):163–167

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431(7008):569–573

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci U S A 97(12):6603–6607

Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3(6):e181

Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303(5664):1626–1632

Kimura Y, Tosa Y, Shimada S, Sogo R, Kusaba M, Sunaga T, Betsuyaku S, Eto Y, Nakayashiki H, Mayama S (2001) OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. Plant Cell Physiol 42(12):1345–1354

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304(5673):982

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220

Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H (2007) Transposase-derived transcription factors regulate light signaling in Arabidopsis. Science 318(5854):1302–1305

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 2009(60):43–66

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A (2011). The gypsy database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 2011 Jan 39(Database issue):D70-4

Luteijn MJ, Ketting RF (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet 14(8):523-34. doi:10.1038/nrg3495. Epub 2013 Jun 25. Review

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ, Grützner F, Bauersachs S, Graf A, Young SL, Lieb JD, DeMayo FJ, Feschotte C, Wagner GP

(2015) Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep 10(4):551–561

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A 101(34):12404–12410 Epub 2004 Jul 7

McClintock B (1953) Induction of Instability at Selected Loci in Maize. Genetics 38(6):579–599

Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol 33(2):257–266

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet. Sept 37(9):997-1002

Mower JP, Stefanović S, Young GJ, Palmer JD (2004) Plant genetics: gene transfer from parasitic to host plants. Nature 432(7014):165–166

Neumann P, Yan H, Jiang J (2007) The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics 176(2):749–761

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson A, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497(7451):579–584

Ortiz MF, Wallau GL, Graichen DÂ, Loreto EL (2015) An evaluation of the ecological relationship between *Drosophila* species and their parasitoid wasps as an opportunity for horizontal transposon transfer. Mol Genet Genomics 290(1):67–78

Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol 15(3):293–302

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. Science 287(5455):1060–1062

Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Res 16(10):1262–1269

Rebollo R, Romanish MT, Mager DL (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 46:21–42. Review

Rocheta M, Carvalho L, Viegas W, Morais-Cecilio M (2012) Corky, a gypsy-like retrotransposon is differentially transcribed in Quercus suber tissues. BMC Res Notes 5:432

Ross RJ, Weiner MM, Lin H (2014) PIWI proteins and PIWI-interacting RNAs in the soma. Nature 505(7483):353–359

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274(5288):765–768

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20(1):43–45

Sanyal A, Ammiraju JS, Lu F, Yu Y, Rambo T, Currie J, Kollura K, Kim HR, Chen J, Ma J, San Miguel P, Mingsheng C, Wing RA, Jackson SA (2010) Orthologous comparisons of the Hd1 region across genera reveal Hd1 gene lability within diploid Oryza species and disruptions to microsynteny in Sorghum. Mol Biol Evol 27(11):2487–2506

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25(9):537–546

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Slotkin RK, Martienssen R (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8(4):272–285. Review

Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL (2011) LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol 4(2):168–183

Thomas CA Jr (1971) The genetic organization of chromosomes. Annu Rev Genet 5:237–256. Review

Vitte C, Panaud O (2003) Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. Mol Biol Evol 20(4):528–540

Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110(1–4):91–107. Review

Vitte C, Panaud O, Quesneville H (2007) LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss. BMC Genom 8:218

Wallau GL, Ortiz MF, Loreto EL (2012) Horizontal transposon transfer in eukarya: detection, bias, and perspectives. Genome Biol Evol 4(8):689–699

Weil C, Martienssen R (2008). Epigenetic interactions between transposons and genes: lessons from plants. Curr Opin Genet Dev 18(2):188–192. Review

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982. Review

Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res 20(9):1229–1237

Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, Von Dassow P, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist EA, Lucas S, Mayer KF, Moreau H, Not F, Otillar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov A, Robbens S, Schmutz J, Toulza E, Wyss T, Zelensky A, Zhou K, Armbrust EV, Bhattacharya D, Goodenough UW, Van de Peer Y, Grigoriev IV (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science 324(5924):268–272

Yamada K, Kawanishi Y, Yamada A, Tokuda G, Gurung RD, Sasaki T, Nakajima Y, Maekawa H (2014) A novel cluster of mariner-like elements belonging to mellifera subfamily from spiders and insects: implications of recent horizontal transfer on the South-West Islands of Japan. Genetica 142(2):149–160

Zhang HH, Feschotte C, Han MJ, Zhang Z (2014) Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. Genome Biol Evol 6(6):1375–1386

Chapter 6
The Impact of Transposable Elements in the Evolution of Plant Genomes: From Selfish Elements to Key Players

Beatriz Contreras, Cristina Vives, Roger Castells and Josep M. Casacuberta

Abstract Transposable elements (TEs) are major components of all eukaryote genomes, and in particular of plant genomes. Whereas these elements have long been considered as selfish 'junk DNA without function', the data accumulated over the years have shown that they are essential components of the genome structure and key players of genome evolution. Here, we summarize the recent advancement in the field and we discuss the role of TEs in the light of the new data coming from whole plant genome sequences and next-generation sequencing (NGS) data on resequencing of plant varieties and lines.

6.1 Transposable Elements, a Major Component of Plant Genome

Transposable elements (TEs) are mobile genetic elements that account for an important fraction of virtually all eukaryote genomes. TEs can be classified into two major classes, class I (retrotransposons) and class II (DNA transposons). Class I elements transpose through an RNA intermediate used as a template in a reverse transcription reaction leading to a new DNA copy that can integrate back into the genome. Therefore, class I TEs do not excise during transposition and their copy number increases as a result of their movement. Whereas the transcription of the element is catalysed by the host's polymerase (Pol II), its reverse transcription and integration are catalysed by enzymatic activities encoded by the retrotransposon itself, in case of autonomous elements, or by a related element, in case of non-autonomous elements. Class II elements transpose via a DNA intermediate, which results from the excision of the element from its chromosomal location and that can be integrated elsewhere into the genome. Both the excision and integration reactions are catalysed by a transposase which is encoded by the mobilized TE in

B. Contreras · C. Vives · R. Castells · J.M. Casacuberta (✉)
Centre de Recerca En Agrigenòmica, CRAG (CSIC-IRTA-UAB-UB), Barcelona, Spain
e-mail: josep.casacuberta@cragenomica.es

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_6

case of autonomous elements or by a related element in case of a defective TE copy. There are, however, some DNA transposons that move through a different mechanism. This is the case of *Helitrons*, which transpose via a rolling-circle mechanism similar to that of some bacterial TEs. Both class I and class II TEs can be further classified into families and subfamilies depending on their structure, encoded proteins and mechanism of transposition (Wicker et al. 2007).

Whereas TEs are commonplace in eukaryotes, and most eukaryotes contain elements belonging to all major types and classes, their prevalence differs from genome to genome. TEs account for a major but variable fraction of plant genomes (Bennetzen and Wang 2014), with LTR retrotransposons and miniature inverted-repeat transposable elements (MITEs) tending to be the most represented types of TEs (Casacuberta and Santiago 2003). The variability in TE content is huge in plants. For instance, as much as 85 % of maize genome or 70 % of Norway spruce genome (Nystedt et al. 2013) has been annotated as transposons, whereas transposon annotations make only the 21 % of the more compact *Arabidopsis thaliana* genome (Ahmed et al. 2011). These numbers are not directly comparable as the methods and the parameters used to perform the annotations are different, and this may have an important impact on the sensibility and specificity of the detection. Indeed, analyses in *A. thaliana* have shown that there is a continuum between repetitive elements and unannotated genomic dark matter, making it somehow arbitrary to define a frontier (Maumus and Quesneville 2014). However, in spite of these limitations, there seem to be a direct relationship between genome size and percentage of TEs within the genome. Analyses of closely related species, for example of the *Oryza* genus (Chénais et al. 2012), suggest that TE activity and polyploidization are the two main mechanisms responsible for genome size increase during evolution (Panaud et al. 2014). The relationship between genome duplication and transposition is interesting. On the one hand, gene duplication can allow genomes to tolerate a higher TE activity, as their mutagenic capacity is buffered by having extra copies of essential genes, but on the other hand, the lack of gene duplications may force the genome to explore other sources of innovations such as transposition. In this respect, it is interesting to note that gymnosperms, that in contrast to angiosperms do not seem to have suffered recent whole-genome duplications, present extremely big genomes with a very high content of TEs (De La Torre et al. 2014).

The effect of TE activity in genome size may be quite dramatic over short periods of time, as suggested by the high activity of TEs associated to the genome size doubling of *Oryza australiensis,* a wild relative of rice, during the last three million years (Zhao and Ma 2013). However, although TEs may be responsible for rapid genome size changes, their activity is not constant during evolution. Indeed, TEs seem to alternate periods where they are relatively quiescent with burst of transposition where their copy number increases significantly (Vitte et al. 2014). This evolutionary behaviour of transposons as a whole can be in part explained by the results obtained analysing the regulation of particular transposons and genomes. All the data accumulated so far indicate that transposons are heavily silenced in genomes by different mechanisms, and in particular by epigenetic mechanisms

(Ito and Kakutani 2014). Silent TEs of different classes, including both DNA transposons and retrotransposons, can be reactivated in mutated genetic backgrounds showing reduced DNA methylation (Ito and Kakutani 2014), which shows that the silenced TEs retain their capacity to be activated. In fact, TEs can be activated in wild-type plants in particular situations or developmental stages. TEs are de-repressed in the gametophytes and their expression may allow the production of sRNAs to ensure the maintenance of the epigenetic silencing of TEs in the following generation, although alternative explanations of this phenomenon are also possible (Martínez and Slotkin 2012). In addition, over the years, data have accumulated on the stress-related activation of different TEs. This includes the well-studied activation of the tobacco retrotransposon *Tnt1* by biotic and abiotic stresses (Grandbastien et al. 2005), the cold and salt activation of the rice MITE *mPing* (Naito et al. 2009) and the heat activation of the Arabidopsis *ONSEN* retrotransposons (Cavrak et al. 2014). Similarly, it is known that in vitro culture, which can be considered as a complex stress, can reactivate TEs in rice and maize (Hirochika 1997; Kaeppler et al. 2000). Plants are subjected to stress in nature, and this may lead to reactivation of TEs in certain cells. In most cases, the somatic activation of TEs will not lead to germinal transpositions and therefore will not be inherited by the successive generations. However, in particular situations, a general release of the control mechanism may lead to a general activation of TEs leading to a burst of transposition. It is interesting to note that it has been shown that interspecific crosses or polyploidization events may lead to global epigenetic changes and activation of TEs (Parisod et al. 2009; Yaakov and Kashkush 2011). As these phenomena are commonplace in plant evolution, this may give the opportunity to TE amplification bursts to occur and accompany speciation events.

6.2 Transposable Elements in Genome Structure

TEs are usually not homogeneously distributed along chromosomes. They concentrate in pericentromeric regions, while they are less abundant in chromosome arms, in a pattern that is usually complementary to that of genes. These pattern of TEs can be the consequence of both a preferential insertion into these regions, as demonstrated for yeast retroelements, or the effect of selection cleaning up the more frequently deleterious TE insertions in gene-rich regions (Neumann et al. 2011; Peterson-Burch et al. 2004). Selection against insertion within genes, which are not homogeneously distributed along chromosomes, and the recombination rate, which is also different in different chromosomal regions and greatly influences TE elimination, explains in part the distribution of TEs (Bennetzen and Wang 2014). However, it has been shown that some TEs indeed have a preferential insertion into certain genomic regions. In general, *Copia*-like TEs show some preference for gene-rich regions, whereas *Gypsy*-like TEs are supposed to target preferentially the heterochromatic pericentromeric regions (Peterson-Burch et al. 2004). As an example, the tobacco *Tnt1* and the rice *Tos17 Copia* elements preferentially insert

into gene-rich regions (Miyao et al. 2003; Le et al. 2007), whereas in cereals, there are some families of *Gypsy* retrotransposons that are almost exclusively located in the centromeres, suggesting a high preference for insertion into these regions (Gao et al. 2009; Wolfgruber et al. 2009; Langdon et al. 2000; Li et al. 2013; Jiang et al. 2003). However, there are exceptions to this rule, and some *Gypsy* elements such as the low-copy-number *LORE1* retrotransposon from *Lotus japonicus* seem to target gene-rich regions (Madsen et al. 2005) and some *Copia*-like retrotransposons such as the *Tal1* element from *Arabidopsis lyrata* target the centromere for integration (Tsukahara et al. 2012).

The fact that TEs, and in particular high-copy-number retrotransposons, tend to concentrate in gene-poor heterochromatic regions, does not imply that they do not impact on genome function. Indeed, TE insertions in the pericentromeric regions probably have a profound impact on the structure and dynamics of genomes. The main mechanism to control the activity of TEs is their epigenetic silencing. As a consequence of their silencing, TE sequences tend to be heavily methylated and are associated with expression-repressive histone modifications (Ito and Kakutani 2014). Therefore, the concentration of TEs in the centromere also concentrates certain epigenetic marks in these regions, leading to a particular chromatin structure that is essential for heterochromatin compaction and function in the centromeres (Wong and Choo 2004). It has been proposed that TEs, and in particular LTR retrotransposons sitting in the centromere, may transcribe flanking repeats and other centromeric sequences leading to the production of double-stranded RNA which would direct their particular heterochromatic structure (Lippman et al. 2004). In fact, studies on the formation of neocentromeres have shown that it is the epigenetic nature of centromere elements, and not their sequence, which ensures its functionality (Zhang et al. 2013). Therefore, there is probably a dynamic interplay between retrotransposons and heterochromatin where some TEs target heterochromatin for integration (in the case of *Gypsy*-like elements through the chromodomains of their integrases that are known to interact with some heterochromatic epigenetic marks) and help thereafter to maintain heterochromatin by directing their epigenetic modification (Gao et al. 2008).

6.3 Transposable Elements as a Source of New Functions

TEs impact on genome and gene evolution in many ways. Perhaps, the most obvious is the generation of null mutations by transposing into a gene. Some of these null mutations have been selected by humans during plant domestication such as the waxy and sticky varieties of foxtail millet (*Setaria italica*), or Mendel's wrinkled peas (Lisch 2013). For TEs that transpose by a cut-and-paste mechanism (e.g. most class II TEs), the excision of the element may result in function recovery giving rise to mosaic phenotypes as exemplified by the kernel colour of maize cobs. Nevertheless, in some cases the excision may leave behind parts of the element that are not removed and can modify the coding capacity of the gene, and in some cases

provide new gene functions (Lisch 2013; Oliver et al. 2013). This process by which a TE, or a part of it, is established in a specific region and gains a cellular function is known as molecular domestication (Kajihara et al. 2012).

There is an important number of plant genes with a transposon origin (Oliver et al. 2013; Bennetzen and Wang 2014). In particular, several important transcription factors derive from class II transposases. For example, *Daysleeper*, a transcription factor that regulates the morphogenetic development in *A. thaliana,* is derived from a *hAT* transposase (Bundock and Hooykaas 2005), or the light response FHY3 and FAR1 transcription factors that are ancient *Mutator* transposases (Hudson et al. 2003; Lin et al. 2007).

Transposons can also capture, duplicate and mobilize genes or gene fragments, creating new opportunities for gene evolution. Retrotransposons duplicate host genes or gene fragments through the reverse transcription of their mRNAs generating what is called a retrogene. The retroposed gene fragments can be fused to host genes to generate new chimeric proteins (Elrouby and Bureau 2010), and retroposed retrogenes can be regulated differently to the original genes (Abdelsamad and Pecinka 2014), which can be a source of gene innovation. Class II transposons can also transduplicate genes. Pack-MULEs, for example, are *Mutator*-like TEs that carry fragments of genes in different plants and were proposed as important mediators of gene evolution in plants (Jiang et al. 2004). The fact that an important fraction of rice Pack-MULEs is transcribed and show signs of purifying selection suggested that indeed these elements have a role in gene evolution in plants (Hanada et al. 2009). A part from MULEs, other class II TEs, such as *CACTA* elements, have been shown to transduplicate host gene fragments in different plants (Benjak et al. 2008; Morgante 2006). But probably the TEs that seem to capture more actively, amplify and mobilize gene fragments are the rolling-circle transposing elements *Helitrons*. More than one-third of the thousands *Helitrons* of maize genome carry at least one host gene fragment (Du et al. 2009). Therefore, TEs have a great potential to generate new gene structures by shuffling host genome sequences (Bennetzen 2005; Morgante 2006).

6.4 Impact of Transposable Elements in Gene Regulation

In addition to their effect on the coding capacity of the host genome, TEs can impact on host genes in many ways. As already explained, the expression of TEs is tightly regulated, both because they are the main target of the silencing machinery and also because they usually have stress-related promoters that are only active under particular situations. For this reason, in addition to being able to modify host gene expression by interrupting gene regulatory regions upon insertion, for example in the case of the *Vgt1* regulatory locus of maize (Salvi et al. 2007), TEs can modify the expression of host genes located nearby by contributing their own regulatory elements or by attracting the silencing machinery.

There are several examples of insertions of TEs that induce new transcriptional regulations to host genes. This is the case of the insertion of a *Hopscotch* TE some 50 Kb upstream of the *theosinte branched* 1 (*tb1*) gene, which represses branching in maize, which results in its overexpression and the apical dominant phenotype of modern maize (Studer et al. 2011) or the insertion of an LTR retrotransposon upstream of the *Ruby* gene in oranges which confers to this gene a developmental regulation and cold inducibility resulting in the blood orange phenotype (Butelli et al. 2012).

MITEs are a particular type of transposons present in high copy numbers in plant genomes (Casacuberta and Santiago 2003). They are relatively small, which may help them avoiding to generate complete knockout phenotypes, and although they do not need to be expressed to transpose, they can contain transcriptional regulatory sequences. For example the rice *mPing* MITE contains stress-responsible transcriptional regulatory elements that upregulate neighbouring genes under cold and salt stress conditions (Yasuda et al. 2013; Naito et al. 2009). The high copy number of MITEs makes them particularly suited to modify the expression of groups of genes, making it possible to create, or to extend, transcriptional regulatory networks. The fact that some transcription factors derive from transposases (see above), and that the sequences bound by transposases (e.g. the TIRs) can be mobilized throughout the genome, was proposed as a potential mechanisms to create and modify transcriptional regulatory networks (Feschotte 2008). In the recent years, evidences that TEs can mobilize transcription factor binding sites and rewire transcriptional networks have accumulated (Rebollo et al. 2012). In plants, a recent report from our laboratory has shown that different families of MITEs have amplified and redistributed the binding sites for the E2F transcription factor during *Brassica* evolution, and the insertion of some of these MITEs may have wired new genes into the E2F transcriptional network (Hénaff et al. 2014).

In spite of the examples explained above that illustrate the potential of TEs to bring new regulatory sequences to host genes, the most frequent effect of a TE insertion within or close a gene promoter is its inactivation. As already explained, TEs are controlled by epigenetic mechanisms that silence them tightly. For this reason, most TEs are heavily methylated and are associated to inactive chromatin, and this can influence genes located nearby that can become silenced by the presence of the TE. A well-studied example of such an effect is the epigenetic silencing of a sex determination gene in melon linked to a TE insertion in its upstream region (Martin et al. 2009). Similarly, the necessary repression of the flowering regulator *FWA* gene in *A. thaliana* is a consequence of the epigenetic silencing of a SINE transposon located in its promoter (Kinoshita et al. 2007). Genome-wide analyses suggest that these effects may be highly relevant. As an example, it has been shown that about 300 genes differentially expressed in maize populations have changes in DNA methylation, and many of these regions are associated with transposons (Eichten et al. 2013). This suggests that polymorphic TE insertions modify the pattern of genome methylation which translates into changes in gene expression.

Silencing of TEs is mediated by siRNAs that target TE sequences which probably originate from the expression of particular TE structures (e.g. inverse repeated elements). Whereas the main target of these siRNAs are TEs, in some cases TEs may produce siRNAs that target host genes (Bennetzen and Wang 2014; McCue and Slotkin 2012). In fact, it has been proposed that TE can be the source of both siRNAs and miRNAs (Li et al. 2011; Piriyapongsa and Jordan 2008) which suggests that the genome has evolved a new layer of gene regulation from its defence mechanisms against TEs.

The expression of TEs may also interfere with host genes creating sense or antisense transcripts that may result in their specific silencing. It has been shown that read-through transcription, due to a leaky transcriptional terminator, is relatively frequent in plant retrotransposons, and this could result in the inclusion of flanking sequences into retrotransposon transcripts. As a consequence, as it has been shown in tobacco (Hernández-Pinzón et al. 2009), the convergent transcription of a retrotransposon located downstream of a host gene could result in the formation of dsRNAs which may potentially regulate the host gene. In addition, TEs insertions in 5' leader region, 3' trailer sequence or introns can modify the sites of RNA processing or polyadenylation affecting gene expression (Bennetzen and Wang 2014).

6.5 Transposable Elements Dynamics and Evolution of Crop Plants

We have seen in the previous sections that TEs can impact on genomes in many ways, from providing new genes or modifying the existing ones or alter their expression, to modify genome or chromosome structure. Because of that TEs are an extraordinary source of novelty useful for evolution (Lisch 2013). In particular, in the last few years, a number of examples of TE insertions leading to important agronomic traits that have been selected during evolution and breeding have accumulated (Lisch 2013). These include the different flesh fruit colour in blood orange (Butelli et al. 2012), the different skin colours in grapevine (This et al. 2007), the nectarine phenotype in peaches (Vendramin et al. 2014) or the seedless phenotype in apples (Yao et al. 2001) (see Fig. 6.1). However, evaluating the impact of TEs in the evolution of eukaryote genomes is not an easy task. In spite of the examples listed above on TEs that gave rise to mutations that have been selected during evolution, a general evaluation is still lacking. There are several reasons for that, as previously pointed out (Vitte et al. 2014). Although the number of plant genomes sequenced is growing rapidly, the quality of the published genomes is not always good enough to allow a proper analysis of the TE content. Indeed, most published genomes contain a variable, and usually important, fraction of unassembled reads which are usually enriched in repetitive sequences including TEs. This precludes a complete genome-wide TE analysis. In addition to the quality of

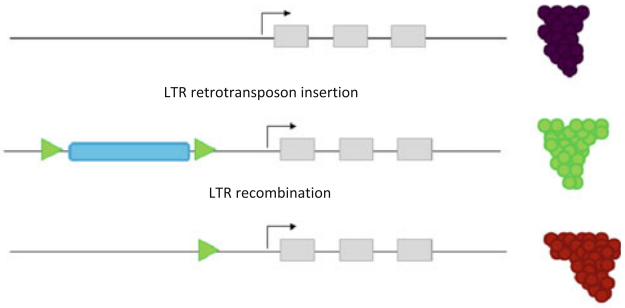
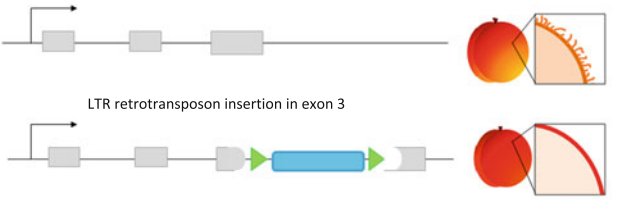
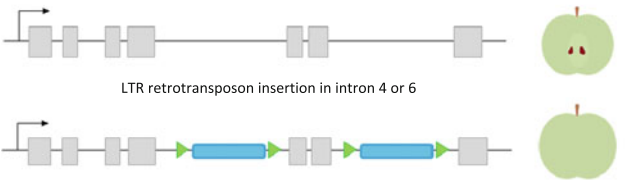
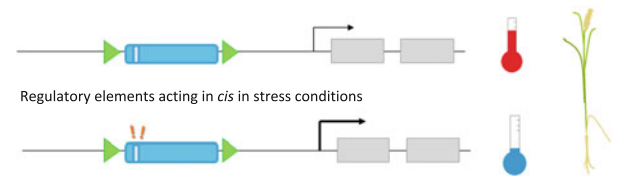
TE related traits in crops	Representation
TE insertion in promoter sequence confers different grape skin colours. (This et al. 2007)	
Peach and nectarine skin phenotype due to a *Copia* retrotransposon insertion in exon sequence. (Vendramin et al. 2014)	
Seedless phenotype in some apple varieties is due to a retrotransposon insertion in an intron. (Yao et al. 2001)	
A TE insertion in regulatory region confers cold specificity regulation in rice. (Yasuda et al. 2013)	

Fig. 6.1 Representation of different important agronomic traits that are due to transposable element insertions. *Grey boxes* represent exons, *blue boxes* represent TE coding region, and *green triangles* represent LTRs

the sequence and assembly, the annotation of the TE content is also highly variable among the sequenced genomes. There are several reasons for that, including the use of different bioinformatics tools and pipelines as well as the thresholds set which determine the sensitivity and specificity of the annotation tools. This makes comparisons of the TE content between genomes a very difficult exercise, and different voices claim that there is a need for an international effort to standardize the methods used for annotating TEs (Hoen, Bureau, Bourke and Blanchette, in

preparation). But even with good genome sequences and TE annotation, reference genomes are only a snapshot, a fixed image, of a genome and analysing the impact of TEs in genome evolution will require sequence variability analysis within a species or among different related species. In the last few years, an important amount of resequencing data of crop varieties and landraces has being accumulated. As an example, 3000 rice varieties have already been sequenced and offer an unprecedented opportunity to search for the genetic bases of a wide range of phenotypic differences (Li et al. 2014). However, in most cases, the analyses of variability are restricted to SNPs, and TE insertion polymorphisms are not analysed. The reason for that is that detecting TE polymorphisms, and in particular TE insertions with respect to the reference genome is far from trivial. There are a number of recent tools that allow detecting TE insertion polymorphisms using paired-end resequencing data, including TEA (Lee et al. 2012), RetroSeq (Keane et al. 2013), VariationHunter (Hormozdiari et al. 2010), TEMP (Zhuang et al. 2014) and Jitterbug (Hénaff et al. submitted), but they are only starting to be used to determine the role of TEs in plant genome evolution (see for example Sanseverino et al., submitted). The use of these tools on the growing amount of resequencing data on plant varieties and accessions will probably allow us in the next future to have a more global and complete view of the impact of TEs in plant genome evolution. In particular, the analysis of crop genomes and the comparison of crop reference genomes with that of, on the one hand, their wild ancestors, and on the other hand, domesticated landraces or elite varieties will shed light on the role of TEs on the evolution of plant genomes during domestication and breeding. In addition, as crop domestication is an excellent model to study genome evolution at large, as it has already been said (Olsen and Wendel 2013), these analyses will probably allow us to better understand the structure and evolution of plant genomes and the key role played by TEs, who once were called junk DNA and now are rediscovered as key factors for genetic innovation.

References

Abdelsamad A, Pecinka A (2014) Pollen-specific activation of Arabidopsis retrogenes is associated with global transcriptional reprogramming. Plant Cell 26:3299–3313. doi:10.1105/tpc.114.126011

Ahmed I, Sarazin A, Bowler C et al (2011) Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res 39:6919–6931. doi:10.1093/nar/gkr324

Benjak A, Forneck A, Casacuberta JM (2008) Genome-wide analysis of the "cut-and-paste" transposons of grapevine. PLoS ONE 3:e3107. doi:10.1371/journal.pone.0003107

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627. doi:10.1016/j.gde.2005.09.010

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. doi:10.1146/annurev-arplant-050213-035811

Bundock P, Hooykaas P (2005) An Arabidopsis hAT-like transposase is essential for plant development. Nature 436:282–284. doi:10.1038/nature03667

Butelli E, Licciardello C, Zhang Y et al (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 24:1242–1255. doi:10.1105/tpc.111.095232

Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1–11. doi:10.1016/S0378-1119(03)00557-2

Cavrak VV, Lettner N, Jamge S et al (2014) How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet 10:e1004115. doi:10.1371/journal.pgen.1004115

Chénais B, Caruso A, Hiard S, Casse N (2012) The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 509:7–15. doi:10.1016/j.gene.2012.07.042

De La Torre AR, Birol I, Bousquet J et al (2014) Insights into conifer giga-genomes. Plant Physiol 166:1724–1732. doi:10.1104/pp.114.248708

Du C, Fefelova N, Caronna J et al (2009) The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci U S A 106:19916–19921. doi:10.1073/pnas.0904742106

Eichten SR, Briskine R, Song J et al (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25:2783–2797. doi:10.1105/tpc.113.114793

Elrouby N, Bureau TE (2010) Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol 153:1413–1424. doi:10.1104/pp.110.157420

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397–405. doi:10.1038/nrg2337

Gao D, Gill N, Kim H-R et al (2009) A lineage-specific centromere retrotransposon in Oryza brachyantha. Plant J 60:820–831. doi:10.1111/j.1365-313X.2009.04005.x

Gao X, Hou Y, Ebina H, et al (2008) Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res 359–369. doi:10.1101/gr.7146408.1

Grandbastien M, Audeon C, Bonnivard E et al (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet Genome Res 110:229–241. doi:10.1159/000084957

Hanada K, Vallejo V, Nobuta K et al (2009) The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell 21:25–38. doi:10.1105/tpc.108.063206

Hénaff E, Vives C, Desvoyes B et al (2014) Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. Plant J 77:852–862. doi:10.1111/tpj.12434

Hénaff E, Zapata L, Casacuberta JM, Ossowski S. (Submitted) Jitterbug: somatic and germline transposon insertion detection at single-nucleotide resolution

Hernández-Pinzón I, de Jesús E, Santiago N, Casacuberta JM (2009) The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. J Mol Evol 68:269–278. doi:10.1007/s00239-009-9204-y

Hirochika H (1997) Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol 35:231–240

Hormozdiari F, Hajirasouliha I, Dao P et al (2010) Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery. Bioinformatics 26:i350–i357. doi:10.1093/bioinformatics/btq216

Hudson M, Lisch D, Quail P (2003) The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J 453–471

Ito H, Kakutani T (2014) Control of transposable elements in Arabidopsis thaliana. Chromosome Res 22:217–223. doi:10.1007/s10577-014-9417-9

Jiang J, Birchler J a, Parrott W a, Kelly Dawe R (2003) A molecular view of plant centromeres. Trends Plant Sci 8:570–575. doi:10.1016/j.tplants.2003.10.011

Jiang N, Bao Z, Zhang X et al (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573. doi:10.1038/nature02945.1

Kaeppler S, Kaeppler H, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol 179–188

Kajihara D, Godoy F, Hamaji T et al (2012) Functional characterization of sugarcane mustang domesticated transposases and comparative diversity in sugarcane, rice, maize and sorghum. Mol Biol 639:632–639. doi:10.1590/S1415-47572012005000038

Keane TM, Wong K, Adams DJ (2013) RetroSeq: transposable element discovery from next-generation sequencing data. Bioinformatics 29:389–390. doi:10.1093/bioinformatics/bts697

Kinoshita Y, Saze H, Kinoshita T et al (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. Plant J 49:38–45. doi:10.1111/j.1365-313X.2006.02936.x

Langdon T, Seago C, Mende M et al (2000) Retrotransposon evolution in diverse plant genomes. 156(1):313–325

Le QH, Melayah D, Bonnivard E et al (2007) Distribution dynamics of the Tnt1 retrotransposon in tobacco. Mol Genet Genomics 278:639–651. doi:10.1007/s00438-007-0281-6

Lee E, Iskow R, Yang L, Gokcumen O (2012) Landscape of somatic retrotransposition in human cancers. Science 337:967–971. doi:10.1126/science.1222077.Landscape

Li B, Choulet F, Heng Y et al (2013) Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. Plant J 73:952–965. doi:10.1111/tpj.12086

Li J-Y, Wang J, Zeigler RS (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. Gigascience 3:8. doi:10.1186/2047-217X-3-8

Li Y, Li C, Xia J, Jin Y (2011) Domestication of transposable elements into MicroRNA genes in plants. PLoS ONE 6:e19212. doi:10.1371/journal.pone.0019212

Lin R, Ding L, Casola C, Ripoll D (2007) Transposase-derived transcription factors regulate light signaling in Arabidopsis. Science 318:1302–1305

Lippman Z, Gendrel A, Black M (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature. doi:10.1038/nature02724.1

Lisch D (2013) How important are transposons for plant evolution? Nat Rev Genet 14:49–61. doi:10.1038/nrg3374

Madsen LH, Fukai E, Radutoiu S et al (2005) LORE1, an active low-copy-number TY3-gypsy retrotransposon family in the model legume Lotus japonicus. Plant J 44:372–381. doi:10.1111/j.1365-313X.2005.02534.x

Martin A, Troadec C, Boualem A et al (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461:1135–1138. doi:10.1038/nature08498

Martínez G, Slotkin RK (2012) Developmental relaxation of transposable element silencing in plants: functional or byproduct? Curr Opin Plant Biol 15:496–502. doi:10.1016/j.pbi.2012.09.001

Maumus F, Quesneville H (2014) Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS ONE 9:e94101. doi:10.1371/journal.pone.0094101

McCue AD, Slotkin RK (2012) Transposable element small RNAs as regulators of gene expression. Trends Genet 28:616–623. doi:10.1016/j.tig.2012.09.001

Miyao A, Tanaka K, Murata K et al (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780. doi:10.1105/tpc.012559.ements

Morgante M (2006) Plant genome organisation and diversity: the year of the junk! Curr Opin Biotechnol 17:168–173. doi:10.1016/j.copbio.2006.03.001

Naito K, Zhang F, Tsukiyama T et al (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130–1134. doi:10.1038/nature08479

Neumann P, Navrátilová A, Koblížková A et al (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA 2:4. doi:10.1186/1759-8753-2-4

Nystedt B, Street NR, Wetterbom A et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. doi:10.1038/nature12211

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5:1886–1901. doi:10.1093/gbe/evt141

Olsen KM, Wendel JF (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. Annu Rev Plant Biol 64:47–70. doi:10.1146/annurev-arplant-050312-120048

Panaud O, Jackson S, Wendel J (2014) Drivers and dynamics of diversity in plant genomes. New Phytol 202:15–18. doi:10.1111/nph.12633

Parisod C, Salmon A, Zerjal T et al (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in Spartina. New Phytol 184 (4):1003–1015. doi:10.1111/j.1469-8137.2009.03029.x

Peterson-Burch B, Nettleton D (2004) Voytas D (2004) Genomic neighborhoods for Arabidopsis retrotransposons: a role for targeted integration in the distribution of the Metaviridae. Genome Biol 5:R78

Piriyapongsa J, Jordan I (2008) Dual coding of siRNAs and miRNAs by plant transposable elements. Rna 814–821. doi:10.1261/rna.916708.ferred

Rebollo R, Romanish MT, Mager DL (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 46:21–42. doi:10.1146/annurev-genet-110711-155621

Salvi S, Sponza G, Morgante M et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci U S A 104:11376–11381

Sanseverino W, Hénaff E, Vives C, et al. (submitted) The contribution of transposon insertion polymorphisms and nucleotide variability to the evolution of the melon genome

Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43:1160–1163. doi:10.1038/ng.942.Identification

This P, Lacombe T, Cadle-Davidson M, Owens CL (2007) Wine grape (Vitis vinifera L.) color associates with allelic variation in the domestication gene VvmybA1. Theor Appl Genet 114:723–730. doi:10.1007/s00122-006-0472-2

Tsukahara S, Kawabe A, Kobayashi A (2012) Centromere-targeted de novo integrations of an LTR retrotransposon of Arabidopsis lyrata. Genes Dev 26:705–713. doi:10.1101/gad.183871.111 Epub 2012 Mar 19

Vendramin E, Pea G, Dondini L et al (2014) A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. PLoS ONE 9:e90574. doi:10.1371/journal.pone.0090574

Vitte C, Fustier M-A, Alix K, Tenaillon MI (2014) The bright side of transposons in crop evolution. Brief Funct Genomics 13:276–295. doi:10.1093/bfgp/elu002

Wicker T, Sabot F, Hua-van A et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Wolfgruber TK, Sharma A, Schneider KL et al (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet 5:e1000743. doi:10.1371/journal.pgen.1000743

Wong LH, Choo KHA (2004) Evolutionary dynamics of transposable elements at the centromere. Trends Genet 20:611–616. doi:10.1016/j.tig.2004.09.011

Yaakov B, Kashkush K (2011) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54:42–49. doi:10.1139/G10-091

Yao J, Dong Y, Morris B (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. Proc Natl Acad Sci U S A 98:1306–1311

Yasuda K, Ito M, Sugita T et al (2013) Utilization of transposable element mPing as a novel genetic tool for modification of the stress response in rice. Mol Breed 32:505–516. doi:10.1007/s11032-013-9885-1

Zhang B, Lv Z, Pang J et al (2013) Formation of a functional maize centromere after loss of centromeric sequences and gain of ectopic sequences. Plant Cell 25:1979–1989. doi:10.1105/tpc.113.110015

Zhao M, Ma J (2013) Co-evolution of plant LTR-retrotransposons and their host genomes. Protein Cell 4:493–501. doi:10.1007/s13238-013-3037-6

Zhuang J, Wang J, Theurkauf W, Weng Z (2014) TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res 42:6826–6838. doi:10.1093/nar/gku323

Chapter 7
Sympatric Differentiation and Speciation: Insights from Drosophila Studies

Eugenia Rashkovetsky, Zeev Frenkel, Pawel Michalak and Abraham Korol

Abstract One can hardly find a more controversial issue in evolution biology than species concept. Inside this widely debatable area, the most discussable questions concern the driving forces of population differentiation and the role of geographical isolation as a factor of species divergence (allopatric vs. sympatric scenarios). Here, we review the main influential theoretical works and experimental evidence regarding the validity of sympatric model. We also present our empirical data on extensive studies of interslope genetic divergence of *Drosophila* at Nahal Oren canyon (Mount Carmel, Israel). Our results suggest that populations inhabiting opposite slopes exemplify ongoing divergence taking place regardless of high migration. For a long time, sympatric speciation was considered possible, but an extremely rare event that can hardly be observed or proved. S. Via found a precise image-bearing expression for the sympatric model: "The Ugly Duckling." Based on our results and experience, we can say that we see a swan in this ugly duckling!

E. Rashkovetsky · Z. Frenkel · A. Korol (✉)
Institute of Evolution, University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, 3498838 Haifa, Israel
e-mail: korol@research.haifa.ac.il

E. Rashkovetsky
e-mail: eugenia@research.haifa.ac.il

Z. Frenkel
e-mail: vova_frenkel@yahoo.com

P. Michalak
Virginia Bioinformatics Institute, Virginia Tech, 1015 Life Science Circle, Blacksburg, VA 24061, USA
e-mail: pawel@vbi.vt.edu

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_7

7.1 Introduction

There are some branches of knowledge where revealed regularities and basic schemes enabled logical description of the system's elements and even prediction of the existence and properties of additional, missing, elements. Application of the periodic table of the elements in chemistry is an example of such a system. Another example is provided by the history of elementary particles in physics. Similar ambitions can be found in evolutionary biology. Numerous efforts have been made to offer an overall concept of evolution which includes all patterns of species diversity. Investigators constantly suggest new approaches and tools for describing and classifying known species in order to create a precise scheme for systematic biology. Despite high expectations, the genome sequences, as it is becoming increasingly clear, are unlikely to meet the requirements necessary to create such a system (Evans 2000).

Ideas about building some kind of evolutionary "bioperiodic table," using protein fold as a suitable fundamental unit (Torday 2004; Dhar 2007), are among the recent suggestions. New findings supporting the idea of overall evolutionary system (picture) and complementing missing elements in "the great chain of being" appear from time to time. Thus, a special collection of 11 articles focusing on discovery of *Ardipithecus ramidus* (a hominid species that lived 4.4 million years ago) has been published by an international team of 47 scientists in a special issue of Science journal (Science, vol. 326, issue 5949, 2009). These articles provide a huge amount of new empirical data on hominid evolution. The findings clearly indicate that hominids and apes have evolved through different evolutionary pathways. Hence, the missing link in human evolution (early hominid common ancestor that was neither chimp nor human) has been found and scrupulously described. The existence of such a missing element in evolutionary chain was predicted a long time ago.

Organisms' ability to survive in severely changing environments is one of the most amazing issues in natural science. Patterns of initial differentiation in populations subjected to various selective factors are of primary importance for evolutionary biology. We used to think that speciation is an exceptionally slow process and that it is practically impossible to directly observe changes occurring in populations. However, the latest studies suggest that steps of speciation may be detected much easier and earlier than expected (Reed and Markow 2004; Grant and Grant 2014). Recent advances in next-generation sequencing (NGS) technologies and bioinformatics provide a new dimension to the detection of selection, adaptation, and demography signatures at the sequence level. Whole-genome sequencing has become a widely used technology for the discovery of genetic variation and its relationship with the phenotype, in particular for complex traits. Based on existing theoretical framework, available genomic tools promote evolutionary explanations and development of new models. Despite fast-falling prices of NGS, obtaining reliable sequence information on the individual level is still rather expensive, but recent bioinformatics tools for population genomics of pooled sequences (Pool-Seq)

provide a good compromise (Futschik and Schlötterer 2010; Kolaczkowski et al. 2011; Orozco-terWengel et al. 2012; Fabian et al. 2012) to identify alleles favorable for certain environment. Let us look at everything in order.

7.1.1 Ecological Selection in Natural Populations

Using morphological, physiological, and biochemical traits as criteria, Endler (1986) assessed that by 1986, natural selection had been directly demonstrated for 314 traits in 141 species. The majority of mentioned examples were for morphological traits, but some of them presented physiological characters, including stress tolerance. The main criterion that Endler used to include examples of selection in his review was that the traits were heritable. Thus, these data clearly demonstrate that natural selection is not rare in nature. Relevant studies that have been published during the last decades focused on the analysis of extant species existing in sympatry (Schluter 1998; Taylor and McPhail 2000). A review of the strength of phenotypic selection in natural populations (Kingsolver et al. 2001) emphasized that more than 80 % of the estimates were obtained for morphological traits, while data for behavioral or physiological traits are rare. A relatively new meta-analysis provided strong empirical support for the conclusion that natural selection plays a truly general role in speciation (Funk et al. 2006). Adding an ecological dimension to comparative studies exploring reproductive isolation and divergence time, the authors analyzed the relationship between natural selection and the ability to interbreed in hundreds of species (plants, insects, fish, frogs, and birds). Ecological divergence was quantified for more than 500 species pairs, and highly consistent positive association was found between ecological divergence and reproductive isolation across taxa.

In *Drosophila*, ecological selection has attracted much attention, both in artificial selection and in natural populations. The overwhelming majority of these studies dealt with adaptive significance of acclimation, habitat selection, and morphological traits in geographically separated populations (macroscale differentiation) (Huey et al. 2000; Robinson et al. 2000; Gilchrist et al. 2000; Zwaan et al. 2000; Gockel et al. 2002; Kennington et al. 2001). Divergence of *Drosophila* populations in the Hawaiian archipelago is one of the well-known examples of speciation occurring in natural conditions (Carson et al. 1970; Carson and Bryant 1979; Kaneshiro 1988). DeSalle and Giddings (1986) reconstructed the phylogeny of several Hawaiian populations using mitochondrial DNA. The phylogeny resulting from these analyses corresponds well with the geographical distribution of the populations. Further insights into initial steps of speciation were provided by the results of Begun and Aquadro (1993) who showed molecular differentiation between Zimbabwean and US populations of *D. melanogaster*. In particular, strong sexual prezygotic isolation was revealed between the Zimbabwean population and cosmopolitan populations of other continents in behavioral and sperm–egg interactions (Wu et al. 1995; Alipaz et al. 2001). The Zimbabwean and other African populations exhibit assortative

mating in favor of partners of their own origin. In *D. melanogaster* populations, one of the mating signals, interpulse interval (IPI) in "courtship song" (produced by wing vibrations during courtship), plays a significant role in mating success. African flies displayed both a strong sexual isolation from the cosmopolitan flies (Hollocher et al. 1997) and unusually short mean IPI (Colegrave et al. 2000). Interesting findings on early speciation events were published by Laura Reed and Therese Markow (2004): Substantial intraspecific polymorphism for genetic factors contributing to hybrid male sterility was found in geographically separated populations of *D. mojavensis*. All these outcomes suggest that the earliest stages of speciation (incipient speciation) in natural groups are observable in high-precision experiments.

Allopatry is generally considered to be a primary system for reproductive isolation and speciation development (Mayr 1963; Rice and Hostert 1993; Coyne 1994). In this allopatric mode, genetic drift and genetic changes driven by population adaptation to ecological conditions are the main factors affecting the evolution of isolating mechanisms. The idea that sympatric speciation is also feasible has become widespread in the last decades (Bush 1975; Rosenzweig 1978; Jiggins and Mallet 2000; Korol et al. 2000; Mallet 2001; Schluter 2001; Turelli et al. 2001; Via 2001; Gavrilets 2003; Jiggins 2006; Nevo 2011). Natural selection caused by ecological factors plays a central role in the contemporary hypotheses of sympatric speciation. Thus, it was suggested that strong divergent natural selection might override gene flow even at very small spatial scales (Bush 1975; Endler 1977) and could initiate speciation (Barton and Charlesworth 1984; Schilthuizen 2000). Sympatric speciation scenario is supported by several formal models for evolution of isolation promoted by adaptation to divergent ecological conditions and sexual selection (Bateson 1909; Dobzhansky 1936; Muller 1942; Nei et al. 1983, Tregenza and Butlin 1999; Rundle et al. 2000; Via 2001; Gavrilets and Waxman 2002; Doebeli and Dieckmann 2003; Coyne and Orr 2004; Bank et al. 2012). The model by Higashi et al. (1999) demonstrates that sexual selection alone can lead to sympatric speciation, even if the male trait is neutral with respect to fitness and the female preference is costly. Two other well-known models also support this point of view (Kondrashov and Kondrashov 1999; Dieckmann and Doebeli 1999). Although these models indicate that even a single factor may be sufficient to counter gene flow and recombination, some other models indicate that single-factor situations are less promising in overriding the effect of migration and recombination. Thus, Nevo et al. (2000) developed a multiple-factor formal model combining niche viability selection, niche choice, and positive assortative mating (Fig. 7.1) to explain extraordinary multilocus genetic organization in mole crickets. The model involved also a special case of Wahlund effect and inbreeding. Detailed simulations showed that combination of these mechanisms may produce the observed distribution of alleles via multilocus selection and affect the entire genome organization.

One of the best-known examples of sympatric speciation is the divergence of *Rhagoletis pomeonella*, the maggot fly (Feder et al. 1988; Filchak et al. 2000; Hood et al. 2013). This species has diverged into two subspecies due to the relatively recent introduction of apple trees in the northeastern USA. Initially, the flies used

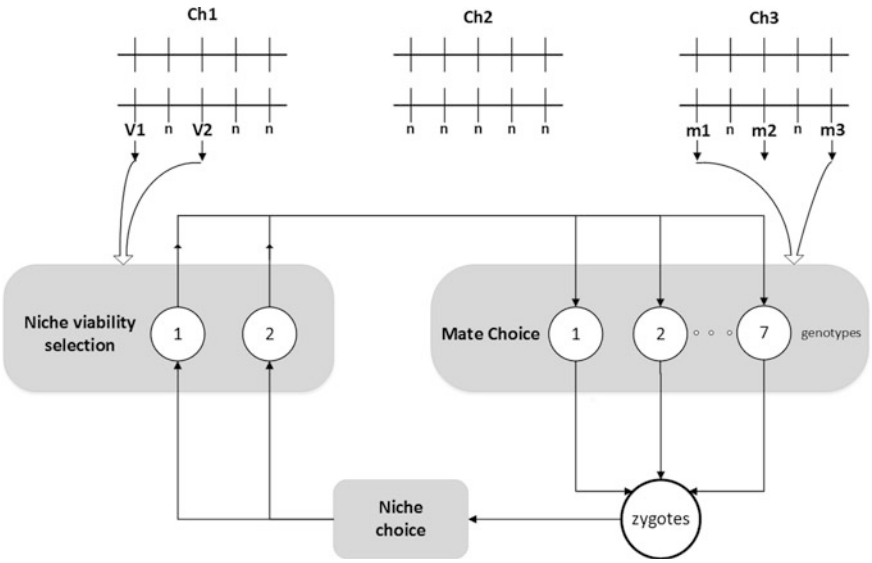


Fig. 7.1 Simulation model for the formation of population structure due to combined effect of niche selection, niche choice, and mate choice (modified from Nevo et al. 2000)

the hawthorn trees as a host plant. Seemingly, the introduced apple trees provided a more attractive food source and an opportunity to escape from common parasites. Difference between the two types of fruits, including maturation timing, allowed for the evolution of some isolation mechanisms in maggot fly and its subsequent divergence into the subspecies. Specialization on different host plants or different habitats and resources may be a driving force for the evolution of reproductive mechanisms in birds and fish (Schliewen et al. 1994; Sorenson et al. 2003; Pereyra et al. 2009). Evidence of sympatric speciation comes from studies investigating evolution of sexual isolation of threespine sticklebacks, inhabiting postglacial lakes in British Columbia (Rundle and Schluter 2004; Boughman et al. 2005; Shluter et al. 2010; Svanbäck and Shluter 2012). In these studies, assortative mating based on preferred body size and male nuptial color was observed in multiple pairs. Ecologically based divergent natural and sexual selections are supposed to initiate speciation in this group. Analogous outcomes were described for nine species of Nicaragua crater lake cichlid fishes in sympatric conditions (Barluenga et al. 2006; Geiger et al. 2010). Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry was recently described in the Eurasian blackcap *Sylvia atricapilla* (Rolshausen et al. 2009). Differential migratory orientation in the birds (arisen due to the improved wintering conditions) facilitated reproductive isolation of sympatric populations in less than 30 generations. In this case, sympatric genetic divergence exceeded that of allopatric blackcap populations separated by 800 km and associated with phenotypic divergence. The authors hypothesize that restricted gene flow accelerates the evolution of adaptive divergence toward the dissimilar selection regimes.

Sympatric speciation was also described in plants. As ploidy changes are rather common in plants, it is considered that polyploid descendants occupying the same ecological niche could be reproductively isolated from parental species. Hybridization followed by chromosome doubling is another possible "differentiation event" in sympatry. The resulting allopolyploids are fertile plants, and there are a lot of species whose appearance can be associated with this mechanism, including crops (hexaploid wheat, triticale, plum *Prunus domestica*, triploid and tetraploid raspberry species, etc.). Besides these known mechanisms of polyploid speciation, other mechanisms may also play a role in differentiation—microsite differences leading to shifts in flowering time (Kelly and Levin 2000; Hall and Willis 2006), different pollinators (Nosil et al. 2005; Lowry et al. 2008), or abiotic microhabitat variables (Rieseberg et al. 1999). However, convincing examples of truly sympatric speciation in plants are very rare (reviewed by Keller and Seehausen 2012). Results published by Savolainen et al. (2006a) on two sister species of palm *Arecaceae* that diverged from each other well after an oceanic island formed caused intense discussion (Stuessy 2006; Savolainen et al. 2006b). The authors argued that sympatric species really do experience diverging selection due to competition for a certain type of soil. In general, there is an obvious lack of studies focusing on the adaptive divergence of plants in sympatry.

Patterns of microscale differentiation (unlike those on the macroscale level reviewed above) are both rarely studied and more controversial. Genetic microscale differentiation between sympatric populations was described in *Bacillus* species (Vilas-Boas et al. 2002; Sikorski and Nevo 2005; Stefanic and Mandic-Mulec 2009), *Formica* ants (Rosengren et al. 1994), Galician *Littorina saxatilis* (Rolan-Alvarez 2007), salt marsh beetle *Pogonus chalceus* from the Guérande salt fields (Dhuyvetter et al. 2007), and other organisms. Significant microscale differentiation (at a few hundred of meters only) was found in *Drosophila persimilis* populations at the Sierra Nevada mountains, California (Taylor and Powell 1977). The observed differentiation for frequencies of inversion and allozymic variants was explained in terms of the habitat choice mechanism. Genotype-specific habitat selection for oviposition sites was demonstrated in the cactophilic species *Drosophila buzzatii* (Barker et al. 1994). Two sympatric populations of *D. melanogaster* derived from the Brazzaville area in Congo (field and brewery populations) were described by Capy et al. (2000) based on the comparisons of several genetically determined traits including morphology, allozymes, microsatellites, cuticular hydrocarbons, and sexual behavior. The authors suggested that reproductive isolation evolves as a by-product of adaptation to certain food resources and environments. Population differentiation along natural ecological (thermal) gradients in *Drosophila* was established by Loeschcke and Hoffman groups (Hoffmann et al. 2005; Sarup et al. 2006), and in some cases, the distance scale was just a few kilometers (Bubliy and Loeschcke 2005).

Genetic divergence was revealed between *D. melanogaster* and (to a lesser extent) *D. simulans* populations inhabiting ecologically contrasting opposite slopes in Nahal Oren canyon, Carmel massif, Israel (Nevo et al. 1998; Harry et al. 1999; Korol et al. 2000, 2006; Iliadi et al. 2001, 2009; Michalak et al. 2001; Lupu et al.

2004; Singh et al. 2005; Rashkovetsky et al. 2000, 2006; Zamorzaeva et al. 2005, 2009; Carmel et al. 2011; Hübner et al. 2013; Kim et al. 2014). The main goal of this multiscale research program (more than 30 publications from different laboratories since 1996) was to characterize the *Drosophila* natural population structure along the sharp microclimatic contrast. *D. melanogaster* from the canyon seems to be one of the most extensively studied natural populations of this species in the world.

We cannot fail to mention here Ernst Mayr's contribution to understanding species concepts (Mayr 1942, 1947, 1963). It is difficult to overestimate his role in the elaboration of the so-called Synthetic Theory of Evolution that synthesized Darwin's views on evolution through natural selection with basics of genetic science. Mayr's influential books and articles integrated knowledge from many disciplines to develop a coherent theory for speciation. The main role in this theory was assigned to geographical isolation as an essential factor for species differentiation. This paradigm was very popular, and Mayr's followers adhered to such a creed for a very long time. Other voices, stating that diverging selection between groups of the same species inhabiting two ecological niches in sympatry could cause reproductive isolation withstanding gene flow, appeared a little later (Smith 1966). A lot of such examples (species divergence without a physical barrier) were given above in this section. Who can better comment the current trend of affairs than Mayr himself? In one of the latest interview to the "Academy of Achievement," he said: "The next thing is sympatric speciation … If you read my work carefully, which most of my opponents don't do, I don't say a sympatric species is impossible. I say, 'No case of sympatric speciation has been proven, has been well documented.' I thought it would be very rare because it would mean simultaneous preference for a given set of characters of the mate, and for the location where the mate is found. Two different things. I said simultaneous preference for two such very different things is impossible. Well, it has now been shown that it occurs very commonly in fishes, particularly cichlid fishes. So here's another thing where my intuition was wrong."

Thus, it is becoming apparent that evolutionary divergence can occur without geographical separation. No matter how convincing the proposed theories would be, empirical evidence on suitable experimental models is mandatory for further progress of this controversial issue (Noor 1999; Korol et al. 2000, 2006a, b; Nevo 2001, 2011; Schluter 2001; Nosil 2008; Schluter et al. 2010, Seehausen et al. 2014).

7.1.2 *Experimental Evolution Perspective: Artificial Ecological Selection and Premating Isolation*

The response to artificial selection for tolerance to various ecological factors may elucidate the genome components essential for adaptation to stress and hint which of the stress-response genes could ensure the evolutionary pathways in nature.

Studies of Huey et al. (1991), James and Partridge (1995), Neat et al. (1995), and Hoffmann and Harshman (1999) strongly support the idea that reproductive isolation develops through pleiotropy and genetic drift, both with and without allopatry. Classical experiments by Thoday (Thoday and Gibson 1962; Thoday 1972) with *D. melanogaster* demonstrated that disruptive selection for morphological quantitative traits (bristle number) within a population could promote sympatric differentiation (positive assortative mating), although numerous attempts (in nearly 20 laboratories) to reproduce these results were unsuccessful. Experiments with *Drosophila* laboratory populations on disruptive selection for a behavioral trait (habitat preference) were more successful, promoting premating sexual isolation despite regulated gene flow (Rice and Salt 1988, 1990). The authors showed that different maintenance conditions may promote partial sexual isolation. The tests conducted on *D. melanogaster* stocks originated from an initial population subjected to different environmental gradients revealed reproductive isolation arisen between the strains. According to their habitat preferences, flies were divided among 24 spatiotemporal habitats related to phototaxis, geotaxis, and chemotaxis. Two strains that had chosen opposite habitats were tested in further experiments for mating preferences. The authors argued in favor of the possibility of sympatric speciation. Divergent selection for morphological traits in some cases also initiated premating sexual isolation regardless of gene flow (reviewed in Rice and Hostert 1993). These results demonstrate that role of geographical isolation in the establishment of sexual preferences was significantly exaggerated in earlier studies.

In the context of the sympatric evolution problem, the attempts to explore reproductive isolation initiated by divergent artificial ecological selection are especially interesting, albeit in most of such experiments, the target populations were maintained under isolation. Dodd (1989) examined eight *D. pseudoobscura* populations reared for one year on starch-based or maltose-based larval medium. Populations raised on different media exhibited assortative mating, with no mating preferences manifested among populations raised on the same medium. Because in this experiment, there was no straight selection for reproductive isolation, the evolved behavioral isolation could be considered as a by-product of adaptation to the two media. Experimental strains of *D. willistoni* adapted to low, intermediate, and high pH substrates were tested for mating preferences. After only several generations, a significant sexual preference was demonstrated in favor of homogamic insemination (de Oliveira and Cordeiro 1980). Kilias et al. (1980) maintained *D. melanogaster* populations for several years in either a cold–dry–dark or warm–damp–light environments and found some degree of sterility and positive assortative mating in crosses between samples raised in different environments. Another series of experiments of the same group showed that selective regimes, even of a short duration (ten generations only), could induce significant adaptive and evolutionary changes: Short-term indirect selection for heat sensitivity and heat tolerance resulted in diverse correlated responses in behavioral, biochemical, and fitness components (Kilias and Alahiotis 1985).

Altogether, experiments with laboratory *Drosophila* populations showed that artificial selection for some quantitative traits and adaptation to diverse ecological

conditions may bring about a tendency of positive assortative mating (incipient differentiation). However, the problem is a lack of confirming evidence in natural populations, in particular if the sympatric scenario is considered. Particularly interesting and controversial is the problem of genetic adaptation to microclimatic heterogeneity.

7.1.3 *Incipient Sympatric Differentiation of* Drosophila *Caused by Natural Microclimatic Gradient*

Two alternative hypotheses of the genetic basis of population differentiation and speciation have been proposed to explain the evolution process. The "gradual divergence model" assumes that adaptation and isolation processes take place due to the accumulation of many additive genes with small effects (Barton and Charlesworth 1984), while the "genetic revolution model" postulates that strong epistatic interactions of fitness genes play the key role (Dobzhansky 1955; Mayr 1963). Analysis of the genetic basis of adaptation processes occurring in nature may be especially successful if a comparison could be done between populations adapted to contrasting conditions, e.g., mild versus stressful environments (Parsons 1993). A convenient opportunity to consider local adaptation is provided by utilization of a natural model, the Lower Nahal Oren canyon (Carmel massif, Israel). This canyon has become an ecological microscale "laboratory" where forces driving adaptation processes of a number of species are being studied within the framework of a large interdisciplinary program (Nevo 2001, 2011, 2012). The opposite slopes of the canyon show strong abiotic contrasts that are consequential for species composition and population genetic structure in diverse organisms, including several *Drosophila* species (Harry et al. 1999). *D. melanogaster* populations on the slopes, separated by 100 and 400 m at the bottom and top, respectively, cope with noticeably different environments, mainly due to the higher insolation on the south-facing slope (SFS) than on the north-facing slope (NFS) (Pavlíček et al. 2003). The slopes also differ in temperature and aridity: NFS has lush vegetation of European origin, while SFS is an open Park Forest or xeric savanna, primarily of African and Asian origin. Material from the canyon auspiciously combines advantages of exhibiting strong differentiation and being of wild origin. Within this system, *Drosophila* is an excellent model for studying the interaction between factors affecting population adaptation and differentiation.

The microtopography of the Nahal Oren canyon permits interslope mixing by easy migration of the flies (Pavlíček et al. 2008). *Drosophila* adults are able to disperse over long distances: 10–15 km overnight (Coyne and Milstead 1987). Therefore, our results describing the existence of slope-specific adaptive complexes that could evolve and escape recombinational collapse were somewhat perplexing and have prompted subsequent investigations. Here, we briefly present the cumulative data on interslope differentiation of *Drosophila* from the canyon.

7.2 The Evidence

Despite easy migration, we found strong interslope *D. melanogaster* population divergence at Nahal Oren canyon involving habitat choice (Nevo et al. 1998), various aspects of induced changes in viability and longevity caused by short-term and lifetime high-temperature treatments (Rashkovetsky et al. 2000, 2006; Lupu et al. 2004), remarkable genetic differentiation of microsatellites (Michalak et al. 2001), Single nucleotide polymorphisms (SNPs) (Hübner et al. 2013), transposable elements (TEs) (Kim et al. 2014), divergence in the regulatory region of *hsp70Ba* which encodes the major inducible heat-shock protein of *Drosophila* (Michalak et al. 2001), sequence polymorphism of candidate genes (Zamorzaeva et al. 2005, 2009), and expression of small heat-shock protein genes (Carmel et al. 2011). The most exciting findings were related to sexual behavior: positive assortative mating and interslope differences in mating propensity, sexual discrimination, reproductive activity, and peculiarities in courtship song patterns (Korol et al. 2000, 2006a, b; Iliadi et al. 2001, 2009; Drake et al. 2005; Singh et al. 2005). Parallel patterns of habitat choice, stress tolerance, and mate choice were also demonstrated in *D. simulans* (Nevo et al. 1998; Singh et al. 2005). The obtained evidence suggests that these populations may represent an example of ongoing divergence taking place regardless of gene flow. However, some tests for interslope genetic differentiation in *Drosophila* gave somewhat conflicting results (Colson 2002; Shlötterer and Agis 2002; Panhuis et al. 2003; Gefen and Brendzel 2011). A possible explanation could be that adaptive differentiation can withstand destructive effects of interslope migration, but it should not necessarily be accompanied by differentiation for selectively neutral markers, unless the latter will be in linkage disequilibrium (LD) with selected loci. Such LD can persist despite migration, but only under very tight linkage and strong selection, bearing in mind generally very steep decay of LD in *D. melanogaster*, sometimes at a distance of a few tens of nucleotides (Mackay et al. 2012). Thus, differentiation for adaptive trait complexes and relevant candidate genes seems to be a better evidence for interslope divergent selection than that displayed by genetic distances estimated using sparsely distributed molecular markers.

We addressed the issue of sympatric differentiation by using a natural model—fruit flies inhabiting a highly heterogeneous microsite—Nahal Oren canyon. Various approaches were employed to characterize the population structure. The accumulated evidence allowed us to suggest that strong interslope microclimatic contrasts could invoke differential selection for stress-related gene complexes, resulting in genetic divergence for stress tolerance traits, accompanied by behavioral elements contributing to the incipient separation of the gene pools.

7.2.1 Migration, Habitat Choice, and Fitness-Related Traits

The analyzed multivariate fitness complex included oviposition temperature preferences, viability, longevity changes caused by short-term and lifetime temperature

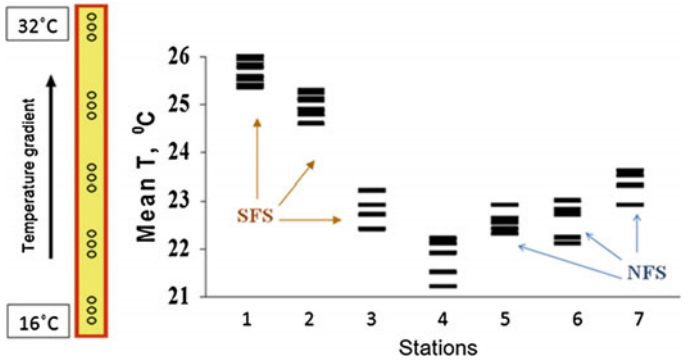


Fig. 7.2 Estimation of preferred oviposition temperature as a proxy for habitat choice testing in flies from the Nahal Oren canyon (modified from Nevo et al. 1998). Flies of isofemale lines from the seven stations across the canyon (*1–3* on the SFS, *5–7* on the NFS, and *4* at the valley bottom) were maintained for 40 h in a tube with a temperature gradient; then, the size of the progeny collected from the vials along the gradient was estimated resulting in average preferred oviposition temperature for each of the tested lines (present by *horizontal bars*)

treatments, and resistance to drought stress at different temperatures (Nevo et al. 1998; Rashkovetsky et al. 2000; Korol et al. 2006a, b). In a large-scale laboratory experiment, including ~50,000 *D. melanogaster* and ~14,000 *D. simulans* flies, we found that flies originating from the opposite slopes displayed clear differences in habitat preferences, including a significantly higher preferred oviposition temperature by females from the SFS and a general increase in the preferred laying temperatures with increased elevation (Fig. 7.2). The revealed differentiation in the preferred oviposition temperature of flies, raised over 1.5 years in standard laboratory conditions, indicates the existence of a genetic basis of this behavioral adaptation to the microclimatic slope-specific environments.

Drosophila adults are very sensitive to desiccation, and survival in a dry and hot environment determines their geographical and microhabitat distribution (Parsons 1983; Hoffmann and Parsons 1993). In our studies, tolerance to the combined desiccation/starvation treatment was significantly higher in SFS flies (Nevo et al. 1998; Korol et al. 2006a; Pechkovsky 2008). SFS flies manifested a much higher stability when the reaction to desiccation was measured under elevated temperature or the reaction to heat treatment was measured under drought stress. Rapid loss of tolerance to desiccation stresses is known to occur throughout adaptation to laboratory conditions (Hoffmann et al. 2001). In our tests, long-term maintenance of flies under standard laboratory conditions has not abolished the differences in reaction to drought treatments implying that the observed dissimilarities reflect genetic adaptations to interslope contrasts.

The interpretation of the interslope divergence strongly depends on whether migration is common between the slopes, or by contrast, if differential selection has resulted in a genetically determined reduction in the migration rate as was predicted by some theoretical models (Wiener and Feldman 1993). The distance between the

slopes is very short; thus, a high migration rate is expected. To evaluate the migration rate, we used the capture–mark–release–recapture method to mark wild adults with a UV fluorescent powder, release them, and after recapture determine the extent to which individuals captured on one slope remain on that slope or visit the opposite one (Pavlíček et al. 2008). Flies from the canyon display substantial, asymmetric interslope migration. The proportion of migrants from the SFS to the NFS was estimated as 9–10 % and from the NFS to the SFS, 1–1.5 %. We also compared the migratory activity between flies from the Nahal Oren canyon and flies collected from an open Forest Park population (northern part of Israel) and flies from the Negev desert. No differences in migratory activity were revealed, suggesting that flies from the canyon are not atypical migrators (Iliadi et al. 2002). Thus, our results indicate that despite the relatively high level of interslope migration, the SFS and NFS populations at Nahal Oren canyon cannot be considered panmictic. The high estimates of migration rates, together with the results on interslope adaptive divergence, suggest that strong microclimatic natural selection may override (presumably together with other behavioral mechanisms) the effect of migration. Knowledge on *Drosophila* interslope migration was of crucial importance for addressing the major question concerning population processes in the canyon: Why interslope migration does not swamp differential adaptation to the contrasting microclimatic regimes on the opposite slopes?

The very first tests revealed interslope divergence for a complex of stress tolerance and adaptive behavioral traits (Nevo et al. 1998). The examined fitness-related traits included viability and longevity reaction to short-term and lifetime temperature treatments. Thus, SFS flies displayed a lower sensitivity of longevity reduction to heat treatment compared to NFS flies, corroborating well with the conclusions reached in artificial selection experiments where selection for stress tolerance increased longevity (Rose et al. 1992; Parsons 1993). Lifespan extension in *D. melanogaster* prompted by hormesis (repeated moderate heat stress) was described by Hercus et al. (2003). Some physiological modifications and/or short-term response to selection, presumably higher in SFS flies (due to their higher genetic heterogeneity caused by more fluctuating environments), could be a source of increased thermotolerance of the SFS flies (Rashkovetsky et al. 2006). These findings initiated further studies of *Drosophila* populations from the canyon. Significant interslope differences were found for the variance in developmental time: At both normal and elevated temperatures, variation among SFS flies was higher than that of the NFS (Korol et al. 2006a). The prolongated developmental period and higher interindividual variation among SFS flies may display an adaptive strategy, reducing the chance that the entire progeny of a female will be eliminated due to a sudden severe stress at a critical period of development. The revealed specific trait complex (increased fecundity, thermotolerance, prolongated pupation, and a higher between-individual variation in the time of development) suggests a more secure survival of flies inhabiting the more climatically variable and stressful SFS.

Extensive experiments tested whether the revealed differences in thermotolerance are lasting over years. Consequent tests included flies sampled in the canyon in

1997–2004 (Lupu et al. 2004; Rashkovetsky et al. 2006; Carmel et al. 2011). The outcomes were in line with our previous studies in many aspects: an inverse correlation between survivorship and heat-shock temperature and male–female differences in basal and inducible thermotolerance. Year of collection, slope, sex, duration of heat-shock temperature, and pretreatment type all affected survival. In the majority of the comparisons, SFS flies exceeded NFS flies in basal and inducible thermotolerance. The correspondence of the results obtained using fresh samples and laboratory stocks after dozens of generations maintained in the laboratory suggests that interslope differences in thermotolerance are ongoing and are genetically based.

7.2.2 Non-random Mating

The revealed interslope divergence for fitness-related traits in the presence of interslope migration suggests the existence of some behavioral mechanisms preventing recombinational reshuffling of the slope-specific trait complexes. And indeed, as shown above, flies from the opposite slopes exhibited clear differences in habitat preferences (Nevo et al. 1998). Another potential factor opposing gene flow could be positive assortative mating. Our result obtained in sexual behavior experiments suggests that population of *D. melanogaster* from the canyon cannot be equivocally referred to as panmictic. Indeed, we found indication of positive assortative mating (Korol et al. 2000; Singh et al. 2005), interslope differences in various elements of sexual behavior (Iliadi et al. 2001), and slope specificity in the courtship song pattern (Iliadi et al. 2009).

7.2.2.1 Mating Preferences

Usual schemes of mate-choice experiments (multiple- and single-choice variants) were used to look for signs of sexual isolation between the populations derived from the opposite slopes. A significant preference of sexual partners originating from the same slope was detected in tests with slope-specific synthetic populations (Korol et al. 2000). Results of single-choice tests corroborated those of multiple-choice tests, although deviation from the expected (50 %) level was not that high. NFS females manifested significant positive assortative mating. The results suggest that despite the extremely small distance between the slopes, interslope ecological contrasts resulted in population differentiation accompanied by certain premating sexual isolation system. However, some discrepant data on non-random mating in populations derived from the Nahal Oren canyon were obtained by Panhuis et al. (2003) and Gefen and Brendzel (2011), inspiring further investigations. Peculiarities in the procedures that could severely affect the results were considered in detail in our study (Korol et al. 2006).

Our group also described some mixed results. Mate-choice experiments, conducted in parallel under a variety of protocols in two laboratories (Simon Fraser University, Burnaby, Canada, and University of Haifa, Israel), revealed that both NFS and SFS females exhibited a significant excess of homotypic mating in four out of six single-choice trials, suggesting that populations from the opposite slopes are behaviorally differentiated in some manner (Drake et al. 2005). But this study failed to detect assortative mating in double-choice tests. Significant positive assortative mating was found in further tests undertaken with fresh collections of *D. melanogaster* and *D. simulans* originated from the opposite slopes of the canyon (Singh et al. 2005). Results of individual crosses (involving nearly 47,000 flies) were analyzed to assess effects of slope, season, and sex. Significant deviation from random mating was revealed in both female-choice and male-choice tests in 152 out of 200 tests with lines collected in spring and fall. The dominating pattern was a significant excess of pairs formed by partners of the same slope. Higher between-line heterogeneity was found for SFS flies. The results obtained in *D. simulans* study correspond well with the previous findings on *D. melanogaster* (Korol et al. 2000; Iliadi et al. 2001) and suggest that differential ecological selection on a microsite scale can induce development of partial premating isolation promoting evolutionary changes.

7.2.2.2 Peculiarities of Sexual and Reproductive Behavior

In *Drosophila*, mating behavior includes a sequence of stimuli produced by males during the courtship and females' responsive reactions. The sequence of elements is usually typical and presents a courtship ritual. Even a slight deviation from the specific pattern of sexual behavior causes a serious decrease in reproductive success (Gleason and Ritchie 1998). Therefore, findings on mate choice in flies inhabiting opposite slopes of the canyon motivated us to analyze the "anatomy" of non-random mating. We studied sexual and reproductive behaviors in a non-choice situation for all possible inter- and intraslope mating combinations (Iliadi et al. 2001). The mating propensity of males was monitored based on the courtship latency and duration of copulation: Generally, SFS males showed higher sexual activity. SFS females received more stimulation from SFS males and demonstrated higher receptivity with SFS males; NFS females did not show differences in receptivity toward males from either slope. Homotypic SFS pairs mated much faster than all other combinations, and the lowest level of mating success was observed in the "NFS female × SFS male" combination. These results most likely reflect effects of the flies' origin. Differences in time of pair formation and specific characters of female courtship behavior suggest that SFS females are more receptive and less discriminating than NFS females. The results of sexual and reproductive behavior in no-choice situation analysis point to two issues: mate choice derives from differences in mating propensity and discrimination; females from the microclimatically milder NFS slope discriminate against males of the SFS slope.

The pattern of courtship song in *Drosophila* plays an essential role in mating success. *Pulse* and *sine* components of the signal, produced by wing vibration during courtship, are considered the main factors affecting female receptivity (Kyriacou and Hall 1984; Ritchie et al. 1999). In our study, significant interslope dissimilarities were observed in flies' courtship song: differences in the IPI between SFS and NFS males courting females of their own origin and a decrease in the IPI of SFS males exposed to females from the opposite slope (Iliadi et al. 2009). SFS males also displayed a remarkable experience-dependent plasticity of the IPI signal. Namely, they modulated IPI in such a manner that it became different from their intrinsic signal, but closer to the IPI typical for NFS males! This plasticity is particularly interesting in light of the accepted view that IPI is a *species-specific feature displaying low phenotypic and genetic variability* (Ewing 1983). These results, characterizing the intensity of stimulation, coincide well with our findings on flies' behavior, in which SFS males manifested higher mating propensity (Iliadi et al. 2001).

The estimation of the relationships between dynamics of egg laying, fecundity, and repeated mating indicated that neither female nor male origin, nor their interaction, had effect on fecundity. SFS females, characterized by increased egg laying speed, also displayed the shortest remating time. The remating time for NFS females was twice that of SFS females, while the origin of males was not important. These peculiarities of sexual and reproductive behaviors could contribute to asymmetric pattern of sexual isolation. The discovered peculiarities of sexual behavior can be considered a part of population adaptive complex arising due to strong spatial variation of ecological conditions on the slopes separated by a very short distance.

7.2.3 Genetic Differentiation

7.2.3.1 Microsatellite Markers

The very first reports of *Drosophila* interslope differentiation in the canyon have generated extensive debate. If the populations on the slopes are indeed distinct, then at least some of their genes should be somehow diverged in sequence. Microsatellite markers proved to be rather useful for evaluation of genetic diversity because of abundant polymorphism and high level of allelic variation, distribution throughout the genome (in both coding and non-coding regions), and codominant inheritance. It is considered that microsatellites may provide a molecular basis for fast adaptation to environmental changes (Li et al. 2004). Primarily genetic changes associated with the local adaptation in the canyon were examined using microsatellite markers (Michalak et al. 2001). Indeed, a very strong interslope divergence was found in this test. The differentiation was as great as between *D. melanogaster* and its sibling species, *D. simulans*. The analysis of microsatellites suggests a limited exchange of migrants and lack of recent population bottlenecks. Despite a

small number of total alleles and low heterozygosity, many of the alleles in both the NFS and SFS populations were private. Microsatellite variability depended on chromosome location, and the number of alleles was correlated with the recombination rate per DNA unit length. In another study with microsatellites (Colson 2002), a much lower interslope genetic differentiation was found for both *D. melanogaster* and *D. simulans*. Similarly, using 48 microsatellite markers, Schlötterer and Agis (2002) reported a low-level interslope genetic differentiation for *D. simulans*. The above results show both large and small degrees of differentiation. Tests for interslope divergence were of low reproducibility when neutral markers were employed, whereas adaptive traits and relevant candidate genes (see below) showed quite consistent patterns among collection years, seasons, and traits. Obviously, the assumed ecological selection driving interslope divergence should not obligatory be accompanied by differentiation of selectively neutral and randomly chosen markers (unless the latter are in LD with loci subjected to differential selection). Under the tight linkage conditions and strong selection, this situation can persist despite migration.

We also cannot exclude from consideration the possibility of seasonal repopulation of the canyon and probable consequences of this phenomenon (i.e., boosting of short-term *Ne*, facilitating recurrent local adaptation complexes and interslope divergence). These factors can be a reasonable cause of the inconsistencies in microsatellite variation and mating discrimination behavior observed in population studies. Apparently, in the "pregenomics" studies of microsatellite variation, we were limited by the number of available marker loci per genome, while in our recent studies, ~65,000 microsatellite loci were used to estimate F_{st} values (Kim et al. 2014). Similarly, hundred thousands of SNPs, both in coding and in non-coding DNA, were involved in the estimation of F_{st} (Hübner et al. 2013). Contradictions and inconsistencies in results, obtained by means of microsatellite markers, highlighted the necessity of further clarification and motivated the changeover from marker loci to relevant candidate genes (genes, presumably involved in selection-driven population divergence of adaptive trait complexes) and studies of sequence polymorphism at the genome-wide level. Such approach should provide much more relevant evidence on the real situation in the targeted natural system.

7.2.3.2 Candidate Genes

To evaluate microgeographical variation of candidate genes, we studied sequence polymorphism of genes related to sexual behavior and known to include polymorphic repeated sequences, indels, or nucleotide substitutions (*period*, *desaturase2*, *no-on-transient A*, etc.) in *D. melanogaster* from the canyon (Zamorzaeva et al. 2005, 2009). These genes proved polymorphic in different geographical regions or environmental conditions. The idea was that polymorphism of the chosen genes might affect behavioral peculiarities in flies living on the opposite slopes. Indeed, interslope differences in the *period* sequence were established. This gene encodes for a transcription cofactor involved in circadian rhythms and is known to

affect also sexual behavior. Variants of the 5th exon repeat encoding for $(Thr–Gly)_n$, $n = 17$ and $n = 20$ that are abundant in natural populations of *D. melanogaster* in Africa and Europe (Kyriacou et al. 1996; Sawyer et al. 1997) were found to predominate in the canyon. The remarkable fact is that the less abundant "European" variant ($n = 20$) was 2.6-fold more frequent on the NFS compared to the SFS. This high proportion may reflect some advantages of the $n = 20$ allele to flies inhabiting the wetter and less warm NFS. A series of female-choice tests showed that NFS females distinguish between males with specific *per* alleles, as well as between males originated from the opposing slopes. SFS females were less discriminating and did not manifest deviation from random mating.

Desat2 gene, responsible for the female cuticular hydrocarbon (CH) synthesis, plays an important role in mate choice (Coyne et al. 1994). Two major isomers of this hydrocarbon, 5, 9-HD and 7, 11-HD, are known, and two types of ratios of their frequencies were established in different geographical populations of *D. melanogaster*: "high" CH (African) type of population having a high ratio (5, 9-HD)/(7, 11-HD) and "low" CH (non-African, or European) type having a low ratio (5, 9-HD)/(7, 11-HD) (Ferveur et al. 1996; Takahashi et al. 2001). CH polymorphism was caused by a 16-bp deletion in the promoter region of *desat2* (Dallerac et al. 2000), which led to the appearance of "low" CH type with inactivated *desat2* expression (Fang et al. 2002). Wu et al. (1995) showed that females of many *D. melanogaster* lines from Africa tend to escape mating with males from other continents. In our analysis of *desaturase2*, all tested Israeli populations of *D. melanogaster* belong to the "low" type (low ratio of 5, 9-/7, 11-heptacosadiene) with a 16-bp deletion in the promoter region (Zamorzaeva et al. 2005). An additional deletion leading to the appearance of a stop codon was found in exon 1 in populations inhabiting opposite slopes of the canyon (Zamorzaeva et al. 2009). To estimate the putative effect of *desat2* status on courtship behavior, we performed a series of mating experiments between carriers of various alleles. The obtained results suggest that the allele with the additional deletion plays certain role in mating success, expressed in a shorter courtship latency and courtship duration. The appearance and long-term (for years) maintenance of this mutant allele in populations inhabiting Nahal Oren canyon may reflect flies' adaptation to peculiar microscale conditions and may be associated with incipient sympatric differentiation.

Interslope divergence in the regulatory region of *Hsp70Ba* gene, which encodes for the major inducible heat-shock protein of *Drosophila*, was found in *D. melanogaster* from the canyon (Michalak et al. 2001; Lerman et al. 2003). Populations in the canyon proved polymorphic for a 1.2-kb *P* element insertion in the *Hsp70Ba* promoter, 28 times more frequent in NFS than in SFS, reducing *Hsp70* expression by about 22 %. Elevated level of *Hsp70* is considered beneficial for inducible thermotolerance, but relatively deleterious for growth and development (Krebs and Feder 1998). This is consistent with our data on the pattern of increased thermotolerance and slower development of SFS-inhabiting flies (Rashkovetsky et al. 2000, 2006; Korol et al. 2006). Again, reduced *Hsp70* levels might be valuable in the absence of repeated extreme stress on the NFS.

Although single candidate gene surveys proved their value in population genetics, genome-wide approaches based on NGS hold promise of new insights into population differentiation (Kolaczkowski et al. 2011; Fabian et al. 2012; Boitard et al. 2012; Pool et al. 2012). We complemented the extensive phenotypic and candidate gene-oriented surveys of Nahal Oren canyon flies with deep Illumina sequencing of entire genomes (see Sects. 7.2.3.4 and 7.2.3.5).

7.2.3.3 Expression of Heat-Shock-Related Candidate Genes

Determining the patterns of expression of genes involved in stress response is among the most relevant targets for microscale comparisons. In our study, tests were focused on heat-shock protein (*Hsp*) genes. The increased synthesis of heat-shock proteins was described as a cellular response to various stresses including elevated temperatures (review Lindquist 1986). Expression of *Hsp* genes is regulated mostly by heat-shock transcription factor gene (*Hsf*) (Pirkkala et al. 2001). Therefore, induced change in transcription of *Hsf* is of special interest in the context of interslope gradient of temperature regime. In our Northern blot tests, the *Hsf* expression level was higher in SFS flies and correlated with survival in the compared populations (SFS vs. NFS) under heat treatments (Baca et al. unpublished results; Fig. 7.3). The higher level of increase in *Hsf* expression in SFS flies as a

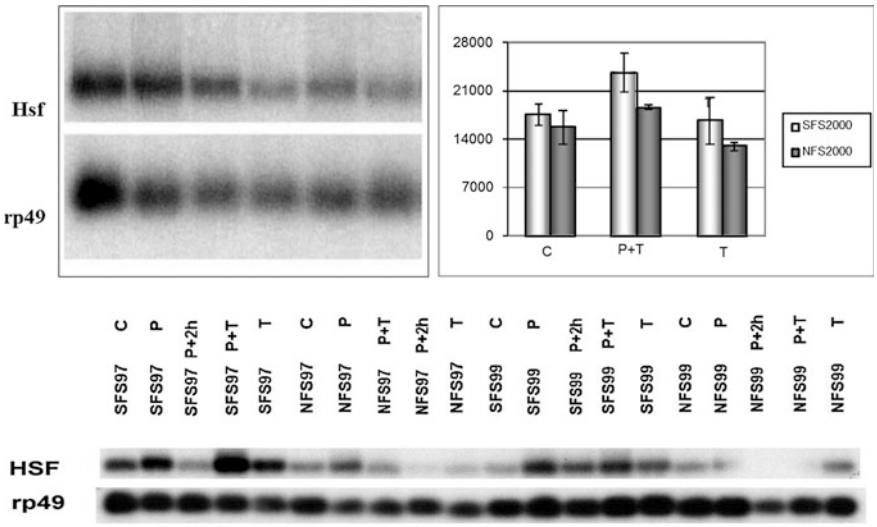


Fig. 7.3 Northern blot analysis of *Hsf* mRNA (*D. melanogaster* flies from Nahal Oren canyon, collections 1997, 1999, and 2000) *C*—control, 25 °C; *P*—pretreatment (36 °C, 1 h); *P + 2h*—pretreatment (36 °C, 1 h) followed by recovery (25 °C, 2 h); *T*—treatment (38.5 °C, 50 min); *P + T*—pretreatment (36 °C, 1 h) followed by recovery (25 °C, 1 h) and severe shock treatment (38.5 °C, 50 min); *rp49* represents RNA loading control. Calculation of *Hsf* mRNA levels in control and treatments was provided relative to *rp49*

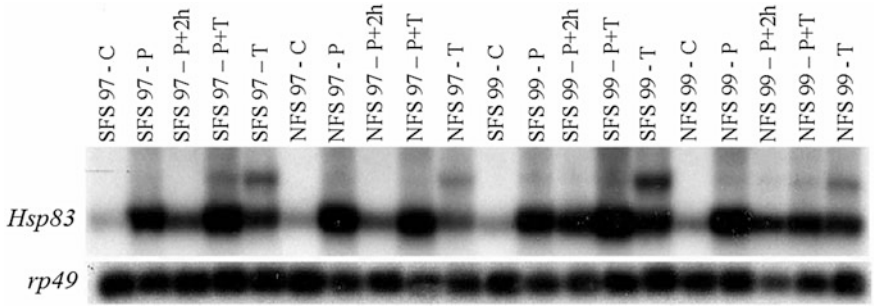


Fig. 7.4 Slope-specific reaction of *Hsp83* gene expression to heat treatments *C* (control)—25 °C; *P* (pretreatment)—36 °C; *T* (severe treatment)—38.5 °C; *P + 2h*—pretreatment followed by 2 h at 25°; *P + T*—pretreatment followed by treatment; *rp49* represents RNA loading control

reaction to severe heat shock or combined heat stress (pretreatment followed by severe heat shock) may contribute to increased basal and acquired thermotolerance, respectively, revealed in SFS populations (Rashkovetsky et al. 2006).

We also tested whether interslope differences in *Drosophila* thermoadaptation are associated with differential expression of small *Hsp* genes (*Hsp23* and *Hsp40*) and heat-shock gene *Hsr-omega* (Carmel et al. 2011). In joint scoring of acquired thermotolerance and expression of *Hsp* genes, higher expression displayed by SFS compared to NFS flies was positively correlated with survivorship under stress. Interesting interslope differences were found in the dynamics of *Hsp83* gene expression after moderate, severe, or combined heat-shock treatments. Expression of the *Hsp83* was higher in NFS flies at moderately elevated temperature (36 °C), but considerably decreased under subsequent more severe treatment. In contrast, the expression of the *Hsp 83* in SFS flies greatly increased when exposed to severe heat shock (Baca et al. unpublished results; Fig. 7.4).

7.2.3.4 Genetic Differentiation at the Genome-Wide Level

Because the revealed divergence between *D. melanogaster* populations inhabiting the opposite slopes persists despite verified high migration rate, this differentiation pattern suggests an idea of a tight linkage and strong selection. To characterize patterns of genetic differentiation at the genome-wide level, we used high-coverage whole-genome sequencing (Hübner et al. 2013). SNP was tested in *D. melanogaster* isofemale lines established from females inseminated in nature and collected on the opposite slopes of the canyon in 2010. Equal amounts of DNA from each line were pooled and sequenced with HiSeq, at $\sim 40\times$ coverage per population. SNPs showed a "non-neutral" pattern upon comparisons of interslope differentiation for different gene regions, as well as density of SNPs significantly differentiated

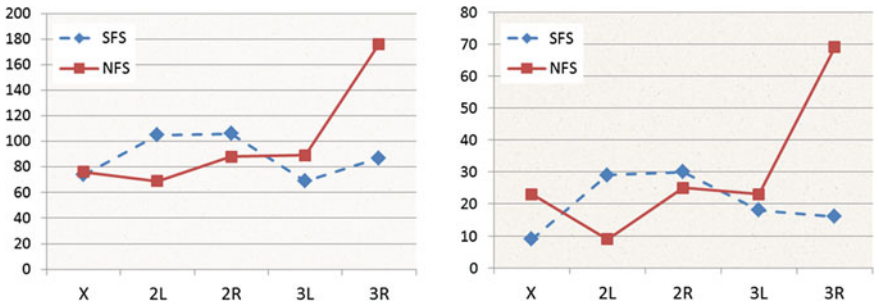


Fig. 7.5 Interslope difference in the distribution of fixed alleles in coding DNA (for genes in the large chromosomes). *Left*—number of genes with SNP alleles fixed on one or both slopes (p-value 2.7×10^{-8}); *Right*—number of genes with different SNP alleles of the same genes fixed on one slope (p-value 1.6×10^{-6})

between the slopes. More than 930 slope-specific SNPs were found, out of which about 250 were fixed on one of the slopes (Fig. 7.5). To test whether significantly differentiated SNPs are evenly distributed along the chromosomes or concentrated in a relatively small number of chromosomal regions, as expected for sympatric populations subjected to differentiating selection (Turner et al. 2005; Michel et al. 2010; Bradbury et al. 2013), we compared the regional distribution of SNPs assigned by the hidden Markov model (HMM) analysis to the highly differentiated state with the genomic distribution of all SNPs polymorphic in the population. Then, we used a finer scale, which enabled us to reveal genomic regions with significantly higher interslope differentiation. There were at least 37 chromosomal "islands" of interslope divergence and low sequence polymorphism, plausible signatures of selective sweeps, more abundant in flies derived from one (north-facing) of the slopes. Chromosomal arm 3R seemed to be the most strongly differentiated region of the genome. A closer examination of genes falling into the three most significant intervals of 3R revealed a number of genes contributing to important biological processes that could be responsible for interslope differentiation. Although introns had the highest density of SNPs across all chromosomes, coding sequences had the highest contribution to genome differentiation between NFS and SFS, suggesting that selection rather than drift might be in play.

A total of 572 genes were significantly different in SNP allele frequencies between the slopes, 106 out of which were associated with 74 significantly over-represented gene ontology (GO) terms, particularly so with response to stimulus and developmental and reproductive processes, thus corroborating previous observations of interslope divergence in stress response, life history, and mating functions. We demonstrate that interslope genetic changes in this species accumulate in a number of chromosomal differentiation "islands" and that gene networks related to adaptive responses and reproductive processes are thus significantly affected.

7.2.3.5 Divergence of Repeatome

Repeat sequences are abundant in eukaryotic genomes and provide enormous evolutionary potential. In *Drosophila*, studies of repeatomes—TEs and tandem array repeats (satellites)—suggested variability between individuals, and application of the next-generation high-throughput sequencing technologies can bring more clarity and certainty. TEs have been implicated in adaptations to environmental changes either through direct remobilization after stress (Capy et al. 2000) or more indirectly as a source of new and preexisting genetic variation (González et al. 2010). Flies derived from the opposing slopes exhibit a significant difference in the contents and distribution of mobile elements, as well as microsatellite allele frequencies (Kim et al. 2014). We identified a total of 14,190 and 15,025 TEs in SFS and NFS populations, respectively. A total of 48 and 51 % TE insertions were positioned uniquely in SFS and NFS populations, respectively, demonstrating the omnipresence of TE-induced polymorphism. In total, 363 TEs in SFS and 518 TEs in NFS were slope unique and fixed (frequency > 95 %). After removing less than 10-read coverage sites and overlapping TE insertions, 9877 and 10,926 TE insertions were identified in SFS and NFS, respectively. Of these filtered TE insertions, 2174 (22 %) and 2452 (22 %) were fixed in SFS and NFS, respectively.

There were at least 20 significant GO term enrichments representing genes with slope-unique TE insertions in their putative promoters, with functional activities related to hydrolases and alternative splicing. Four hundred and twenty-six genes were disrupted by 511 slope-specific TE insertions within coding sequences, presumably leading to the genes' functional inactivation. Cognition, sensory perception of chemical stimuli, and olfaction were among the most significantly overrepresented GO terms among genes with TE-disrupted coding sequences. This slope-divergent transposition of mobile elements may contribute to courtship changes and mating isolation.

Inhabitants of the colder and more humid NFS carried about 6 % more TEs than those from the hot and dry SFS; nearly 50 % of all mobile element insertions were slope unique, with many of them disrupting coding sequences of genes critical for cognition, olfaction, and thermotolerance. These data are consistent with the previously observed phenotypic patterns of stress resistance (thermotolerance) differences and assortative mating in the system. Our results show substantial interslope divergence in repeat sequences, in parallel with differentiation of coding sequences (Hübner et al. 2013). We find no evidence that microsatellites contribute to local adaptations in the canyon, though TEs emerge as potential "players" of adaptive divergence along the sharp microclimate gradient. Purifying selection against TEs or positive selection favoring particular TE insertions may have operated differently on NFS and SFS, dependent on local conditions of the contrasting slopes.

7.3 Interpretation

Our working hypothesis was that strong interslope microclimatic contrasts may cause differential selection for stress-related gene complexes, promoting interslope genetic divergence accompanied/facilitated by evolving behavioral divergence (habitat choice and positive assortative mapping) (Korol et al. 2000, 2006). To assess the efficiency of the proposed mechanism of differentiation in sympatry as a proxy for the situation with *Drosophila* populations in Nahal Oren canyon, we developed and tested a simulation model of a finite diploid population with non-overlapping generations. It is analogous to our model on niche viability selection combined with niche choice and mate choice (see Fig. 7.1) proposed to explain a complex pattern observed in Israeli populations of mole crickets that included high multilocus polymorphism, paucity of heterozygotes, and LD (Nevo et al. 2000). In the new model (Fig. 7.6), individual's genotype is defined by allele states at L diallelic loci consequently situated in a single chromosome. Recombination events in different intervals are independent (no interference), and recombination rates $r_{l,l+1}$ between consequent loci M_l/m_l and M_{l+1}/m_{l+1}, $l = 1,\ldots, L-1$, are the same for all individuals, although the model can easily be extended to include polymorphic modifiers of recombination. Each individual inhabits one of two ecological states, $s = 1$ or 2, corresponding to SFS and NFS. Before selection and mating, individuals may migrate from the state of origin to the other state with genotype-independent probabilities μ_{12} and μ_{21} (from state 1 to state 2 and from state 2 to state 1, respectively). Selection in state s is simulated by choosing N_s survivors out of all non-migrated individuals originated in state s and immigrants from the other state. The probability to survive is modeled as proportional to fitness of the genotype. Viability selection is simulated as a Gaussian function of a single trait controlled by several trait loci (out of the considered L loci of the genotype). In each selected trait locus l, additive and dominant effects d_l and h_l of alleles on the trait value are independent of the environmental state, while optimum trait value T_s and the strength of selection σ_s^2 are state specific. Each individual produces R children as a "female" (if it finds a pair to mate) and can produce any number of children as a "male." To produce children as a female, each individual inhabiting state s (no matter whether it was born here or immigrated before the selection stage) searches for a mating partner from the same state; such a selected individual is

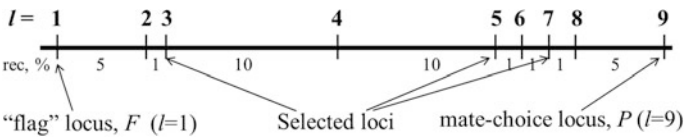


Fig. 7.6 Position of loci in the chromosome of the simulated examples. The number of loci is $L = 9$; loci 3, 5, and 7 define the selected trait value, while loci 2, 4, 6, and 8 are selectively neutral. Locus P ($l = 9$) defines the level of females' assortativeness in mating; flag locus F ($l = 1$) defines the similarity of female and male genotypes affecting mating probability

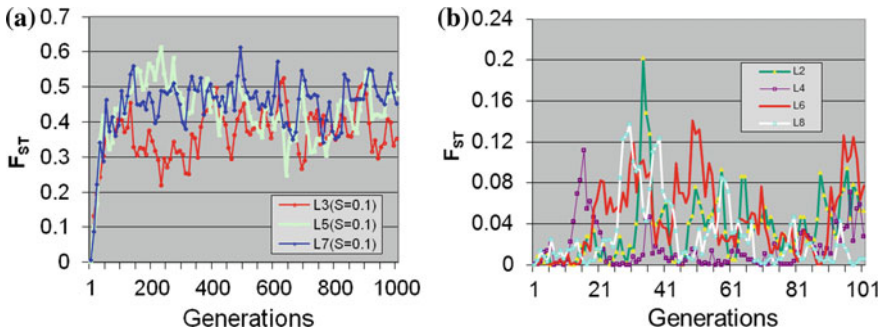


Fig. 7.7 Divergence of sympatric populations with two ecological states. **a** Selected trait loci 3, 5, and 7; **b** selectively neutral loci 2, 4, 6, and 8. One can see that in population with rather liberal selection and high migration rate ($\mu_{12} = \mu_{21} = 0.1$), differentiation of selectively neutral markers is not conserved with time, while differentiation on each of selected loci quite quickly arises and tends to be maintained with time

considered as a male in this mating. The probability to be selected as a male for mating depends on (a) the level of assortativeness in mating that is defined by female's genotype at specific mate-choice locus M_P/m_P and (b) similarity of genotypes of female and candidate male at "flag" locus M_F/m_F (Fig. 7.7). In fact, for each individual, we consequently randomly select K candidates for mating. For each candidate male, we simulate the female's decision "to mate or not to mate" with a probability dependent on M_P/m_P and M_M/m_M genotypes. If her decision is negative, then the next candidate is randomly selected. After K negative decisions, the female is considered as remained unmated and produces no children. The described model enables simulation of evolution of differentiation upon joint effect of differential ecological selection, drift, and assortative mating. It also enables to study evolution of level of assortative mating up to complete premating isolation.

The simulations start from non-differentiated population subdivided between states 1 and 2. Under proper choice of the model parameters, as typical outcome, we could observe divergence of population for viability-selected loci and sexually selected flag locus F (Figs. 7.7 and 7.8). Divergence at selected loci was directly caused by differential selection: low optimal trait value T_1 for state $s = 1$ and high optimal trait value T_2 for state $s = 2$ leading to excess of alleles m_3, m_5, *and* m_7 (at selective loci 3, 5, and 7) in state 1 and M_3, M_5, *and* M_7 in state 2. With time, the differentially selected genotypes become associated by random drift with alternative alleles of flag locus F (M_1 in state 1 and m_1 in state 2, Fig. 7.8). Correspondingly, partial assortative mating evolves, expressed in mate preference by state-selected females (hence better adapted to their state) with males having similar/same genotype in flag locus 1. Hence, on average, males from the same ecological state have flag-induced mating preference. This preference increases mean fitness of the resulted offspring: Individuals with higher level of positive mating assortativeness, on average, produce offspring with higher fitness. Higher fitness of such offspring leads to increasing (in both ecological states) of the frequency of allele M_7 causing

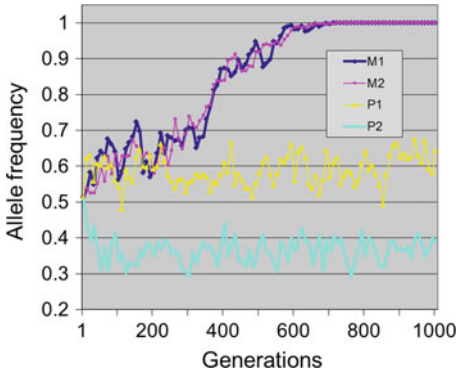


Fig. 7.8 Selection-driven incipient premating isolation between sympatric populations. Coevolution of a sexually selected flag locus F (l = 1) and a modifier of mate choice P (l = 9) (see the scheme in Fig. 7.6). Allele increasing the level of assortativeness in mating invades both subpopulations corresponding to the two simulated ecological states (s = 1 and 2). This differentiation is caused by state-specific viability selection accompanied with differentiation on flag locus F

higher level of positive mating assortativeness (see Fig. 7.8). Interestingly, differentiation at selective and flag loci supported by increased level of mating assortativeness does not induce stable differentiation at selectively neutral marker loci (2, 4, 6, and 8) linked to the selected loci (see Fig. 7.7a). To obtain differentiation at these markers, extremely strong selection ($\sigma^2 \ll 1$) with absolute preference for alleles m_3, m_5, and m_7 in state 1 and for M_3, M_5, and M_7 at state 2 was needed (not shown). With except of cases with fixation, such differentiation is not stable and becomes disrupted by drift.

The accumulated evidence on *Drosophila* from the Nahal Oren canyon suggests that even partial isolation between the populations from the opposite slopes is important in preserving the evolving adaptive complexes from recombinational breakdown. The comparison of fitness-related traits revealed interslope differences in other parameters: higher thermotolerance, prolongated pupation, and higher level of between-individual variation in the time of development in SFS flies, suggesting a more secure survival under less predictable conditions (Fig. 7.9; Rashkovetsky et al. 2000, 2006). These findings probably reflect differences in the adaptive life strategy of flies inhabiting ecologically contrasting slopes. The demonstrated indications of population divergence between the canyon slopes are not a trivial matter, bearing in mind the very short interslope distance and verified migration (Pavlicek et al. 2008). Our analysis confirmed that the described pattern either persisted or re-evolved over multiple years.

We examined patterns of genetic differentiation at the genome-wide level (Hübner et al. 2013; Kim et al. 2014) to identify the loci pertaining to divergence and/or partial sexual isolation. Genome analysis reveals a number of chromosomal regions of interslope divergence and low sequence polymorphism, suggestive of

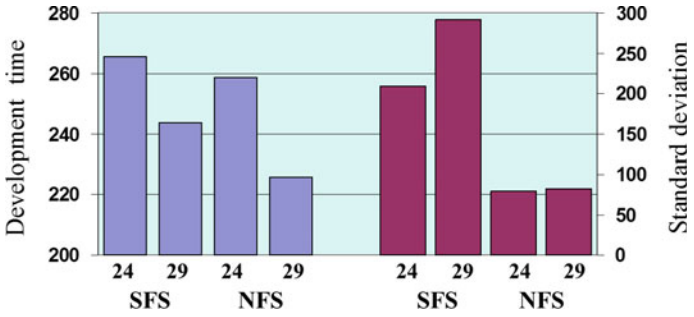


Fig. 7.9 Differential effect of temperature on development time (in hours) of flies from Nahal Oren canyon

selective sweeps. The list of corresponding affected genes was enriched for functions related to stimulus responses, developmental and reproductive processes, remarkably well suited with our previous findings on the phenotypic patterns of stress responses, life history, and mating functions. It seems that the microclimatic contrast in the canyon may create divergent or disruptive selection despite ongoing demographic processes.

Most of the current statistical approaches for the study of adaptation rely on the expected signatures of hard sweeps (Sella et al. 2009; Burke et al. 2010). Selective sweeps can be "hard" (a single adaptive allele sweeps through the population), or "soft," when multiple adaptive alleles at the same locus sweep through the population at the same time (Hermisson and Pennings 2005). *Drosophila* belong to organisms that undergo recurrent boom–bust cycles allowing adaptation during the boom years to occur in populations of large short-term N_e, making short-term evolution act primarily on preexisting intermediate-frequency genetic variants that are swept the remainder of the way to fixation (Karasov et al. 2010; Burke et al. 2010; Messer and Petrov 2013). Therefore, adaptation in *Drosophila* may not be limited by waiting for mutations at single sites. Complex adaptive alleles can be generated quickly without intermediate fixation. Garud et al. (2015) described a new statistical method that can identify both hard and soft sweeps. The authors applied developed approach to a *D. melanogaster* population genomic dataset of 145 sequenced strains collected in North Carolina and found that majority of the strongest and most recent sweeps show patterns that are more consistent with soft rather than hard sweeps. By now, only few studies have presented the sequencing data of differentially selected populations (Orozco-terWengel et al. 2012). One of these studies compared populations from a long-term selection experiment (separating fast growers *D. melanogaster* with short life spans from slow growers with longer life spans) and concluded that even after 600 generations of artificial selection new mutations did not contribute to adaptation (Burke et al. 2010). Divergence of the tested populations was based on allele frequency changes from standing variation. In another study, flies were selected for hypoxia tolerance (Zhou

et al. 2011); the outcomes included a large number of selected SNPs, and some of the identified candidate genes were functionally validated. Turner et al. (2011) examined the results obtained on flies selected for body size and also reported about large number of selected SNPs, validating the possibility of precise identification of candidate genes.

In our case, in addition to differential selection regimes, it is possible that demographic processes may have starkly different dynamics on the opposite slopes of the canyon. Such events as droughts and forest fires, like that one in December of 2010 that ravaged the Carmel mountains, touching also canyon's habitats, presumably drive local *Drosophila* populations to very low numbers, followed by rapid repopulations and expansions. Even without environmental disasters, *D. melanogaster* populations have been known to be subject to boom and bust cycles (Harry et al. 1999), boosting short-term N_e and enabling short-term evolution act primarily on preexisting intermediate-frequency genetic variants that are swept the remainder of the way to fixation (Pool et al. 2012).

Thus, our interpretation is that the discovered sympatric differentiation is caused by genetic adaptation to the strong interslope ecological contrasts, facilitating preservation of the evolving genetic complexes from recombination collapse. Based on the obtained data, we suggest that *Drosophila* populations at the canyon represent a rare example, demonstrating how selection may override migration. We explain the accumulated data by strong ecological selection and complex behavioral mechanism restricting gene flow between the slopes, despite migration, that includes habitat choice and assortative mating. It is also possible that the discovered differentiation reflects a dynamic balance between migration, selection, recombination, habitat choice, and partial isolation rather than ongoing speciation (Korol et al. 2006). Thus, the revealed interslope divergence calls for additional experiments using other *Drosophila* species and populations from other canyons to test the obtained patterns and critically compare explanatory scenarios. The divergence in adaptive traits within a population at a microsite makes this system a good model for an in-depth study of adaptation under heterogeneous conditions. The *Drosophila* system from the canyon has the potential to advance our understanding of how local adaptations and reproductive isolation can originate despite presumed ongoing genetic exchange. As was noted by Schlötterer et al. (2006), "*D. melanogaster* is particularly well suited for the identification of ecologically relevant alleles." A large number of important long-standing questions related to sympatric speciation still remain open and should be addressed in further studies utilizing new tools of structural and functional genomics combined with experimental evolution.

Acknowledgments We thank our colleagues and collaborators E. Nevo, K. Iliadi, T. Pavlíček, I. Zamorzaeva, S.-R. Singh, I. Baca, A. Pechkovsky, J. Carmel, and D. Weiner for the inspiring cooperations, helpful discussions, and participation in fieldwork and experimental studies. This project was supported by Israel Science Foundation Grant 601/03-17.3, United States–Israel Binational Science Foundation Grants 9800443 and 2011438, the Ancell–Teicher Research Foundation for Genetics and Molecular Evolution, and the Israeli Ministry of Absorption.

References

Alipaz JA, Wu C-I, Karr TL (2001) Sperm-egg incompatibility between races of *Drosophila*. Proc Roy Soc 268:789–795

Bank C, Bürger R, Hermisson J (2012) The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-island model. Genetics 191(3):845–863

Barker JSF, Starmer WT, Fogleman JC (1994) Genotype-specific habitat selection for oviposition sites in the cactophilic species *Drosophila buzzatii*. Heredity 72:384–395

Barluenga M, Stölting K, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439(7077):719–723

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects and speciation. Annu Rev Ecol Syst 15:133–164

Bateson W (1909) Heredity and variation in modern lights. In: Seward AC (ed) Darwin and modern science. Cambridge University Press, Cambridge, pp 85–101

Begun DJ, Aquadro CF (1993) African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365(6446):548–550

Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting selective sweeps from pooled next-generation sequencing samples. Mol Biol Evol 29(9):2177–2186

Boughman JW, Rundle HD, Schluter D (2005) Parallel evolution of sexual isolation in sticklebacks. Evolution 59(2):361–373

Bradbury IR, Hubert S, Higgins B, Bowman S, Borza T et al (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. Evol Appl 6(3):450–461

Bubliy OA, Loeschcke V (2005) Variation of life-history and morphometrical traits in *Drosophila buzzatii* and *D. simulans* collected along an altitudinal gradient from a Canary island. Biol J Linn Soc 84:119–136

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467(7315):587–590

Bush GL (1975) Modes of animal speciation. Annu Rev Ecol Syst 6:339–364

Capy P, Veuille M, Paillette M, Jallon JM, Vouidibio J, David JR (2000) Sexual isolation of genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: the first step towards speciation? Heredity 84:468–475

Carmel J, Rashkovetsky E, Nevo E, Korol A (2011) Differential expression of small heat shock protein genes *Hsp23* and *Hsp40*, and heat shock gene *Hsr-omega* in fruit flies (*Drosophila melanogaster*) along a microclimatic gradient. J Hered 102(5):593–603

Carson HL, Bryant PJ (1979) Genetic variation in Hawaiian *Drosophila*. VI. Change in a secondary sexual character as evidence of incipient speciation in *Drosophila*. Proc Natl Acad Sci USA 76:1929–1932

Carson HL, Hardy DE, Spieth HT, Stone WS (1970) The evolutionary biology of the Hawaiian Drosophilidae. In: Hecht MK, Steere WC (eds) Essays in evolution and genetics in honor of Theodosius Dobzhansky. Appleton-Century Crofts, New York, pp 437–543

Colegrave N, Hollocher H, Hinton K, Ritchie M (2000) The courtship song of African *Drosophila melanogaster*. J Evol Biol 13:143–150

Colson I (2002) Selection and gene flow between microenvironments: the case of *Drosophila* at Lower Nahal Oren, Mount Carmel, Israel. Mol Ecol 11:1311–1316

Coyne JA (1994) Ernst Mayr and the origin of species. Evolution 48:19–30

Coyne J, Milstead B (1987) Long-distance migration of *Drosophila*. Am Nat 130:70–82

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland

Coyne J, Crittenden AP, Mah K (1994) Genetics of a pheromonal difference contributing to reproductive isolation in *D. melanogaster*. Science 265:1461–1464

Dallerac R, Labeur C, Jallon J-M, Knipple DC, Roelofs WL, Wicker-Thomas C (2000) A Δ9 *desaturase* gene with different substrate specificity is responsible for the cuticular diene

hydrocarbon polymorphism in *Drosophila melanogaster*. Proc Natl Acad Sci USA 97: 9449–9454
de Oliveira AK, Cordeiro AR (1980) Adaptation of *Drosophila willistoni* experimental populations to extreme pH medium. Heredity 44:123–130
DeSalle R, Giddings LV (1986) Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian *Drosophila*. Proc Natl Acad Sci USA 83(18):6902–6906
Dhar PK (2007) The next step in biology: a periodic table? J Biosci 32:1005–1008
Dhuyvetter H, Hendrickx F, Gaublomme E, Desender K (2007) Differentiation between two salt marsh beetle ecotypes: evidence for ongoing speciation. Evolution 61(1):184–193
Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. Nature 400:354–357
Dobzhansky T (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21:113–135
Dobzhansky T (1955) Evolution, genetics and man. Wiley, New York
Dodd DMB (1989) Reproductive isolation as a consequence of adaptive divergence in *Drosophila melanogaster*. Evolution 43:1308–1311
Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. Nature 421:259–264
Drake AE, Rashkovetsky E, Wong D, Rundle HD, Mooers AO (2005) Variable assortative mating in replicate mating trials using *Drosophila melanogaster* populations derived from contrasting opposing slopes of 'Evolution Canyon', Israel. J Evol Biol 18:1123–1129
Endler JA (1977) Geographic variation, speciation and clines. Princeton University Press, Princeton
Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton
Evans GA (2000) The periodic table of biology. Microb Comp Genomics 5(4):179–180
Ewing AW (1983) Functional aspects of *Drosophila* courtship. Biol Rev 58:275–292
Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. Molec Ecol 21(19):4748–4769
Fang S, Takahashi A, Wu C-I (2002) A mutation in the promoter of *desaturase 2* is correlated with sexual isolation between *Drosophila* behavioral races. Genetics 162:781–784
Feder JL, Chilcote CA, Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Nature 336:61–64
Ferveur J-F, Cobb M, Bourella H, Jallon J-M (1996) World-wide variation in *Drosophila melanogaster* sex pheromone: behavioral effects, genetic bases and potential evolutionary consequences. Genetica 97:73–80
Filchak KE, Roethele JB, Feder JL (2000) Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. Nature 407(6805):739–742
Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc Natl Acad Sci USA 103 (9):3209–3213
Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics 186:207–218
Garud NR, Messer PW, Buzbas EO, Petrov DA (2015) Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. PLOS Genet 11: e1005004. doi:10.1371/journal.pgen.1005004
Gavrilets S (2003) Models of speciation: what have we learned in 40 years? Evolution 57:2197–2215
Gavrilets S, Waxman D (2002) Sympatric speciation by sexual conflict. Proc Natl Acad Sci USA 99:10533–10538
Gefen E, Brendzel O (2011) Desiccation resistance and mating behavior in laboratory populations of *Drosophila simulans* originating from the opposing slopes of Lower Nahal Oren (Israel). J Evol Biol 24:2110–2117

Geiger MF, McCrary JK, Schliewen UK (2010) Not a simple case—a first comprehensive phylogenetic hypothesis for the Midas cichlid complex in Nicaragua (Teleostei: Cichlidae: *Amphilophus*). Mol Phylogenet Evol 56(3):1011–1024

Gilchrist AS, Azevedo RBR, Partridge L, O'Higins P (2000) Adaptation and constraint in the evolution of *Drosophila melanogaster* wing shape. Evol Dev 2:114–124

Gleason JM, Ritchie MG (1998) Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: do sexual signals diverge the most quickly? Evolution 1493–1500

Gockel J, Robinson SJW, Kennington WJ, Goldstein DB, Partridge L (2002) Quantitative genetic analysis of natural variation in body size in *Drosophila melanogaster*. Heredity 89:145–153

González J, Karasov TL, Messer PW, Petrov DA (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. PLoS Genet 6 (4):e1000905

Grant PR, Grant BR (2014) 40 Years of evolution. Darwin's finches on Daphne Major Island. Princeton University Press, Princeton

Hall MC, Willis JH (2006) Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. Evolution 60:2466–2477

Harry M, Rashkovetsky E, Pavlicek T, Baker S, Derzhavets E, Capy P, Cariou ML, Lachaise D, Asada N, Nevo E (1999) Fine-scale biodiversity of *Drosophila* in 'Evolution Canyon' at the Lower Nahal Oren Microsite, Israel. Biologia (Bratislava) 54:683–703

Hercus MJ, Loeschcke V, Rattan SI (2003) Lifespan extension of *Drosophila melanogaster* through hormesis by repeated mild heat stress. Biogerontology 4:149–156

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352

Higashi M, Takimoto G, Yamamura N (1999) Sympatric speciation by sexual selection. Nature 402:523–526

Hoffmann AA, Harshman LG (1999) Desiccation and starvation resistance in *Drosophila*: patterns of variation at the species, population and intrapopulation levels. Heredity 83:637–643

Hoffmann AA, Parsons PA (1993) Direct and correlated responses to selection for desiccation resistance: a comparison of *D. melanogaster* and *D. simulans*. J Evol Biol 6:643–657

Hoffmann AA, Hallas R, Sinclair C, Partrige L (2001) Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. Evolution 55:436–438

Hoffmann AA, Shirriffs J, Scott M (2005) Relative importance of plastic versus genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. Funct Ecol 19:222–227

Hollocher H, Ting C-T, Pollack F, Wu C-I (1997) Incipient speciation by sexual isolation in *Drosophila melanogaster*: variation in mating preference and correlation between sexes. Evolution 51:1175–1181

Hood GR, Yee W, Goughnour RB, Sim SB, Egan SP, Arcella T, Saint-Jean G, Powell THQ, Xu CCY, Feder JL (2013) The geographic distribution of *Rhagoletis pomonella* (Diptera: Tephritidae) in the Western United States: introduced species or native population? Ann Entomol Soc Am 106(1):59–65

Hübner S, Rashkovetsky E, Kim YB, Oh JH, Michalak K, Weiner D, Korol A, Nevo E, Michalak P (2013) Genome differentiation of *Drosophila melanogaster* from a microclimate contrast in Evolution Canyon, Israel. Proc Natl Acad Sci USA 110(52):1059–21064

Huey RB, Patridge L, Fowler K (1991) Thermal sensitivity of *Drosophila melanogaster* responds rapidly to laboratory natural selection. Evolution 45:751–756

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. Science 287:308–309

Iliadi K, Iliadi N, Rashkovetsky E, Minkov I, Nevo E, Korol A (2001) Sexual and reproductive behavior of *Drosophila melanogaster* from a microclimatically interslope differentiated population of 'Evolution Canyon' (Mount Carmel, Israel). Proc R Soc London B 268: 2365–2374

Iliadi K, Iliadi N, Rashkovetsky E, Girin S, Nevo E, Korol A (2002) Sexual differences for emigration behavior in natural populations of *Drosophila melanogaster*. Behav Genet 32: 173–180

Iliadi KG, Kamyshev NG, Popov AV, Iliadi NN, Rashkovetsky EL, Nevo E, Korol AB (2009) Peculiarities of the courtship song in the *Drosophila melanogaster* populations adapted to gradient of microecological conditions. J Evol Biochem Physiol 45:579–589

James AC, Partridge L (1995) Thermal evolution of rate of larval development in *Drosophila melanogaster* in laboratory and field populations. J Evol Biol 8:315–330

Jiggins CD (2006) Sympatric speciation: why the controversy? Curr Biol 16:333–334

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. Trends Ecol Evol 15:250–255

Kaneshiro KY (1988) Speciation in the Hawaiian *Drosophila*: sexual selection appears to play an important role. Bioscience 38(4):258–263

Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PLoS Genet 6(6):e1000924

Keller I, Seehausen O (2012) Thermal adaptation and ecological speciation. Molec Ecol 21: 782–799

Kelly MG, Levin DA (2000) Directional selection on initial flowering date in *Phlox drummondii* (Polemoniaceae). Am J Bot 87:382–391

Kennington WJ, Gilchrist AS, Goldstein DB, Partridge L (2001) The genetic basis of divergence in desiccation and starvation resistance among tropical and temperate populations of *Drosophila melanogaster*. Heredity 87:363–372

Kilias G, Alahiotis SN (1985) Indirect thermal selection in *Drosophila melanogaster* and adaptive consequences. Theor Appl Genet 69:645–650

Kilias G, Alahiotis SN, Delecanos M (1980) A multifactorial investigation of speciation theory using *Drosophila melanogaster*. Evolution 34:730–737

Kim YB, Oh JH, McIver LJ, Rashkovetsky E, Michalak K, Garner HR, Kang L, Nevo E, Korol AB, Michalak P (2014) Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. Proc Natl Acad Sci USA 111 (29):10630–10635

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CH, Hoang A, Gibert P, Beerli P (2001) The strength of phenotypic selection in natural populations. Am Nat 157: 245–261

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187: 245–260

Kondrashov A, Kondrashov F (1999) Interactions among quantitative traits in the course of sympatric speciation. Nature 400:351–354

Korol A, Rashkovetsky E, Iliadi K, Michalak P, Ronin Y, Nevo E (2000) Nonrandom mating in *Drosophila melanogaster* laboratory populations derived from closely adjacent ecologically contrasting slopes at 'Evolution Canyon'. Proc Natl Acad Sci USA 97:12637–12642

Korol A, Rashkovetsky E, Iliadi K, Nevo E (2006a) *Drosophila* flies in "Evolution Canyon" as a model for incipient sympatric speciation. Proc Natl Acad Sci USA 103:18184–18189

Korol A, Rashkovetsky E, Nevo E (2006b) Enigmatic flies: is *Drosophila* in the 'Evolution canyon' a model for incipient speciation? Isr J Ecol Evol 52:507–525

Krebs RA, Feder ME (1998) Experimental manipulation of the cost of thermal acclimation in *Drosophila melanogaster*. Biol J Linn Soc 63:593–601

Kyriacou CP, Hall JC (1984) Learning and memory mutations impair acoustic priming of mating behaviour in *Drosophila*. Nature 308:56–65

Kyriacou CP, Sawyer LA, Piccin A, Couchman ME, Chalmers D (1996) Evolution and population biology of the period gene. Cell Dev Biol 7:803–810

Lerman DN, Michalak P, Helin AB, Bettencourt BR, Feder ME (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. Mol Biol Evol 20:135–144

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 6:991–1007
Lindquist S (1986) The heat-shock response. Annu Rev Biochem 55:1151–1191
Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. Phil Trans R Soc B 363:3009–3021
Lupu A, Pechkovskaya A, Rashkovetsky E, Nevo E, Korol A (2004) DNA repair efficiency and thermotolerance in *Drosophila melanogaster* from 'Evolution Canyon'. Mutagenesis 19: 383–390
Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF et al (2012) The *Drosophila melanogaster* genetic reference panel. Nature 482:173–178
Mallet J (2001) The speciation revolution. J Evol Biol 14:887–888
Mayr E (1942) Systematics and the origins of species. Columbia University Press, New York
Mayr E (1947) Ecological factors in speciation. Evolution 1(4):263–288
Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge
Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol Evol 28(11):659–669
Michalak P, Minkov I, Helin A, Lerman DN, Bettencourt B, Feder ME et al. (2001) Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along a microclimatic contrast in 'Evolution Canyon', Israel. Proc Natl Acad Sci USA 98: 13195–13200
Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL (2010) Widespread genomic divergence during sympatric speciation. Proc Natl Acad Sci 107:9724–9729
Muller H (1942) Isolating mechanisms, evolution and temperature. Biol Symp 6:71–125
Neat F, Fowler K, French V, Partridge L (1995) Thermal evolution of growth efficiency in *Drosophila melanogaster*. Proc R Soc B 260:73–78
Nei M, Maruyama T, Wu C-I (1983) Models of evolution of reproductive isolation. Genetics 103:557–579
Nevo E (2001) Evolution of genome–phenome diversity under environmental stress. Proc Natl Acad Sci USA 98:6233–6240
Nevo E (2011) Evolution under environmental stress at macro-and microscales. Genome Biol Evol 2:1039–1052
Nevo E (2012) "Evolution Canyon", a potential microscale monitor of global warming across life. Proc Natl Acad Sci USA 109(8):2960–2965
Nevo E, Rashkovetsky E, Pavliceck T, Korol A (1998) A complex adaptive syndrome in *Drosophila* caused by microclimatic contrasts. Heredity 80:9–16
Nevo E, Beiles A, Korol AB, Ronin YI, Pavlicek T, Hamilton W (2000) Extraordinary multilocus genetic organization in mole crickets, *Gryllotalpidae*. Evolution 54(2):586–605
Noor MAF (1999) Reinforcement and other consequences of sympatry. Heredity 83:503–508
Nosil P (2008) Speciation with gene flow could be common. Mol Ecol 17(9):2103–2106
Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719
Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C (2012) Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Molec Ecol 21:4931–4941
Panhuis TM, Swanson WJ, Nunney L (2003) Population genetics of accessory gland proteins and sexual behavior in *Drosophila melanogaster* populations from 'Evolution Canyon'. Evolution 57:2785–2791
Parsons PA (1983) Ecobehavioral genetics: habitats and colonists. Ann Rev Ecol Syst 14:35–55
Parsons PA (1993) Evolutionary adaptation and stress: energy budgets and habitats preferred. Behav Genet 23:231–238
Pavlíček T, Sharon D, Kravchenko V, Saaroni H, Nevo E (2003) Microclimatic interslope differences underlying biodiversity contrasts in 'Evolution Canyon', Mt. Carmel, Israel. Isr J Earth Sci 52:1–9

Pavlíček T, Frenkel Z, Korol A et al (2008) *Drosophila* at the "Evolution Canyon" microsite, Mt. Carmel, Israel: selection overrules migration. Isr J Ecol Evol 54:165–180

Pechkovsky A (2008) Genetic structure of *Drosophila* populations adapted to contrasting microclimatic gradients: evidence from "EC" and laboratory experiments. Thesis for the degree "Doctor of Philosophy", University of Haifa, Israel

Pereyra RT, Bergström L, Kautsky L, Johannesson K (2009) Rapid speciation in a newly opened postglacial marine environment, the Baltic Sea. BMC Evol Biol 9:70

Pirkkala L, Nykanen P, Sistonen L (2001) Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. FASEB J 15:1118–1131

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ, Langley CH (2012) Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet 8(12): e1003080

Rashkovetsky E, Iliadi K, Nevo E, Korol A (2000) Fitness related traits in *Drosophila melanogaster* subpopulations from the opposite slopes of 'Evolution Canyon', lower Nahal Oren, Mount Carmel, Israel. Dros Inf Serv 83:138–140

Rashkovetsky E, Iliadi K, Michalak P, Lupu A, Nevo E, Feder M, Korol A (2006) Adaptive differentiation of thermotolerance in *Drosophila* along a microclimatic gradient. Heredity 96 (5):353–359

Reed L, Markow TA (2004) Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. Proc Natl Acad Sci USA 101(24):9009–9012

Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in forty years? Evolution 47:1637–1653

Rice WR, Salt GW (1988) Speciation via disruptive selection on habitat preference: experimental evidence. Am Nat 131:911–917

Rice WR, Salt GW (1990) The evolution of reproductive isolation as a correlated character under sympatric conditions: experimental evidence. Evolution 44:1140–1152

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity 83:363–372

Ritchie MG, Halsey EJ, Gleason JM (1999) *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. Anim Behav 58:649–657

Robinson SJW, Zwaan B, Partridge L (2000) Starvation resistance and adult body composition in a latitudinal cline of *Drosophila melanogaster*. Evolution 54:1819–1824

Rolán-Alvarez E (2007) Sympatric speciation as a by-product of ecological adaptation in the Galician *Littorina saxatilis* hybrid zone. J Mollus Stud 73(1):1–10

Rolshausen G, Segelbacher G, Hobson KA, Schaefer HM (2009) Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. Cur Biol 19(24):2097–2101

Rose MR, Vu LN, Park SU, Graves JL (1992) Selection on stress resistance increases longevity in *Drosophila melanogaster*. Exp Gerontol 27:241–250

Rosengren R, Chautems D, Cherix D, Fortelius W, Keller L (1994) Separation of two sympatric spesies of *Formica* L. ants by a behavioral choice test based on brood discrimination. Memorabilia Zoolog 48:237–249

Rosenzweig ML (1978) Competitive speciation. Biol J Linn Soc 10:275–289

Rundle HD, Schluter D (2004) Natural selection and ecological speciation in Sticklebacks. In: Dieckmann U, Doebeli M, Metz JAJ, Tautz D (eds) Adaptive speciation. Cambridge University Press, Cambridge, pp 192–209

Rundle HD, Nagel L, Boughman JW, Schluter D (2000) Natural selection and parallel speciation in sympatric Sticklebacks. Science 287(5451):306–308

Sarup P, Sorensen JG, Dimitrov K, Barker JSF, Loeschcke V (2006) Climatic adaptation of *Drosophila buzzatii* populations in southeast Australia. Heredity 96:479–486

Savolainen V, Anstett M-C, Lexer C, Hutton J, Clarkson JC, Noroup MV, Powell MP, Springate D, Salamin N, Baker WJ (2006a) Sympatric speciation in palms on an oceanic island. Nature 441:210–213

Savolainen V, Lexer C, Anstett M-C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ (2006b) Evolutionary biology: sympatric plant speciation in islands? Nature 443:E12–E13

Sawyer LA, Hennesy JM, Peixoto AA, Rosato E, Parkinson H, Costa R et al (1997) Natural variation in a *Drosophila* clock gene and temperature compensation. Science 278:2117–2120

Schilthuizen M (2000) Dualism and conflicts in understanding speciation. BioEssays 22: 1134–1141

Schliewen UK, Tautz D, Paabo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368:629–632

Schlötterer C, Agis M (2002) Microsatellite analysis of *Drosophila melanogaster* populations along a microclimatic contrast at lower Nahel Oren canyon, Mount Carmel, Israel. Mol Biol Evol 19:563–568

Schlötterer C, Neumeier H, Sousa C, Nolte V (2006) Highly structured Asian *Drosophila melanogaster* populations: a new tool for hitchhiking mapping? Genetics 172:287–292

Schluter D (1998) Ecological speciation in postglacial fishes. In: Grant PR (ed) Evolution on islands. Oxford University Press, New York, pp 163–180

Schluter D (2001) Ecology and the origin of species. Trends Ecol Evol 16:372–380

Schluter D, Marchinko KB, Barrett RDH, Rogers SM (2010) Natural selection and the genetics of adaptation in threespine stickleback. Phil Trans R Soc B 365:2479–2486

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P et al (2014) Genomics and the origin of species. Nat Rev Genet 15:176–192

Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the *Drosophila* genome? PLoS Genet 5(6):e1000495

Sikorski J, Nevo E (2005) Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. Proc Natl Acad Sci USA 102 (44):15924–15929

Singh SR, Rashkovetsky E, Iliadi K, Nevo E, Korol A (2005) Assortative mating in *Drosophila* adapted to a microsite ecological gradient. Behav Genet 35:753–764

Smith JM (1966) Sympatric speciation. Am Nat 100(916):637–650

Sorenson MD, Sefc KM, Payne RB (2003) Speciation by host switch in brood parasitic indigobirds. Nature 424:928–931

Stefanic P, Mandic-Mulec I (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. J Bacteriol 191(6):1756–1764

Stuessy TF (2006) Evolutionary biology: sympatric plant speciation in islands? Nature 443:E12

Svanbäck R, Schluter D (2012) Niche specialization influences adaptive phenotypic plasticity in the *Threespine Stickleback*. Am Nat 180(1):50–59

Takahashi A, Tsaur S-C, Coyne J, Wu C-I (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc Natl Acad Sci USA 98:3920–3925

Taylor EB, McPhail JD (2000). Historical contingency and ecological determinism interact to prime speciation in sticklebacks. Proc R Soc London Biol Sci 267:2375–2384

Taylor CE, Powell JR (1977) Microgeographic differentiation of chromosomal and enzyme polymorphisms in *Drosophila persimilis*. Genetics 85:681–695

Thoday JM (1972) Disruptive selection. Proc Roy Soc B. 182:109–143

Thoday JM, Gibson JB (1962) Isolation by disruptive selection. Nature 193:1164–1166

Torday JS (2004) A periodic table for biology. Scientist 18(12):32–33

Tregenza T, Butlin RK (1999) Speciation without isolation. Nature 400:311–312

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. Trends Ecol Evol 16:330–343

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol 3:e285

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM (2011) Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. PLoS Genet 7(3):e1001336

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol Evol 16:381–390

Vilas-Boas G, Sanchis V, Lereclus D, Lemos MVF, Bourguet D (2002) Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. Appl Environ Microbiol 68(3):1414–1424

Wiener P, Feldman MW (1993) The effect of the mating system on the evolution of migration in a spatially heterogeneous population. Evol Ecol 7:251–269

Wu C-I, Hollocher H, Begun DJ, Aquadro CF, Xu Y (1995) Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. Proc Natl Acad Sci USA 92:2519–2523

Zamorzaeva I, Rashkovetsky E, Nevo E, Korol A (2005) Sequence polymorphism of candidate behavioral genes in *Drosophila melanogaster* flies from 'Evolution Canyon'. Mol Ecol 14:3235–3245

Zamorzaeva I, Rashkovetsky E, Nevo E, Korol A (2009) Deletions in the *desaturase2* gene in Israeli populations of *Drosophila melanogaster*: possible role in mate recognition. Insect Mol Biol 18(1):61–69

Zhou D, Udpa N, Gersten M, Visk DW, Bashir A et al (2011) Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. Proc Natl Acad Sci USA 108:2349–2354

Zwaan BJ, Azevedo RBR, James AC, van't Land J, Partridge L (2000) Cellular basis of wing size variation in *Drosophila melanogaster*: a comparison of two continents. Heredity 84:338–347

Chapter 8
Mutation-Driven Evolution: Microsatellite Instability Drives Speciation in a Mammalian Taxon

Paul Sequeira, Yen-Shan Chen and Michael A. Weiss

Abstract The concept of *evolvability*—from its evolutionary origins to molecular mechanisms—defines a fundamental problem at the intersection of biochemistry, genetics, and developmental biology. An emerging paradigm, *mutation-driven evolution*, posits that chromosomal dynamics (including changes in ploidy, chromosome loss, aberrant recombination, and mechanics of DNA damage and repair) underlie the origin of genetic variation as a precondition for selection. A model is provided by microsatellite instability. Although widely exploited as an experimental marker of evolutionary change with application to human disease, the potential contribution of such instability to evolvability itself is less well understood. Here, we propose that microsatellite instability within a vertebrate sex-determining gene can drive rapid adaptation and speciation. Our analysis focuses on superfamily Muroidea (order Rodentia) wherein four anomalies are observed: (1) This superfamily is unusually speciose, indeed the most species rich in Mammalia; (2) speciation has occurred rapidly (i.e., within the past 25 million years) and apparently in overlapping ranges; (3) inherited XY sex reversal has evolved independently within multiple genera; and (4) uniquely among therian mammals, male sex-determining mechanisms not dependent on Y-encoded testis-determining factor Sry have emerged. A unifying hypothesis is presented whereby these anomalies have a single molecular basis, to wit the dynamics of a Muroidea-specific microsatellite-encoded transcriptional activation domain. An ancestral microsatellite in this taxon has functioned as a "genetic capacitor" to

Amino acids are designated by standard three-letter code.

P. Sequeira
Thermalin Diabetes, LLC, 10000 Cedar Rd, Cleveland, OH 44106, USA
e-mail: pxs175@case.edu

P. Sequeira · Y.-S. Chen · M.A. Weiss (✉)
Department of Biochemistry, Case Western Reserve University School of Medicine,
10900 Euclid Avenue, Cleveland, OH 44106, USA
e-mail: maw21@case.edu

Y.-S. Chen
e-mail: yxc113@case.edu

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_8

enable cryptic variation to accumulate within *Sry*. On discharge, this capacitor provided a recurring source of reproductive isolation and thus enabled rapid evolution of biological novelty at the edge of sexual ambiguity.

Abbreviations

ARD	Alanine-rich domain
DDA	Distance distribution analysis
FRET	Fluorescence resonance energy transfer
GRD	Glutamine-rich domain
GRN	Gene regulatory network
GRT	Gln-rich tract
HMG	High mobility group
PCR	Polymerase chain reaction
q-PCR	Quantitative reverse-transcription PCR
Sox	Sry-related HMG box
Sry	Sex-determining region of the Y chromosome
TAD	Transcriptional activation domain
TDF	Testis-determining factor
TES	Testis-specific enhancer
TF	Transcription factor
tr-FRET	Time-resolved FRET

> My own field of paleontology has strongly challenged the Darwinian premise that life's major transformations can be explained by adding up, through the immensity of geological time, the successive tiny changes produced generation after generation by natural selection.
>
> Stephen Jay Gould, "Darwinian Fundamentalism"
> *New York Review of Books* (1997)

8.1 Introduction

Evolutionary biology in the pre-molecular era, from Darwin (1872) through the "modern synthesis" of the mid-twentieth century (Huxley 1942), focused on natural selection and its implications for changes in gene frequencies in a population. The origins of genetic variation, unknown to Darwin, were often assumed to represent small genetic changes occurring over long timescales. In this population-based paradigm, sufficient genetic diversity was assumed to exist, i.e., such that gradual changes in allele frequencies may occur through natural selection. Reproductive isolation and hence barriers to gene flow were considered primarily in relation to geographical barriers or isolated ecological niches (Mayr 1947). Peripheral to this paradigm were possible molecular mechanisms of saltatory genetic change, such as gene duplication, gene deletion, and changes in ploidy (Hodgkin 2005). The Modern Synthesis thus provided an incomplete view of the origins of biological innovation and its relative pace among taxa (Nei 2013).

The advent of molecular biology, its application to metazoan development (Robert 2004), and integration within evolutionary theory (Carroll 2008; Bolnick and Fitzpatrick 2007) have by contrast demonstrated the importance of chromosomal dynamics, including "mutation" in its broadest sense (Nei 2013). Indeed, molecular discontinuities in a single generation can generate significant phenotypic effects, including alterations in body plan through the unmasking of preexisting cryptic variation (Rutherford and Lindquist 1998) and even effect reproductive isolation (Bolnick and Fitzpatrick 2007). Respective examples are provided by the role of heat-shock proteins in molecular compensation of "neutral" mutations in proteins ("genetic capacitors"; Rutherford and Lindquist 1998) and changes in ploidy in plants (whose discovery by De Vries in the early 1900s foreshadowed present concepts; De Vries 1909). Medical implications of chromosomal dynamics are at the frontier of cancer biology (Loeb 2011) and rationalize clinical observations of "anticipation" in the generational inheritance of neurodegenerative diseases associated with unstable DNA triplet repeats (microsatellites; Kimpara et al. 1997; Gövert and Schneider 2013). Insights of molecular genetics and enzymology of DNA transactions complement the punctuated view of the fossil record offered by Eldredge and Gould (Eldredge and Gould 1972) as reflected in the opening quote.

In this chapter, we highlight a well-known—yet often overlooked—anomaly in the natural history of mammals (Fig. 8.1a): the peculiar radiation of muroid rodents (Fig. 8.1b). Four features of this radiation are distinctive and together appear unique to this taxon. First, Rodentia is the most speciose mammalian order. Although this might merely reflect successful exploration and adaption to diverse niches, speciation has predominantly occurred within a single superfamily, Muroidea (Fig. 8.1c). Indeed, two extant families in Muroidea (Muridae and Cricetidae) together contain more genera and species than the rest of Rodentia combined (Wilson and Reeder 2005). Second, radiation within Muroidea (and even within individual genera) has been rapid on the geological timescale (Steppan et al. 2004). Such an accelerated pace of speciation cannot be entirely ascribed to a short reproductive life span or an accelerated mutational clock (as other mammalian taxa include species similarly short lived and with similar clock rates; Martin and Palumbi 1993). Third, several genera in Muroidea contain species whose populations are characterized by a significant fraction of XY females ("chromosomal sex reversal") as well as XX females and XY males (Jiménez et al. 2012; Veyrunes et al. 2010). Of particular interest, the genus *Akodon* has experienced, at least six times within (at most) the past eight million years (Hoekstra and Edwards 2000), the independent evolution of fertile XY females as a normal population component (Bianchi 2002). Finally, unique to Muroidea (among mammals) are species (spanning four genera) in which maleness is not specified by Sry (Jiménez et al. 2012), an architectural transcription factor (TF) encoded by the Y chromosome (Wallis et al. 2008) that otherwise functions to initiate male development in therian mammals (Koopman et al. 1991).

The goal of this chapter is to present a unifying hypothesis that suggests a molecular explanation for the above anomalies in the radiation of Rodentia. A framework is provided by a specific mechanism of sudden change in chromosomal dynamics: Microsatellite instability due to replicative DNA slippage as DNA

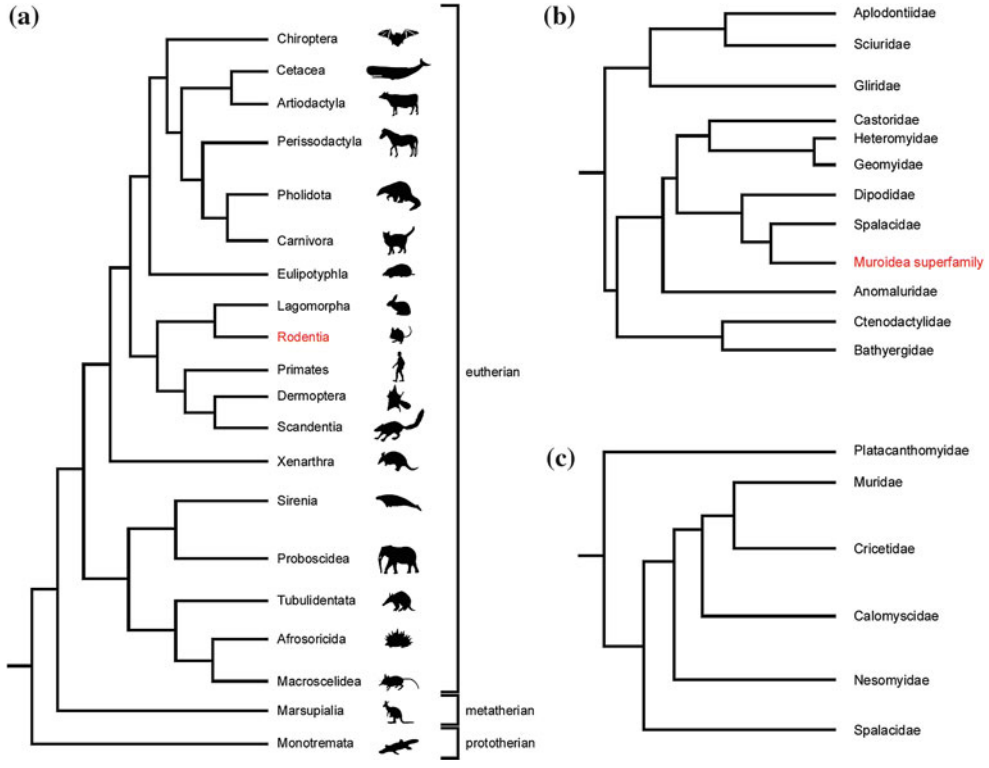


Fig. 8.1 Radiation of Mammalia**. a** Orders include eutherian (placental), metatherian (marsupial), and prototherian (monotreme) mammals. Rodentia is highlighted in *red*. **b** Radiation of Rodentia, including superfamily Muroidea (*red*). **c** Radiation of Muroidea

polymerase I encounters triplet repeats (Chatterjee et al. 2013). We posit that a CAG DNA microsatellite invaded the ancestral *Sry* gene at the root of Muroidea—unlike all other extant mammalian taxa—enabling an evolutionary process that underlies each of the four distinctive features of this superfamily now observed. Our model illuminates how microsatellite instability, broadly known for its role in human disease (Kimpara et al. 1997; Brouwer et al. 2009; Gövert and Schneider 2013), may function as a source of biological innovation through the molecular operation of a developmental switch at the edge of ambiguity (Kauffman and Johnsen 1991) as envisaged in testis determination (Chen et al. 2013a).

8.2 SRY and Mammalian Male Sex Determination

The molecular genetics of sex determination and gonadogenesis have attracted broad interest in relation to both principles of development (Goodfellow and Lovell-Badge 1993) and mechanisms of reproductive isolation (Kocher 2004). Given the direct relationship between sex and reproduction (and hence fitness), a surprising diversity of sex-determining systems has evolved in different phyla (Kraak 2002), including chromosomal and environmental mechanisms. Whereas in birds, females represent the heterogametic sex (ZW vs. ZZ; Kraak 2002), for example, in mammals males are ordinarily XY and females XX (Goodfellow and Lovell-Badge 1993). Such diversity stands in seeming contrast to the fundamental conservation of patterning genes (as exemplified by the *Hox* cluster; Ferrier and Holland 2001) and of specific master TFs in organogenesis (such as in eye development; Gehring and Ikeo 1999). Resolution of this bewildering paradox has been proposed through the concept of sex-determining pathways "growing backward," i.e., recruitment of novel upstream regulatory genes with progressive conservation of downstream elements (Graves et al. 1995; Graves 2013). In this view, the traditional focus of molecular genetics on primary determinants of sex in representative taxa has highlighted such upstream diversity at the expense of downstream commonalities. The underlying theme of this chapter is whether such backward-evolving pathways may provide an opportunity to probe the interplay between evolvability (upstream) and robustness (downstream; Brookfield 2009).

Male sex determination in therian mammals is regulated by Sry, a Y-encoded architectural TF that functions in late embryogenesis as the long-postulated testis-determining factor (TDF; Koopman et al. 1991). Although *Sry* is only present in therian mammals (and not in other vertebrates), the protein's sequence-specific DNA-binding domain (a high-mobility-group (HMG) box; Goodfellow and Lovell-Badge 1993) is an L-shaped α-helical domain as shown in Fig. 8.4a. This box provides a molecular signature of a large and more ancestral family of TFs (designated Sox; Sry-related HMG box), which are broadly conserved in metazoa (Jager et al. 2006). *Sry*, which likely arose from an ancestral autosomal *Sox3* on proto-autosomal-derived sex chromosomes (Stevanovic et al. 1993; Lahn and Page 1999; Sutton et al. 2011), has diverged from other *Sox* genes through the loss of

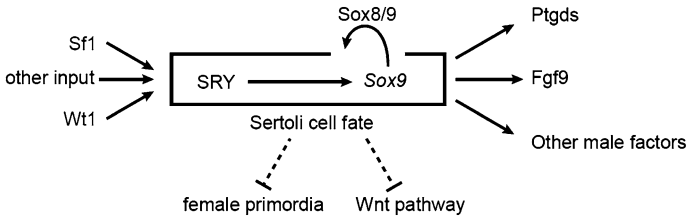


Fig. 8.2 Canonical male sex determination in therian mammals. Sry initiates the male developmental program by activating *Sox9*, the first step in a complex gene regulatory network that leads to Sertoli cell differentiation and testicular development. Activation of *Sox9* also serves to repress the default female pathway (*dashed lines*)

other encoded protein domains, such as coiled-coil dimerization elements (Wilson and Koopman 2002) and discrete motifs of transcriptional activation or repression (Lefebvre et al. 2007). The principal target of Sry in the differentiating gonadal ridge is *Sox9*, a downstream autosomal element that also operates in the male program of birds (Kent et al. 1996) and possibly of reptiles and amphibians (Pieau et al. 1999; Valleley et al. 2001). Sry binds to specific DNA sites in the testis-specific enhancer (TES) element of *Sox9* to activate its lineage- and stage-specific expression in the XY gonadal ridge (Fig. 8.2; Sekido and Lovell-Badge 2008). Mutations in human SRY or SOX9 are associated with XY gonadal dysgenesis and somatic sex reversal (Wagner et al. 1994).[1] Clinical mutations in SRY predominantly occur in its HMG box (Knower et al. 2011), whereas mutations in SOX9 are broadly distributed among its various protein domains (Gordon et al. 2009). This distinction, together with the marked divergence among the non-box regions of Sry, has motivated the hypothesis that, unlike Sox factors, Sry functions as "just a box" (Canning and Lovell-Badge 2002).

In apparent conflict with the "just a box" hypothesis is an exceptional set of *Sry* genes in Rodentia, including those of laboratory strains of mice commonly employed as genetic model organisms: The *Sry* genes of the expansive Muroidea superfamily differ from those of other mammals (Fig. 8.3; Chen et al. 2013b). First, the muroid alleles are remarkable for insertion of a CAG microsatellite (or its remnant) within the coding region downstream of the HMG box. This DNA segment encodes a Gln- or Ala-rich domain (GRD or ARD, respectively) of variable length (depending on reading frame; Bowles et al. 1999; Zhao et al. 2014). Second, the HMG boxes of muroid Sry exhibit more marked sequence variation than do the Sry boxes of other mammals. Remarkably, the range of variation is more extensive than that spanned by the large family of mammalian SOX factors, essentially in its entirety. Such variation is not neutral at the level of protein stability or functions as

[1]SOX9 also functions in the specification of cartilage and morphogenesis of bone, and so its mutation gives rise to the syndrome of *campomelic dysplasia* (Foster et al. 1994). Related phenotypes occur in association with deletion of upstream regulatory elements in the *SOX9* gene (Sekido and Lovell-Badge 2008).

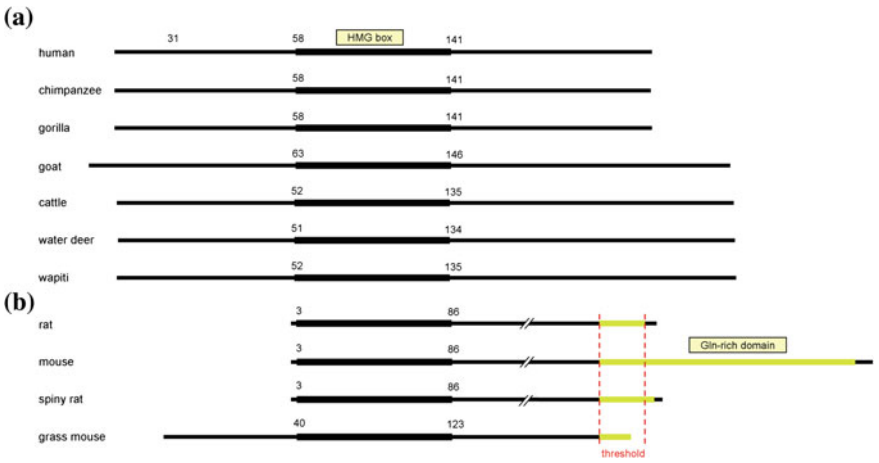


Fig. 8.3 Domain organization of Sry. **a** Non-rodent Sry alleles. **b** Divergence of rodent Sry proteins. HMG box is highlighted in *black* and the GRD in chartreuse

biochemical studies of the mouse Sry domain have shown the following indicators of decreased "molecular fitness": (i) loss of the free domain's thermodynamic stability, (ii) foreshortened lifetime of the protein–DNA complex, and (iii) impaired degree and precision of DNA bending (Chen et al. 2013b) with attenuated sequence specificity (Phillips et al. 2004). Such multifaceted biophysical "degeneration" was unexpectedly observed in assays of chemical protein denaturation (Fig. 8.4b), in kinetic studies based on stopped-flow fluorescence (Fig. 8.4c), and time-resolved fluorescence analysis of protein-directed DNA bending (Fig. 8.4d). Although the physiological implications of such biophysical properties are not a priori obvious, the above findings may address a biological puzzle that whereas in XX transgenic mice, human SRY can function as "just a box" to direct testicular differentiation, the function of mouse Sry (as an analogous transgene) requires both its box and C-terminal GRD (Bowles et al. 1999; Zhao et al. 2014).

Might it be possible that biophysical degeneration of the muroid HMG-box motif is a consequence of a contingent biochemical function of the GRD?[2] This question has motivated consideration of natural history of the *Sry* microsatellite in Muroidea in relation to general properties of such repeats (Chatterjee et al. 2013) and the specific biochemical function of Sry in initiation of a testis-specific gene regulatory network (GRN; Koopman et al. 1991). We address these considerations in turn.

[2]Proof of principle was provided in studies of chimeric mouse *Sry* genes containing the human SRY HMG box (Chen et al. 2013b). In this chimera the mouse GRD compensates for a human clinical mutation associated with partial loss of specific DNA-binding activity.

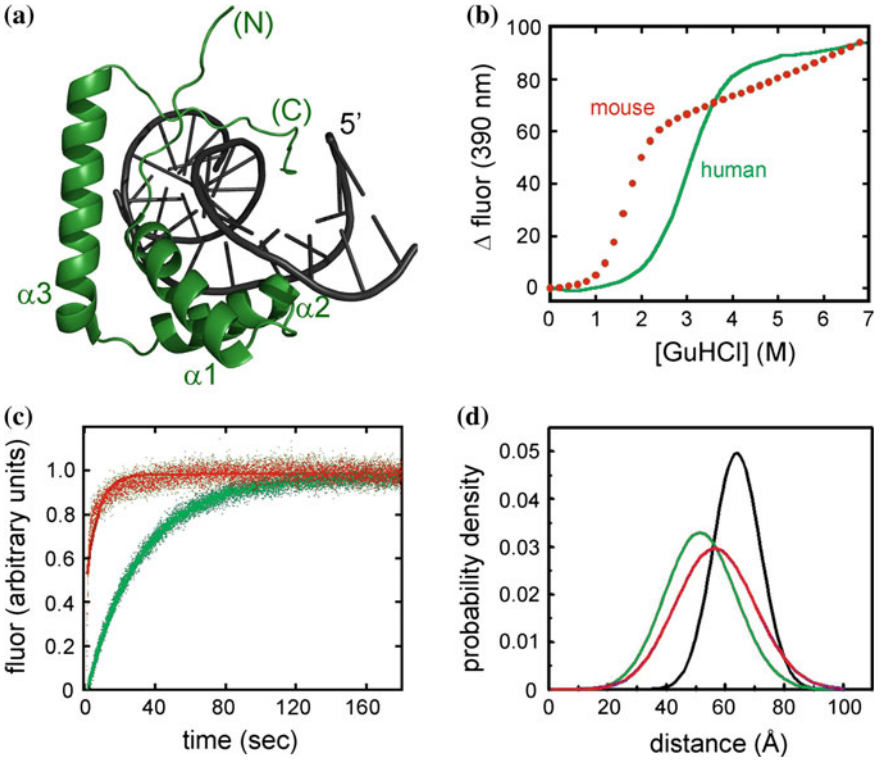


Fig. 8.4 Biochemical and biophysical degeneration of the mouse HMG box. **a** Ribbon model of L-shaped SRY HMG box (*green*) bound to DNA (*black*). Coordinates were obtained from PDB entry 1J46 (Murphy et al. 2001). **b** The murine HMG box (*red circles*) shows increased sensitivity to guanidine–HCl denaturation relative to the human domain (*green line*). Assay was evaluated by intrinsic tryptophan fluorescence as described in Chen et al. 2013b. **c** Stopped-flow FRET-based dissociation kinetic assay of SRY/Sry HMG box–DNA complexes; data and fitted lines showing increased donor fluorescence of FRET-labeled DNA due to dissociation of the labeled domain–DNA complex. Dissociation of the murine complex (*red*) is more rapid than that of the human complex (*green*). **d** End-to-end distance distributions of FRET-labeled free DNA (*black*) versus bent domain–DNA complexes: mouse (*red*) and human (*green*). Specific DNA bending by the human SRY HMG box is sharper (decreased mean end-to-end distance) and more precise (narrower distance distribution) than that by the murine Sry HMG box. Panels B and C were adapted from Chen et al. (2013b) and panel D from Phillips et al. (2004)

8.3 DNA Replication and Microsatellite Instability

In both prokaryotic and eukaryotic genomes (Chatterjee et al. 2013), microsatellite repeats (defined by repeat elements containing up to nine nucleotides) exhibit a general propensity to gain or lose tracts at a frequency much higher than point mutations (either within the satellite or in non-repetitive DNA regions; Rando and

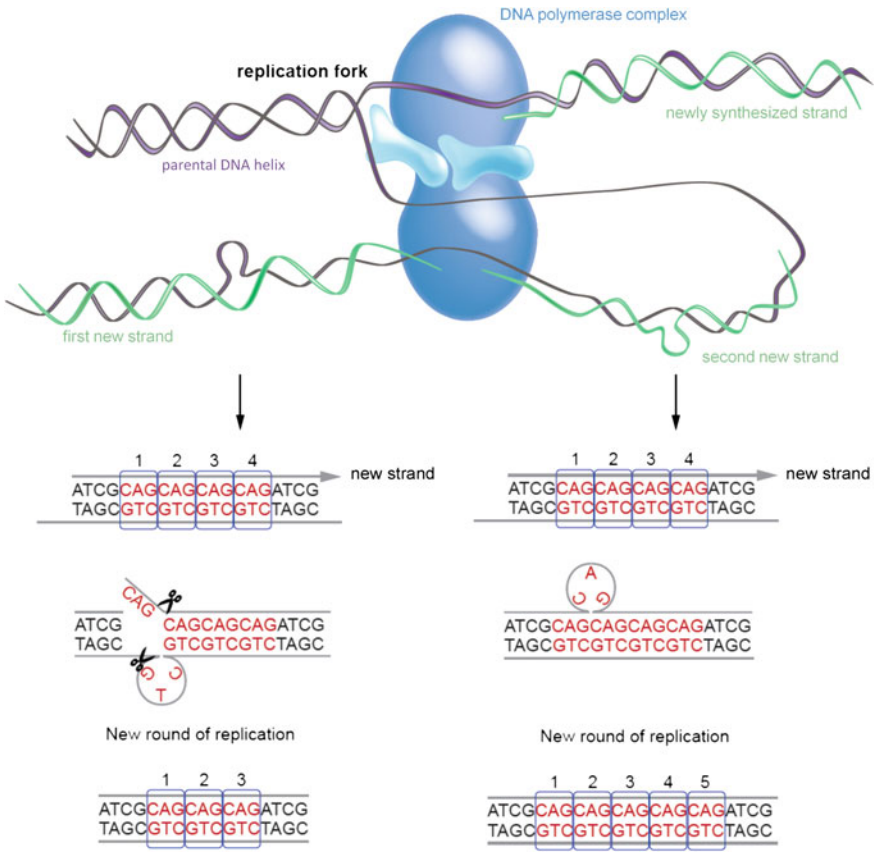


Fig. 8.5 Mechanism of microsatellite instability. Replication fork with leading and lagging DNA strands (*upper panel*). Slippage of replication fork within microsatellite DNA segments can lead to loss (*bottom left*) or gain (*bottom right*) of DNA repeat units. DNA polymerase complexes (*blue*) unzip and duplicate both parental strands (*purple*) during each round of DNA replication. Decrease (or increase) in number of microsatellite repeats occurs when the parental (or new strand; *green*) slips down one repeat in binding to the template DNA strand (*purple*). Such slippage predisposes the DNA polymerase to delete (or add) a repeat unit in the new strand

Verstrepen 2007). Although such gain or loss may arise through several molecular mechanisms, the predominant cause is thought to be mispairing of newly synthesized DNA strands during replication (Fig. 8.5) or analogous misalignment during recombination. Indeed, the repetitive nature of a DNA microsatellite lends itself to such misalignment, leading to both growth and shrinkage (Richard et al. 2008). In addition, trinucleotide repeat regions can be more susceptible to chromosomal breakage with subsequent expansion during repair (Freudenreich et al. 1998). Such repeat-associated properties of DNA, considered as a class of heteropolymers, represent an independent layer of heritable "information" that is distinct from the formal genetic properties of protein- or RNA-encoding elements. Microsatellite

instability has thus emerged as a general mechanism of genetic variation and adaptation in the progression of cancer (Loeb 1994), leading to analogies between its natural history (such as the transition from the primary tumor to metastasis) and speciation (Greaves and Maley 2012). The structural properties of DNA microsatellites have also rationalized clinical observations of genetic "anticipation" (phenotypes of progressive severity in successive generations) in the inheritance of several neurodegenerative diseases associated with triplet repeats (Li et al. 2004; Kimpara et al. 1997), including the paradigmatic Huntington's disease (Landles and Bates 2004).

The microsatellite within muroid *Sry* functions as a key component of its sex-determining mechanism and not as a disease allele. In accordance with the general instability of microsatellites, however, marked variation in its length has been observed among species in this taxon, in turn leading to Sry proteins containing GRDs (or ARDs) of highly variable length. Because *Sry* is contained within the non-recombining region of the Y chromosome and has no X homolog (Goodfellow and Lovell-Badge 1993), such variation in repeat length is presumably due primarily to DNA replicative slippage, whereas in *Mus musculus,* the GRD spans 224 residues, for example (including 20 blocks of Gln residues; Gubbay et al. 1992), in *Rattus norvegicus* only 25 residues are present (3 blocks) (Jäger et al. 1990). These differences in Gln-rich tracts (GRTs) motivated comparative functional studies of mouse Sry containing successive C-terminal deletions (Fig. 8.6a). Whereas the transcriptional potency of the transfected protein in cell culture (defined based on the transcriptional activation of *Sox9* in an embryonic rat XY cell line microdissected from the differentiating gonadal ridge just prior to the onset of *Sry* expression; Haqq and Donahoe 1998) is robust to the deletion of up to 17 of 20 GRTs, further deletion (beyond the limit of the rat gene) impaired activity (Fig. 8.6b) (Chen et al. 2013b). Interestingly, species in the genus *Akodon*, which contain high proportions of sex-reversed XY females, have only a single GRT in their respective Sry proteins (Sánchez et al. 2010), offering such a foreshortened GRD as a potential explanation for chromosomal sex reversal (Fig. 8.7a). These findings led to the hypothesis that the mouse GRD (and its foreshortened counterpart in the rat) functions as transcriptional activation domains and further that this ancillary function can compensate for deleterious (and not just neutral) mutations in the HMG box itself (Chen et al. 2013b). The in vivo relevance of these findings has been corroborated in transgenic XX mice, in which the 20-repeat mouse GRD also preserves the proteolytic stability of the divergent HMG box (Zhao et al. 2014).

Together, these considerations have motivated a fresh look at the natural history of Muroidea, including the fossil record of its radiation and molecular correlates (Steppan et al. 2004). Such studies have placed the origins of Muroidea in the early Miocene ($\sim$25 million years ago; Steppan et al. 2004). This period saw the evolution of >1300 extant muroids, which account for nearly one-third of all mammalian species, with additional (and presumably many) extinct lineages (Michaux et al. 2001; Schenk et al. 2013). Although several researchers have attributed such rapid speciation to geographic radiation in both the New- and Old Worlds (Steppan et al. 2004; Jansa et al. 2009)—amplified by an accelerated mutational rate

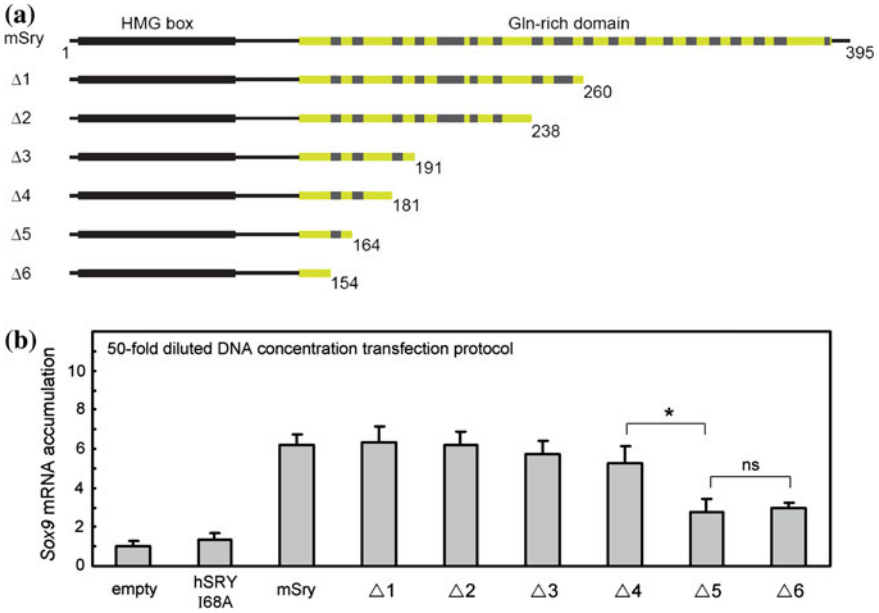


Fig. 8.6 Threshold repeat length in mouse Sry. **a** Deletion analysis of the Gln-rich domain in murine Sry. GRD is shown in chartreuse and intervening non-Gln-rich elements in *gray*. Substitution I68A in full-length human SRY blocks specific DNA binding (Weiss et al. 1997). Murine Sry contains 20 GRTs (chartreuse blocks). The following constructs were designed: Δ1, 10 GRTs; Δ2, 8 GRTs; Δ3, 4 GRTs; Δ4, 3 GRTs (which corresponds to the number in rat Sry); Δ5, 2 GRTs; and Δ6, 1 GRT. **b** Functional assessment of deletion Sry constructs in rat embryonic gonadal ridge cell line CH34. Data were obtained from Chen et al. (2013b)

and short generation time (Martin and Palumbi 1993)—other families within Rodentia share these characteristics and yet are significantly less speciose (Wilson and Reeder 2005). In particular, the breakneck pace of evolution within the Sigmodontine subfamily of New World muroid rodents, which contains the genus *Akodon* (superfamily Muroidea; family Cricetidae), has long puzzled paleontologists and evolutionary biologists; dating their origins and dispersal remains uncertain (for review, see Leite et al. 2014). Branches of this taxon migrated from Central to South America, presumably requiring a land bridge. According to geological evidence, however, this bridge (the Isthmus of Panama) became fully formed only ∼4 million years ago (Coates et al. 1992).[3] Is it plausible that such a marked radiation from a Sigmodontine ancestor could have occurred in this time period? Indeed, some paleontologists have suggested contrariwise that, given the

[3]Since submission of this chapter, a study by Montes et al has suggested an earlier date for the formation of the Panamanian land bridge. Based on uranium-lead dating of sedimentary-river deposits in the Andes, the authors have concluded that the Isthmus of Panama existed in the middle Miocene, approximately 15 million years ago (Montes et al. 2015). This revised geological estimate promises in part to explain the broad radiation of Muroid rodents in South America.

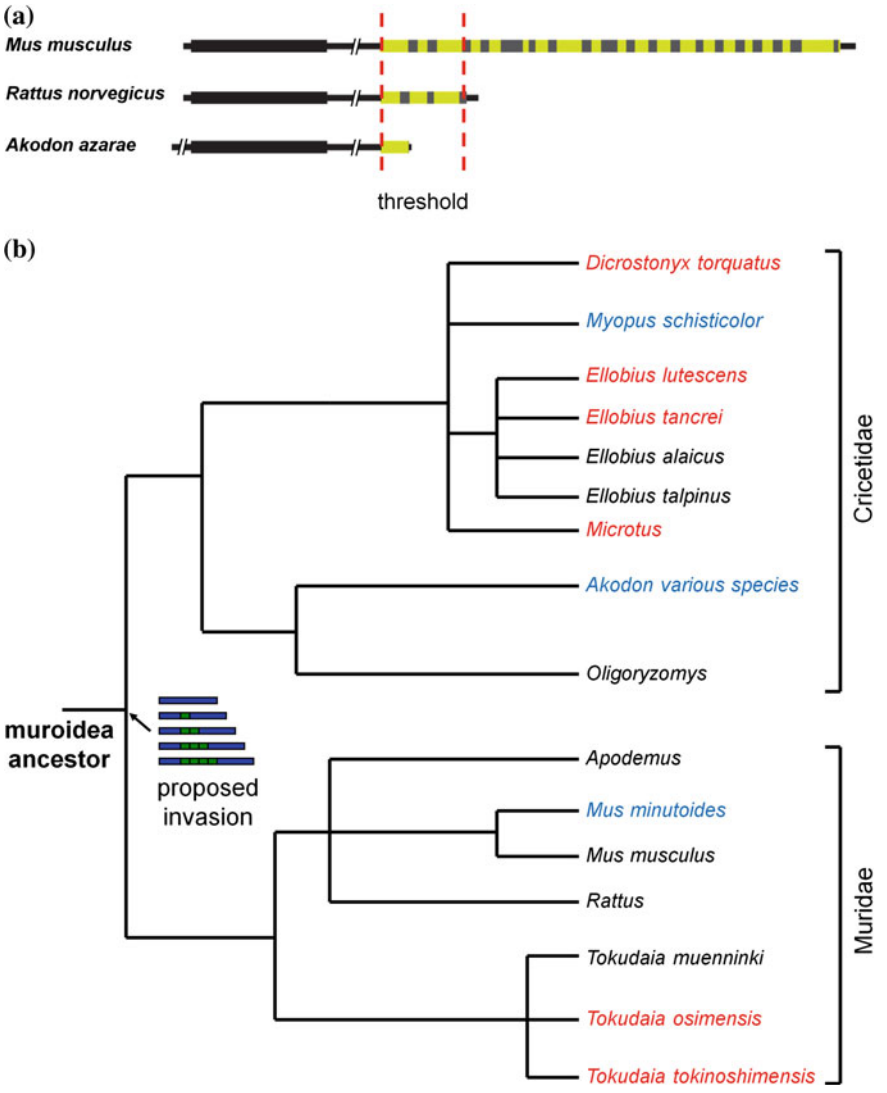


Fig. 8.7 Evolution of biological novelty in Muroidea. **a** Domain organization of Sry in a representative *Akodon* species exhibiting fertile XY females is shown in relation to mouse and rat Sry. Respective GRDs are shown in chartreuse and non-Gln-rich linkers in *gray*. The Akodon GRD contains only a single Gln-rich tract and thus lies below the 3-repeat threshold established in Fig. 8.6 (above), suggesting attenuation of its transcriptional activity. **b** Radiation of Muroidea highlighting the anomalous emergence of non-Y-dependent sex determination (*red*) and XY sex reversal (*blue*)

recency of this date, an ancestral migration must have preceded the land bridge. Contingent waif dispersal (i.e., overwater rafting), for example, has been proposed to have occurred at least 20 million years ago (Hershkovitz 1969; Savage 1974),

and if so, it would have enabled diversification of Sigmodontine rodents in South America in accordance with general estimates of speciation rates (Leite et al. 2014). Although (or because) these competing views are difficult to distinguish based on an incomplete fossil record, this question has engendered continuing debate.

We envisage that mutation-driven evolution (Nei 2013) provides a parsimonious explanation for the seeming discrepancies between geologic events (such as the formation of the Isthmus of Panama) and anomalous speciation rates in Muroidea. If microsatellite instability within the *Sry* gene of a muroid ancestor might indeed have provided a mechanism of reproductive isolation and at a rate higher than that predicted by the general mutational clock, then there would be no need to invoke the "skyhook" of distant waif dispersal [in the sense of Dennett (1995)], a presumption not readily supported by the available fossil record. The Sigmodontine radiation may thus have occurred without paradox or contradiction. We anticipate that molecular mechanisms of mutation-driven evolution, as discussed below, will provide an insight into the temporal dynamics of the Great American Biotic Interchange (Savage 1974).

8.4 Genetic Capacitors and the Evolution of Evolvability

"There is intrinsic interest," observed S.J. Steppan and colleagues in relation to the radiation of Muroidea, "in understanding why this group is so much more diverse than any comparable mammalian clade, and more broadly, has sustained one of the highest net speciation rates among land vertebrates" (Steppan et al. 2004). Molecular insight into the generation of biological novelty and its pace has motivated the general notion of "genetic capacitors" (Rutherford and Lindquist 1998). Formulated by analogy to electrical capacitors, which store charge in a circuit in a form amenable to sudden discharge, the biological analog is posited to store genetic variation in a form susceptible to sudden unmasking. This analogy was first proposed by Rutherford and Lindquist in studies of mutations in a heat-shock protein (HSP90), a chaperone involved in protein folding and misfolding. Strikingly, in *Drosophila melanogaster,* such mutations were found to unmask preexisting but cryptic genetic variation in morphogenetic pathways (Rutherford and Lindquist 1998). Because HSP90 has multiple and diverse protein "clients" (Chatterjee et al. 2013), each in principle sensitive to graded levels of chaperone activity, genetic variation in the function of HSP90 itself was proposed to underlie discontinuities in the fossil record (Cossins 1998; Baker 2006). The dual role of HSP90—both to ensure phenotypic stability despite genotypic variation and to unmask such variation in the face of environmental change—has led to its designation as "Waddington's widget" (Sollars et al. 2003), i.e., a molecular device that enables a transition from one canalized developmental plan to another in a rugged fitness landscape (Waddington 1959).

The concept of a genetic capacitor is more general than its particular molecular embodiment in HSP90. We have highlighted the extension of this metaphor to a strategic DNA microsatellite (Chen et al. 2013b). Three properties of microsatellites

may in principle contribute to a capacitor function: (i) Expansion and contraction of the DNA repeat, if occurring in gametogenesis, can engender significant genetic change in one generation at rates higher than the overall mutation rate (Rando and Verstrepen 2007); (ii) such change can alter thresholds of gene expression (if in regulatory regions of the genome) or protein function (if within the coding region of a gene); and (iii) linkage between extent of microsatellite instability and environmental stress (Chatterjee et al. 2013) can provide a mechanism by which cryptic variation is unmasked to enable adaptation. These properties are well known in cancer biology (wherein adaptation pertains to evolution of malignant traits such as invasiveness and metastatic spread to new "niches"; Greaves and Maley 2012). Extension of Waddington's widget to the muroid *Sry* microsatellite is immediate. Not only do muroid *Sry* alleles exhibit extensive variation in repeat length, but the function of the encoded GRD exhibits a threshold dependence in its ability to buffer protein stability and transcriptional activity. Like HSP90, the microsatellite-encoded capacitor of muroid Sry may modify the properties of multiple proteins (i.e., "clients"), including not only its tethered DNA-binding motif (the HMG box; Bowles et al. 1999) but also putative partner proteins in the enhanceosome complex. Yet the critical role of this tethered client in male sex determination (and hence reproductive fitness) would magnify the impact of any unmasked variation. In the face of the biophysical degeneration of the Sry HMG box (above), we imagine that recruitment of novel regulatory genes would be an evolutionary necessity. Although the phenotypic consequences of microsatellite contraction in an individual XY embryo would most likely be an evolutionary dead end (as exemplified by infertile XY human females in Swyer Syndrome; Michala et al. 2008), the geological timescale would provide an opportunity for favorable combinations of genetic variation to occur (as exemplified by the rare fertile fathers of sterile XY daughters; Chen et al. 2013a). The analogous phenomena of "inherited human sex reversal" (due to Sry variants of partial function; Chen et al. 2013a) and murine intersexual phenotypes associated with Y chromosome/autosome incompatibility (due in part to differences in degree of Sry expression; Albrecht et al. 2003; Bullejos and Koopman 2005) highlight the contribution of genetic background to the phenotypic output of a complex gene regulatory network. Thus, in the muroid radiation (25 million years) or its Sigmodontine subradiation (4–8 million years), the stark selective pressure of male fertility could rationalize emergence of fertile XY females within genera *Akodon*, *Myopus*, and *Mus* (Jiménez et al. 2012; Veyrunes et al. 2010) and recruitment of non-Sry-dependent upstream signals (within genera *Tokudaia*, *Ellobius*, and *Microtus*; Jiménez et al. 2012) as outlined in Fig. 8.7b.

8.5 Mutation-Driven Evolution and "Darwinian Heresies"

"The intellectual landscape of Darwinism for the last 150 years," observed Abigail Lustig, "bears a certain resemblance to Germany during the Thirty years' War [in the seventeenth century]. Sects and churches, preachers and dissenters, holy

warriors and theocrats vie with each other for the hearts of the faithful and the minds of the unconverted, all too often leaving scorched earth behind" (Lustig 2004). The model of microsatellite-driven radiation of Muroidea, as outlined in this chapter, bears upon these and other enduring controversies in evolutionary theory.

Sympatric speciation. Defined as occurring within overlapping ranges, this process remains controversial and is most commonly accepted in plants and insects (Mayr 1947; Wolinsky 2010). Sudden change in ploidy, for example, more common among plants, can cause immediate reproductive isolation from cognate neighbors (Wolinsky 2010). Similarly, larvae of herbivorous insects can imprint upon the plant species of hatching, enabling alternative imprinting on a novel plant species, in turn creating reproductively isolated insect populations (Berlocher and Feder 2002). In most animals, however, sufficient gene flow occurs between sympatric populations to prevent speciation, casting doubt on the generality of this mode (Mayr 1947). Although skepticism was widely shared in the era of the Modern Synthesis, the debate has been reinvigorated by studies of African cichlid fish and their recent radiation (Schliewen et al. 1994). Like changes in ploidy, contraction of the *Sry* microsatellite can rapidly introduce a reproductive barrier between sympatric subpopulations of a given rodent species: the original population could then be divided by genetic background into one group in which the contracted *Sry* allele yields fertile XY males and fertile XY females and another in which the contracted allele yields infertile progeny.[4] The fitness cost of even partial impairment of fertility could lead to loss of that Y chromosome from the latter population in a small number of generations. Accordingly, we speculate that the radiation of Muroidea, accelerated by discharge of the *Sry* capacitor, occurred at least in part among overlapping rodent populations (Erlinge et al. 1990; Brown and Kurzius 1987; Patton et al. 1996).

Lamarckian speciation. Environmental stress can have an important impact on the internal workings of cells, including DNA transactions (Chatterjee et al. 2013). Cellular stress responses can in particular enhance microsatellite instability (Chatterjee et al. 2013). Perhaps for this reason, many of the genes involved in stress response contain microsatellites (Kozlowski et al. 2010), raising the possibility that such instability underlies aspects of the cellular response to stressful environments. Linkage of a DNA transaction to environmental stress thus provides a potential mechanism for inheritance of acquired characteristics, a pre-Darwinian evolutionary idea (Lamarck 1809). Although Lamarck's theory has been widely ignored since Darwin, recent investigations into inheritance of epigenetic changes have shed new light on this idea as an ancillary feature of evolutionary biology. In the Muroid radiation, promotion of *Sry*-linked microsatellite instability in response

[4]Infertility may ordinarily arise in such cases due to antagonistic effects of other Y-chromosome linked genes. Indeed, the presence of Y-linked gene *Zfy2* (near *Sry* on the short arm) was recently found to cause loss of fertility in outbred (Sry^-) sex-reversed XY female mice as its deletion partially restored fertility (Vernet et al. 2014). Analogous incompatibility in genetic background may contribute to human 46, XY gonadal dysgenesis in association with SRY mutations as a mechanism of infertility in Swyer's syndrome (Michala et al. 2008).

to environmental stress (whether imposed by inter-species competition, drought, change in habitat, or sexual selection) could confer reproductive barriers as described above. Such "Lamarckian speciation" could underlie the observed acceleration of allopatric or sympatric divergence.

Enhanced evolvability within a taxon has, as a seeming paradox, its origins in the non-robustness of DNA microsatellites and the tenuous threshold of the *Sry*-encoded developmental switch. Since the amplification or contraction of a microsatellite DNA domain is both more common and more reversible than is a point mutation (Kashi and King 2006), such genomic changes can allow evolution to "test the waters" in response to external stress and so make large-scale adaptations in a small number of generations. The increased evolvability provided by a microsatellite domain in a key regulator of development (like Sry) may be more general than this specific case. This model stands in contrast to the view that upstream TFs are more likely to exhibit broad conservation than downstream TFs. That this pattern and its opposite may both be observed, depending on the genetic program, highlights the intrinsic tension between evolvability and robustness (Brookfield 2009). Taken to an extreme, a robust gene regulatory network may be refractory to adaptive change. Such considerations have highlighted the evolution of evolvability itself as a frontier of current research (Wagner 2013). Families in Rodentia may differ in evolvability and *so in a measure of fitness defined at a taxonomic level above the species*. The survival of a genus (as distinct from a given species or individual genotype within that species) may depend on adaptive speciation to changes in environment broadly associated with geological or climactic events.

These reflections motivate consideration of a specific heresy: sex ratio biasing as a Lamarckian adaptation. In the genus *Akodon*, eight species exhibit polymorphic populations of fertile XY females (Bianchi 2002). This phenomenon is shared by other groups in Muroidea, including *Myopus Schistocolor* (Jiménez et al. 2012) and *Mus minutoides* (Veyrunes et al. 2010). In accordance with the Trivers–Willard hypothesis (Trivers and Willard 1973), it is intriguing to imagine a scenario in which a stressful environment leads to *Sry* microsatellite instability (as above) and in turn to female skewing of the sex ratio. Indeed, in Muroidea are found many examples of female-biased sex ratios (Jiménez et al. 2012). Although such skewing may initially be a spandrel of DNA mechanics (and unstable in classical population theoretic models; Fisher 1930), potential advantages of maternal investment in daughters could render such skewing advantageous, thereby aligning mutation-driven evolution with selection.

8.6 Testable Predictions

Our view of Muroidea and its anomalous radiation make predictions that may soon be testable given advances in genetic and genomic technologies (Mardis 2008). The next frontier will highlight the integration of classical taxonomy, ecology and the fossil record with molecular genetics (Steppan et al. 2004), including application of

transgenic technology in laboratory strains of mice to recapitulate divergent sex-determining systems as uncovered among *Akodon* species (with XY female subpopulations; Bianchi 2002) or as presumed to function in muroids that lack *Sry* entirely (Jiménez et al. 2012). The following hypotheses may motivate such studies.

(i) Sry *genes in Muroidea will be found to exhibit a broad range of microsatellite repeat lengths 3' to the HMG box.* Such variation may even occur within a species as Y chromosomal strains with paternal inheritance (analogous to mitochondrial DNA strains with maternal inheritance; Moritz et al. 1987). Corresponding Sry proteins will exhibit a commensurate range of C-terminal GRDs or, in the case of a frameshift, ARDs. Repeat lengths less than the critical threshold (3 tracts as defined above in a rat embryonic gonadal ridge cell line; Chen et al. 2013b) are by contrast predicted in species that contain a subset of fertile XY females.

(ii) *Muroid Sry HMG-box sequences will be found to exhibit greater sequence variation than within any other order of placental mammals and perhaps even greater than that spanned* in toto *by all such orders.* Amino acid substitutions in Muroidea, but not in other therian mammals, may be either neutral or deleterious. In species with strict XY male sex-determining systems, severity would be limited by extent of biochemical compensation provided by the microsatellite-encoded domain. More dramatic substitutions (such as those that prevent specific DNA binding entirely) may occur only in species with fertile XY females. We predict that this divergence will be seen only in Sry and not in any other upstream elements of gene-regulatory networks within these Muroid species.

(iii) *In species with XY females, variant Sry factors may either be without transcriptional activity (if a distinct Y chromosomal strain encodes a functional Sry) or with activity poised at the threshold of function*. In the first case, XY females and XY males would contain different Y chromosomes (Marın and Baker 1998), whereas in the second case, the male switch would depend on autosomal or X-linked variation (Ortiz et al. 1998); in laboratory mice, the importance of autosomal genetic background is well established (Yoshiki and Moriwaki 2006). *Sry* alleles with partial function as a TDF are likely to provide informative probes with which to define mechanistic boundaries of molecular function, such as allowed ranges of specific DNA affinity constants, DNA bend angles, or protein stability.[5]

[5]In genus *Akodon,* distinct *Sry* alleles coexist within the same species. In *A. azarae*, for example, the SRY sequences obtained from two XY females differ at consensus position 13 of the HMG box (i.e., corresponding to position 13 in mouse Sry and position 68 in human SRY): Met or Val (Sánchez et al. 2010). This side chain, designated the *cantilever*, inserts between DNA base pairs at a site of sharp DNA bending (King and Weiss 1993; Haqq et al. 1994; Murphy et al. 2001) and is conserved as Met, Phe, Ile, or Leu among therian Querymammals (substitution by Ala or Thr blocks specific DNA binding (Weiss et al. 1997). We speculate that the Val cantilever is non-functional and underlies, at least in part, the high percentage of XY sex reversal in this population.

(iv) *Non-Sry-dependent sex-determining systems emerged in those lineages wherein Sry first degenerated to the edge of function*. Systematic genomic sequencing of neighboring clades in the phylogeny of Muroidea, guided by "proband" species lacking Y chromosomes (e.g., *Tokudaia osimensis*; Soullier 1998), is likely to (a) provide molecular evidence of rapid divergence among extant Y chromosomes and (b) suggest candidate genes whose co-optation enabled emergence of a novel TDF. The latter may be investigated through studies of TES and its divergence (Kimura et al. 2014), since alternative mechanisms of *Sox9* activation in the gonadal ridge presumably reflect recruitment of a new upstream TES-binding TF or quantitative changes in expression of such canonical upstream TFs WT1 or SF1 included in Fig. 8.2 (Matsuzawa-Watanabe et al. 2003; Sekido and Lovell-Badge 2008).

8.7 Concluding Remarks

"The mystery of the beginning of all things is insoluble by us," Darwin remarked, "and I for one must be content to remain an agnostic" (Darwin 1887). Beginnings in the modern evolutionary sense—the origins of biological innovation are inextricably tied to an intrinsic interplay between DNA as bearer of information and DNA as a molecular entity subject to physical and chemical change. This chapter has sought to highlight a seeming paradox of natural history, the anomalous radiation of Muroidea, as an opportunity to resolve this tension.

Akodon and other genera in Muroidea in which non-canonical sex-determining mechanisms are evolving provide systems amenable to both molecular genetic and ecological studies. We envisage that our microsatellite hypothesis will provide a testable framework for the evolutionary impact of saltatory changes in DNA structure. Possible consequences include the rapid emergence of reproductive isolation among sympatric populations (Schliewen et al. 1994) and even (through stress-induced microsatellite instability; Chatterjee et al. 2013) to a Lamarckian mechanism of speciation. In this view, the contingent invasion of a trinucleotide repeat into the sex-determining region of an ancestral muroid Y chromosome set the stage for the accumulation of cryptic genetic variation in the DNA-binding domain of a master TF. Unmasking of such variation through microsatellite instability hastened the tempo of evolutionary change in the *Sry*-related gene regulatory network, leading even to recruitment of another TDF. The explosive speciation and biological novelty observed in Muroidea thus reflect the precarious missteps of repeat-associated replicative slippage in the molecular choreography of DNA replication.

Central to our proposed framework is the tenuous character of Sry as a developmental switch (Polanco and Koopman 2007; Chen et al. 2013a). Whereas a

considerable literature pertains to the origins of robustness (Waddington 1959) as general features of gene regulatory networks (Siegel and Bergman 2002), the existence and advantages of tenuousness have not been sufficiently recognized. Yet as evidenced in studies of murine Y chromosome/autosome incompatibility in male sex determination (Albrecht et al. 2003; Bullejos and Koopman 2005) and further supported by studies of human SRY variants associated with "inherited" XY sex reversal (i.e., variant SRY alleles shared by a fertile father and sterile daughter; Chen et al. 2013a), the transcriptional threshold of wild-type Sry lies close to the border of ambiguity. This vulnerable threshold is a critical feature of the microsatellite model as it would magnify the evolutionary opportunities (and dangers) of subtle changes in the biochemical properties of Sry, such as those associated with contraction of its microsatellite-encoded protein domain (Zhao et al. 2014). Such a model also rationalizes the preservation of the Sry-TDF system (with conservation of its HMG box) in mammalian orders and rodent families lacking the muroid microsatellite.

The heightened evolvability of muroid rodents and their successful radiation may thus reflect the side consequence of a molecular accident in chromosomal dynamics. The ability of a genetic capacitor (irrespective of its molecular embodiment) to enhance evolvability presumably depends on quantitative relationships between variation in biochemical function and downstream consequences for cellular differentiation and organismal development (Wagner 2012). The natural history of Muroidea thus provides a model in which to investigate whether a master switch in mammalian development may be tenuous—and if so, why. We propose that a developmental switch may evolve to the "edge of chaos"[6] (Kauffman and Johnsen 1991; Shmulevich et al. 2005) as a mechanism to generate phenotypic diversity (Chen et al. 2013a). Application to male sex determination is of special interest. In the nascent testis, quantitative variation in fetal Leydig cell function and testosterone secretion may lead to a spectrum of hormone-dependent male brain patterning (Carrer and Cambiasso 2009) and ultimately to variation in postnatal behaviors (Auyeung et al. 2009). Therein lies the essence of Darwin's heresy: Whether this intricate chain of consequences—extending from a tenuous male signal in embryogenesis to the dynamics of groups, tribes, and populations—underlies the evolution of social mammals as a case study in multilevel selection (Wilson et al. 2007). The natural history of mammals, perhaps including aspects of human behavioral diversity, may thus connect molecular principles of mutation-driven evolution (Nei 2013) to the precepts and predictions of sociobiology (Wilson 1980).

[6]The term *edge of chaos* originally pertained to transition phenomena in cellular automata capable of universal computation (Langton 1990). Stimulated by innovative evolutionary models (Kauffman and Johnsen 1991), this notion provides a general metaphor for critical boundaries between organization and dysgenesis in biological systems, including cellular differentiation (Shmulevich et al. 2005). Dynamic competition between male- and female-specific nonlinear GRNs in the nascent gonad encompasses alternative *basins of attraction*, respectively, leading to Sertoli- or granulosa cell fates.

Acknowledgments We thank JD Racca and NB Phillips (CWRU) for discussion and communication of results prior to publication; E Haas (Bar-Ilan University) for collaborative fluorescence studies; C-Y King, JD Radek, and A Jansco-Radek for participation in early stages of this work; J Graves (University of Melbourne) for her encouragement and scientific example; Robin Lovell-Badge for his astute critique and gracious advice; and PK Donahoe (Massachusetts General Hospital and Harvard Medical School) for cell line CH34 and advice throughout these studies. MAW is grateful to B. Baker (HHMI Janelia Farms) and the late FA Jenkins Jr (Harvard University) for inspiration. The present studies were supported in part by a grant from the US National Institutes of Health (GM080505).

References

Albrecht KH, Young M, Washburn LL, Eicher EM (2003) Sry expression level and protein isoform differences play a role in abnormal testis development in C57BL/6 J mice carrying certain *Sry* alleles. Genetics 164:277–288

Auyeung B, Baron-Cohen S, Ashwin E, Knickmeyer R, Taylor K, Hackett G, Hines M (2009) Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. Psychol Sci 20:144–148

Baker ME (2006) The genetic response to snowball earth: role of HSP90 in the Cambrian explosion. Geobiol 4:11–14

Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? Ann Rev Entomol 47:773–815

Bianchi NO (2002) Akodon sex reversed females: the never ending story. Cytogen Gen Res 96:60–65

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. Ann Rev Ecol Evol System 38:459–487

Bowles J, Cooper L, Berkman J, Koopman P (1999) SRY requires a CAG repeat domain for male determination in *Mus musculus*. Nat Gen 22:405–408

Brouwer JR, Willemsen R, Oostra BA (2009) Microsatellite repeat instability and neurological disease. BioEssays 31:71–83

Brookfield JF (2009) Evolution and evolvability: celebrating Darwin 200. Biol Lett 5:44–46

Brown JH, Kurzius MA (1987) Composition of desert rodent faunas: combinations of coexisting species. Ann Zool Fennici 24:227–237

Bullejos M, Koopman P (2005) Delayed *Sry* and *Sox9* expression in developing mouse gonads underlies B6-Y^{DOM} sex reversal. Dev Biol 278:473–481

Canning CA, Lovell-Badge R (2002) Sry and sex determination: how lazy can it be. Trends Gen 18:111–113

Carrer HF, Cambiasso MJ (2009) Sexual differentiation of the brain: genetic, hormonal, and trophic factors. In: Janigro D (ed) Mammalian brain development. Humana Press, pp 1–15

Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25–36

Chatterjee N, Santillan BA, Wilson JH (2013) Microsatellite repeats: canaries in the coalmine. In: Mittelman D (ed) Stress-Induced Mutagenesis. Springer, New York, pp 119–150

Chen YS, Racca JD, Phillips NB, Weiss MA (2013a) Inherited human sex reversal due to impaired nucleocytoplasmic trafficking of SRY defines a male transcriptional threshold. Proc Natl Acad Sci USA 110:E3567–E3576

Chen YS, Racca JD, Sequiera PW, Phillips NB, Weiss MA (2013b) Microsatellite-encoded domain in rodent Sry functions as a genetic capacitor to enable the rapid evolution of biological novelty. Proc Natl Acad Sci USA 110:E3061–E3070

Coates AG, Jackson JB, Collins LS, Cronin TM, Dowsett HJ, Bybell LM, Jung P, Obando JA (1992) Closure of the Isthmus of Panama: the near-shore marine record of costa rica and western Panama. Geol Soc Am Bull 104:814–828
Cossins A (1998) Cryptic clues revealed. Nature 396:309–310
Darwin C (1872) The origin of species, 6th edn. John Murry, London
Darwin C (1887) The autobiography of Charles Darwin. In: Regal B (ed) Barnes and Noble, 2005 edn, p 70
De Vries H (1909) The mutation theory, vol 1. Open Court Publishing Company, Chicago
Dennett DC (1995) Darwin's dangerous idea. The Sciences 35:34–40
Eldredge N, Gould, SJ (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM (ed) Models in Paleobiology. Freeman Cooper, San Francisco, pp 82–115
Erlinge S, Hoogenboom I, Agrell J, Nelson J, Sandell M (1990) Density-related home-range size and overlap in adult field voles (Microtus agrestis) in southern Sweden. J Mammal 71:597–603
Ferrier DE, Holland PW (2001) Ancient origin of the Hox gene cluster. Nature Rev Gen 2:33–38
Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford
Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 372:525–529
Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. Science 279:853–856
Gehring WJ, Ikeo K (1999) Pax 6: mastering eye morphogenesis and eye evolution. Trends Gen 15:371–377
Goodfellow PN, Lovell-Badge R (1993) SRY and sex determination in mammals. Ann Rev Gen 27:71–92
Gordon CT, Tan TY, Benko S, Fitzpatrick D, Lyonnet S, Farlie PG (2009) Long-range regulation at the SOX9 locus in development and disease. J Med Genet 46:649–656
Gövert F, Schneider S (2013) Huntington's disease and Huntington's disease-like syndromes: an overview. Curr Opin Neurol 26:420–427
Graves JA (2013) How to evolve new vertebrate sex determining genes. Dev Dyn 242:354–359
Graves JAM, Ferguson-Smith M, McLaren A, Mittwoch U, Renfree M, Burgoyne P (1995) The evolution of mammalian sex chromosomes and the origin of sex determining genes. Philo Trans Royal Soc B (Biol Sci) 350:305–312
Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481:306–313
Gubbay J, Vivian N, Economou A, Jackson D, Goodfellow P, Lovell-Badge R (1992) Inverted repeat structure of the Sry locus in mice. Proc Natl Acad Sci USA 89:7953–7957
Haqq CM, Donahoe PK (1998) Regulation of sexual dimorphism in mammals. Physiol Rev 78:1–33
Haqq CM, King CY, Ukiyama E, Falsafi S, Haqq TN, Donahoe PK, Weiss MA (1944) Molecular basis of mammalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY. Science 266:1494–1500
Hershkovitz P (1969) The recent mammals of the Neotropical region: a zoogeographic and ecological review. Q Rev Biol 44:1–70
Hodgkin J (2005) Karyotype, ploidy, and gene dosage. WormBook, pp 1–9
Hoekstra HE, Edwards SV (2000) Multiple origins of XY female mice (genus Akodon): phylogenetic and chromosomal evidence. Proc Roy Soc Lond B (Biol Sci) 267:1825–1831
Huxley J (1942) Evolution. The modern synthesis. Allen & Unwin, London
Jager M, Quéinnec E, Houliston E, Manuel M (2006) Expansion of the SOX gene family predated the emergence of the Bilateria. Mol Phylogenet Evol 39:468–477
Jäger RJ, Anvret M, Hall K, Scherer G (1990) A human XY female with a frame shift mutation in the candidate testis-determining gene *SRY*. Nature 348:452–454
Jansa SA, Giarla TC, Lim BK (2009) The phylogenetic position of the rodent genus *Typhlomys* and the geographic origin of Muroidea. J Mammal 90:1083–1094
Jiménez R, Barrionuevo FJ, Burgos M (2012) Natural exceptions to normal gonad development in mammals. Sex Dev 7:147–162

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends Gen 22:253–259

Kauffman SA, Johnsen S (1991) Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. J Theor Biol 149:467–505

Kent J, Wheatley SC, Andrews JE, Sinclair AH, Koopman P (1996) A male-specific role for SOX9 in vertebrate sex determination. Dev 122:2813–2822

Kimpara T, Takeda A, Watanabe K, Itoyama Y, Ikawa S, Watanabe M, Arai H, Sasaki H, Higuchi S, Okita N (1997) Microsatellite polymorphism in the human heme oxygenase-1 gene promoter and its application in association studies with Alzheimer and Parkinson disease. Hum Gen 100:145–147

Kimura R, Murata C, Kuroki Y, Kuroiwa A (2014) Mutations in the testis-specific enhancer of SOX9 in the SRY independent sex-determining mechanism in the genus *Tokudaia*. PLoS ONE 9:e108779

King CY, Weiss MA (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. Proc natl Acad Sci USA 90:11990–11994

Knower KC, Kelly S, Ludbrook LM, Bagheri-Fam S, Sim H, Bernard P, Sekido R, Lovell-Badge R, Harley VR (2011) Failure of *SOX9* regulation in 46XY disorders of sex development with SRY, SOX9 and SF1 mutations. PLoS ONE 6:e17751

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288–298

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R (1991) Male development of chromosomally female mice transgenic for Sry. Nature 351:117–121

Kozlowski P, de Mezer M, Krzyzosiak WJ (2010) Trinucleotide repeats in human genome and exome. Nucleic Acids Res 38:4027–4039

Kraak SBM (2002) Sex-determining mechanisms in vertebrates. In: Hardy ICW (ed) Sex ratios. Campbridge University Press, Campbridge, pp 158–177

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964−967

Lamarck JB (1809) *Zoological Philosophy*, Trans. Hugh Elliot (New York: Hafner, 1963) pp 109–111

Landles C, Bates GP (2004) Huntingtin and the molecular pathogenesis of Huntington's disease. EMBO Rep 5:958–963

Langton CG (1990) Computation at the edge of chaos. Physica D 42:12–37

Lefebvre V, Dumitriu B, Penzo-Méndez A, Han Y, Pallavi B (2007) Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol 39:2195–2214

Leite RN, Kolokotronis S-O, Almeida FC, Werneck FP, Rogers DS, Weksler M (2014) In the wake of invasion: tracing the historical biogeography of the South American cricetid radiation (Rodentia, Sigmodontinae). PLoS ONE 9:e100687

Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Loeb LA (1994) Microsatellite instability: marker of a mutator phenotype in cancer. Cancer Res 54:5059–5063

Loeb LA (2011) Human cancers express mutator phenotypes: origin, consequences and targeting. Nat Rev Cancer 11:450–457

Lustig A (2004) Biologists on a crusade in *Darwinian Heresie*. In: Lustig A, Richards RJ, Ruse M (eds). Cambridge University Press, Cambridge, pp 1–13

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Gen 24:133–141

Marın I, Baker BS (1998) The evolutionary dynamics of sex determination. Science 281:1990–1994

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA 90:4087–4091

Matsuzawa-Watanabe Y, Inoue J-I, Semba K (2003) Transcriptional activity of testis-determining factor SRY is modulated by the Wilms' tumor 1 gene product, WT1. Oncogene 22:7900–7904
Mayr E (1947) Ecological factors in speciation. Evolution 1:263–288
Michala L, Goswami D, Creighton SM, Conway GS (2008) Swyer syndrome: presentation and outcomes. BJOG 115:737–741
Michaux J, Reyes A, Catzeflis F (2001) Evolutionary history of the most speciose mammals: molecular phylogeny of muroid rodents. Mol Biol Evol 18:2017–2031
Montes C, Cardona A, Jaramillo C, Pardo A, Silva JC, Valencia V, Ayala C, Pérez-Angel LC, Rodriguez-Parra LA, Ramirez V, Niño H (2015) Middle Miocene closure of the Central American Seaway. Science 348:226–229
Moritz C, Dowling T, Brown W (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann Rev Ecol Syst 18:269–292
Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM (2001) Structural basis for SRY-dependent 46-X, Y sex reversal: modulation of DNA bending by a naturally occuring point mutation. J Mol Biol 312:481–499
Nei M (2013) Mutation-driven evolution. Oxford University Press, Oxford
Ortiz MI, Dalmasso G, Dezi R, Senn EP, Lisanti J (1998) A C-band polymorphism of the X Chromosome in *Akodon azarae* (Rodentia, Cricetidae). Cytologia 63:365–369
Patton J, Silva MD, Malcolm J (1996) Hierarchical genetic structure and gene flow in three sympatric species of Amazonian rodents. Mol Ecol 5:229–238
Phillips NB, Nikolskaya T, Jancso-Radek A, Ittah V, Jiang F, Singh R, Haas E, Weiss MA (2004) Sry-directed sex reversal in transgenic mice is robust to enhanced DNA bending: comparison of human and murine in HMG boxes. Biochemistry 43:7066–7081
Pieau C, Dorizzi M, Richard-Mercier N (1999) Temperature-dependent sex determination and gonadal differentiation in reptiles. Cell Mol Life Sci 55:887–900
Polanco JC, Koopman P (2007) Sry and the hesitant beginnings of male development. Dev Biol 302:13–24
Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell 128:655–668
Richard GF, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Micro Mol Biol Rev 72:686–727
Robert JS (2004) Embryology, epigenesis and evolution: taking development seriously. Cambridge University Press, Cambridge
Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. Nature 396:336–342
Sánchez A, Marchal JA, Romero-Fernández I, Pinna-Senn E, Ortiz MI, Bella JL, Lisanti JA (2010) No differences in the *Sry* gene between males and XY females in *Akodon* (Rodentia, Cricetidae). Sex Dev 4:155–161
Savage JM (1974) The isthmian link and the evolution of Neotropical mammals. Contrib Sci Nat Hist Mus Los Angeles County 260:1–51
Schenk JJ, Rowe KC, Steppan SJ (2013) Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. Syst Biol 62:837–864
Schliewen UK, Tautz D, Pääbo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368:629–632
Sekido R, Lovell-Badge R (2008) Sex determination involves synergistic action of SRY and SF1 on a specific *Sox9* enhancer. Nature 453:930–934
Shmulevich I, Kauffman SA, Aldana M (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. Proc Natl Acad Sci USA 102:13439–13444
Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci USA 99:10528–10532
Soullier S, Hanni C, Catzeflis F, Berta P, Laudet V (1998) Male sex determination in the spiny rat Tokudaia osimensis (Rodentia: Muridae) is not Sry dependent. Mammal Genome 9:590–592
Sollars V, Lu X, Xiao L, Wang X, Garfinkel MD, Ruden DM (2003) Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat Genet 33:70–74

Steppan SJ, Adkins RM, Anderson J (2004) Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol 53:533–553

Stevanovic M, Lovell-Badge R, Collignon J, Goodfellow PN (1993) *SOX3* is an X-linked gene related to SRY. Hum Mol Genet 2:2013–2018

Sutton E, Hughes J, White S, Sekido R, Tan J, Arboleda V, Rogers N, Knower K, Rowley L, Eyre H, Rizzoti K, McAninch D, Goncalves J, Slee J, Turbitt E, Bruno D, Bengtsson H, Harley V, Vilain E, Sinclair A, Lovell-Badge R, Thomas P (2011) Identification of SOX3 as an XX male sex reversal gene in mice and humans. J Clin Invest 121:328–341

Trivers RL, Willard DE (1973) Natural selection of parental ability to vary the sex ratio of offspring. Science 179:90–92

Valleley E, Cartwright EJ, Croft NJ, Markham AF, Coletta PL (2001) Characterisation and expression of *Sox9* in the leopard gecko, Eublepharis macularius. J Exp Zoology 291:85–91

Vernet N, Szot M, Mahadevaiah SK, Ellis PJ, Decarpentrie F, Ojarikre OA, Rattigan A, taketo T, Burgoyne PS (2014) The expression of Y-linked Zfy2 in XY mouse oocytes leads to frequent meiosis 2 defects, a high incidence of subsequent early cleavage stage arrest and infertility. Development 141:855–866

Veyrunes F, Chevret P, Catalan J, Castiglia R, Watson J, Dobigny G, Robinson TJ, Britton-Davidian J (2010) A novel sex determination system in a close relative of the house mouse. Proc Biol Sci 277:1049–1056

Waddington CH (1959) Canalization of development and genetic assimilation of acquired characters. Nature 183:1654–1655

Wagner A (2012) The role of robustness in phenotypic adaptation and innovation. Proc Biol Sci 279:1249–1258

Wagner A (2013) Robustness and evolvability in living systems. Princeton University Press, Princeton, pp 217–228

Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene *SOX9*. Cell 79:1111–1120

Wallis MC, Waters PD, Graves JA (2008) Sex determination in mammals-before and after the evolution of SRY. Cell Mol Life Sci 65:3182–3195

Weiss MA, Ukiyama E, King CY (1997) The SRY cantilever motif discriminates between sequence-and structure- specific DNA recognition: alanine mutagenesis of an HMG box. J Biomol Struct Dyn 15:177–184

Wilson DE, Reeder DM (2005) Mammal species of the world: a taxonomic and geographic teference, vol 1. Johns Hopkins University Press, Baltimore

Wilson DS, Vugt MV, O'Gorman R (2007) Multilevel selection theory and major evolutionary transitions. Curr Dir Psychol Sci 17:4

Wilson EO (1980) Sociobiology. Belknap Press of Harvard University Press, Cambridge

Wilson M, Koopman P (2002) Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Gen Dev 12:441–446

Wolinsky H (2010) The puzzle of sympatry. EMBO Rep 11:830–833

Yoshiki A, Moriwaki K (2006) Mouse phenome research: implications of genetic background. ILAR J 47:94–102

Zhao L, Ng ET, Davidson T-L, Longmuss E, Urschitz J, Elston M, Moisyadi S, Bowles J, Koopman P (2014) Structure–function analysis of mouse Sry reveals dual essential roles of the C-terminal polyglutamine tract in sex determination. Proc Natl Acad Sci USA 111:11768–11773

Part III
Evolutionary Mechanisms

Chapter 9
Adaptive Diversification in Coevolutionary Systems

C. Hui, H.O. Minoarivelo, S. Nuwagaba and A. Ramanantoanina

Abstract Coevolution can trigger frequency-dependent selection by reciprocal effects on the fitness of involved species. Through directional and disruptive selection, coevolution can lead to rich evolutionary possibilities. It can be classified by major types of biotic interactions (mutualism and antagonism) or by the number of species involved (specific, diffuse and escape-and-radiate coevolution). Using two mainstream methods for studying the evolution of quantitative traits [adaptive dynamics (AD) based on canonical equations and evolutionary distribution (ED) based on trait diffusion], we examine three coevolutionary systems, including those driven by mutualistic and antagonistic interactions, as well as food webs. Results highlight the importance of trait-mediated competition, assortative cross-trophic interactions and consumption niche width (dietary width) on adaptive diversification in these coevolutionary systems. Interactions between two species can lead to diffuse and escape-and-radiate coevolution, making coevolutionary networks an ideal model for studying complex adaptive systems.

9.1 Introduction

Evolutionary adaptation is traditionally viewed as a hill-climbing and niche-filling process in a static fitness landscape, and the potential diversification from such adaptation often occurs allopatrically along an environmental gradient or through the restriction of gene flows by geographical barriers. Consequently, the number of species that a local ecosystem can hold depends on the intensity of niche competition and the carrying capacity of the environment. Coevolution, in contrast, often

C. Hui (✉) · H.O. Minoarivelo · S. Nuwagaba · A. Ramanantoanina
Department of Mathematical Sciences, Stellenbosch University, Matieland 7602, South Africa
e-mail: chui@sun.ac.za

C. Hui · H.O. Minoarivelo · S. Nuwagaba · A. Ramanantoanina
African Institute for Mathematical Sciences, Cape Town 7945, South Africa

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_9

triggers frequency-dependent selection where the evolutionary change in one species can lead to a reciprocal change in another species to balance their fitness. Such a dynamic fitness landscape then allows species to converge and diverge, respectively, through directional and disruptive selection, resulting in a wide variety of evolutionary possibilities.

The phenomenon of coevolution was first speculated by Darwin in 1862 and related to a moth with a 30-cm-long proboscis that pollinates an orchid of similar shape. The former was later discovered in 1903 to be the sphinx moth *Xanthopan morganii praedicta* and the latter, the Madagascan orchid *Angraecum sesquipedale*. The two coevolving species, as Darwin (1862) put it, 'might slowly become, either simultaneously or one after the other, modified and adapted in the most perfect manner, by the continued preservation of individuals presenting mutual and slightly favourable deviation of structure'. Coevolution can typically be classified according to the type of biotic interactions: largely mutualistic interactions (e.g. pollination and seed dispersal networks) and antagonistic interactions (e.g. predation and parasitism networks, as well as food webs).

Species maintain a mutualistic interaction by providing each other with benefits (fitness gains) as is the case of the Madagascan orchid and the sphinx moth. In this pollination syndrome, the moth pollinates the flower and in return is rewarded the sugar-rich nectar from the orchid. By changing the interacting functional traits through evolution (e.g. the proboscis of pollinators and the floral tube of flowers), mutualistic interactions can lead to intriguing coevolutionary games. Some long-standing mutualistic interactions can lead to the symbioses of both partners such as the symbiotic mycorrhizas formed in many plants with glomeromycete fungi. The fungus helps the plant obtain water and phosphates and is rewarded in turn with carbohydrates from the plant. To this end, some have argued that an organism is better understood as a compound system together with its symbionts (Gilbert et al. 2012), such as the extreme case in the evolution of mitochondria from purple bacteria and chloroplasts from cyanobacteria (Moran 2007). Maintaining a symbiotic or mutualistic relationship can be costly. As such, a major challenge is to unveil the protective mechanism that the involved partners have adopted for discerning and correcting the cheating behaviour which can be disastrous to the functioning of the system (Pellmyr and Leebens-Mack 1999).

Antagonistic interactions often occur through the mediation between the foraging traits of predators and the anti-foraging traits of their prey, such as between the speed of cheetahs and the agility of gazelles, between the fish stock and fishery policies (Landi et al. 2015) and between the toxicity of rough skinned newts (*Taricha granulosa*) and the resistance of garter snakes (*Thamnophis sirtalis*) (Hanifin et al. 2008). The drastic antagonistic warfare between plants and herbivores has resulted in the syntheses of diverse secondary compounds by plants as a defence mechanism against herbivores (Fraenkel 1959). Coevolution via antagonistic interactions can also lead to interesting phenomena of aposematism and mimicry (Mallet 2010), such as the coloration in the poison dart frog *Ranitomeya imitator* (Chouteau and Angers 2012). Again, the key to elucidating an antagonistic

interaction is to identify the interacting traits that affect the predator's energy intake and the prey's survival.

Antagonism is also typical between the arms race of hosts and their parasites or pathogens. Reed warblers distinguish artificial eggs closely resembling their own, while brood parasitic cuckoos *Cuculus canorus* produce eggs that are increasingly difficult for host warblers to recognize (Rothstein and Robinson 1998). Examples of host-parasite coevolution abound in many infectious diseases. Planktonic crustacean *Daphnia magna* can control the infectivity of the parasitic bacterium *Pasteuria ramose* while facing an ever-increasing virulence of the parasite (Decaestecker et al. 2007). Of course, the coevolution between the host and pathogens does not necessarily lead to an ever higher level of virulence as many pathogens require the well-being of their hosts for vertical transmission. The weakening virulence of human immunodeficiency virus (HIV) can be considered an example of reduced virulence from the antagonistic coevolution between the virulence and the host's immunity (Payne et al. 2014).

Although coevolution, by definition, involves a specific pair of species, it can be expanded to include multiple interacting species in interacting guilds or lineages (Futuyma and Slatkin 1983; Thompson 1994; Futuyma 2013). As aforementioned, two species in *specific coevolution* are engaging an evolutionary arms race through the interaction of their functional traits that affect each other's fitness. Such specific coevolution can typically lead to matched traits through convergence evolution in mutualistic systems and evolutionary cycles known as the Red Queen dynamics in antagonistic systems. In *diffuse coevolution*, several species from a functional guild affect each other's fitness by their own evolutionary changes (Zhang et al. 2011). In *escape-and-radiate coevolution*, the interaction between species enables one or both species to radiate into a diverse clade (Hui and McGeoch 2006; Rezende et al. 2007; Minoarivelo et al. 2014). Importantly, the coevolution between two species could lead to diffusive and then escape-and-radiate coevolution, through the process of adaptive diversification triggered by repeatedly occurring disruptive selection in the system (Brännström et al. 2011). Such adaptive diversification from coevolution is the concern here.

To date, direct examples of adaptive diversification from coevolution have been rare. One such case has been detected in the Darwin's race between a long-proboscid fly, *Moegistorhynchus longirostris*, of the Nemestinidae family and a long-tubed iris, *Lapeirousia anceps*, of the Iridaceae family (Fig. 9.1). In this arms race, effective feeding occurs when proboscis length exceeds floral tube length because the pollinator is then able to drain all the nectar from the flower; in contrast, effective pollination occurs when floral tube length exceeds proboscis length because this ensures sufficient contact with the stigma and anthers near the entrance of the floral tube (Pauw et al. 2009). These two coevolving traits thus impose reciprocal directional selection on each other, leading to an escalating arms race. Imbalanced costs to trait elongation, constrained by physiological constraints and related to environmental variation, trigger the divergent selection and the trait dimorphism in the high-cost species (specifically in some iris populations; Zhang et al. 2013).

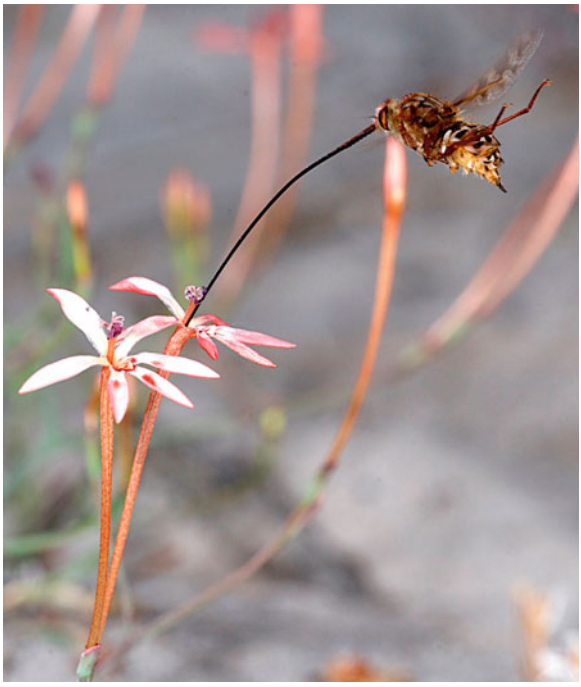


Fig. 9.1 Darwin's race between the long-proboscid fly, *Moegistorhynchus longirostris*, and the long-tubed iris, *Lapeirousia anceps*, in South Africa. (Photograph courtesy of A. Pauw)

Phylogenetic evidence of adaptive diversification from escape-and-radiate coevolution is common, suggesting coevolution also a potential source of clade diversification. The mutualistic interaction of seed dispersal by ants, known as myrmecochory, could have promoted diversification in flowering plants (Lengyel et al. 2009). Pollination syndrome between insects and flowers could explain why angiosperms of flowering plants are more diverse than gymnosperms. Escape-and-radiate coevolution could also be common between plants and herbivores, such as between the leaf beetles Blepharida and their host trees Bursera (Becerra and Venable 1999) and between endosymbiotic bacteria *Buchnera aphidicola* and aphids (Moran and Baumann 1994). Moreover, when multiple species are closely involved in a community, they often form an adaptive coevolutionary network (Zhang et al. 2011), with a mixture of mutualistic and antagonistic interactions affecting each other's fitness. For instance, leaf-cutter ants nourish the actinomycete bacteria *Pseudonocardia* by their gland secretions and use the antibiotic produced by the bacteria to inhibit the growth of the unpalatable fungi Escovopsis that competes with their food fungus Lepiotaceae in their fungal garden (Futuyma 2013).

All these clues have suggested that coevolution can potentially lead to rich evolutionary trajectories via frequency-dependent selection, in particular the possibility of diversification and polymorphism via evolutionary branching by disruptive selection in the system. These clues have further triggered abundant theoretical studies, attempting to understand how these trait-mediated interactions

in coevolutionary systems trigger disruptive selection and adaptive diversification (Doebeli and Dieckmann 2000). Using phylogenies as the record of evolutionary history, studies have shown that coevolution could explain, to a certain degree, contemporary structures of many ecological networks (Rezende et al. 2007; Minoarivelo et al. 2014). Here, we use two numerical approaches of evolutionary invasion analysis, known as AD and ED, to explore the patterns and conditions of adaptive diversification and evolutionary branching in generic models of mutualism, antagonism and food webs. Specifically, we explore under what conditions a pair of interacting species can potentially trigger disruptive selection and diversify through specific, diffuse and even escape-and-radiate coevolution.

9.2 Evolutionary Invasion Analysis

Evolutionary trajectory is traditionally regarded as the process of organisms attempting to maximize their fitness via optimizing its life-history strategies (or loosely defined as traits). Such a perspective involves two assumptions. First, there is a fitness measure that can be maximized in the attainable trait set. Second, this optimal trait can be reached, from the current stand, through incremental evolutionary changes. The first assumption leads to the definition of the evolutionarily stable strategy (ESS): there exists a trait that has competitive advantage over all other attainable traits; in other words, it can resist the invasion of all other traits. The second assumption refers to the convergence (asymptotic) stability of this optimal trait; that is, a trait close to the optimal trait can be invaded/replaced by a trait even closer to the optimal trait through directional selection. A convergence stable ESS is called a continuously stable strategy (CSS). Evolutionary invasion analysis is a set of quantitative techniques designed to address these two assumptions: conditions for the existence of an ESS and for a rare mutant trait to invade a resident population (Otta and Day 2007). Notably, the invading trait normally is considered not far from the resident one; that is, we are looking for a local CSS, strategies that are convergence stable and cannot be invaded by local traits. However, with the onslaught of global environmental changes, many non-indigenous species or genotypes are constantly being introduced to native ecosystems, suggesting an increasing relevance of searching for the global CSS in an evolutionary system. In the following, we first introduce two approaches for evolutionary invasion analysis and then apply these approaches to coevolutionary models of mutualism, antagonism and food webs

9.2.1 Adaptive Dynamics

Adaptive dynamics (AD) is a powerful analytical tool for studying the evolution of quantitative traits or phenotypic characters, developed in the 1990s by game

theorists (e.g. Nowak and Sigmund 1990), population geneticists (e.g. Abrams et al. 1993) and theoretical ecologists (e.g. Metz et al. 1992; Dieckmann and Law 1996). It studies evolutionary changes induced by rare and small mutations when fitness is density or frequency dependent (Waxman and Gavrilets 2005). As individuals interact within a community, their fitness not only depends on their own traits but also depends on the frequency or density of different traits among individuals. The evolution of traits can be evaluated by examining the survival of rare mutants in a community dominated by resident populations at their stable equilibriums. To this end, the *canonical equation* of AD describes the evolution of traits under directional selection through the continuous invasion of rare mutants into resident populations.

We illustrate here the standard procedure of using AD in a resource competition model. For a given set of n traits, changes in population densities $u_i (i = 1, 2, \ldots, n)$ are described by the Lotka–Volterra model,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = ru_i\left(1 - \frac{\sum_l \alpha(x_i, x_l)u_l}{k(x_i)}\right), \tag{9.1}$$

where r is the intrinsic population growth rate, $k(x_i)$ the trait-dependent carrying capacity and $\alpha(x_i, x_k)$ the competition strength between individuals with trait value x_i and x_k. Because mutations only occur at a low rate, the population densities are considered to be already at their equilibriums when a mutation happens. In this regard, we need to distinguish two different timescales in the concept of AD: a slow evolutionary timescale (including the slow trait shift by directional selection and the even slower evolutionary branching by disruptive selection) and a fast ecological timescale. Let x' be the trait value of a rare mutant, $x = (x_1, x_2, \ldots, x_n)$ the resident traits and u_i^* the population density at equilibrium. The invasion fitness of the mutant can be described as its per capita growth rate when setting its initial density to be negligible:$f(x, x') = r(1 - \sum_l \alpha(x', x_l)u_l^*/k(x'))$. The selection gradient of population i, $g(x_i) = \partial f(x, x')/\partial x'|_{x'=x_i}$, determines the speed of directional selection. The evolutionary dynamics of trait x_i can be depicted by the *canonical equation* as being proportional to the selection gradient (Dieckmann and Law 1996), $\dot{x}_i = \varepsilon \cdot u_i^* g(x_i)$, where ε is a parameter related to the rate and variation of mutation. If the directional selection pushes the traits to become unfeasible (i.e. the population density at equilibrium becomes equals or less than zero), it is termed an evolutionary suicide (Gyllenberg et al. 2002).

Let x_i^* indicate the trait when the selection gradients of all resident traits disappear, termed an evolutionary singularity. The singularity is convergence stable if all eigenvalues of the Jacobian of the canonical equations have negative real parts (Doebeli and Dieckmann 2000); in this case, $\partial g/\partial x_i|_{x_i=x_i^*} < 0$. The singularity represents a fitness minimum, an indication of disruptive selection, if the curvature of fitness landscape is greater than zero, $\partial^2 f/\partial x'^2|_{x'=x_i^*} > 0$, allowing traits other than the singularity to invade (Geritz et al. 1998); intuitively, the curvature is also a measure of the strength of disruptive selection. To have an evolutionary branching,

not only the singularity needs to be a fitness minimum and under disruptive selection, but also the two morphs emerged from the evolutionary branching need to be protected (Geritz et al. 1998); that is, the two morphs (x' and x'') can invade each other: $(\partial^2 f/\partial x'^2 + \partial^2 f/\partial x''^2)|_{x'=x''=x_i^*} > 0$. If the singularity represents a fitness maximum (i.e. an ESS) and convergence stable (i.e. a CSS) but the dimorphism cannot be protected, it is called an evolutionary trap (Zhang et al. 2013).

9.2.2 Evolutionary Distribution

For simplicity, common approaches of evolutionary invasion analysis, such as AD, often ignore the variation of traits and only consider the evolution of average traits (Dieckmann and Law 1996; Champagnat et al. 2001, but see Barton and Turelli 1987; Sasaki and Dieckmann 2011). However, studying only the evolution of average traits may overlook many important ecological and evolutionary features. Bolnick et al. (2011) have identified different mechanisms by which trait diversity can affect the outcome of ecological interactions. Ignoring trait variation can also lead to an underestimation of the spreading velocity in many invasive species (Ramanantoanina et al. 2014). Furthermore, numerical analyses of mean traits often rely on the separation of ecological and evolutionary timescales. This assumption is inconsistent with recent observations that ecological and evolutionary processes can occur at similar timescales (Yoshida et al. 2003; Jones et al. 2009).

Following the initial proposition of Levin and Segel (1985), Cohen (2003) coined the term of ED that studies the evolution of trait distribution in a continuous space. Cohen (2009) further suggested that considering only the mean phenotypic trait may mislead studies to traits that might not be adopted by any individuals in real populations. The concept of evolutionary distribution (ED) was initiated by Levin and Segel (1985), though the term was coined by Cohen (2003). ED studies the evolution of trait distributions in a continuous space. Reaction diffusion models are derived from ecological and evolutionary principles. While the reaction term is used to capture ecological processes such as competition and predation, the diffusion term represents the process of mutation that allows the phenotypic traits to drift on the trait space.

Using the framework of ED, the eco-evolutionary dynamics of species undergoing resource competition can be modelled by (Doebeli and Ispolatov 2010),

$$\frac{\partial u}{\partial t} = ru\left(1 - \frac{\int \alpha(x,y)u(y,t)\mathrm{d}y}{k(x)}\right) + \eta\frac{\partial^2 u}{\partial x^2}, \tag{9.2}$$

where η is the trait diffusion rate. A *morph* is defined as a trait value where the ED reaches a local maximum of frequency, and the diversity can be quantified by the number of *morphs,* i.e. the number of local maxima as well as the variance of the trait distribution around each morph (Cohen 2009; Doebeli and Ispolatov 2010).

Stable ED represents a set of ESS because all possible mutants are included in the ED (Cohen 2009). However, the study of stable ED can also be more complicated as the stability theory of partial differential equations, especially of nonlinear systems, is far from complete. Here, all initial conditions for the ED models correspond to a Dirac mass at the peak of resource distribution. A branching event is identified numerically when a local maximum emerges beside the previous one, or two local maxima appear around the previous one. To minimize the risk of a false branching (numerical fluctuations can be mistakenly taken as local maxima), branching is detected only every 20 time steps, and further fine-scale fluctuations are removed by the moving average algorithm.

9.3 Mutualistic Coevolution

9.3.1 Modelling Mutualistic Coevolution

Since Darwin's coevolutionary hypothesis between flower traits and the features of their pollinators, some patterns of mutualistic communities have been attributed to coevolution. For instance, the yucca moth (*Tegeticula synthetica*) is the only pollinator of the Joshua tree (*Yucca brevifolia*), while the seed of Joshua tree is the only food source for the yucca moth. The speciation in the moth has resulted in the radiation of the flower shape in Joshua trees. However, as species in a community are simultaneously under different and often conflicting selection pressures (such as also from predation and intra-specific competition), mutualism may not be the main driver of adaptive diversification (Raimumdo et al. 2014).

To study the role played by mutualistic interactions in generating diversification, we expand the Lotka–Volterra model with a Holling (1959) type II functional response. The population dynamics is governed by the demography, including intrinsic population growth and density dependence, and the additional contribution from the mutualistic interaction. For the AD approach, let there be n functional morphs of animals and m functional morphs of plants. Each functional morph, indexed by i for animals and j for plants, is characterized by its population density u_i and v_j, respectively. In a pollination system, the functional trait of each morph could represent the proboscis length of the pollinator, or the length of pollen tube of the flowering plant. We denote the trait of animal morph i by x_i and the trait of plant morph j by y_j. The population dynamics of the AD system thus is given by,

$$\begin{aligned}\frac{\mathrm{d}u_i}{dt} &= r' u_i \left(1 - \frac{\sum_l \alpha'_{il} u_l}{k_1(x_i)}\right) + \frac{u_i \sum_l \gamma_{il} \omega'_{il} v_l}{1 + h \sum_l \omega'_{il} v_l} \\ \frac{\mathrm{d}v_j}{dt} &= r'' v_j \left(1 - \frac{\sum_l \alpha''_{jl} v_l}{k_2(y_j)}\right) + \frac{v_j \sum \gamma_{lj} \omega''_{lj} u_l}{1 + h \sum_l \omega''_{lj} u_l}\end{aligned} \tag{9.3}$$

where h is the handling time. The equivalent ED model for the coevolution of mutualistic species can be written as the following:

$$\begin{aligned}\frac{\partial u}{\partial t} &= r'u\left(1-\frac{\int\alpha'(z)u(x-z)\mathrm{d}z}{k_1(x)}\right)\\ &\quad+\frac{u\int\gamma(z)\omega'(z)v(z)\mathrm{d}z}{1+h\int\omega'(z')v(z')\mathrm{d}z'}+\eta\frac{\partial^2 u}{\partial x^2}\\ \frac{\partial v}{\partial t} &= r''v\left(1-\frac{\int\alpha''(z)v(y-z)\mathrm{d}z}{k_2(y)}\right)\\ &\quad+\frac{v\int\gamma(z)\omega''(z)u(z)\mathrm{d}z}{1+h\int\omega''(z')u(z')\mathrm{d}z'}+\eta\frac{\partial^2 v}{\partial y^2}\end{aligned} \tag{9.4}$$

Specifically, we assign the trait-dependent carrying capacity, $k_1(x_i) = K_1 N(x^{\max}, \delta_1, x_i)$, where $N(\mu, \sigma, x)$ is a Gaussian density function at x with the mean μ and standard deviation σ, and K_1 is the carrying capacity for optimal trait $x^{\max}$, and the standard deviation of the Gaussian function, δ_1, represents the resource niche width accessible to the animals. The carrying capacity for plants, $k_2(y_j)$, is similarly defined. The intra-trophic competition kernels (α' and α") are set to let more similar morphs suffer stronger competition, (Bürger et al. 2006; Doebeli and Dieckmann 2000; Raimumdo et al. 2014): $\alpha'_{il} = N(x_i, \sigma_1, x_l)$ or $\alpha'(z) = N(x, \sigma_1, z)$, where σ_1 controls the width of the competition kernel. The cross-trophic mutualistic benefit, $\gamma_{ij} = \lambda N(x_i, \sigma_m, y_j)$ or $\gamma(z) = \lambda N(x, \sigma_m, z)$, reflects the assumption that matching traits bring high profit to each other, where λ is a parameter controlling the magnitude of the mutualistic support, and the parameter σ_m controls the tolerance level of successful interactions to the trait difference of involved traits (Nuismer et al. 2010). The interaction preference (ω'_{ij} and ω''_{ij}) of the two morphs determines the possibility of interaction after an encounter and is assumed following the adaptive foraging strategy, depending on both the benefit and abundance of the involved morphs (Doebeli and Dieckmann 2000): $\omega'_{ij} = \gamma_{ij}\sum_l u_l / \sum_l (\gamma_{lj} u_l)$, where the summation term $\sum_l u_l$ in the numerator is for normalization, or $\omega'(z) = \gamma(z)\int u(z')dz / \int \gamma(z'')u(z'')dz''$. The two approaches, AD and ED, were numerically solved with an initially monomorphic population, with a unit density for both plants and animals. Under the AD approach, the three conditions for evolutionary branching are examined once the system reaches its singularity.

9.3.2 Diversification by Mutualism

We focused on three key parameters in the system and examined their effects on the evolutionary dynamics, including the standard deviations of competition (σ_1 and σ_2) and the tolerance to trait difference (σ_m). The widths of carrying capacity (δ_1 and δ_2) are kept equal for simplicity (=1.65). Other parameters were fixed throughout the

chapter unless specified ($r' = r'' = 1; h = 0.1; \lambda = 0.25; x^{\max} = 3; K_1 = 300; y^{\max} = 2; K_2 = 400$). As illustrated in Fig. 9.2, mutualistic interactions between a monomorphic animal population and a monomorphic plant population, inserted in a resource competition model (Eqs. 9.3 and 9.4), can trigger disruptive selection and lead to diffuse and even escape-and-radiate coevolution.

Evolutionary branching is more likely to happen for stronger tolerance to trait difference (larger σ_m) and narrower competition kernel (smaller σ_1 and σ_2) (Fig. 9.3a). Adaptive diversification only happens to one species when its competition kernel is narrower than the competition kernel of its mutualistic partner species. Narrow competition kernel suggests an intense trait-specific competition, i.e. strong negative frequency dependence, which is a common condition for diversification (Day and Young 2004; Doebeli and Ispalatov 2010).

When the tolerance to trait difference is low (small σ_m), the trait value of a morph needs to become more similar to the trait value of its interacting morph to take advantage of the benefit from the mutualistic interaction (Fig. 9.3b), leading to matched traits. It is worth noting that such matched traits from low tolerance are often not the end point of coevolution, as evolutionary branching normally occurs after the trait matching. In contrast, when species have strong tolerance to trait difference (large σ_m), as in many generalists, the reciprocal selection for trait convergence is not strong, leading to bias in trait matching, or mismatched traits (Fig. 9.3b). In addition, patterns of trait matching or mismatching, as depicted by the trait value ratio in Fig. 9.3, are predominantly governed by the mutualistic term in the model and nearly independent of competition (i.e. insensitive to σ_1 and σ_2).

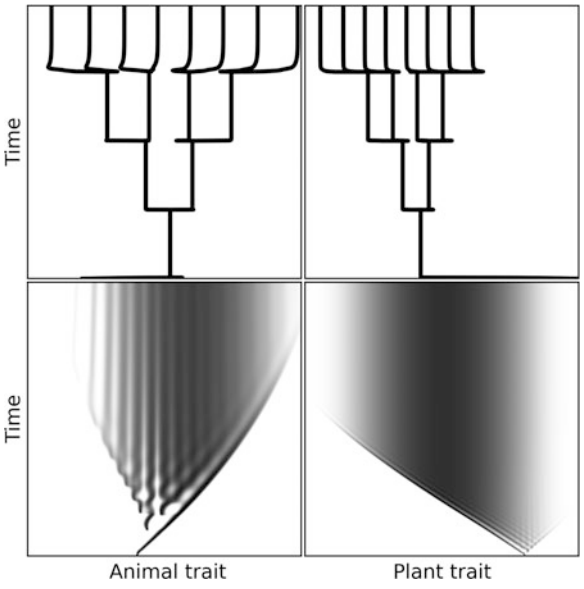


Fig. 9.2 Adaptive diversification triggered by mutualistic interactions. Parameters: $\sigma_1 = 0.14$; $\sigma_2 = 0.08$; $\sigma_m = 1$. *Top panel* is generated by the adaptive dynamics method; *bottom panel* is generated by the evolutionary distribution method

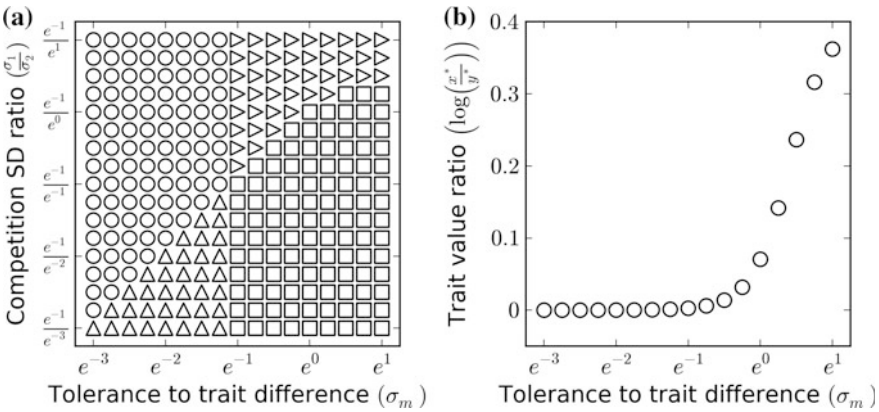


Fig. 9.3 Effects of competition kernel and tolerance to trait difference on evolutionary branching in mutualistic systems. **a** Evolutionary branching scenarios with respect to the tolerance to trait difference (σ_m) and competition standard deviations (σ_1 and σ_2) for mutualistic interactions. *Squares* represent branching in both animals and plants; *right-facing triangles* represent branching only in animals; *up-right triangles* represent branching only in plants; *circles* represent no branching. **b** The trait value ratio (animal over plant) at the first branching point as a function of the tolerance to trait difference. Both figures were obtained using the adaptive dynamics method

9.4 Antagonistic Coevolution

9.4.1 Modelling Antagonistic Coevolution

Many laboratory experiments have been conducted for observing the effect of antagonistic interactions on the diversification in coevolutionary systems. Specifically, the coevolution between hosts and their perspective parasites has been extensively studied. Results suggest that although hosts often develop resistances against their parasites, this often triggers the adaptive diversification in the parasites which in turn diversifies the resistance strategies of hosts, commonly termed as the arms race dynamics (Marston et al. 2012). In what follows, we once again make use of the Lotka–Volterra model for depicting the dynamics of predator densities (u_i) and prey densities (v_i):

$$\begin{aligned}\frac{\mathrm{d}u_i}{\mathrm{d}t} &= -r' u_i \left(1 + \frac{\sum_l \alpha'_{il} u_l}{k_1(x_i)}\right) + \frac{\lambda u_i \sum_l a\gamma_{il} v_l}{1 + h\sum_l a\gamma_{il} v_l} \\ \frac{\mathrm{d}v_j}{\mathrm{d}t} &= r'' v_j \left(1 - \frac{\sum_l \alpha''_{jl} v_l}{k_2(y_j)}\right) - \frac{v_j \sum_l a\gamma_{lj} u_l}{1 + h\sum_l a\gamma_{il} v_l}\end{aligned} \tag{9.5}$$

where functions for intra-trophic competition are similar to those in the mutualistic model. The attack rate of the prey j with trait y_j by the predator i with trait x_i is governed by a Gaussian function of trait difference, $a\gamma_{ij} = aN(x_i\text{-}\mu_p, \sigma_p, y_j)$. The

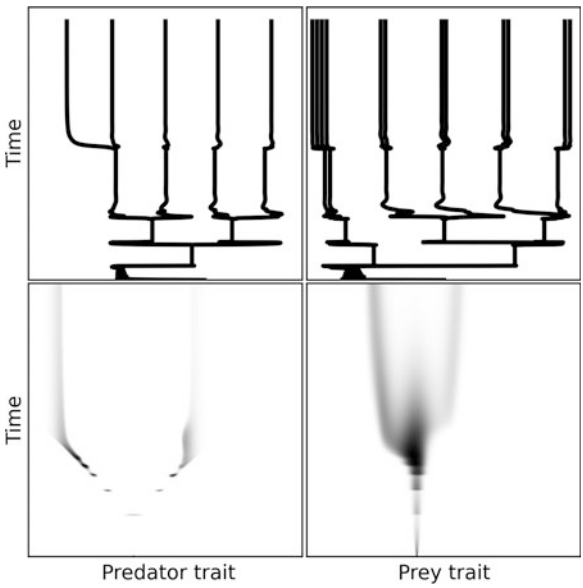


Fig. 9.4 Adaptive diversification triggered by antagonistic interactions. Parameters: $\sigma_1 = \sigma_2 = e^{-1}; \sigma_p = e^{0.25}$. *Top panel* is generated by the adaptive dynamics method; *bottom panel* by the evolutionary distribution method

attack rate is maximal when the prey trait value (y_j) is the predator trait value (x_i) minus μ_p. As above, the evolutionary dynamics of the traits can be derived from the canonical equation of the AD. Of course, we can also model the evolutionary dynamics of the predator–prey interactions using the ED framework.

9.4.2 Diversification by Antagonism

We focused here on the effects of competition kernels (σ_1 and σ_2) on adaptive diversification. Another key parameter that can affect diversification is the dietary width of predators (σ_p) which will be explored in the next section on food webs. The widths of carrying capacity were kept equal for simplicity ($\delta_1 = \delta_2 = 2$). Other parameters were fixed to be the same as in the model for mutualism, except that $r' = 0.01; \lambda = 0.3; a = 0.5; \mu_p = 1/3$. First, it is clear that the antagonistic interaction can lead to disruptive selection and evolutionary branching, using both methods (Fig. 9.4). Evolutionary branching is more likely to occur in predators, especially when the competition kernel of predators (σ_1) is relatively high ($>e$) where the branching happens exclusively to predators (Fig. 9.5a). In other words, prey cannot diversify if the competitive interference between predators is strong (large σ_1). Strong competitive interference between predators also generates larger foraging traits in predators than the anti-predation traits of prey at the first branching event (Fig. 9.5b). Moreover, when the competition between predators is strong but that between prey is weak (the bottom right corner in Fig. 9.5), the system becomes unstable, suggesting that the increased mortality due to

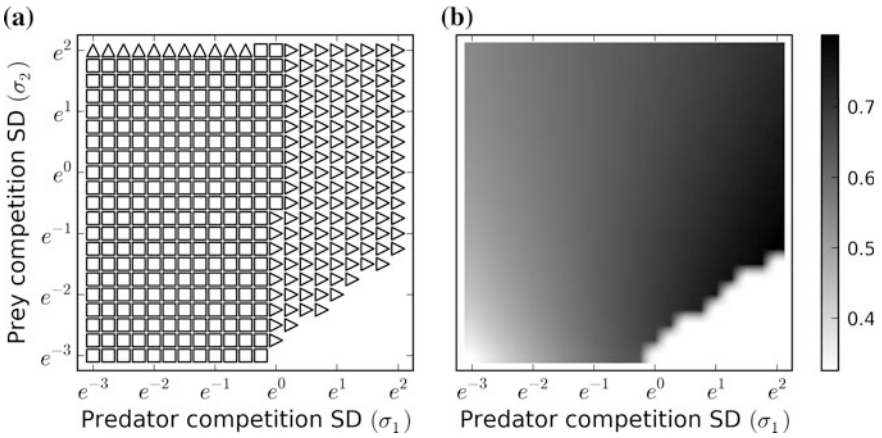


Fig. 9.5 Effects of competition kernels on evolutionary branching in predator–prey systems. **a** Evolutionary branching scenarios with respect to the competition kernel of the predator and prey. *Squares* represent branching in both predators and prey; *right-facing triangles* represent branching only in predators; *up-right triangles* represent branching only in the prey. The *empty area* represents the extinction of either species. **b** The trait value ratio (predator over prey) at the first branching point. Both figures were obtained using the adaptive dynamics method

intensive intra-trophic competition has exceeded the capacity that the cross-trophic energy flow can support, producing a zone of evolutionary suicide if convergence occurs.

9.5 Food Webs

9.5.1 Modelling Food Webs

Food webs exhibit more complex dynamics as they encompass a variety of interactions such as antagonism and competition across multiple trophic levels. For such a complex system, mathematical models of coevolution can provide insights as to the conditions that foster diversification within and cross-trophic levels (Cattin et al. 2004; Loeuille and Loreau 2005; Brännström et al. 2011, 2012). In particular, Brännström et al. (2011) have explored the role of body size as the key functional trait in initiating, structuring and maintaining food web biodiversity. Here, we use a similar model but with a type II functional response to explore the conditions that promote diversification in a food web, with specific emphasis on the role of the consumption kernel (explained below).

Consider a basal autotrophic resource ($i = 0$) and n heterotrophic morphs with population densities (u_i) such that each morph is associated with its average body size s_i. While defining the trait value of each morph as the body size relative to the autotroph, $x_i = \ln(s_i/s_0)$, we can describe the dynamics of heterotrophic morphs by the following Lotka–Volterra equations:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -d_i u_i + \sum_{j=0}^{n} \lambda \frac{s_j}{s_i} \frac{a\gamma_{ij} u_j u_i}{1 + \sum_{k=0}^{n} h_{ik} a \gamma_{ik} u_k} - \sum_{j=1}^{n} \frac{a\gamma_{ji} u_i u_j}{1 + \sum_{k=0}^{n} h_{jk} a \gamma_{jk} u_k} - \sum_{j=1}^{n} \frac{1}{k_1} \alpha_{ij} u_i u_j \quad , \tag{9.6}$$

where the intrinsic death rate $d_i = \exp(-qx_i)$, following Brännström et al.(2011); a is the attack rate; α_{ij} describes the mortality rate as a result of interference competition between morphs i and j, while $1/k_1$ defines the intensity of competition. Here, the competition kernel α_{ij} follows a Gaussian distribution as defined above. The consumption kernel γ_{ij} describes the probability of a morph i individual successfully hunting and consuming a morph j individual after the encounter and is assumed to follow a normal distribution, $\gamma_{ij} = N\left(\mu, \sigma_p, x_i - x_j\right)$, where μ defines the optimal consumer to resource body size ratio at which the consumer can make the most successful attacks, and σ_p describes the dietary niche width (i.e. the standard deviation of the consumption kernel). Conversion parameter λ is the fraction of captured resources that a consumer uses for its reproduction. The handling time h_{ij} is the time a consumer morph i spends handling one individual of morph j; following Kalinkat et al. (2013), we let $h_{ij} = h_0 s_j s_i^{-3/4}$. The demographic dynamics of the autotrophic morph can be described as follows:

$$\frac{du_0}{dt} = r u_0 \left(1 - \frac{u_0}{k_2}\right) - \sum_{j=1}^{n} \frac{a\gamma_{j0} u_0 u_j}{1 + \sum_{k=0}^{n} h_{jk} a \gamma_{jk} u_k}, \tag{9.7}$$

where r is the intrinsic growth rate of the autotrophic resource; k_2 is the carrying capacity such that r/k_2 depicts the strength of density dependence in the resource.

To study the emergence of a food web using the ED framework, we consider the integro-partial differential equation for the heterotrophic morphs:

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\mathrm{d}(x)u - \frac{1}{k_1} \int_{\mathrm{H}} \alpha(y) u(x-y) \mathrm{d}y \\ &\quad - u \int_{\mathrm{H}} \frac{\gamma(y,x) u(y)}{1 + \int_{\mathrm{A} \cup \mathrm{H}} h(y,z) \gamma(y,z) u(z) \mathrm{d}z} \mathrm{d}y \\ &\quad + u \int_{\mathrm{A} \cup \mathrm{H}} (\lambda e^{y-x}) \frac{\gamma(x,y) u(y)}{1 + \int_{\mathrm{A} \cup \mathrm{H}} h(x,z) \gamma(x,z) u(z) \mathrm{d}z} \mathrm{d}y + \eta \frac{\partial^2 u}{\partial x^2} \\ \frac{\mathrm{d}u(0,t)}{\mathrm{d}t} &= r\left(1 - \frac{u(0,t)}{k_2}\right) - u(0,t) \int_{\mathrm{H}} \frac{\gamma(y,0) u(y)}{1 + \int_{\mathrm{A} \cup \mathrm{H}} h(y,z) \gamma(y,z) u(z) \mathrm{d}z} \mathrm{d}y \end{aligned} \tag{9.8}$$

The first term on the right-hand side of the heterotrophic dynamics is the body-size-dependent intrinsic death rate. The second term captures the interspecific competition. The third term models the loss of biomass due to predation. The fourth term represents captured biomass used for reproduction. The fifth term depicts the trait diffusion. Subscript H indicates that the integration is performed over the heterotrophic morphs only, while subscript A U H represents that the integration is performed over both the heterotrophic and autotrophic morphs. The dynamics of autotrophic resource (the second equation) is assumed to follow a logistic growth in the absence of the heterotrophic morphs, with additional mortality of the autotrophic resource caused by predation by all heterotrophic morphs.

9.5.2 Diversification in Food Webs

In this section, we investigate the extent to which the dietary niche width of predators, depicted by the standard deviation of consumption kernel, σ_p, influences the first evolutionary branching event during the emergence of a food web. For all simulations, we set the initial density of one for both the autotrophic and heterotrophic morphs, with relative trait values 0 and 3, respectively. The top predator (largest trait value) gradually increases its body size, while the body size gap between the top predator and the autotroph is gradually filled up by meso-predators (Fig. 9.6). With the elapse of time, the morph richness is increasing, but the total biomass declines.

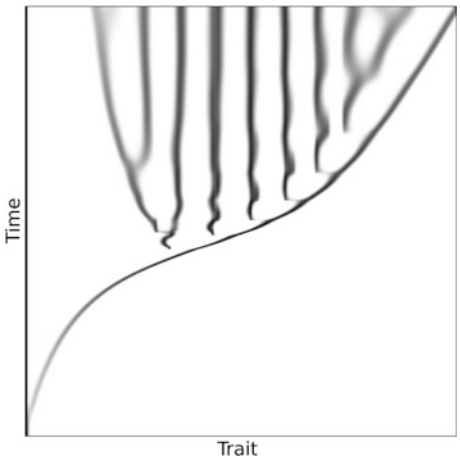


Fig. 9.6 Emergence of a food web from a single heterotrophic morph, obtained by using the evolutionary distribution method. Parameters: $r = 10, \lambda = 0.3, h_0 = 0.01, \eta = 0.001, q = 0.25, \sigma = 0.6$ (competition kernel), $a = 10, \mu = 3, \sigma_p = 1.5, s_0 = 1, k_1 = 300$ and $k_2 = 400$.

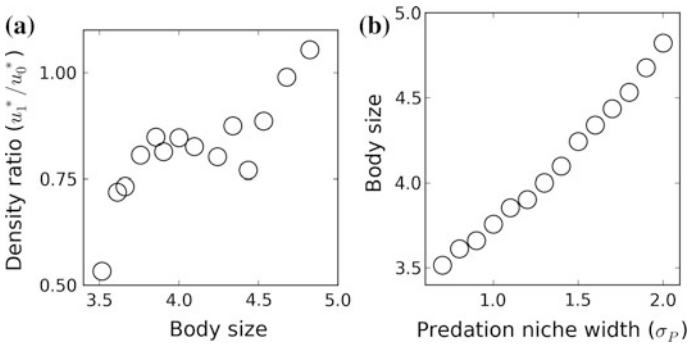


Fig. 9.7 Relationships between body size, density ratio and predation niche width at the first branching point in a food web. **a** Density ratio with respect to heterotroph body size at the first branching point. **b** Heterotroph body size as a function of the predation niche width. Parameters are the same as in Fig. 9.6

The strength of disruptive selection, measured by the curvature of fitness landscape at the singularity, increases with the increase of dietary width (σ_p) and the decrease of competition strength. This suggests that diversification, at least the first evolutionary branching, is easier in communities with more generalists than specialists. Importantly, parameters that foster the first evolutionary branching are not necessarily suitable for biodiversity maintenance (Brännström et al. 2011). Although many laboratory experiments have been designed to determine factors that favour the initial diversification (Buckling and Rainey 2002; Friesen et al. 2004; Nosil and Crespi 2006), more research is needed to understand how diversity is maintained along the evolutionary trajectory. Other factors may play an increasingly critical role for biodiversity maintenance with the increase of species richness but have only trivial effects when the system is species poor.

There is an overall positive relationship between the ratio of heterotroph to autotroph density and the heterotroph body size at the first singularity (Fig. 9.7a), with a local peak when the heterotrophy is at the optimal size for predation ($s_1 = 4$). The wider the predation niche (i.e. a diverse diet with large σ_p), the larger the body size of the heterotroph can become (Fig. 9.7b). Moreover, with the increase of dietary width, the body size ratio between adjacent morphs declines, and the food chains become longer as the mean predator–prey body size ratios decline (Jennings and Warr 2003). This is also true here since an increase in the standard deviation of the consumption kernel (σ_p) increases the strength of disruptive selection and hence supports high trophic levels. Since there is a strong correlation between body size and trophic level (Loeuille and Loreau 2005), a generalist top predator often has a larger body size than a specialist.

9.6 Conclusion: Complex Adaptive Networks

Coevolution is a major source of adaptive diversification. Mutualistic and antagonistic interactions between species can strongly affect each other's fitness and trigger frequency-dependent selection which is essential for both evolutionary branching and diversity maintenance (Genieys et al. 2006; Doebeli and Ispolatov 2010; Biktashev 2014). As a species often has multiple functions in a community, e.g. as prey, predator, pollinator, etc., whether a specific biotic interaction drives adaptive diversification is often context based (Raimumdo et al. 2014). Resource competition has been shown to trigger niche-filling diversification, with a narrower competition kernel supporting easier diversification and higher species richness. Intra-trophic competition plays the same role in mutualistic and antagonistic coevolution, with narrower competition kernel (weaker trait-specific competition) more easily triggering disruptive selection and evolutionary branching.

A newly discovered factor in coevolution is the cross-trophic interaction, between flowers and pollinators and between predators and prey. Such bipartite interactions form a divide between the two functional groups. In mutualistic systems, adaptive diversification only happens to the group with a narrower competition kernel, indicating stronger negative frequency dependence (Day and Young 2004; Doebeli and Ispolatov 2011). Low tolerance to cross-trophic trait difference (σ_p) leads to matched traits but could then lead to diversification when competition is relatively strong. High tolerance as in many generalists often leads to bias between interacting traits. Strong cross-trophic interactions often lead to convergence evolution towards an ESS, while species involving weak cross-trophic interactions behave independently as resource competition within its own functional group. Mutualistic interactions can trigger diversification when the cross-trophic interaction is moderate so that asymmetric fitness between the two groups often triggers the diversification in the less fit group.

The two functional groups in antagonistic systems are not symmetrical as in mutualistic systems. Consequently, predators are more susceptible to disruptive selection and diversification, although competition within each group also plays a role in adaptive diversification. Food webs, a more generic antagonistic system than the bipartite network, behave rather similarly. Disruptive selection is strengthened when species are dietary generalists, and wider diets also support top predators with larger body size. Of course, factors for initial diversification may be different from those that influence eventual diversity maintenance, similar to the case of community succession where pioneer species are often have distinct traits from climax species at later succession stages.

Coevolutionary networks provide an ideal model of complex adaptive systems. In this system, it is important to choose adaptively with whom to interact (habitat and diet selection) or to avoid (anti-predation strategies) (Zhang et al. 2011; Nuwagaba et al. 2015). Such interactions are often assortative as modelled by the function of α and γ used in above models. Assortative mating is important for evolutionary branching in sexual populations, while assortative cross-trophic interactions are

essential for adaptive diversification in coevolutionary systems. Such preferential interactions could simply arise from optimal or adaptive foraging where species aim to maximize their energy intake rate (Zhang and Hui 2014), while being undermined by others during their maximization. This is a grand multiplayer game. To survive in such a game, species often have to have multiple contingency plans with which to handle ecological or evolutionary selection pressures. Ecologically, species can adjust the extent and structure of their geographical range, or simply shifting its range (Roura-Pascual et al. 2011), forming different aggregation patterns of biodiversity (e.g. Hui and McGeoch 2014). They can also invoke different population dynamic strategies to release the pressure, e.g. population cycles (Ramanantoanina et al. 2011; Zhang and Hui 2011). For evolutionary pressures, species can modify their functional traits convergently or divergently (e.g. Berthouly-Salazar et al. 2012, 2013). They can change their morphology, phenology, tolerance, performance and plasticity, which are reinforced by heritable genotypes, leading to diverse evolutionary trajectories.

Acknowledgments We are grateful to the following colleagues for inspiration discussions on the topic: Ulf Dieckmann, Åke Brännström, Feng Zhang, Anton Pauw, Pietro Landi, Jaco Le Roux and John Terblanche. This work is supported by the National Research Foundation of South Africa (grants 76912, 81825 and 89967) and the DST-NRF Centre of Excellence for Invasion Biology.

References

Abrams PA, Harada Y, Matsuda H (1993) On the relationship between quantitative genetic and ESS models. Evolution 47:982–985

Barton NH, Turelli M (1987) Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet Res 49:157–173

Becerra JX, Venable DL (1999) Macroevolution of insect-plant associations: the relevance of host biogeography to host affiliation. Proc Natl Acad Sci USA 96:12626–12631

Berthouly-Salazar C, van Rensburg BJ, Le Roux JJ, van Vuuren BJ, Hui C (2012) Spatial sorting drives morphological variation in the invasive bird, *Acridotheris tristis*. PLoS One 7:e38145

Berthouly-Salazar C et al (2013) Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. Mol Ecol 22:5793–5804

Biktashev VN (2014) A simple mathematical model of gradual Darwinian evolution: emergence of a Gaussian trait distribution in adaptation along a fitness gradient. J Math Biol 68:1225–1248

Bolnick DI et al (2011) Why intraspecific trait variation matters in community ecology. Trends Ecol Evol 26:183–192

Brännström Å et al (2012) Modelling the ecology and evolution of communities: a review of past achievements, current efforts, and future promises. Evol Ecol 14:601–625

Brännström Å, Loeuille N, Loreau M, Dieckmann U (2011) Emergence and maintenance of biodiversity in an evolutionary food-web model. Theor Ecol 4:467–478

Buckling A, Rainey PB (2002) The role of parasites on sympatric and allopatric diversification of hosts. Nature 420:496–499

Bürger R, Schneider KA, Willensdorfer M (2006) The conditions for speciation through intraspecific competition. Evolution 60:2185–2206

Cattin MF et al (2004) Phylogenetic constraints and adaptation explain food-web structure. Nature 427:835–838

Champagnat N, Ferriere R, Arous GB (2001) The canonical equation of adaptive dynamics: a mathematical view. Selection 2:73–83

Chouteau M, Angers B (2012) Wright's shifting balance theory and the diversification of aposematic signals. PLoS ONE 7:e34028
Cohen Y (2003) Distributed predator-prey co-evolution. Evol Ecol Res 5:819–834
Cohen Y (2009) Evolutionary distributions. Evol Ecol Res 11:611–635
Darwin C (1862) On the various contrivances by which British and foreign orchids are fertilised by insects, and on the good effects of intercrossing. John Murray, London
Day T, Young KA (2004) Competitive and facilitative evolutionary diversification. Bioscience 54:101–109
Decaestecker E et al (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. Nature 450:870–873
Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivative from stochastic ecological processes. J Math Boil 34:579–612
Doebeli M, Dieckmann U (2000) Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am Nat 156:S77–S101
Doebeli M, Ispolatov I (2010) Complexity and diversity. Science 328:494–497
Fraenkel G (1959) The *raison d'être* of secondary plant substances. Science 129:1466–1470
Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. Evolution 58:245–260
Futuyma DJ (2013) Evolution, 3rd edn. Sinauer, Sunderland, MA
Futuyma DJ, Slatkin M (1983) Coevolution. Sinauer, Sunderland
Genieys S, Volpert V, Auger P (2006) Pattern and waves for a model in population dynamics with non-local consumption of resources. Math Model Nat Phen 1:65–82
Geritz SAH, Kisdi É, Meszéna G, Metz JAJ (1998) Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57
Gilbert SF, Sapp J, Tauber AI (2012) A symbiotic view of life: we have never been individuals. Q Rev Biol 87:325–341
Gyllenberg M, Parvinen K, Dieckmann U (2002) Evolutioanry suicide and evolution of dispersal in structured populations. J Math Biol 45:79–105
Hanifin CT, Brodie ED Jr, Brodie ED III (2008) Phenotypic mismatches reveal escape from arms-race coevolution. PLoS Biol 6:471–481
Holling CS (1959) Some characteristics of simple types of predation and parasitism. Canad Entomol 91:385–398
Hui C, McGeoch MA (2006) Evolution of body size, range size, and food composition in a predator-prey metapopulation. Ecol Complex 3:148–159
Hui C, McGeoch MA (2014) Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. Am Nat 184:684–694
Jennings S, Warr KJ (2003) Smaller predator-prey body size ratios in longer food chains. Proc R Soc Lond B 270:1413–1417
Jones LE et al (2009) Rapid contemporary evolution and clonal food web dynamics. Phil Trans R Soc Lond B 364:1579–1591
Kalinkat G et al (2013) Body masses, functional responses and predator-prey stability. Ecol Lett 16:1126–1134
Landi P, Hui C, Dieckmann U (2015) Fisheries-induced disruptive selection. J Theor Biol 365:204–216
Lengyel S, Gove AD, Latimer AM, Majer JD, Dunn RR (2009) Ants sow the seeds of global diversification in flowering plants. PLoS ONE 4:e5480
Levin SA, Segel LA (1985) Pattern generation in space and aspect. SIAM Rev 27:45–67
Loeuille N, Loreau M (2005) Evolutionary emergence size-structured food webs. Proc Natl Acad Sci USA 102:5761–5766
Mallet J (2010) Shift happens! Shift balance and the evolution of diversity in warning color and mimicry. Ecol Entomol 35:90–104
Marston M et al (2012) Rapid diversification of coevolving marine synechococcus and a virus. Proc Natl Acad Sci USA 109:4544–4549

Metz JAJ, Nisbet R, Geritz SAH (1992) How should we define 'fitness' for general ecological scenarios? Trends Ecol Evol 7:198–202
Minoarivelo HO, Hui C, Terblanche JS, Kosakovsky Pond SL, Scheffler K (2014) Detecting phylogenetic signal in mutualistic interaction networks using a Markov process model. Oikos 123:1250–1260
Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci USA 104:8627–8633
Moran NA, Baumann P (1994) Phylogenetics of cytoplasmically inherited microorganisms of arthropods. Trends Ecol Evol 9:15–20
Nosil P, Crespi BJ (2006) Experimental evidence that predation promotes divergence in adaptive radiation. Proc Natl Acad Sci USA 103:S9090–S9095
Nowak M, Sigmund K (1990) The evolution of stochastic strategies in the prisoner's dilemma. Acta Appl Math 20:247–265
Nuismer SL, Gomeulkiewiez R, Ridenhour BJ (2010) When is correlation coevolution? Am Nat 175:525–537
Nuwagaba S, Zhang F, Hui C (2015) A hybrid behavioural rule of adaptation and drift explains the emergent architecture of antagonistic networks. Proc R Soc B 282:20150320
Otta SP, Day T (2007) A biologist's guide to mathematical modelling in ecology and evolution. Princeton University Press, Princeton
Pauw A, Stofberg J, Waterman RJ (2009) Flies and flowers in Darwin's race. Evolution 63:268–279
Payne R et al (2014) Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. Proc Natl Acad Sci USA 111:E5393–E5400
Pellmyr O, Leebens-Mack J (1999) Forty million years of mutualism: evidence for Eocene origin of the yucca-yucca moth association. Proc Natl Acad Sci USA 96:9178–9183
Raimumdo LG, Gilbert GP, Hembry DH, Guimarães PR (2014) Conflicting selection in the course of adaptive diversification: the interplay between mutualism and intraspecific competition. Am Nat 183:363–375
Ramanantoanina A, Hui C, Ouhinou A (2011) Effects of density-dependent dispersal behaviours on the speed and spatial patterns of range expansion in predator-prey metapopulations. Ecol Model 222:3524–3530
Ramanantoanina A, Ouhinou A, Hui C (2014) Spatial assortment of mixed propagules explains the acceleration of range expansion. PLoS ONE 9:e103409
Rezende EL, Jordano P, Bascompte J (2007) Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. Oikos 116:1919–1929
Rothstein SI, Robinson SK (1998) Parasitic birds and their hosts: studies in coevolution. Oxford University Press, New York
Roura-Pascual N et al (2011) Relative roles of climatic suitability and anthropogenic influence in determining the pattern of spread in a global invader. Proc Natl Acad Sci USA 108:220–225
Sasaki A, Dieckmann U (2011) Oligomorphic dynamics for analyzing the quantitative genetics of adaptive speciation. J Math Biol 63:601–635
Thompson JN (1994) The coevolutionary process. University of Chicago Press, Chicago
Waxman D, Gavrilets S (2005) 20 questions on adaptive dynamics. J Evol Biol 18:1139–1154
Yoshida T, Jones LE, Ellner SP, Fussmann GF, Hairston NG (2003) Rapid evolution drives ecological dynamics in a predator-prey system. Nature 424:303–306
Zhang F, Hui C (2011) Eco-evolutionary feedback and the invasion of cooperation in Prisoner's dilemma games. PLoS ONE 6:e27523
Zhang F, Hui C (2014) Recent experience-driven behaviour optimizes foraging. Anim Behav 88:13–19
Zhang F, Hui C, Terblance JS (2011) An interaction switch predicts the nested architecture of mutualistic networks. Ecol Lett 14:797–803
Zhang F, Hui C, Pauw A (2013) Adaptive divergence in Darwin's race: how coevolution can generate trait diversity in a pollination system. Evolution 67:548–560

Chapter 10
Structure, Interaction, and Evolution: Reflections on the Natural History of Proteins

Gavin C. Conant

Abstract Protein interactions are critical to many cell functions and yet studying their evolution is challenging because comparative protein interaction data are vanishingly scarce. I discuss what we know so far, including the effect of structure, expression, and protein interactions on the rate of protein evolution, the interplay of gene duplication and protein interaction, neutral and selective forces that may shape interaction evolution, and what is known about how conserved protein interactions are over evolutionary time. I then discuss future directions where I think new insights may be found, including how new approaches to protein structure modeling may help us resolve the question of how well interactions are conserved over evolutionary time and whether we need to expand our intuition about how protein interactions direct cell function.

Keywords Protein-protein interaction · Molecular evolution · Structural Biology · Gene duplication · Gene dosage

Abbreviations

PPI	Protein–protein interaction
WGD	Whole Genome Duplication
SSD	Small-scale duplication
DBH	Dosage Balance Hypothesis

G.C. Conant (✉)
Division of Animal Sciences, University of Missouri, Columbia, MO, USA
e-mail: conantg@missouri.edu

G.C. Conant
Informatics Institute, University of Missouri, Columbia, MO, USA

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_10

10.1 Protein Interactions in Health and Disease

Protein–protein interactions, also known as PPIs, play a key role in many cellular functions, building biological complexes (Gavin et al. 2002; Krogan et al. 2006), organizing signal transduction (Pawson and Scott 1997), and being involved in various aspects of disease (Ryan and Matthews 2005). In this last case, protein interactions can be exploited by infectious agents (Murphy 2001) but are also involved in immune system protection from such infections (Roitt et al. 1998). Well-known examples of protein interactions where we have an understanding of the both functional role of the interaction and some of the nuances of its evolution include the hemoglobins (Storz et al. 2009), histones (Piontkivska et al. 2002), and the glycolytic enzyme phosphofructokinase (Poorman et al. 1984; Heillisch et al. 1989; Rodicio et al. 2000).

One particularly intriguing area of PPI research is in the role that altered or disrupted PPIs play in genetic disease (Schuster-Bockler and Bateman 2008; Vidal et al. 2012). There are several facets to this association: For instance, researchers have identified a tendency for disease-associated polymorphisms to be found at protein interaction interfaces (Schuster-Bockler and Bateman 2008; David et al. 2011; Wang et al. 2012). In other cases, the association between disease and interaction is even more clear: The neurodegenerative diseases appear to be characterized by aberrant protein interactions (Bucciantini et al. 2002; Ryan and Matthews 2005); researchers have found that even "normal" proteins can induce cell damage if forced to aggregate (Guijarro et al. 1998; Fändrich et al. 2001; Bucciantini et al. 2002). I will discuss an interesting hypothesis as to why this toxicity is observed in the section "*Protein interactions and spatial organization*" below (Ovádi et al. 2004).

10.2 The Protein Interaction Network

Given the importance of these interactions, there was naturally a great deal of excitement among biologists when it became possible to survey the space of protein interactions in a high-throughput manner, first using yeast two-hybrid approaches and later mass spectrometry (Uetz et al. 2000; Ito et al. 2001; Gavin et al. 2002, 2006; Ho et al. 2002; Krogan et al. 2006). While such large-scale experimental approaches suffer somewhat from a tendency toward both false positives and negatives, they do represent an unbiased sampling of the interaction space, as opposed to literature-derived interactions, which may have stronger experimental evidence but represent potentially limited regions of the interaction network. The early high-throughput experiments in yeast were then augmented with datasets from other model organisms including mammals, plants, and insects (Giot et al. 2003; Rual et al. 2005; *Arabidopsis* Interactome Mapping Consortium 2011).

Because these types of global interaction data do not give a fully quantitative assessment of the nature of the interaction (e.g., they lack binding kinetics and

interaction orientations), one natural approach to analyzing them is with network approaches (Zhu et al. 2007). A network perspective abstracts out some of the details of a system to allow for a global perspective on it. Thus, proteins are represented as *nodes* and the PPIs between them as *edges*. As it turns out, the protein interaction network has some of the same small-world properties as networks from other domains, such as neural networks or even the network of movie stars (Watts and Strogatz 1998; Jeong et al. 2001). In particular, the long-tailed scaling of the number of interactions per protein means that while most proteins interact with one or a few other proteins, a few proteins have dozens to even hundreds of interaction partners. This uneven distribution results in cellular robustness to loss of random proteins but sensitivity to removal of those proteins with higher numbers of interactions (Jeong et al. 2001). On the other hand, while low-level network structures, including the node degree, clustering coefficient (how "cliquish" the network is), and path length (the number of interactions one needs to traverse to connect two arbitrary proteins), are relatively well understood, there remain questions regarding the higher level structures in the network. In particular, whether specific modules can be identified within the networks and whether there are different types of highly interacting "hub" proteins appears to depend strongly on subtle details of how the networks are analyzed, suggesting caution is warranted in interpreting more complex patterns in the protein interaction network (Han et al. 2004; Batada et al. 2006).

10.3 Protein Interaction, Protein Structure, and Selection

The availability of protein interaction data immediately inspired questions about the nature of interaction evolution. Because data were initially available only in a few distantly related species, the first analyses focused on the divergence in interactions among duplicated proteins in the same genome and concluded that duplicated genes rapidly diverged in their interaction partners (Wagner 2001; Teichmann and Babu 2004). Later work sought to instead explore the rate of interaction gain and loss in genome comparisons, essentially considering how quickly orthologous interactions might be gained or lost. These approaches often used the presence of orthologs of a known pair of interacting genes in outgroup genomes to identify the most ancient point at which a particular PPI could have originated. The range of comparisons of this type was varied: encompassing studies using species across the three domains of life (Kunin et al. 2004; Kim and Marcotte 2008), within the eukaryotes (Saeed and Deane 2006), or within the fungal–animal clade (Beltrao and Serrano 2007). The conclusions from these studies were rather similar: Most protein interactions appear well conserved. These studies were somewhat preliminary due to the poor coverage of global interaction data and the indirect nature of using ortholog presence as a proxy for interaction presence. Nonetheless, Beltrao and Serrano's estimated rate of 10^{-5} protein interaction changes per protein per million years (2007) was only a factor of 10 too low when compared to estimates from a very

well-controlled and ground-breaking experimental study of protein interactions in distantly related yeast species (Qian et al. 2011).

In addition to exploring the evolution of the network itself, there were also studies assessing how network position altered (or did not alter) the evolution of the proteins making up the network. An initial study found that possessing a large number of protein interactions was predictive of a stronger selective constraint acting on the protein-coding gene in question (Fraser et al. 2002). However, later analyses were able to show that any such effect was at best quite small, and might in fact be an artifact of how differing types of interaction data were collected and analyzed (Bloom and Adami 2003; Jordan et al. 2003; Hahn et al. 2004). Work on the more general question of what determines the selective constraint of a given protein has also tended to confirm this relative lack of influence of the network on evolutionary rate, although it is true that interacting amino acids are subject to stronger constraint than other surface residues (Franzosa and Xia 2009). Instead, it appears that gene expression level is the critical predictor of selective constraint (Bloom et al. 2006; Drummond et al. 2006). What is most interesting about this finding is the idea that it is driven by selection for avoiding incorrectly folded proteins. As already mentioned, misfolded proteins are implicated in a number of cellular pathologies, and natural selection can favor protein sequences with a reduced tendency to such misfolding. Such selection is naturally strongest if the protein is highly expressed, and hence, the potential incidence of misfolded examples is high (Drummond et al. 2005).

A recent proposal by Fernandez and Lynch (2011) serves to tie such selection for stably folding proteins back to the interaction network: These authors present evidence that, at least in organisms with a small effective population sizes such as humans, interactions may evolve to stabilize proteins into their folded state, with interactions serving to mask nonpolar regions of the folded protein from the solvent by burying them in an interaction interface. In addition to linking to Drummond et al.,'s work in yeast, where population sizes are instead presumably large enough for selection stabilize folding directly through sequence selection, this origin for PPIs poses a warning to those too eager to ascribe functional significance to every protein interaction.

10.4 Gene Duplication and Protein Interactions

While protein interactions are relatively poor predictors of how the sequence of a protein evolves, they are actually quite important in understanding a second aspect of gene evolution: Which genes are duplicated and survive over evolutionary time. Gene duplication is a key source of evolutionary innovations and critical for studying genome evolution (Ohno 1970; Lynch and Conery 2000; Taylor and Raes 2004). But of course, not all gene duplications are destined to survive in a population. Understanding why some survive and others are lost is an active area of research in molecular evolution (Hahn 2009). Several analyses have pointed to a preference for the duplication of genes encoding proteins with few interactions (He and Zhang 2006;

Li et al. 2006). However, it has more recently been realized that the relationship of interaction abundance and duplicability is strongly confounded by the mechanism of duplication. Thus, in bakers' yeast, the ancient whole genome duplication (WGD) produced roughly 550 surviving duplicate gene pairs (from an initial genome of about 5000 genes) that are intermixed in the genome with duplicated genes produced by small-scale duplications, or SSDs (Wolfe and Shields 1997; Byrne and Wolfe 2005). It is important to recall that the other 4500 duplicate pairs initially created by the WGD were returned to single copy through *fractionation* (Scannell et al. 2007; Sémon and Wolfe 2007; Woodhouse et al. 2010), a process similar to, but not identical to the failure to fix an SSD in the population. With this information in hand, it became clear that while small-scale duplications do indeed tend to involve genes whose proteins have low interaction counts, duplicates produced by WGD tend on average to have more interactions (Hakes et al. 2007b).

10.5 The Dosage Balance Hypothesis

This slightly counterintuitive difference in interaction structure between the types of duplicates has actually been theoretically explained as a prediction of the *dosage balance hypothesis* (DBH). This hypothesis is an attempt to synthesize a large body of experiments that all concern the effects of changes in relative gene copy number in genomes (Papp et al. 2003; Edger and Pires 2009; Freeling 2009; Birchler 2010; Makino and McLysaght 2010; Birchler and Veitia 2012; Veitia et al. 2013). In the particular case of gene and genome duplications, the DBH argues that cells where there is an imbalance in gene copy number between members of highly interacting modules (such as protein complexes) will be at a selective disadvantage relative to their more balanced relatives. There are several potential reasons for this, including the kinetics of how these complexes assemble (Veitia et al. 2008). For cases of SSD, this propensity will tend to favor duplications at the periphery of the interaction network where the potential to make unbalanced complexes is lower. On the other hand, in the case of a WGD, the highly interacting core of the network may be generally maintained in duplicate to avoid the imbalance that would result from returning some, but not all, members of the complex to single copy. This hypothesis is in clear accord with considerable data on genome duplications, explaining patterns such as the classes of duplicates retained after WGD (Blanc and Wolfe 2004; Maere et al. 2005; Freeling 2009; Carretero-Paulet and Fares 2012). I recently used a time-resolved model of the fractionation of the yeast WGD to further explore and validate the DBH (Conant 2014). Because this model considers the full history of the WGD, not merely the set of surviving duplicates, it is able to detect evidence for the DBH not evident in extant genomes. In fact, I found two such lines of evidence. First, I was able to show that proteins with many interactions were retained in duplicate more frequently than expected in the early stages of post-WGD evolution, as predicted by the DBH—importantly, this pattern is rather difficult to perceive among the extant duplicates (Zhu et al. 2013). Second and even more strikingly,

pairs of genes whose proteins interact were more likely to both be returned to single copy at the same point in the post-WGD history of these genomes—a pattern not explicable assuming that losses were independent.

10.6 Coevolution of Interacting Proteins

Once an interaction is fixed and maintained by selection, another complication kicks in. The selection to maintain the interaction, even if relatively strong, may not result in sequence conservation, since it is the interaction, and not the sequences, that is conserved. As a result, the two sequences may coevolve over evolutionary time, diverging in sequence without the interaction being lost. A number of examples of coevolving interactions are known (Korber et al. 1993; Travers and Fares 2007; Travers et al. 2007). However, caution should be exercised in conflating coevolution and interaction: while directly interacting residues can coevolve (Travers and Fares 2007; Lovell and Robertson 2010), such coevolution is not universal (Yeang and Haussler 2007), and proteins show coevolution among residues that do not physically interact (Hakes et al. 2007a; Lovell and Robertson 2010). Instead, factors such as coexpression can yield signals of coevolution (Fraser et al. 2004; Hakes et al. 2007a; Codoner and Fares 2008). Such coevolution has relevance for other fields as well: For instance, it can interfere with sequence divergence in such a manner as to give misleading phylogenetic signals (Campos et al. 2004).

10.7 Protein Interactions and Spatial Organization

Protein interactions may also play an unexpected role in other cellular processes. The organization of biological elements in space is critical to organismal complexity at scales ranging from the arrangement of proteins in the ribosome, through that of cells in tissues and transport vessels in an organism, up to individuals in an ecosystem. However, one area where there is perhaps a slight lack of appreciation for the role of spatial structure is at what one might term the "mesoscale" of cellular interior. Protein structure (Zhang 2008) and protein complex membership (Gavin et al. 2006) have both been explored in some detail. Similarly, there is a history of research into the arrangement of organelles in the cell (Hoffmann and Avers 1973). However, there is ample evidence that a mesostructure, describing how hundreds to thousands of proteins are arranged in the cell, exists and has important functional consequences (Ovádi and Srere 1996). For instance, even after plasma membrane disruption, many of the cell's proteins maintain their positions rather than diffusing away (Hudder et al. 2003). Functionally, spatial organization contributes to metabolic channeling, whereby metabolites are spatially confined so that they move between sequential enzymes without being released into the bulk solvent (Verkman 2002). Channeling makes metabolism less diffusion-limited, and macromolecular complexes formed

from several enzymes of a single pathway that would facilitate channeling are known from several organisms (Brandina et al. 2006; Graham et al. 2007). In addition to overcoming diffusion limitations, another function of channeling is to isolate metabolites such as ATP, so that local macromolecular machines, such as ion pumps, "observe" the local and not the global concentration of that metabolite. A similar situation is found in signal transduction, where the colocalization of members of a signal cascade helps to both increase efficiency and minimize cross talk (Pawson and Scott 1997). The colocalization in question is partly mediated by scaffolding proteins that place signaling proteins in the correct physical location.

Despite this long history of interest in spatial organization within the cell, it is somewhat unusual for it to be discussed in the context of protein interactions, which is surprising, since one of the more obvious ways to organize proteins in space is through selective physical interactions (Huthmacher et al. 2007, 2008; Durek and Walther 2008; Pérez-Bercoff et al. 2011). Indeed, in our previous work, we were not only able to show an association between high-flux enzymes and potentially channeling protein interactions but also to show that some of the nonenzymatic "scaffolding" proteins involved in inducing such organization had annotations such as "localization" (Pérez-Bercoff et al. 2011).

This notion of protein interactions being used for metabolic channeling led us to explore the role of such interactions in spatial organization of the cell more generally. While biologists are used to thinking about protein interactions as forming discrete and distinct complexes, the cellular interior is more complex (Ellis 2001), and the interaction network is not nearly so uniform as a complex-centric view would suggest. We asked the simple question of what would happen if we combined known protein abundances with known PPIs into a spatially explicit, albeit highly simplified, model (Dhroso et al. 2014). Our model organized protein molecules into a discrete lattice, allowing every point in that space to be occupied by no more than one molecule. In Fig. 10.1, we show one of our lattice simulations, with different enzyme proteins shown in different colors. Using this model, we found that a protein's number of interactions alone was an imperfect representation of its role in the interaction network, because that number fails to reflect the relative copy number of the two interacting partners. We likewise found that these lattice models showed robustness to changes in protein expression despite no such robustness having been included in the model specification. Finally, we found that enzyme proteins tended to appear on the surface of protein aggregates in the model, suggesting another mechanism by which interaction might improve the efficiency of metabolism. Our model was of course highly simplified, but other researchers have shown how such spatially explicit and large-scale models may begin to describe the organization of cellular interiors, namely by adding to our approach a protein structure component (McGuffee and Elcock 2010). In the future, it should be possible to combine the structurally explicit features of McGuffee et al.,'s model with the much larger number of protein types used in our lattice model to better understand how the organization of the cellular components in space alters (or even drives) their function.

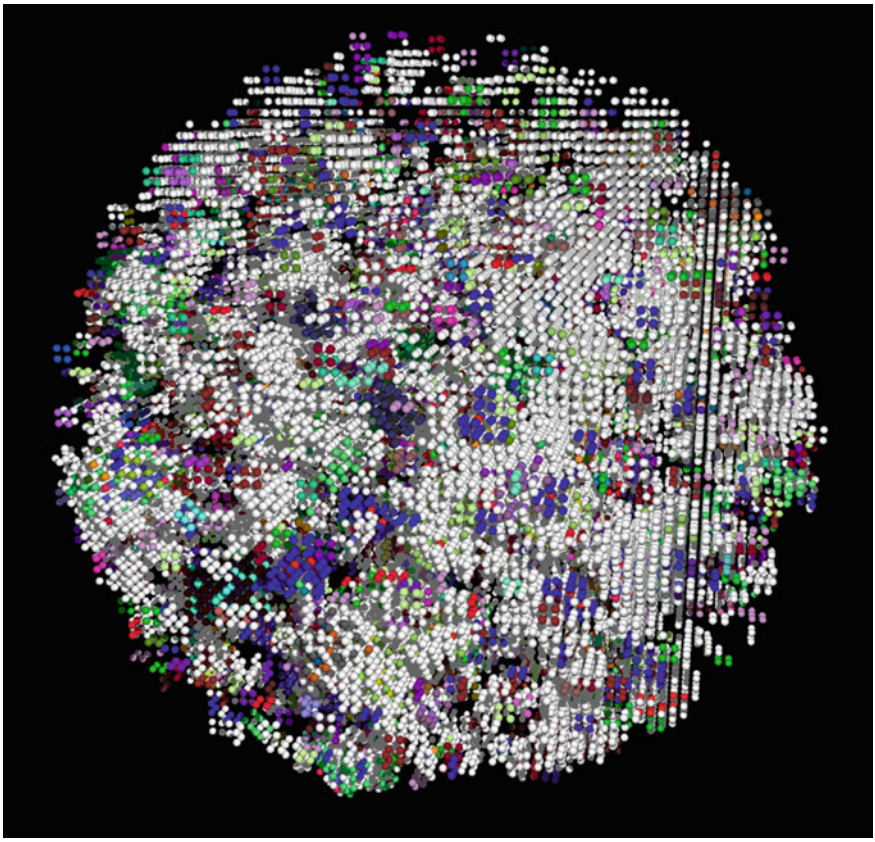


Fig. 10.1 A lattice view of yeast protein interaction network. Using the known yeast protein abundances (Ghaemmaghami et al. 2003) and interactions (Stark et al. 2011), we placed each protein molecule in a discrete lattice and optimized that lattice to give as many neighbors with interactions as possible (Dhroso et al. 2014). This lattice has radius 39, corresponding to 248,439 distinct positions. Each dot represents a protein and colored dots represent enzymes (differing colors imply catalysis of a different reaction). Our algorithm found 124,521 pairwise interactions among these 74,489 protein molecules

10.8 Channeling and Neurodegeneration

That such spatial organization may be of critical importance in cell biology is suggested indirectly by a number of features of the neurodegenerative diseases such Alzheimer's and Parkinson's. These diseases are accompanied by aggregations of toxic, misfolded, proteins (Taylor et al. 2002), but there is also a somewhat poorly understood link between neurodegeneration and metabolic defects (Lin and Beal 2006; Liang et al. 2008). For example, defects in glucose metabolism are diagnostic of even early stage Alzheimer's (Mosconi 2005), possibly because the amyloid β (Aβ) aggregates interfere with glucose import (Mark et al. 1997). Likewise, Aβ

inhibits mitochondrial enzymes (Lustbader et al. 2004). These observations suggest that it is protein aggregation that induces the metabolic defects responsible for Alzheimer's characteristic oxidative stresses (Lin and Beal 2006). Note also that these aggregations are nonspecific: Researchers have produced cytotoxic fibrils from other proteins (Bucciantini et al. 2002). Combining these observations, Ovádi et al. (2004) have thus argued that these degenerative conditions are diseases of damage to cellular organization, where protein aggregation disrupts the cell's spatial organization. These disruptions have the effect of weakening metabolic channeling, inducing many of the metabolic defects (diffusion limitations and spread of toxic intermediates such as oxygen radicals) described above. While still unproven, this hypothesis is exciting because it proposes a mechanism explaining how one of the features common to the various neurodegenerative diseases, namely misfolded proteins, might lead to disease even though the identity of these misfolded proteins differs from disease to disease.

10.9 Structural Biology, Computation, and Evolutionary Protein Interactions

The work of McGuffee and colleagues (2010) also reminds us that we are entering an era when structural biology will be able to inform our understanding of much of protein biology and evolution (Liberles et al. 2012). On the evolutionary side, it has long been known that residues more exposed to the solvent (surface residues) experience less selective constraint than do interior ones (Thorne et al. 1996; Goldman et al. 1998; Bustamante et al. 2000; Mintseris and Weng 2005; Bloom et al. 2006): The interior residues likely experience stronger selection due to their importance in folding. Early analyses differentiated surface and interior residues (Thorne et al. 1996; Goldman et al. 1998), but of course there is a continuum of how much a given residue is exposed to the aqueous cellular interior (Bustamante et al. 2000). More recent work has shown that there is a correlation between selective constraint (nonsynonymous to synonymous substitution rate or K_a/K_s) and exposure (Franzosa and Xia 2009; Ramsey et al. 2011; Scherrer et al. 2012). In our own work, we have found a strong selective constraint against introducing charged residues into the interior of a folded protein (Conant and Stadler 2009) and made the somewhat surprising observation that many mammalian proteins show no apparent selection acting on the resides on the surfaces of their proteins (Conant 2009). We interpret this second result not as strict neutrality, but rather as an observation that most residues are at least partly buried in the structure and selection is primarily acting on how the residue fits in that structure, not on its exposed portion.

In parallel with these evolutionary questions, computational protein structure prediction and analysis has, after a long incubation period, matured such that at least in a reasonable minority of cases, useful structural predictions for proteins

without a close relative of known structure can be deduced and applied to biological questions (Dill and MacCallum 2012). Such tools are a first step on the road toward the computational prediction of the structure of interacting proteins. In that vein, initially, machine-learning approaches were applied to the problem of simply predicting whether an interaction was likely to exist between two proteins in an arbitrary organism. A variety of types of evidence were used for these predictions, ranging from open-reading frame fissions and fusions, to protein orthology co-occurrence across a range of genomes, to an assessment of whether coevolution was observed between the proposed interactors (Shoemaker and Panchenko 2007; Liu et al. 2013). This linkage to coevolution, and the success of that metric in assessing interactions (Liu et al. 2013), is obviously an encouraging observation that may suggest a route toward better understanding of interaction evolution.

Even more recently, structural biology approaches have been integrated into the interaction prediction tools (Zhang et al. 2010). These tools go beyond simple structure homology modeling, which is limited by the number of protein complex structures available in the Protein Database (Keshava Prasad et al. 2009), to use data such as the factors in interaction prediction just mentioned and geometric models, to predict interactions. The approach appears quite successful, allowing large-scale predictions of protein interactions in yeast and humans (Zhang et al. 2012).

10.10 Future Prospects

I believe that new approaches and datasets will soon greatly improve our understanding of protein and interaction evolution. But the key to such success will be to embrace interdisciplinary ideas that will integrate the expertise of many different kinds of scientists. The revolution in sequencing costs has naturally captured a great deal of attention in biological circles, and the use of large-scale comparative genomics from the many new genomes coming online will be an important part of progress in this area. However, equally important will be structural approaches to protein interaction modeling (Zhang et al. 2010, 2012) and the incorporation of structural and physical details into our models of molecular evolution (Liberles et al. 2012). Such studies will necessarily be computationally expensive and so close collaborations with computer scientists and experts in parallel computing will be the third leg of the analysis "stool" required. What remains, beyond the technical requirements, is to develop the habit and skills of crossdisciplinary communication to make this work feasible.

References

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an Arabidopsis interactome map. Science 333:601–607

Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hurst LD, Tyers M (2006) Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol 4:e317

Beltrao P, Serrano L (2007) Specificity and evolvability in eukaryotic protein interaction networks. PLoS Comput Biol 3:e25

Birchler JA (2010) Reflections on studies of gene expression in aneuploids. Biochem J 426:119–123

Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci USA 109:14746–14753

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1679–1691

Bloom JD, Adami C (2003) Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol 3:21

Bloom JD, Drummond DA, Arnold FH, Wilke CO (2006) Structural determinants of the rate of protein evolution in yeast. Mol Biol Evol 23:1751–1761

Brandina I, Graham J, Lemaitre-Guillier C et al (2006) Enolase takes part in a macromolecular complex associated to mitochondria in yeast. Biochim Biophys Acta 1757:1217–1228

Bucciantini M, Giannoni E, Chiti F et al (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416:507–511

Bustamante CD, Townsend TM, Hartl DL (2000) Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*. Mol Biol Evol 17:301–308

Byrne KP, Wolfe KH (2005) The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res 15:1456–1461

Campos PRA, Olivera VM, Wagner GP, Stadler PF (2004) Gene phylogenies and protein-protein interactions: possible artifacts resulting from shared protein interaction partners. J Theor Biol 231:197–202

Carretero-Paulet L, Fares MA (2012) Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. Mol Biol Evol 29:3541–3551

Codoner FM, Fares MA (2008) Why should we care about molecular coevolution? Evol Bioinform 4:29–38

Conant GC (2009) Neutral evolution on mammalian protein surfaces. Trends Genet 25:377–381

Conant GC (2014) Comparative genomics as a time machine: how relative gene dosage and metabolic requirements shaped the time-dependent resolution of yeast polyploidy. Mol Biol Evol 31:3184–3193

Conant GC, Stadler PF (2009) Solvent exposure imparts similar selective pressures across a range of yeast proteins. Mol Biol Evol 26:1155–1161

David A, Razali R, Wass MN, Sternberg MJ (2011) Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum Mutat 33:359–363

Dhroso A, Korkin D, Conant GC (2014) The yeast protein interaction network has a capacity for self-organization. FEBS J 281:3420–3432

Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. Science 338:1042–1046

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc Natl Acad Sci USA 102:14338–14343

Drummond DA, Raval A, Wilke CO (2006) A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol 23:327–337

Durek P, Walther D (2008) The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. BMC Syst Biol 2:100

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 17:699–717

Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26:597–604
Fändrich M, Fletcher MA, Dobson CM (2001) Amyloid fibrils from muscle myoglobin. Nature 410:165–166
Fernandez A, Lynch M (2011) Non-adaptive origins of interactome complexity. Nature 474:502–505
Franzosa EA, Xia Y (2009) Structural determinants of protein evolution are context-sensitive at the residue level. Mol Biol Evol 26:2387–2395
Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science 296:750–752
Fraser HB, Hirsh AE, Wall DP, Eisen MB (2004) Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci USA 101:9033–9038
Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Ann Rev Plant Biol 60:433–453
Gavin AC, Aloy P, Grandi P et al (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440:631–636
Gavin AC, Bosche M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415:141–147
Ghaemmaghami S, Huh W-K, Bower K et al (2003) Global analysis of protein expression in yeast. Nature 425:737–741
Giot L, Bader JS, Brouwer C et al (2003) A protein interaction map of Drosophila melanogaster. Science 302:1727–1736
Goldman N, Thorne JL, Jones DT (1998) Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics 149:445–458
Graham JW, Williams TC, Morgan M, Fernie AR, Ratcliffe RG, Sweetlove LJ (2007) Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. Plant Cell 19:3723–3738
Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM (1998) Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci USA 95:4224–4228
Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered 100:605–617
Hahn MW, Conant GC, Wagner A (2004) Molecular evolution in large genetic networks: connectivity does not equal constraint. J Mol Evol 58:203–211
Hakes L, Lovell SC, Oliver SG, Robertson DL (2007a) Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proc Natl Acad Sci USA 104:7999–8004
Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL (2007b) All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol 8:R209
Han JD, Bertin N, Hao T et al (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430:88–93
He X, Zhang J (2006) Higher duplicability of less important genes in yeast genomes. Mol Biol Evol 23:144–151
Heillisch J, Ritzel RG, Von Borstel R, Aguilera A, Rodicio R, Zimmermann FK (1989) The phosphofructokinase genes of yeast evolved from two duplication events. Gene 78:309–321
Ho Y, Gruhler A, Heilbut A et al (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415:180–183
Hoffmann HP, Avers CJ (1973) Mitochondrion of yeast: ultrastructural evidence for one giant, branched organelle per cell. Science 181:749–751
Hudder A, Nathanson L, Deutscher MP (2003) Organization of mammalian cytoplasm. Mol Cell Biol 23:9318–9326
Huthmacher C, Gille C, Holzhutter HG (2007) Computational analysis of protein-protein interactions in metabolic networks of Escherichia coli and yeast. Genome Inform 18:162–172
Huthmacher C, Gille C, Holzhutter HG (2008) A computational analysis of protein interactions in metabolic networks reveals novel enzyme pairs potentially involved in metabolic channeling. J Theor Biol 252:456–464

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98:4569–4574
Jeong H, Mason SP, Barabási A-L, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411:41–42
Jordan IK, Wolf YI, Koonin EV (2003) No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol 3:1
Keshava Prasad TS, Goel R, Kandasamy K et al (2009) Human protein reference database–2009 update. Nucleic Acids Res 37:D767–D772
Kim WK, Marcotte EM (2008) Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence. PLoS Comput Biol 4:e1000232
Korber BT, Farber RM, Wolpert DH, Lapedes AS (1993) Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc Natl Acad Sci USA 90:7176–7180
Krogan NJ, Cagney G, Yu H et al (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440:637–643
Kunin V, Pereira-Leal JB, Ouzounis CA (2004) Functional evolution of the yeast protein interaction network. Mol Biol Evol 21:1171–1176
Li L, Huang Y, Xia X, Sun Z (2006) Preferential duplication in the sparse part of yeast protein interaction network. Mol Biol Evol 23:2467–2473
Liang WS, Reiman EM, Valla J et al (2008) Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. Proc Natl Acad Sci USA 105:4441–4446
Liberles DA, Teichmann SA, Bahar I et al (2012) The interface of protein structure, protein biophysics, and molecular evolution. Protein Sci 21:769–785
Lin MT, Beal MF (2006) Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature 443:787–795
Liu CH, Li K-C, Yuan S (2013) Human protein–protein interaction prediction by a novel sequence-based co-evolution method: co-evolutionary divergence. Bioinformatics 29:92–98
Lovell SC, Robertson DL (2010) An integrated view of molecular coevolution in protein-protein interactions. Mol Biol Evol 27:2567–2575
Lustbader JW, Cirilli M, Lin C et al (2004) ABAD directly links Aβ to mitochondrial toxicity in Alzheimer's disease. Science 304:448–452
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA 102:5454–5459
Makino T, McLysaght A (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci USA 107:9270–9274
Mark RJ, Pang Z, Geddes JW, Uchida K, Mattson MP (1997) Amyloid beta-peptide impairs glucose transport in hippocampal and cortical neurons: involvement of membrane lipid peroxidation. J Neurosci 17:1046–1054
McGuffee SR, Elcock AH (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol 6:e1000694
Mintseris J, Weng Z (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci USA 102:10930–10935
Mosconi L (2005) Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD. Eur J Nucl Med Mol Imaging 32:486–510
Murphy PM (2001) Viral exploitation and subversion of the immune system through chemokine mimicry. Nat Immunol 2:116–122
Ohno S (1970) Evolution by gene duplication. Springer, New York

Ovádi J, Orosz F, Hollán S (2004) Functional aspects of cellular microcompartmentation in the development of neurodegeneration: mutation induced aberrant protein-protein associations. Mol Cell Biochem 256–257:83–93
Ovádi J, Srere PA (1996) Metabolic consequences of enzyme interactions. Cell Biochem Funct 14:249–258
Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197
Pawson T, Scott JD (1997) Signaling through scaffold, anchoring, and adaptor proteins. Science 278:2075–2080
Pérez-Bercoff Å, McLysaght A, Conant GC (2011) Patterns of indirect protein interactions suggest a spatial organization to metabolism. Mol BioSyst 7:3056–3064
Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. Mol Biol Evol 19:689–697
Poorman RA, Randolph A, Kemp RG, Heinrikson RL (1984) Evolution of phosphofructokinase—gene duplication and creation of new effector sites. Nature 309:467–469
Qian W, He X, Chan E, Xu H, Zhang J (2011) Measuring the evolutionary rate of protein-protein interaction. Proc Natl Acad Sci USA 108:8725–8730
Ramsey DC, Scherrer MP, Zhou T, Wilke CO (2011) The relationship between relative solvent accessibility and evolutionary rate in protein evolution. Genetics 188:479–488
Rodicio R, Strauß A, Heinisch JJ (2000) Single point mutations in either gene encoding the subunits of the heterooctameric yeast phosphofructokinase abolish allosteric inhibition by ATP. J Biol Chem 275:40952–40960
Roitt I, Brostoff J, Male D (1998) Immunology. Mosby, London
Rual J-F, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437:1173–1178
Ryan DP, Matthews JM (2005) Protein-protein interactions in human disease. Curr Opin Struct Biol 15:441–446
Saeed R, Deane CM (2006) Protein protein interactions, evolutionary rate, abundance and age. BMC Bioinform 7:128
Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Natl Acad Sci USA 104:8397–8402
Scherrer MP, Meyer AG, Wilke CO (2012) Modeling coding-sequence evolution within the context of residue solvent accessibility. BMC Evol Biol 12:179
Schuster-Bockler B, Bateman A (2008) Protein interactions in human genetic diseases. Genome Biol 9:R9
Sémon M, Wolfe KH (2007) Consequences of genome duplication. Curr Opin Genet Dev 17:505–512
Shoemaker BA, Panchenko AR (2007) Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol 3:e43
Stark C, Breitkreutz BJ, Chatr-Aryamontri A et al (2011) The bioGRID interaction database: 2011 update. Nucleic Acids Res 39:D698–D704
Storz JF, Runck AM, Sabatino SJ et al (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proc Natl Acad Sci 106:14450–14455
Taylor JP, Hardy J, Fischbeck KH (2002) Toxic proteins in neurodegenerative disease. Science 296:1991–1995
Taylor JS, Raes J (2004) Duplication and divergence: The evolution of new genes and old ideas. Ann Rev Genet 38:615–643
Teichmann S, Babu MM (2004) Gene regulatory network growth by duplication. Nat Genet 36:492–496
Thorne JL, Goldman N, Jones DT (1996) Combining protein evolution and secondary structure. Mol Biol Evol 13:666–673

Travers SA, Fares MA (2007) Functional coevolutionary networks of the Hsp70-Hop-Hsp90 system revealed through computational analyses. Mol Biol Evol 24:1032–1044
Travers SA, Tully DC, McCormack GP, Fares MA (2007) A study of the coevolutionary patterns operating within the env gene of the HIV-1 group M subtypes. Mol Biol Evol 24:2787–2801
Uetz P, Giot L, Cagney G et al (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403:623–627
Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet 24:390–397
Veitia RA, Bottani S, Birchler JA (2013) Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. Trends Genet 29:385–393
Verkman AS (2002) Solute and macromolecule diffusion in cellular aqueous compartments. Trends Biochem Sci 27:27–33
Vidal M, Cusick ME, Barabasi AL (2012) Interactome networks and human disease. Cell 144:986–998
Wagner A (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol Biol Evol 18:1283–1292
Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol 30:159–164
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713
Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol 8:e1000409
Yeang CH, Haussler D (2007) Detecting coevolution in and among protein domains. PLoS Comput Biol 3:e211
Zhang QC, Petrey D, Deng L et al (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature. 490: 556–560
Zhang QC, Petrey D, Norel R, Honig BH (2010) Protein interface conservation across structure space. Proc Natl Acad Sci 107:10896–10901
Zhang Y (2008) Progress and challenges in protein structure prediction. Curr Opin Struct Biol 18:342–348
Zhu X, Gerstein M, Snyder M (2007) Getting connected: analysis and principles of biological networks. Genes Dev 21:1010–1024
Zhu Y, Lin Z, Nakhleh L (2013) Evolution after whole-genome duplication: a network perspective. G3 (Bethesda) 3:2049–2057

Chapter 11
Pollination Syndromes: A Global Pattern of Convergent Evolution Driven by the Most Effective Pollinator

Lorena Ashworth, Ramiro Aguilar, Silvana Martén-Rodríguez, Martha Lopezaraiza-Mikel, Germán Avila-Sakar, Víctor Rosas-Guerrero and Mauricio Quesada

Abstract Convergent evolution of floral traits driven by pollinators has resulted in floral syndromes shared among different plant lineages. However, the flowers of many plant species are often visited by different pollinator groups, which apparently

L. Ashworth · R. Aguilar
Instituto Multidisciplinario de Biología Vegetal, Universidad Nacional de Córdoba - CONICET, C.C. 495, 5000 Córdoba, Argentina
e-mail: loashworth@gmail.com

R. Aguilar
e-mail: rams.aguilar@gmail.com

S. Martén-Rodríguez
Instituto de Ecología, A. C., Centro Regional Del Bajío, Pátzcuaro, Michoacán, Mexico
e-mail: smartenr@gmail.com

S. Martén-Rodríguez · M. Quesada (✉)
Laboratorio Nacional de Análisis y Síntesis Ecológica para la Conservación de Recursos Genéticos, Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autónoma de México, Apartado Postal 27-3 (Xangari), Morelia, Michoacán 58089, Mexico
e-mail: mquesada@cieco.unam.mx

M. Lopezaraiza-Mikel · V. Rosas-Guerrero
Unidad Académica en Desarrollo Sustentable, Universidad Autónoma de Guerrero, Carretera Nacional Acapulco Zihuatanejo Km 106+900, , Tecpán de Galeana, Guerrero 40900, Mexico
e-mail: me.lopezaraiza@gmail.com

V. Rosas-Guerrero
e-mail: victor_rosas@yahoo.com

G. Avila-Sakar
Department of Biology, The University of Winnipeg, Winnipeg, Manitoba, Canada
e-mail: gasakar@gmail.com

M. Quesada
Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Apartado Postal 27-3 (Xangari), Morelia, Michoacán 58089, Mexico

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_11

contradict the idea of syndromes. Here, we demonstrate that the most efficient pollinators consistently correspond to the ones predicted by the syndrome, and the predictive accuracy of the syndrome tends to be higher for species pollinated exclusively by one functional group than for species pollinated by more than one functional group. Secondary pollinator functional groups affected differentially the relative efficiency of the primary pollinator depending of the syndrome. The most frequent secondary pollinator group of a given syndrome is also the least efficient one. Floral symmetry did not influence predictability of pollination syndromes. Except for the bee-syndrome plants, pollination syndromes were more effective on plants that depend strongly on animal pollination than on less dependent plants. Last, effective pollinators for each floral syndrome were better predicted for plants from tropical regions, particularly for the bat, bee, and bird syndromes. Our results have implications on the effects of global change on floral evolution and suggest that current suites of floral traits in most plant species have the potential to adapt to new conditions under changing selective pollination environments.

11.1 Introduction

Angiosperms are the most diverse group of living plants with more than 350,000 species distributed across all major ecosystems of the earth (The Plant List 2013). This plant group, which rapidly diversified in the Cretaceous period, is characterized by two important features that include a great diversity of flowers and an unparalleled diversity of pollination and reproductive systems. Darwin described the sudden and rapid radiation of the angiosperms as an abominable mystery and recognized Gaston de Saporta's idea that early interactions with pollinating insects favored outcrossing promoting diversification (Friedman 2009). Darwin (1862) also described the great variety of floral traits and breeding systems in the angiosperms as examples of adaptation to promote cross-pollination by animal vectors. Today, pollinator-mediated selection is considered one of the major evolutionary processes underlying floral diversification (Harder and Johnson 2009; van der Niet and Johnson 2012). The specific combinations of floral traits, including flower morphology, color, scent, type, and amount of reward that have independently evolved to attract specific groups of animal pollinators, are known as pollination syndromes.

The idea that floral traits should be associated with particular pollinating agents was first proposed by Delpino (1867) and later elaborated by Kunth (1906), Vogel (1954), and Faegri and van der Pijl (1979). In accordance with this hypothesis, Stebbins (1970) proposed that floral traits reflect adaptation to the pollinators that visit flowers most frequently and effectively, an idea that was later acknowledged as the "most effective pollinator principle." Effective pollinators are differentiated from other floral visitors in their ability to effect fruit set and are expected to have a direct impact on plant fitness. Thus, according to Stebbins´ principle, flowers may receive visits by different pollinator groups, but floral phenotypes should correspond to the

most effective ones. The concept of pollination syndromes was later complemented with the idea that pollinators can be clustered into functional groups that have similar behavior and exert similar selection on flowers (Fenster et al. 2004).

Phenotypic selection exerted by pollinators on single reproductive traits or trait combinations of plants is the main evidence to assert that most flowers reflect specialization for pollination by particular animal groups (Stebbins 1970; Fenster et al. 2015). However, it is also evident that more than one pollinator species visit the flowers of many plant species (Waser et al. 1996). These observations have stimulated discussion in the literature on the premises that most pollination systems are generalized in nature (Waser et al. 1996) and that pollinators do not always correspond to those predicted by floral traits (e.g., Ollerton et al. 2009). However, other studies have shown that floral traits are associated with particular functional groups of pollinators (Fenster et al. 2004), and that pollination syndromes do predict the most important pollinators of plants (Martén-Rodríguez et al. 2009; Reynolds et al. 2009). The most recent comprehensive quantitative review on pollination syndromes demonstrates that syndromes predict the most effective pollinators of plant species even when there are secondary pollinators (i.e., pollinators not according to the syndrome) within the pollinator assemblage of plants (Rosas-Guerrero et al. 2014).

Here, we further analyze the database of Rosas-Guerrero et al. (2014) focusing on testing hypotheses within and across pollination syndromes. In the context of each pollination syndrome, various traits might influence the association between pollinators and floral traits, but not all syndromes may respond identically. In this chapter, we expand our previous findings and test five hypotheses relating the level of specialization, the identity of secondary pollinators, floral morphology, breeding systems, and geographic location to each pollination syndrome.

In the first hypothesis, we consider that the level of pollination specialization should determine the predictability of syndromes. In species with more generalized pollination systems, selection on floral traits may be disruptive or more relaxed (Gómez et al. 2014). Thus, we expect that species pollinated exclusively by one functional group should show greater predictive accuracy of pollination syndromes than species pollinated by more than one functional group.

A second hypothesis proposes that within a syndrome the efficiency of secondary pollinators varies. Therefore, the relative efficiency of the primary pollinator within any given syndrome will be differentially affected depending on the identity of the secondary pollinator group. For example, in bat-syndrome flowers, typically with highly exerted stamens, birds may be more efficient secondary pollinators than bees (e.g., *Gesneria pedunculosa*, Martén-Rodríguez and Fenster 2008). In such case, we might expect birds to exert stronger selection on floral traits than bees, modifying more the relative efficiency of bats.

In a third hypothesis, we propose that floral symmetry may act as a pollinator-filtering agent. Bilateral flowers can restrict the directionality of approach and movement of pollinators within flowers (Sargent 2004), with a consequent decrease in the type of visitors that may access these flowers (Huang and Gong 2009). Indeed, the idea that the origin of bilateral symmetry is a consequence of

strong selection exerted by specialized pollinators has been recently supported by empirical data (Gómez et al. 2006). Therefore, we expect greater predictability of pollination syndromes in plants with bilateral flowers than with radial flowers. Since sensory abilities differ among pollinator functional groups, we explored floral symmetry within each syndrome.

In a fourth hypothesis, we suggest that within each syndrome, pollinator-dependent species should experience more consistent selection on floral traits than species that have the ability to set seeds via autonomous self-pollination. We previously documented that regardless of their syndrome, pollinator-dependent species, such as dioecious, monoecious, and self-incompatible species, show greater predictability of pollination syndromes than self-compatible species (Rosas-Guerrero et al. 2014). Here, we explore whether this finding is consistent within each pollination syndrome.

The fifth hypothesis proposes that within each syndrome, the predictability of pollination syndromes will differ between tropical and extra-tropical plant species. Because the strength of biotic interactions is expected to increase with decreasing latitude, tropical species should have narrower niches, facilitating coexistence and promoting diversification (Schemske et al. 2009; Moya-Laraño 2010). In tropical species, the pollinators expected by the syndrome are more efficient than in extra-tropical species (Rosas-Guerrero et al. 2014). Here, we further explore the relationship between geographic distribution and predictability by making such comparisons within each pollination syndrome.

By means of ordinary and phylogenetic meta-analyses, we previously tested whether pollination syndromes can predict the most effective pollinator of plants (Rosas-Guerrero et al. 2014). The analysis was based on a complete and systematic literature review of detailed pollination studies throughout the world that quantified the efficiency of the entire pollinator assemblages of plant species (Rosas-Guerrero et al. 2014). From a total of 1990 studies in the literature search, we considered 213 suitable publications including 370 plant species, and 47 species of our own studies, that were conducted under natural conditions and quantitatively assessed pollination effectiveness of all floral visitors of plants. Pollination effectiveness measures considered were pollen on pollinator's body, contact of pollinator with the flower's reproductive organs, pollen deposited on stigmas, pollen removed from anthers, or fruit and/or seed production by specific functional groups. These pollination effectiveness measures (Ne'eman et al. 2010) did not significantly differ in their ability to detect differences in pollination syndromes accuracy to predict the effective pollinator functional groups (Rosas-Guerrero et al. 2014).

We assigned one of 11 pollination syndromes to each plant species based on the presence or absence of character states of floral traits (Rosas-Guerrero et al. 2014). Each syndrome was assigned to each plant species without previous knowledge of its assemblage of floral visitors. Pollination syndromes were characterized according to Faegri and van der Pijl (1979), Proctor et al. (1996), Ollerton et al. (2009), and Willmer's (2011) descriptions.

In our previous synthesis, we found that both phylogenetically independent and traditional meta-analyses produced almost identical response patterns

(Rosas-Guerrero et al. 2014). Such homogeneity of responses between both types of meta-analyses implies that calculated effect sizes are not conserved across the phylogeny of the sample of species included in our review, i.e., there is no phylogenetic signal in the relative efficiency of pollinators according and not according to any particular syndrome. When effect sizes are not conserved within the phylogeny (i.e., there is weak or null phylogenetic signal), any phylogenetic correction may have a trivial effect on meta-analytical results, as effect sizes are fundamentally independent across the phylogeny (Chamberlain et al. 2012). Therefore, in this chapter, we only conduct traditional meta-analyses using the entire database to gain power in effect size estimations.

For the meta-analyses, we used the standardized unbiased mean difference (Hedges' *d*) as a measure of effect size that expresses the difference in pollination effectiveness between two pollinator groups:

$$d = \frac{X_s - X_{ns}}{S_{\text{within}}} J$$

where X_s is the mean value of pollination effectiveness of the expected pollinator functional group according to the syndrome, X_{ns} is the pollination effectiveness of the pollinator functional group not according to the syndrome, S_{within} is the within-groups standard deviation, pooled across groups, and J is a correction factor for small sample sizes (see Gurevitch and Hedges 2001 for calculation details).

In cases where mean values, standard deviations, and/or sample sizes were not provided by a study, we calculated a different effect size; the Odds Ratio (OR, Cooper et al. 2009):

$$\text{OR} = \frac{AD}{BC}$$

where A is the number of effective pollination events of the expected pollinator, B is the number of effective pollination events of the non-expected pollinator, C is the number of ineffective pollination events of the expected pollinator, and D is the number of ineffective pollination events of the non-expected pollinator. When one of the pollinator groups was not observed, we added 0.5 to each cell to be able to calculate OR (Cooper et al. 2009). To unify effect size metrics and be able to run the meta-analyses, we converted log (OR) values and their variance into Hedges' *d* and its corresponding variance values through mathematical transformations (Cooper et al. 2009).

When effect sizes Hedges' *d* are positive, it implies that pollinators expected by the syndrome are more efficient (i.e., there is support for the pollination syndrome hypothesis), whereas when effect sizes *d* are negative, pollinators not matching the syndrome are more efficient. In the cases that a plant species had two or more effective pollinator functional groups besides the one predicted by the syndrome,

we calculated an effect size for each syndrome versus each alternative group combination for that plant species. Due to these situations, we ended up with 517 data points from 417 unique plant species. We used MetaWin 2.0 to run the traditional meta-analyses (Rosenberg et al. 2000). Confidence intervals of effect sizes were calculated using bootstrap resampling procedures (Adams et al. 1997). An effect was considered significant if the 95 % biased-corrected bootstrap confidence intervals (CI) of the effect size (*d*) did not overlap zero. Data were analyzed using random-effect models, which assume that studies differ not only by sampling error, as fixed-effects models do, but also by a random component in effect sizes (Gurevitch and Hedges 1999), which is the expectation in ecological and evolutionary studies.

To test our hypotheses, we analyzed whether the following moderator or predictor variables influenced differentially the effect sizes within each of the syndromes with large sample sizes (bat, bee, bird, fly, moth, and wasp): breeding system (pollinator-dependent: self-incompatible/monoecious/dioecious species versus non-pollinator-dependent: self-compatible species), floral symmetry (bilateral versus radial flowers), diversity of functional groups (plants pollinated by only one functional group versus pollinated by two or more functional groups), identity of secondary pollinators functional group (expected pollinator´s functional group versus each alternative pollinator functional group), and geographical region (tropical versus extra-tropical plants). To examine whether each of these comparisons were statistically significant, we used Q statistics, examining the P values associated with $Q_{between}$ statistics, which describe the heterogeneity in effect sizes that can be ascribed to differences between each of these categories (Cooper et al. 2009). We found no publication bias in our meta-analysis (Rosas-Guerrero et al. 2014), which implies that studies with significant results were not systematically more published than non-significant studies.

11.2 Overall Prevalence of Pollination Syndromes Across Angiosperms

The 417 plant species of our database represent 217 genera from 81 plant families of angiosperms from all continents around the world except the Antarctica. The most represented pollination syndromes in order of importance were bee, bird, bat, fly, wasp, and moth. The least represented groups were butterfly, long-tongued fly, beetle, carrion fly, and non-flying mammal (Table 11.1; Fig. 11.1).

Most of the studies were conducted in Africa, followed by Mesoamerica (Mexico and Central America), Asia, South America, and North America. The least represented continents were Europe and Oceania (Fig. 11.1). In our sample, the bee syndrome is predominant in North and South America, Asia, and Europe. Bat syndrome predominates only in Mesoamerica. Moreover, in Asia, Africa, and

Table 11.1 Results of traditional meta-analyses. *K* is the number of pooled effects; effect size values (Hedges' *d*) are given for the overall effect and for each pollination syndrome; LCI and UCI are lower and upper confidence intervals around effect sizes, respectively. *Q*between test evaluates differences in effect size among pollination syndromes

	k	Hedges' *d*	LCI	UCI
Overall effect	517	0.5937	0.4964	0.6915
		Qb = 14.53, *d.f.* = 10, *P* = 0.175		
Pollination syndromes	*N*			
Bat	58	0.7173	0.4462	1.079
Bee	184	0.6199	0.4854	0.7444
Beetle	12	0.1189	−0.5199	0.7577
Bird	96	0.6985	0.4466	0.9692
Butterfly	14	0.2297	−0.2426	0.9054
Carrion fly	5	0.3385	−0.9425	0.7096
Fly	58	0.3775	0.0719	0.6302
Long-tongued fly	9	0.6572	−0.2214	1.5326
Moth	33	0.6689	0.3280	0.9868
Non-flying mammal	7	1.2131	0.1302	2.1998
Wasp	41	0.5508	0.2395	0.8634

N = Data points

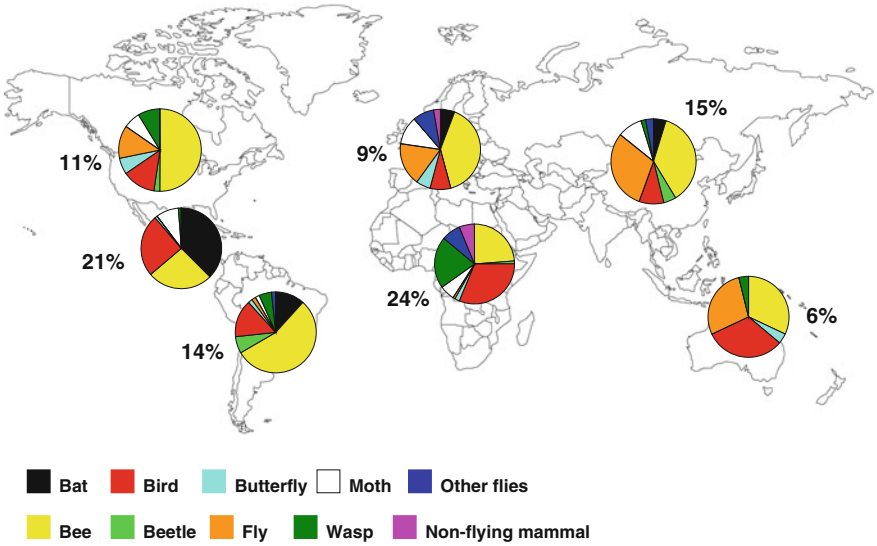


Fig. 11.1 Distribution of studies assessing pollination effectiveness of entire pollination assemblages in 417 plant species throughout the world. Within each circle, we show the relative proportion of pollination syndromes of the plant species studied in each region: North America, Meso America, South America, Europe, Africa, Asia, and Oceania. *Percentages given in numbers* refer to the relative representation of plant species for each region to the total

Oceania, there is no unique dominant syndrome, and two or three syndromes are similarly frequent (bee, bird, fly, and wasp syndromes, Fig. 11.1).

In most of the syndromes, the effect sizes were positive and significantly different from zero (Table 11.1), meaning that pollinators that matched the floral syndrome were significantly more efficient than pollinators that did not match the syndrome. The beetle, butterfly, and carrion fly syndromes had positive effect sizes but were not different from zero, though they had small sample sizes and statistical power. In these syndromes, a non-significant effect indicates at most that pollinators predicted by the syndrome were not more efficient than pollinators not predicted by the syndrome.

Our overall results indicate that particular suites of floral traits do correlate with particular effective functional groups of pollinators across a set of taxonomically widespread angiosperm species. Thus, our results suggest that adaptation to the most effective pollinator functional group drives the convergent evolution of floral traits, supporting Stebbins' most effective pollinator principle (Stebbins 1970). We, however, stress the current scarcity of studies on syndromes such as beetle, butterfly, carrion fly, long-tongued fly, and non-flying mammal across different regions of the world.

11.3 First Hypothesis: The Level of Pollination Specialization Within Each Syndrome Should Determine Its Predictability

The general pattern found here shows that within each syndrome, the predictive accuracy of the syndrome tends to be higher when the primary pollinator is alone (Syndrome-None, Fig. 11.2) than when one or more secondary pollinators are present (Syndrome-Alternative, Fig. 11.2). Therefore, results obtained here agree with our first hypothesis.

Interestingly, secondary pollinators were common regardless of the pollination syndrome (see Syndrome-Alternative sample sizes in Fig. 11.2) and may play an important role in the evolution of plant reproduction. However, with the exception of the fly syndrome, the presence of secondary pollinators did not imply the rejection of the pollination syndrome hypothesis, as pollinators predicted by the syndromes were the most efficient (i.e., effect sizes are positive for Syndrome-Alternative groups of plants). Thus, we argue that the concept of pollination syndromes does not necessarily imply the absence of secondary pollinators (Fig. 11.2). The interaction of plants with secondary pollinators might be expected to reduce the strength of selection exerted by primary pollinators, given that the relative efficiency of primary pollinators is decreased by the presence of secondary pollinators (Fig. 11.2). This does not imply a widespread generalization of pollination systems and the absence of pollination syndromes as proposed by Waser

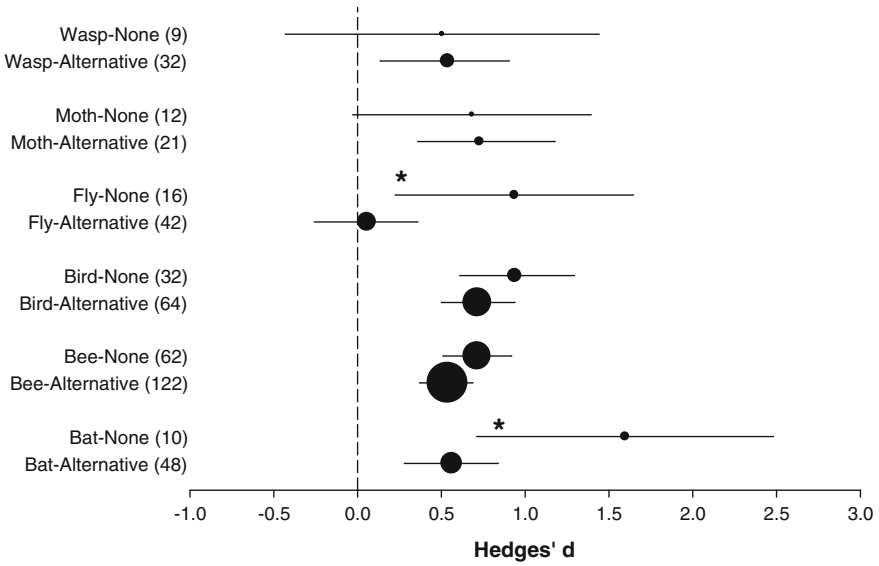


Fig. 11.2 Weighted-mean effect sizes and 95 % bias-corrected confidence intervals of the predictability of pollination syndromes on the most effective pollinators when each pollinator predicted by syndromes is alone (i.e., no other pollinator functional group was registered within the plant´s pollinator assemblage) and when there are alternative pollinator functional groups within the plant´s pollinator assemblage. Sample sizes for each category are shown in parentheses. The size of each dot represents the proportional weight or contribution to the overall mean calculation. Dotted lines show Hedges' $d = 0$. When confidence intervals overlap zero, the effect sizes are not significantly different from zero. *Asterisks* indicate significance level at $p < 0.05$ associated with $Q_{between}$-values for each group comparison

et al. (1996). Instead, this study demonstrated that syndromes were correctly predicted even in the presence of secondary pollinators.

11.4 Second Hypothesis: Within Each Syndrome the Efficiency of Secondary Pollinators Varies Depending on Their Identity

We found great variation in the level of generalization of plant species among syndromes. Bat-, bird-, moth-, and wasp-syndrome flowers were pollinated by up to three functional groups of secondary pollinators while fly and bee flowers by five or six alternative functional groups of pollinators, respectively (Rosas-Guerrero et al. 2014). Our results show that within each syndrome, the relative efficiency of primary pollinators can be affected differentially by the identity of the secondary functional group of pollinators. In the case of bee-syndrome flowers, the efficiency of bees tends to be lower when birds are present (Bee–Bird), than when bees are

alone (Bee–NONE, Fig. 11.3). Such result implies that birds are rather efficient pollinators of bee-syndrome flowers. In contrast, for bee-syndrome flowers, the efficiency of bees is not affected when butterflies or flies are present as secondary pollinators (i.e., similar effect size between Bee–Butterfly or Bee–Fly and Bee–NONE, Fig. 11.3); thus, butterflies and flies would not be efficient pollinators in bee-syndrome flowers.

For bat-syndrome flowers, the presence of birds, bees, or moths as secondary pollinators (Bat–Bird or Bat–Moth) tends to reduce the efficiency of bats, as compared to bats alone (Bat–NONE, Fig. 11.3). Between these three groups, the relative efficiency of bats decreases more when moths or bees are the secondary pollinators (Fig. 11.3). For fly-syndrome flowers, the presence of beetles and bees as secondary pollinators tends to reduce the efficiency of flies, as compared to flies alone (Fig. 11.3). It should be noticed that the relative effect of a given secondary pollinator on the efficiency of a primary pollinator is not reciprocal. For example,

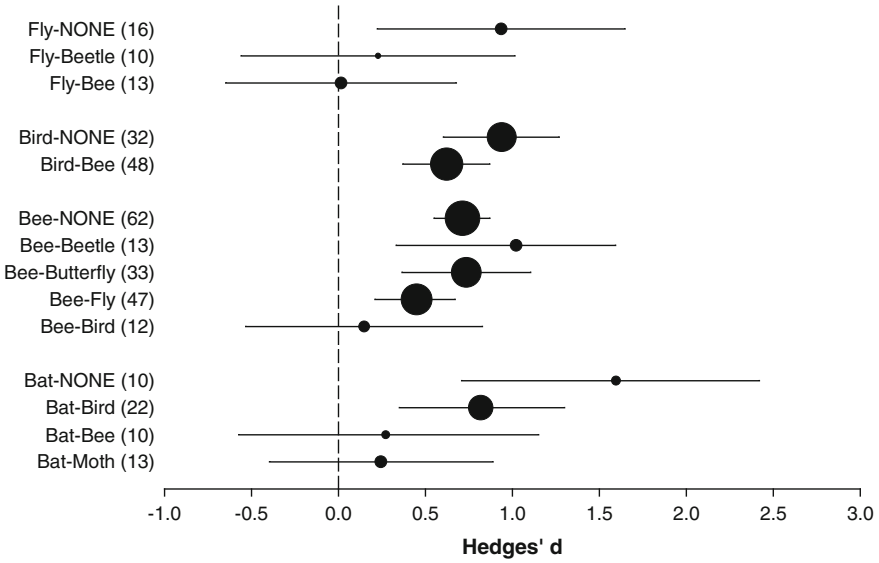


Fig. 11.3 Weighted-mean effect sizes and 95 % bias-corrected confidence intervals of the predictability of pollination syndromes on the most effective pollinators when each pollinator predicted by syndromes is alone (i.e., no other pollinator functional group was registered within the plant´s pollinator assemblage) and when another pollinator functional group is also present within the plant´s pollinator assemblage. Here, we only make comparisons when alternative pollinator functional groups were observed in 10 or more plant species. While differences in mean effect sizes are observed, they were not statistically significant following the omnibus $Q_{between}$ test among categories within each pollination syndrome. Sample sizes for each category are shown in parentheses. The size of each dot represents the proportional weight or contribution to the overall mean calculation. *Dotted lines* show Hedges' $d = 0$. When confidence intervals overlap zero, the effect sizes are not significantly different from zero

for bee-syndrome flowers, the efficiency of birds (Bee–Bird in Fig. 11.3) is higher than the efficiency of bees on bird-syndrome flowers (Bird–Bee in Fig. 11.3).

Pollination networks showed that certain functional groups are more commonly found as secondary pollinators of particular syndromes (Rosas-Guerrero et al. 2014). The most frequent associations between primary and secondary pollinators were birds for bat-syndrome flowers, butterflies and flies for bee-syndrome flowers, bees for bird-syndrome flowers, and bees and beetles for fly-syndrome flowers (Fig. 11.3). Here, we show that this non-random association is related to the efficiency of the secondary pollinator; that is, the most frequent secondary group of pollinators of each syndrome is the one with the lowest impact on the relative efficiency of the primary pollinator. Under this situation, there would be no conflicting selection between primary and secondary pollinators and the primary pollinator would drive the evolution of floral traits. On the other hand, if secondary pollinators reduced the relative efficiency of the primary pollinators, they would have the potential to drive pollination syndrome transitions. For example, we found that birds are quite efficient secondary pollinators of bee-syndrome flowers, thus birds would have the potential to exert selection and drive floral transitions on bee-syndrome flowers; however, the opposite would not occur because bees are not efficient pollinators of bird-syndrome flowers. These differences in pollination efficiency may explain the asymmetry in evolutionary transitions showed in a recent review, where 43 transitions were registered from bee to bird, but only 13 transitions from bird to bee (van der Niet and Johnson 2012). Following the same reasoning, our data also show that bat-syndrome flowers would have more chances to evolve to moth- or bee-syndrome flowers than to bird-syndrome flowers. Similarly, bee-syndrome flowers would have a better chance to evolve to bird-syndrome flowers than to fly-, butterfly-, or beetle-syndrome flowers. Transitions from bat to other functional groups have been scantly registered in the literature (van der Niet and Johnson 2012), which may be due to the fact that the evolution of this pollinator group is relatively recent (see Table 1 in Rosas-Guerrero et al. 2014).

11.5 Third Hypothesis: Floral Symmetry Can Act as a Pollinator-Filtering Agent

Bilateral and radial flowers occurred in all pollination syndromes (i.e., bat, bee, bird, fly, moth, and wasp) and were equally frequent for bat, bird, and bee syndromes. Radial flowers were more common for fly, moth, and wasp syndromes. The predictive power of pollination syndromes did not differ between bilateral and radial flowers neither overall ($Q_{\text{between}} = 19.83$; $P = 0.098$), nor within pollination syndromes ($Q_{\text{between}} \leq 2.99$; $P \geq 0.138$). The traditional view is that the evolution of bilateral symmetry is associated with the increased levels of specialization (Wolfe and Krstolic 1999; Fenster et al. 2004). However, our results suggest that symmetry by itself is not directly related to levels of pollination specialization or to particular pollination syndromes. For example, in *Ipomoea* (Convolvulaceae), where flowers

are radial, there is a great variation in specialization and in pollination systems (Rosas-Guerrero et al. 2011). Similarly, in the radial flowers of *Aquilegia* (Ranunculaceae), the length of the nectar spur determines accessibility to effective pollinators and the level of pollination specialization (Whittall and Hodges 2007). In contrast, in the family Gesneriaceae, species with bilateral flowers have different specialized and generalized pollination systems, while radial symmetry occurs in species with specialized bee-pollination (Martén-Rodríguez et al. 2010). Further studies on groups of plants that include both types of floral symmetry should use a phylogenetic approach to assess the evolution of floral symmetry in relation to pollinator shifts. Since the genes that determine floral symmetry are known in various plant groups (e.g., *Antirrhinum*), experimental approaches using symmetry mutants could be useful in evaluating how changes in symmetry may affect the evolution of pollination systems.

11.6 Fourth Hypothesis: Pollinator-Dependent Species Should Experience More Consistent Selection on Floral Traits than Species Less Dependent on Pollinators

In general, pollinators expected by the syndrome are more efficient on pollinator-dependent plants (dioecious, monoecious, and self-incompatible hermaphrodite species) than on non-dependent species (self-compatible species) (Rosas-Guerrero et al. 2014). This pattern was statistically significant for the bat and bird syndromes only (Fig. 11.4). This result indicates that the fitness of plant species of these two syndromes is maintained by high outcrossing rates that, in turn, are apparently maintained by the effective pollination and long gene flow distances via pollen provided by bats and birds (e.g., Aldrich and Hamrick 1998; Quesada et al. 2004). On the other hand, for these groups, non-dependent species are less efficiently pollinated by the pollinator expected by the syndrome. Here, autogamy, either attained through autonomous self-pollination or effected by less mobile secondary pollinators, would have more chances to contribute to the reproduction of non-dependent than dependent plant species, and thus, we might expect more relaxed selection by primary pollinators on floral traits. For example, Lobo et al. (2005) found for the bat-pollinated tree *Ceiba pentandra* that high outcrossing rates predominate in regions with high pollinator visitation, while in environments with low pollinator visitation, trees changed to a mixed mating system with high levels of self-pollination.

Differences in efficiency among groups might be associated with differences in floral display, visitation rates, and pollinator behavior. It may be expected that bat- and bird-pollinated species allocate more resources per flower than bee-pollinated species (i.e., flowers are larger and have higher amounts of reward). At the same

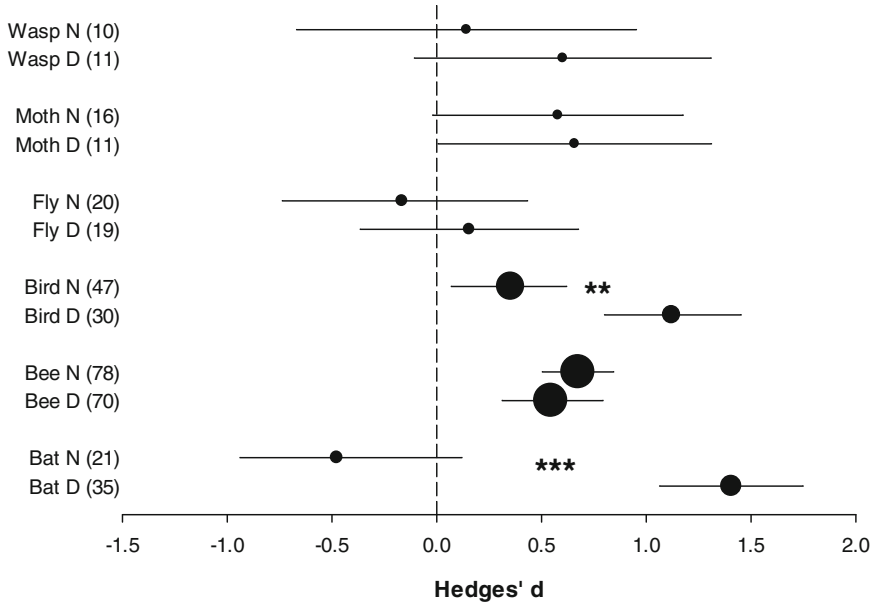


Fig. 11.4 Weighted-mean effect sizes and 95 % bias-corrected confidence intervals of the predictability of pollination syndromes on the most effective pollinators for plants with traits associated with higher predictability of pollination syndromes. Dependent (*D* self-incompatible, monoecious, and dioecious) and non-dependent (*N* self-compatible) species. Sample sizes for each category are shown in parentheses. The size of each dot represents the proportional weight or contribution to the overall mean calculation. *Dotted lines* show Hedges' $d = 0$. When confidence intervals overlap zero, the effect sizes are not significantly different from zero. Asterisks indicate significance level at $p < 0.05$ associated with Qbetween-values for each group comparison

time, due to the smaller population sizes of birds and bats compared to bees, in general, visitation frequency of bat and bird flowers should be lower compared to bees which in turn might be related to lower fruit set (e.g., *Costus*, Kay and Schemske 2003; bee-pollinated Malvaceae, Spira et al. 1992, bat-pollinated Malvaceae, Quesada et al. 2004). Given the cost of producing large flowers and low pollinator visitation rates, a high efficiency per visit would be expected in pollinator-dependent plants pollinated by bats and birds. In non-dependent species, autonomous self-pollination as a reproductive assurance mechanism is expected to be more important. In contrast, bee-floral syndrome species should have larger floral displays and higher visitation rates, but since bees collect pollen to feed their brood, bee visits may be equally or less efficient than those of bats and birds (Thomson and Wilson 2008). Therefore, for bee-syndrome species, both dependent and non-dependent plants should have similar chances of getting pollinated and selection on floral traits may not differ between different breeding systems. Overall, our results suggest that pollinator-mediated selection on suites of floral traits may be stronger on outcrossing species pollinated by highly mobile organisms, in which fitness would be highly dependent on effective visits by pollinators.

11.7 Fifth Hypothesis: Predictability of Pollination Syndromes Differs Between Tropical and Extra-Tropical Plant Species

Floral syndromes predicted the most effective pollinators in species from tropical regions than in species from other regions, suggesting that interactions with effective pollinators generate stronger selection on floral traits in the tropics (Rosas-Guerrero et al. 2014). Some associations between pollination syndromes and geographical region in our dataset are worth mentioning. Plant species with bat syndrome were mainly from tropical regions, while species with fly and wasp syndromes were mostly found in extra-tropical regions (Figs. 11.1 and 11.5).

Bats, birds, and bees were significantly more effective pollinators of their predicted floral syndromes in plants from the tropics than in plants from temperate regions, but there were no differences for the wasp, moth, and fly syndromes (Fig. 11.5). Many species from the former pollinator groups have evolved specialized relations with their host plants in tropical regions. Interestingly, the areas with the highest number of chiropterophilous columnar cacti and of bat pollinator species in Mexico overlap (Valiente-Banuet et al. 1996). Additionally, specific evolutionary relationships have

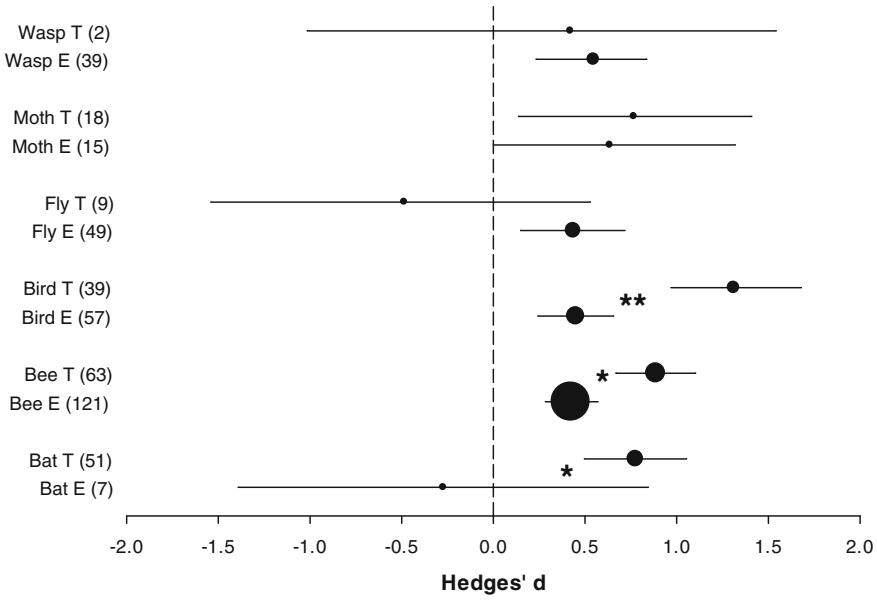


Fig. 11.5 Weighted-mean effect sizes and 95 % bias-corrected confidence intervals of the predictability of pollination syndromes on the most effective pollinators for plants belonging to tropical (*T*) and extra-tropical (*E*) regions. The size of each dot represents the proportional weight or contribution to the overall mean calculation. *Dotted lines* show Hedges' $d = 0$. When confidence intervals overlap zero, the effect sizes are not significantly different from zero. *Asterisks* indicate significance level at $p < 0.05$ associated with $Q_{between}$-values for each group comparison

arisen between euglossine bee species and fragrance-producing orchid species (Ramírez et al. 2011), and oil- and resin-producing flowers and bees (e.g., *Dalechampia* spp. pollinated by *Eulaema* and *Eufriesea* bees, Armbruster 1993; Malpighiaceae pollinated by *Centris* bees, Sigrist and Sazima 2004).

The fact that floral syndromes best predicted the most effective pollinators in the tropics may be attributed to stronger biotic interactions and narrower niches (Mittelbach et al. 2007; Schemske et al. 2009). Selection to reduce niche overlap may be reflected in the non-overlapping flowering phenologies of various groups of tropical plants. For example, Bombacaceous trees at different tropical sites maintain similar flowering phenologies and sequential flowering at each site, which possibly maintains a steady supply of floral resources for pollinators promoting pollinator fidelity (Lobo et al. 2003; Rosas-Guerrero et al. 2014). These phenological patterns may allow selection on floral traits associated with effective and constant functional groups of pollinators (Janzen 1967).

Another possible explanation for the stronger association between floral syndromes and the most effective pollinators is that, in general, tropical plant taxa have had a longer evolutionary history than temperate taxa (Hawkins et al. 2011; Kerkhoff et al. 2014). For example, some temperate plant taxa (e.g., originally placed in families Apocynaceae, Apiaceae) are derived from plant lineages that originated in the tropics (e.g., Asclepiadaceae and Araliaceae, respectively; Judd et al. 1994). Therefore, tropical plant species may have had more time to experience selection by particular pollinators than their temperate counterparts.

Studies based on visitor assemblages that analyze specialization between geographical regions (e.g., Ollerton et al. 2009; Schleuning et al. 2012) are not directly comparable with our review, because they did not quantify the effectiveness of all floral visitors and frequent floral visitors are often poor pollinators (Fenster et al. 2004; King et al. 2013). Our approach allowed demonstrating that globally, pollinators expected by the syndromes are indeed the most efficient, among other pollinators. Our results significantly contrast with the findings of Ollerton et al. (2009) who found support for pollination syndromes for around 30 % of plants of six communities and stated that tropical communities did not exhibit greater predictability of pollination syndromes than temperate communities. Ollerton et al.'s conclusions are rather limited because they consider floral visitors of a subset of plant species from each community. Furthermore, their approach resulted in the evident mis-assignment of syndromes to many species of known floral syndromes, such as a fly syndrome to specialized resin-producing *Dalechampia*; and bee syndrome to *Heliconia* spp. The appropriateness of assignments can not be assessed due to the lack of taxonomic resolution of their dataset. Additionally, the lack of nocturnal pollinator observations in Ollerton et al.´s study would undoubtedly have caused disagreement between assigned moth or bat pollination syndromes and the observed floral visitors. While these authors argue to present a worldwide review of pollination syndromes, they only analyzed three tropical communities, mainly under disturbed conditions, with biased sampling effort against tropical communities that resulted in a large number of species without pollinator observations, and an extremely limited plant species taxonomic identification effort, which unfortunately precludes a real comparison between tropical

and temperate plant species. Last, Ollerton et al. (2009) did not include the major tropical biomes such as mature tropical rain or tropical dry forests and had an extremely limited sampling of important tropical life forms such as trees, epiphytes, and lianas. Our study is based on a representative sample of tropical and temperate natural communities, and therefore, we are confident in stating that pollinators expected by the syndrome are in fact better predicted in tropical than in temperate plant communities, indicating stronger selection on floral traits in the tropics.

11.8 Conclusions and Future Directions

In this study, we found that within each syndrome, there is greater predictive accuracy of the syndrome when only primary pollinators are involved than when secondary pollinators are present. However, the occurrence of secondary pollinators does not contradict the evolution of pollination syndromes, among other reasons, because the most frequent secondary pollinator was also the least efficient one. All pollination syndromes had species with secondary functional groups of pollinators. These results suggest that current suites of floral traits in most plant species have the potential for adapting to new conditions under changing selective pollination environments (Kay et al. 2005; Whittall and Hodges 2007).

The only group of plants able to filter effectively secondary pollinators was the long-tongued fly-syndrome flowers (Rosas-Guerrero et al. 2014). Nevertheless, about 30 % of plant species were exclusively visited by the syndrome pollinator functional group. Exclusion of secondary pollinators would be favored by greater costs than benefits of secondary pollinators. In terms of the pollen presentation theory (Thomson 2003), the current lack of secondary pollinators in those species indicates that the marginal gain in fitness derived from secondary pollinator visitation is less than the costs to those plants in terms of lost mating opportunities and wasted rewards on visitors that do not contribute to the plant's fitness. However, the fitness landscape is bound to fluctuate with time and across environments. Many factors may influence the reliability of primary pollinators including climatic events and natural and anthropogenic disturbance (e.g., droughts, storms, hurricanes, pesticides) (Aguilar et al. 2006; Winfree et al. 2009; Goulson et al. 2015). An increase in the relative frequency of less vulnerable secondary pollinators could have evolutionary consequences for the plants (Thomson and Wilson 2008).

Widespread current human-induced disturbances often modify species distributions, abundance, composition, and biotic interactions. In mutualistic plant–pollinator relationships developed through evolutionary time, differential species-specific responses to human disturbances may alter original species matching by triggering changes at different levels, involving new temporal and spatial species distributions, and/or new physiological and morphological responses (Aizen and Vázquez 2006). Such new outcomes can modify the strength of interactions and promote novel plant–pollinator relationships (Aizen et al. 2008;

Tylianakis et al. 2008). Thus, under changes driven by human activities, we expect changes in pollinator-mediated selection on floral traits.

Incipient but consistent evidence indicates that human disturbances affect the composition of pollinator assemblages, their visitation rates to flowers, and/or their foraging behavior and efficiency (e.g., Aguilar et al. 2006; González-Varo et al. 2009; Parsche et al. 2011). For most flowering plants, such changes will have direct effects on their mating patterns (Aguilar et al. 2008), which may trigger the development of novel reproductive strategies to cope with new scenarios, which in turn may influence the evolution of floral traits. Plant species unable to exploit or to develop new interactions or alternative reproductive strategies will increase their local extinction probabilities (Biesmeijer et al. 2006; Anderson et al. 2011).

Furthermore, as a consequence of increased human mobility, habitat transformation, and global change, many species of plants and animals are increasing or shifting their distribution ranges, and some of them can become invasive in new habitats (Simberloff et al. 2013). Invasive plants may not only lose their natural enemies, but also some of their natural mutualists. This scenario may increase the importance of alternative pollinators, and depending on the relative rates of migration of the plants and their primary pollinators, it could lead to an evolutionary shift in primary pollinators.

A similar case is that of invasive pollinator species. For instance, as the Africanized bee expanded its range, it could have become an important secondary or even primary pollinator of many species. Apart from bee-syndrome plants, Africanized bees might easily exploit some bird-, fly-, and bat-syndrome plants (Fig. 11.3). Africanized bees have become as or more efficient than native pollinators of the herb *Kallstroemia grandiflora* (Zygophyllaceae), the prominent Amazonian tree *Dimizia excelsa* (Fabaceae), the South American tree *Tibouchina granolas* (Melastomataceae), and certain crops such as tomato (Osorio-Beristain et al. 1997; Dick 2001; Macias-Macias et al. 2009; Brizola-Bonacina et al. 2012).

A requisite for invasiveness of a plant species is its ability to use the resources in the novel environments to its favor in demographic terms. To become invasive, an animal-pollinated plant would have to be preadapted to the pollinator fauna of the novel environment and would be seen as rapidly integrated into that network (Lopezaraiza-Mikel et al. 2007). Further adaptation to a bee syndrome or floral specialization of invasive plants that would augment the efficiency of pollination by Africanized bees has not been documented yet, but cases in which invasive plant species are able to use the services of invasive pollinators have been documented (Beavon and Kelly 2012) and could be common. Neither of the above scenarios is exclusive of one another. Thus, one could hypothesize that the importance of secondary pollinators for a given plant (and even for plants of a given syndrome) is a function of the magnitude by which primary pollinators are differentially negatively affected by factors that fluctuate in time and across environments relative to secondary pollinators.

Here, we have found that the combinations of flower traits known as "floral syndromes" are significantly shaped by the most efficient pollinator functional group. However, most plant species are also visited by several animals with

different pollination efficiencies. Under changing environments sustained in time, the most efficient pollinator functional group may no longer prevail and its role may be taken up by another pollinator functional group. The increased relative contribution to effective pollination exerted by a different functional group can impose a potential venue to drive novel evolutionary changes in floral traits, eventually modifying floral syndromes in novel environments. Such possibility may be more feasible for plant species originally interacting with more than one functional group, that is, around 70 % of the plant species in our study, whereas plants interacting with mainly a single pollinator functional group (e.g., long-tongued fly syndrome) may sustain their syndrome by the presence of one or some species of the same functional group (redundant), and depending on its breeding system (more or less dependent on animal pollination), it will survive or perish in disturbed habitats. Here, we have presented for the first time, evidenced-based clues of the most probable venues or candidates of pollinator transitions for each pollination syndrome in currently changing environments.

Should disturbance-induced extinction probabilities differ among pollinator's functional groups, we may predict decreases in floral syndrome diversity. Certain ecological traits of animal pollinators such as their mobility capacity, niche breadth, life cycle requirements, or reproductive capacity may help to predict which functional types will be more likely to disappear (Alanen et al. 2011). For example, insect pollinators such as beetles, butterflies, and moths require different resources from multiple habitat types during their life cycle, and therefore, habitat disturbance may reduce populations of these animals, in some cases to the point of extinction (Alanen et al. 2011). Absent pollinator functional groups in novel environments will loosen their influence on their plant partners, which in turn may switch their "attention" to other functional groups. In contrast, more mobile or social functional groups such as birds, bats, and certain bees may be more resilient to habitat disturbance (Winfree et al. 2009; Phillips et al. 2010, but see Anderson et al. 2011) and potentially predominate as pollinators and agents of selection. However, a more realistic scenario should consider the likelihood of both, syndrome transitions and extinctions of functional groups of pollinators. Empirical evidence around these ideas is scant, but they certainly deserve research attention to achieve a better understanding of the potential future of floral evolution.

Acknowledgements This work was supported by grants from Universidad Nacional Autónoma de México (UNAM:PAPIIT Project IN201011 to MQ, PEID to GA-S.), Consejo Nacional de Ciencia y Tecnología (CONACyT: Proyecto Laboratorio Nacional de Análisis y Síntesis Ecológica para la Conservación de Recursos Genéticos U-3-2015-2-250996, CB-2009-131008 to MQ, CB-2010-155016 to SM-R and MQ, CB-2014-241696 to ML-M), CONICET (PIP 0019) and FONCyT (PICT 1606) to RA and LA. We thank Gumersindo Sánchez Montoya for assistance in data collection.

References

Adams DC, Gurevitch J, Rosenberg MS (1997) Resampling tests for meta-analysis of ecological data. Ecology 78:1277–1283

Aguilar R, Ashworth L, Galetto L, Aizen MA (2006) Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. Ecol Lett 9:968–980

Aguilar R, Quesada M, Ashworth L et al (2008) Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Mol Ecol 17:5177–5188

Aizen MA, Vázquez DP (2006) Flower performance in human-altered habitats. In: Harder LD, Barrett SCH (eds) Ecology and evolution of flowers. Oxford University Press, Oxford, pp 159–179

Aizen MA, Morales CL, Morales JM (2008) Invasive mutualists erode native pollination webs. PLoS Biol 6:e31. doi:10.1371/journal.pbio.0060031

Alanen EL, Hyvonen T, Lindgren S et al (2011) Differential responses of bumblebees and diurnal Lepidoptera to vegetation succession in long-term set-aside. J App Ecol 48:1251–1259

Aldrich PR, Hamrick JL (1998) Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. Science 281:103–105

Anderson SH, Kelly D, Ladley JJ et al (2011) Cascading effects of bird functional extinction reduce pollination and plant density. Science 331:1068–1070

Armbruster WS (1993) Evolution of plant pollination systems: hypotheses and tests with the neotropical vine Dalechampia. Evolution 47:1480–1505

Beavon MA, Kelly D (2012) Invasional meltdown: pollination of the invasive liana *Passifloratripartita var. mollissima* (Passifloraceae) in New Zealand. N Z J Ecol 36:100–107

Biesmeijer JC, Roberts SPM, Reemer M et al (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313:351–353

Brizola-Bonacina AK, Arruda VM, Alves-Junior VV et al (2012) Bee visitors of quaresmeira flowers (*Tibouchina granulosa* Cogn.) in the region of Dourados (MS-Brasil). Sociobiology 59:1253–1267

Chamberlain SA, Hovick SM, Dibble CJ et al (2012) Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. Ecol Lett 15:627–636

Cooper HM, Hedges LV, Valentine JC (eds) (2009) The handbook of research synthesis and meta-analysis. Russell Sage Foundation Publications, New York

Darwin C (1862) On the various contrivances by which British and foreign orchids are fertilized. Murray, London

Delpino F (1867) Sugli apparecchi della fecondazione nelle piante antocarpee (fanerogame): sommario di osservazioni fattenegli anni 1865-1866. Coi tipi di M. Cellini e C. alla Galileiana

Dick CW (2001) Genetic rescue of remnant tropical trees by an alien pollinator. Proc R Soc Lond B 268:2391–2396

Faegri K, van der Pijl L (1979) Principles of pollination ecology. Pergamon Press, Oxford

Fenster CB, Armbruster WS, Wilson P et al (2004) Pollination syndromes and floral specialization. Annu Rev Ecol Evol Syst 35:375–403

Fenster CB, Reynolds RJ, Williams CW et al (2015) Quantifying hummingbird preference for floral trait combinations: the role of selection on trait interactions in the evolution of pollination syndromes. Evolution. doi:10.1111/evo.12639

Friedman WE (2009) The meaning of Darwin's "abominable mystery". Am J Bot 96:5–21

Gómez JM, Perfectti F, Camacho JPM (2006) Natural selection on *Erysimum mediohispanicum* flower shape: Insights into the evolution of zygomorphy. Am Nat 168:531–554

Gómez JM, Perfectti F, Klingenberg CP (2014) The role of pollinator diversity in the evolution of corolla-shape integration in a pollination-generalist plant clade. Phil Trans R Soc B 369 (2013):0257. doi:10.1098/rstb.2013.0257

González-Varo JP, Arroyo J, Aparicio A (2009) Effects of fragmentation on pollinator assemblage, pollen limitation and seed production of Mediterranean myrtle (*Myrtus communis*). Biol Cons 142:1058–1065

Goulson D, Nicholls E, Botías C et al (2015) Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347. doi:10.1126/science.1255957

Gurevitch J, Hedges LV (1999) Statistical issues in conducting ecological meta-analyses. Ecology 80:1142–1149

Gurevitch J, Hedges LV (2001) Meta-analysis: combining the results of independent experiments. In: Scheiner SM, Gurevitch J (eds) Design and analysis of ecological experiments, 2nd edn, Oxford University Press, New York, pp 378–398

Harder LD, Johnson S (2009) Darwin's beautiful contrivances: evolutionary and functional evidence for floral adaptation. New Phytol 183:530–545

Hawkins BA, Rodriguez MA, Weller SG (2011) Global angiosperm family richness revisited: linking ecology and evolution to climate. J Biogeogr 38:1253–1266

Huang SQ, Gong YB (2009) Floral symmetry: pollinator-mediated stabilizing selection on flower size in bilateral species. Proc R Soc B 276:4013–4020

Janzen D (1967) Synchronization of sexual reproduction of trees within the dry season in Central America. Evolution 21:620–637

Judd WS, Sanders RW, Donoghue MJ (1994) Angiosperm family pairs: preliminary phylogenetic analyses. Harvard Pap Bot 1:1–51

Kay KM, Schemske DW (2003) Pollinator assemblages and visitation rates for eleven species of Neotropical Costus (Costaceae). Biotropica 35:198–207

Kay KM, Reeves PA, Olmstead RG et al (2005) Rapid speciation and the evolution of hummingbird pollination in neotropical *Costus* subgenus *Costus* (Costaceae): evidence from nrDNA ITS and ETS sequences. Am J Bot 92:1899–1910

Kerkhoff AJ, Moriarty PE, Weiser MD (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. Proc Nat Acad Sci 111:8125–8130

King C, Ballantyne G, Willmer PG (2013) Why flower visitation is a poor proxy for pollination: measuring single-visit pollen deposition, with implications for pollination networks and conservation. Methods Ecol Evol 4:811–818

Kunth P (1906) Handbook of flower pollination. (English trans. from German)

Lobo JA, Quesada M, Stoner KE et al (2003) Factors affecting phenological patterns of bombacaceous trees in seasonal forests in Costa Rica and Mexico. Am J Bot 90:1054–1063

Lobo JA, Quesada M, Stoner KE (2005) Effects of pollination by bats on the mating system of Ceiba pentandra (Bombacaceae) populations in two tropical life zones in Costa Rica. Am J Bot 92:370–376

Lopezaraiza-Mikel ME, Hayes RB, Whalley RM et al (2007) The impact of an alien plant on a native plant–pollinator network: an experimental approach. Ecol Lett 10:539–550

Macias-Macias O, Chuc J, Ancona-Xiu P et al (2009) Contribution of native bees and Africanized honey bees (Hymenoptera: Apoidea) to Solanaceae crop pollination in tropical Mexico. J App Entomol 133:456–465

Martén-Rodríguez S, Fenster CB (2008) Pollination ecology and breeding systems of five *Gesneria* species from Puerto Rico. Ann Bot 102:23–30

Martén-Rodríguez S, Almarales-Castro A, Fenster CB et al (2009) Evaluation of pollination syndromes in Antillean Gesneriaceae: evidence for bat, hummingbird and generalized flowers. J Ecol 97:348–359

Martén-Rodríguez S, Fenster CB, Agnarsson I et al (2010) Evolutionary breakdown of pollination specialization in a Caribbean plant radiation. New Phytol 188:403–417

Mittelbach GG, Schemske DW, Cornell HV et al (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol Lett 10:315–331

Moya-Laraño J (2010) Can temperature and water availability contribute to the maintenance of latitudinal diversity by increasing the rate of biotic interactions? Open Ecol J 3:1–13
Ne'eman G, Jurgens A, Newstrom-Lloyd L et al (2010) A framework for comparing pollinator performance: effectiveness and efficiency. Biol Rev 85:435–451
Ollerton J, Alarcón R, Waser NM et al (2009) A global test of the pollination syndrome hypothesis. Ann Bot 103:1471–1480
Osorio-Beristain M, Dominguez CA, Eguiarte LE et al (1997) Pollination efficiency of native and invading Africanized bees in the tropical dry forest annual plant, *Kallstroemia grandiflora* Torr ex Gray. Apidologie 28:11–16
Parsche S, Frund J, Tscharntke T (2011) Experimental environmental change and mutualistic vs. antagonistic plant flower–visitor interactions. Persp Pl Ecol Evol Syst 13:27–35
Phillips RD, Hopper SD, Dixon KW (2010) Pollination ecology and the possible impacts of environmental change in the Southwest Australian Biodiversity Hotspot. Phil Trans R Soc B 365:517–528
Proctor M, Yeo P, Lack A (1996) The Natural History of Pollination. Harper Collins, London
Quesada M, Stoner KE, Lobo JA et al (2004) Effects of forest fragmentation on pollinator activity and consequences for plant reproductive success and mating patterns in bat-pollinated bombacaceous trees. Biotropica 36:131–138
Ramírez SR, Eltz T, Fujiwara MK et al (2011) Asynchronous diversification in a specialized plant-pollinator mutualism. Science 333:1742–1746
Reynolds RJ, Westbrook MJ, Rohde AS et al (2009) Pollinator specialization and pollination syndromes of three related North American *Silene*. Ecology 90:2077–2087
Rosas-Guerrero V, Quesada M, Armbruster WS et al (2011) Influence of pollination specialization and breeding system on floral integration and phenotypic variation in *Ipomoea*. Evolution 65:350–364
Rosas-Guerrero V, Aguilar R, Martén-Rodríguez S, Ashworth L, Lopezaraiza-Mikel M, Bastida JM, Quesada M (2014) A quantitative review of pollination syndromes: do floral traits predict effective pollinators? Ecol Lett 17:388–400
Rosenberg MS, Adams DC, Gurevitch J (2000) MetaWin: Statistical software for meta-analysis: version 2.1.5. Sinauer Associates, Sunderland, Massachusetts
Sargent RD (2004) Floral symmetry affects speciation rates in angiosperms. Proc R Soc B 271:603–608
Schemske DW, Mittelbach GG, Cornell HV et al (2009) Is there a latitudinal gradient in the importance of biotic interactions? Annu Rev Ecol Evol Syst 40:245–269
Schleuning M, Fründ J, Klein AM et al (2012) Specialization of mutualistic interaction networks decreases toward tropical latitudes. Curr Biol 22:1925–1931
Sigrist MR, Sazima M (2004) Pollination and reproductive biology of twelve species of neotropical Malpighiaceae: stigma morphology and its implications for the breeding system. Ann Bot 94:33–41
Simberloff D, Martin JL, Genovesi P et al (2013) Impacts of biological invasions: what's what and the way forward. Trends Ecol Evol 28:58–66
Spira TP, Snow AA, Whigham DF et al (1992) Flower visitation, pollen deposition, and pollen-tube competition in *Hibiscus moscheutos* (Malvaceae). Am J Bot 4:428–433
Stebbins GL (1970) Adaptive radiation of reproductive characteristics in angiosperms. I: pollination mechanisms. Annu Rev Ecol Evol Syst 1:307–326
The Plant List (2013) Version 1.1. Published on the internet. http://www.theplantlist.org. Accessed 26 March 2015
Thomson JD (2003) When is it mutualism? Am Nat 162:S1–S9
Thomson JD, Wilson P (2008) Explaining evolutionary shifts between bee and hummingbird pollination: convergence, divergence, and directionality. Int J Plant Sci 169:23–38
Tylianakis JM, Didham RK, Bascompte J et al (2008) Global change and species interactions in terrestrial ecosystems. Ecol Lett 11:1351–1363

Valiente-Banuet A, Arizmendi MC, Rojas-Martinez A et al (1996) Ecological relationships between columnar cacti and nectar-feeding bats in Mexico. J Trop Ecol 12:103–119
van der Niet TA, Johnson SD (2012) Phylogenetic evidence for pollinator-driven diversification of angiosperms. Trends Ecol Evol 27:353–361
Vogel S (1954) Blütenbiologische typen als elemente der sippengliederung: dargestellt anhand der Flora Südafrikas. Bot Stud (Fischer-Jena)
Waser NM, Chittka L, Price MV et al (1996) Generalization in pollination systems, and why it matters. Ecology 77:1043–1060
Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature 447:706–709
Willmer P (2011) Pollination and floral ecology. Princeton University Press, Princeton
Winfree R, Aguilar R, Vázquez D et al (2009) A meta-analysis of bees' responses to anthropogenic disturbance. Ecology 90:2068–2076
Wolfe LM, Krstolic JL (1999) Floral symmetry and its influence on variance in flower size. Am Nat 154:484–488

Chapter 12
Altruism, Religion, and Self-Enhancement in a Framework of Ad Hoc Evolutionary Adaptation

Florian Habermacher

Abstract We review evolutionary explanations for three major puzzles of the human mind: altruism, religiosity, and self-enhancement. Human altruism reaches beyond reciprocity or close-kin care readily explained by game theory and genetic kin selection. Group selection is widely seen as too weak to lead to substantial altruism as it struggles to contain selfishness favored by within-group selection. Yet, reciprocity and punishment leverage the effectiveness of altruism within a group, and genuine altruism is testified to be weak, leaving scope for explanation even by a force as weak as group selection. Moralistic religious culture appears tightly linked to altruism, yet the fitness advantage of a defector within a religious society makes it difficult to conceive religion or related genetic predisposition as an evolutionarily stable strategy. Self-enhancement has direct links to altruism and religiosity, leading to warm-glow altruistic contributions and increased receptiveness to comforting narratives of heavenly justice. Suggested intrapersonal and interpersonal benefits of self-enhancement do not detail how the trait should be competitive against more direct behavioral adjustments that yield similar personal benefits but avoid the fitness costs of misperception. We explain religion and bias as imperfect ad hoc evolutionary adaptations rather than perfect evolutionarily stable strategies (ESS). The scant time available for fine-tuning the mind since the emergence of higher cognitive capabilities means near-perfect traits were unlikely to emerge. Instead, the extraordinary evolutionary pressure induced by the rapidly evolving environment favored a broad range of genetic novelties despite extra costs.

F. Habermacher (✉)
Swiss Institute of International Economics and Applied Economic Research,
Department of Economics, University of St. Gallen, Bodanstrasse 8,
9000 St. Gallen, Switzerland
e-mail: florian.habermacher@unisg.ch

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_12

12.1 Introduction

Applied to the individual, the first principle of evolution, 'survival of the fittest' (Spencer 1864; Darwin 1869), would seem to favor a rationally calculating mind that acts in perfect accordance with its capabilities and with evidence weighted by statistical significance. Others—with the exception of close relatives if the principle was broadened to include kin—would be valued strictly according to the personal benefit one could expect to derive from them. It is difficult to imagine a world populated entirely with humans exhibiting such characteristics, notably the psychopathic substitution of empathy with purely manipulative rationality. A glimpse at our surroundings, as well as a host of scientific studies, confirms that the human mind differs in various ways from such a characterization. Three striking (categories of) variations fall under the complex concepts—for simplicity called *traits*—of altruism, religiosity, and self-enhancement. All three traits are puzzling in the sense that they have fitness costs but unclear (individual) benefits. The following provides an—necessarily incomplete—overview of proposed evolutionary explanations for the interrelated phenomena. Some implicit assumptions in the proposed theories seem difficult to reconcile with reality and we propose amendments to address these difficulties.

We consider altruism in the strict biological sense of an individual deliberately undergoing net (lifetime) fitness costs in order to increase the welfare of others (e.g., West et al. 2007; Okasha 2013). Evolutionarily, such behavior is, almost by definition, difficult to sustain in a simple world. The personal fitness cost means that genes of bearers of an allele for non-discriminatory altruism are underrepresented in the offspring, leading to a crowding out. The exception is altruism targeted at close kin: Kin share an elevated fraction of genes, so that helping them increases 'inclusive fitness,' the likelihood of *copies* of one's own genes propagating (Hamilton 1964; Grafen 1984). It is important to delineate altruism from purely strategic cooperation between individuals to their mutual (longer term) advantage, two often confounded concepts (West et al. 2007). Cooperation, often sustained by direct and indirect reciprocity in repeated interactions with reputational concerns and/or punishment (e.g., Fehr and Fischbacher 2003, 2004; Nowak 2006), can improve the fitness of all involved. The first principle of evolution thus naturally explains strategic cooperation when cognitive capabilities allow tracking historic behavior of individuals (Bshary and Bergmueller 2008). Altruism observed in daily life, laboratories, and hunter-gatherer societies goes well beyond what kin-regarding preferences and strategic cooperation imply (e.g., Fehr and Fischbacher 2003; Bowles 2006), requiring further explanation. Group-selection forces, simplistic interpretations of which dominated the literature earlier in the twentieth century, were largely, and rather categorically, dismissed as inconsequential since Williams (1966) debunked a particularly naïve 'good-for-the-species' version of group selection among animals. The literature seems, however, to lack compelling alternatives to group-level explanations for the long reach of deep-rooted human altruism beyond kin-aid and strategic interaction. Section 12.2

considers evidence in favor of a plausible role of group-level forces in shaping human other-regarding preferences, drawing attention notably to the fact that on a general level, observed human altruism has to be understood as *being* of a very modest level itself (e.g., Kirchgässner 2008, 2010), so that the weakness of a force does not a priori dismiss it as a relevant explanatory factor.

Evolutionary biology mainly focused on genetic evolution of altruism, and experimental evidence suggests that altruism is deeply rooted in the physiology of the mind (Warneken and Tomasello 2009). Yet, no agreement has been reached as to the degree to which genes directly or culture instead has shaped human altruism: 'Some would say that culture is so important that genes, whether selfish or not, are virtually irrelevant to the understanding of human nature. Others would disagree. It all depends where you stand in the debate over "nature versus nurture"' (Dawkins 1989, p. 3).

Undeniably, culture plays a crucial role in shaping human altruistic behavior. Throughout the world, an extraordinary influence seems to stem from religions. Even if the popular saying 'Without God, everything is permitted' is hardly reconcilable with observations of atheistic societies, we find an impressive literature showing links between religion and moral behavior, and it is striking how closely the evolution of moralistic religions is linked to the enlargement of human group sizes in the past millennia (e.g., Wright 2010). While prehistoric religions and superstitions of bands and small tribes were largely unmoral, the emergence of life within larger groups was accompanied by a shift of the focus of supernatural agents from rather random behavior to the enforcement of more 'moral' behavior, beneficial for the provision of public goods in the progressively more anonymous groups (Wright 2010). A basic link between exhibited altruistic behavior,[1] or morality, and religion is straightforward, with a supernatural, possibly omnipresent punisher (or rewarder) enforcing moral norms when respecting them would otherwise not be in an individual's self-interest. Evolutionarily, religion has thus been explained as spreading through cultural group selection, increasing group fitness by lowering the cost for the enforcement of moral norms. From a group perspective, the high cost of rites and distortionary rules opens the question of why more cost-effective ways to enforce moral norms, such as genuine altruism, did not evolve instead. Furthermore, from an individual perspective, defecting individuals with immunity against internal adoption of religious belief should enjoy a fitness advantage as they would be able to exploit the system. According to the first law of evolution, such immunity should thus spread over time within religious societies. Section 12.3 considers proposed solutions to the evolutionary dilemmas regarding the emergence of superstition and religions based notably on patternicity—the tendency to causally link independent random events (Shermer 2008, 2009; Foster and Kokko 2009)—and on costly signaling theory (Henrich 2009). Contrary to what has been inferred,

[1]If helping others is based on piety, classification as altruism may seem counterintuitive, as no genuine other-(human)-regarding preference may be involved. However, from an evolutionary, biological perspective, the behavior does indeed imply the trade-off of a reduction in own fitness against an increase of others' fitness.

we do not find these propositions convincing as (long-run) evolutionarily stable strategies (ESS) (Smith and Price 1973). Instead, we explain the observed traits as the result of *ad hoc evolutionary selection*, favoring adaptations that are imperfect yet broadly in line with the dominant evolutionary pressure, relevant when this pressure is strong and time for adaptation scarce on a genetic time scale, as was likely the case for the fine-tuning of the human mind since the emergence of our higher cognitive capacities.

Self-enhancement, or self-serving bias, denotes a class of perceptual biases that involve taking a tendentiously positive view about one's own personality, capabilities, and things associated with the self (Alicke and Sedikides 2011; Heine et al. 1999; Babcock and Loewenstein 1995; Codol 1995; Sedikies and Gregg 2008). This comprises tendencies such as that of substantially more than half of the population typically judging themselves as above average in various skills, attributing personal success to our ability but failure to bad luck, or judging favorably information in support of our own position while disregarding equally strong evidence against it (Codol 1995). Section 12.5 considers direct links of such bias to altruism and religiosity. Explanations the literature proposes for the evolution of self-enhancement can be classified as intrapersonal or interpersonal. Intrapersonally, self-enhancement is found to prevent evolutionarily unfavorable states of the mind such as depression (Greenwald 1980). Interpersonally, the personal bias lowers the cognitive burden of enhancing the self in front of others, helping to instill confidence and to achieve evolutionarily relevant social goals such as high status. Both dimensions have natural benefits, but an impressive list of overhead costs of the bias has been found, and Paulhus (1998) concluded 'self-enhancement is best viewed as a mixed blessing' (p. 1207). Fundamentally, it has been argued that the intrapersonal aspects are evolutionarily inconsequential as, instead of distorting perception in a costly manner, an ESS would more efficiently optimize directly the decision rule rather than the underlying perceptions whose distortion implies undesirable side effects (e.g., Pinker 2011). We explain that this critique strictly speaking extends to the case of interindividual effects, as observers perceive an individual's perceptions only indirectly, through implied actions and expressions of emotions. Updating the relationship between perceptions and the induced actions and expressions would be the efficient strategy, rather than biasing perceptions that are relevant for various other aspects of life. Nevertheless, biased perceptions as a pancultural human trait (Sedikides et al. 2003; Markus et al. 1991) are a well-documented fact that requires explanation. Section 12.4 suggests viewing also self-enhancement as the product of *ad hoc evolutionary selection*. When early hominids and humans emerged, attempts for other-deception would in many cases have been largely futile, owing to the intimacy of small band living, where direct observation offered a primary source of evidence on which judgment of others could largely rely. As group sizes increased over time (Aiello and Dunbar 1993; Dunbar and Shultz 2007; Grove 2013), secondary evidence, notably oral accounts, gained in importance, increasing the scope for other-deception (e.g., Wright 2010). This sudden emergence (or increase) of possibilities for personal fitness advantage through regular exaggerations about personal virtues and deeds must have

represented a strong pressure and a formidable playground for numerous genetic traits that could aid individual humans toward this deceptive aim. Ideally, from the individual's point of view, adaptations would have made of human a great, fully conscious storyteller that himself perfectly knows reality apart from his tales. This may not have been achievable in the short time available. Yet, a lot was gained with a possibly simpler rewiring of subconscious functions, leading to self-enhancement as an indirect way to influence others. Evidence of group-level biases derived from self-enhancement suggests that self-enhancement could have further played a role in containing a rather general level of altruism to group members.

Sections 12.2–12.4 discuss such evolutionary aspects of altruism, religiosity, and self-enhancement in detail. Section 12.5 brings the pieces together. It also considers how the interplay of flexible religiosity and self-enhancement could have outperformed a more stable innate altruism in an evolutionary setting with intermittent intergroup pressure (Habermacher 2014).

Only where necessary, the discussion on altruism refers to the tangent issue of purely strategic cooperation. By leveraging the effect of altruistic contributions from strong reciprocators (Fehr and Fischbacher 2003), reciprocal cooperation with punishment suggests that the effectiveness of altruism on group fitness can be large.

12.2 Altruism

12.2.1 Group and Kin Selection

In *The Descent of Man*, Darwin (1871) advanced the idea of group selection to rationalize notably human parochial altruism. By the 1930s, group selection was commonly used among biologists to explain all sorts of traits for which a benefit for groups or species seemed intuitive. Quantitative subtleties relevant to answer whether the force of occasional selection at group level could plausibly override more immediate forces of intragroup competitiveness were largely ignored. During the 1960s, this changed dramatically. Hamilton's inclusive fitness theory (Hamilton 1964), applied notably as 'kin selection,' explained many phenomena originally ascribed to group-level forces, as the natural outcome of selection of genes without direct invocation of explicit group boundaries (Smith 1964; Smith and Price 1973). The critics had good points. After all, genes for helping close-kin are readily explained by relatives sharing an elevated fraction of genes (Hamilton 1964); competing individuals may restrain aggression in their rivalry, preventing serious injuries, not so much for the benefit of the population but simply to their individual benefit (Smith and Price 1973); and empirical work found individuals reproduce at a rate that maximizes their longer run reproductive success rather than to practice reproductive restraint (Lack 1966; Krebs and Davies 1993). The categorical rejection went as far as the claim by Williams (1966) that whenever selection at a lower level could offer an alternative explanation, it be *impermissible* to invoke

group selection (Lack 1966, pp. 130–131). Such strong backlash was arguably an overshoot. Smith's (1964) haystack model, seemingly explaining the impossibility of effective group selection, has been found to artificially inflate the intragroup pressure (Wilson 1987), and the early haplodiploidy hypothesis, which seemed to explain eusociality as a trivial result of kin selection in various insects, had neglected longer run genetic dynamics whose recognition today means the original kin-based explanation requires revision (West and Gardner 2010). The idea of group selection, framed in modern multi-level selection models, is now understood as mathematically equivalent to kin selection (Frank 2012), yet the empirical relevance of the selective pressure at the group level for the shaping of traits within species remains debated (Wilson 2012; Traulsen et al. 2008; Sober and Wilson 1998; West et al. 2008).

12.2.2 Human Altruism: Qualitatively Broad, Quantitatively Weak

The long reach of human altruism is unparalleled in the animal kingdom and requires refined explanation. Eusocial animals seem more perfect altruists within their hive than humans among their entourage, but human altruistic behavior extends—at a however low level—to entirely unrelated members in experiments and natural settings where own material or reputational reward is unlikely, such as in many anonymous donations (Fehr and Fischbacher 2003; Bowles 2006; Sen 1977). So much of seemingly altruistic behavior has been found to be subtly self-serving that it is regularly held that genuine altruism would not be real. This position is hardly tenable and instead altruism at a low level is found to be so important that the functioning of modern society without it is hardly imaginable (e.g., Sen 1977; Kirchgässner 2010). Kin selection offers direct explanation only for altruism toward closer kin. The *big mistake hypothesis* proposes to see the extension as based on imprecise kin recognition (Henrich 2004; Boyd and Richerson 2002b). Subconscious kin recognition is a noisy process, and the closer physical (transport), informational (writing and new media), and economic (extended non-zero sum relations with extended trade) integration (e.g., Wright 2010) of society relative to prehistory could support an unconscious misinterpretation of strangers as more closely related than they really are. However, toward any lower-than-average-related individual, pure-kin-selection absent of group-selection pressure should lead to a weakly positive level of *spite*, rather than any altruism (Hamilton 1970), and it remains unclear how our subconscious mind should lead us to overestimate all strangers as close enough to us such as to not only compensate for non-kin spite, but to even attribute them a discernable amount of altruism. Henrich (2004) and Boyd and Richerson (2002b) detail further reasons why the big mistake hypothesis is unlikely to realistically explain the extension of kin selection to the global population. In this respect, preferences derived from group-selection forces

seem a more plausible explanation. Group selection naturally results in the valuation of the welfare of genetically unrelated group members, and it seems more plausible that the increased interconnectedness has led humans to implicitly think of strangers as what group members used to be. Modern trade, labor division, and communication mean that we share many things in a way our prehistoric ancestors used to do only with band or tribe members.

J.B.S. Haldane's 'I will jump into the river to save two brothers or eight cousins' (Nowak and Sigmund 2007) conveys the strength of kin-selection forces, observed in families and readily explained by Hamilton's rule: For a known genetic relatedness, sacrifices are evolutionarily stable whenever the targeted kin's benefit multiplied by the relatedness exceeds the personal costs. The conditions for group selection to support a *high* level of altruism are severe (Smith 1964; Levin and Kilmer 1974; West et al. 2007). However, as Kirchgässner (1992, 2010) explains, human exhibited altruism toward non-kin is quantitatively very low in most cases. How low a level of altruism interdemic selection could support among humans has not been systematically studied. Yet recent evidence about prehistoric population structures suggests group dynamics may have supported the emergence of a limited albeit significant level of within-group altruism (Bowles 2006, 2009; Weibull and Salomonsson 2006; Sober and Wilson 1998; Traulsen et al. 2008). Group selection is thus conceivably in line with both the qualitatively broad scope and the quantitatively limited level of observed, possibly innate, altruism.[2]

Section 12.5 further details how effects of self-enhancement may help explaining the extension of a group-selection-rooted altruism beyond any specific group borders in the modern world.

12.2.3 Cultural and Indirect Sources of Altruism

Human altruism is a highly complex phenomenon, and disagreement prevails regarding its innateness. Undeniably, cultural norms have a tremendous influence. Reciprocity leverages the altruism of strong reciprocators in norm enforcement (e.g., Fehr and Fischbacher 2003, 2004), and religion is found strongly related to morality. Many altruistic acts can be seen as conspicuous on a personal level, with a low cost but an often even incomparably lower impact, satisfying the personal desire to feel good about oneself: a 'warm-glow' effect related to self-enhancement (Andreoni 1990; Crumpler and Grossman 2008). Religiosity and self-enhancement thus linked to altruism are themselves evolutionarily complex, partly puzzling phenomena, discussed in the next two sections.

[2]Warneken and Tomasello (2009) propose experimental evidence showing innate altruism in young humans.

12.3 Religion

12.3.1 Omnipresent Spirituality

Spirituality accompanied the modern human species throughout its history (e.g., Frazer 1922). Our banding ancestors saw the elements of nature as animated and used past observations to infer rules how to win them over, hopping to elicit benign weather and to avoid calamities (for a detailed account, see Wright 2010). Not dissimilar to pigeons in Skinner's famous boxes (Skinner 1948), they drew links between their own behavior and observations that appear rather random, and importantly, their animistic beliefs were essentially unmoral (Wright 2010; Tylor 1958).[3] In larger tribes and chiefdoms, spirits or Gods with moral concerns of increasing complexity emerged (Wright 2010; Stark 2001). How can such a development be rationalized within an evolutionary framework?

12.3.1.1 Theory of Asymmetric Payoffs: A False Positive

Measured against the criterion of ESS, early animistic beliefs seem puzzling. Their unmoral character is seen as a natural consequence of the low importance of supernatural norm enforcement in small band living, where intimate contact between all group members sufficed to limit the scope of hideous free-riding on public goods (Wright 2010). Yet, it makes the significant resources devoted to please the spirits even more puzzling, as the practices have no obvious fitness advantage for such groups or their individuals. The standard explanation refers to the asymmetry between fitness costs of statistical errors of type I (believing a falsehood) and type II (rejecting a truth) (Shermer 1998, 2008, 2009; Foster and Kokko 2009; Beck and Forstmeier 2007), and the basic idea is easily understood with an example: You hear a rustling in the grass that may be from wind or from a predator. In the majority of cases, a harmless gust is the cause. But uselessly running away has little cost, while staying when it was the predator can be lethal. In this case, the theory maintains, it is evolutionarily favorable to *interpret* the rustling as predator's deed; after all, with asymmetric enough payoffs in the sense of an overly high risk from ignoring the sign, we definitely fare better running away! It is claimed that the ubiquity of such asymmetric payoff situations means that superstition as the false attribution of causality (and intent) is '*an inevitable feature of adaptive behavior in all organisms, including ourselves*' (Foster and Kokko 2009), that is, necessarily a long-run ESS, and that the superstitious tendencies would today still be beneficial even for modern humans, rather than just an inherited trait (Beck and Forstmeier 2007). This is perplexing. A modern educated human could obviously enjoy a material benefit from behaving according to adequate statistical inference rather than to live

[3]Despite occasional claims to the contrary, evidence seems clear on that point; cf. Stark (2001) for an overview.

according to superstitious rules. The simple asymmetric payoff-theory ignores the human capability of attributing probabilities to relations (e.g., Vyse 2013, p. 117), going beyond dichotomous linking. Humans act contingent on both, *probabilities* and *payoffs* of outcomes. A great meal awaiting on the other side of the river, we are not limited to reflecting 'there is an alligator in the river and he will kill me' or there 'is no alligator and I can just swim through.' Instead, we will inevitably ponder on the *likelihood* of there being an alligator and *weigh* it against our appetite. Ancestors, and their modern counterparts alike, are prone to cross the river when starving and the river normally found to be rather safe, but will abstain if the region and its waters are populated by dangerous animals and alternative sources of food abound. This puts into perspective simple explanations of patternicity, based on which superstition and early spirituality are maintained as obvious, evolutionarily (long-term) stable strategies. Instead, we propose that superstition in humans is more directly a natural characteristic of a mind whose higher cognitive capabilities were strongly shaped by interaction with indeed 'animated' peers and selected throughout time according to its ability to capture and anticipate what other minds think and do to reach their aims. In this sense, the tendency to find intent in natural happenings may indeed be primarily an artifact of these mental core functions, rather than *itself* a directly selected trait. This seems plausible considering the only recent emergence of higher cognitive capacities of the mind as suggested by the final major step of brain size some 200,000–100,000 years ago (e.g., Donald 1991). The few thousand generations at disposal for selection to improve the structure of the final brain were arguably scant time for genetic fine-tuning. In this case, a rather unrefined, general mental tendency of seeking agency in the world, stemming from a rather *ad hoc* adaptation to the environment in which it evolved, needs not surprise.

12.3.2 Big City Life

The evolutionary challenge for modern moralistic religion typical of evolved human societies such as chiefdoms, kingdoms, and states is distinct from that of unmoral spirituality. The (evolutionarily) costly rites, sacrifices, and taboos (e.g., Atran and Henrich 2010) of moral religion may be justified by implied fitness benefits. The larger groups, characterized by more ephemeral social relationships, offer increased scope for free-riding, jeopardizing the production of public goods crucial for group fitness. A supernatural surveyor and punisher is thus a welcome aid, plausibly decisive for the group's stability and success (Wright 2010; Norenzayan and Shariff 2008), and many studies suggest an empirical link between morality and religiosity (Saroglou 2012, 2013; Atran and Henrich 2010). Some maintain religious *practices* (rather than the underlying *beliefs*) could be major explanatory factors for morality (Bloom 2012; Galen 2012). Stark (2001) considers different types of religions and finds religiosity sustains the moral order only if it is centered on active and morally concerned Gods. This fits well the above evolutionary account with moralistic

elements of religions emerging only when larger group sizes imply an important benefit from heavenly surveillance.

Section 12.2 suggested that innate altruism emerging from group selection would unlikely be strong. Yet, religions are regularly associated with substantial sacrifices by individuals. Boyd and Richerson (2002a) and Henrich (2004) explain that challenges typically associated with genetic group selection need not apply to cultural group selection (see also Fehr and Fischbacher 2003; Atran and Henrich 2010). The need for genetic replacement is dropped; cultural traits, aka memes (Dawkins 1989), can spread by simple imitation. A rational tendency to mimic successful groups could mean that beneficial memes spread and conformist traits stabilize cultural groups when integrating migrants. From the perspective of the cultural *group,* this seemingly explains the evolutionary success of religions.

12.3.3 Uncostly and Costly Signaling

Even cultural evolution must, however, ultimately be understood at the level of the *individual*. Proposed intrapersonal explanations of belief include that religiosity increases fitness by offering relief from terror of death (Vail et al. 2010) and by preventing costly real punishment by others (Johnson and Bering 2006). Such largely intrapersonal explanations leave important questions open. An individual with a psychology evolved so as to cope with fear from death directly fared better than one relying on personally costly religious traits. Johnson and Bering's implicit assumption of an underestimation of the costs of prospective punishment would require explanation in the first place.

This begs the question of what facilitates the mind's acceptance of a belief with an individually costly[4] moral code, supported only by oral delivery (and potentially partly by patternicity). Henrich (2009) proposes that costly, 'conspicuous,' rites act as *credibility enhancing displays* (CRED), belief-enforcing signals to observers.[5] This idea is closely related to the economic concept of costly signaling as a rational means to separate truthful private information from cheap talk (e.g., Akerlof 1970; Farrell and Rabin 1996). However, the following argues that Henrich's concept of CRED (i) does not directly explain why specifically *moralistic* religious beliefs emerged instead of more arbitrary beliefs, and (ii) it ignores that the display of the rites by individual community members does not present any truly costly display; for an unbeliever, it can instead easily be seen as individually rational to respect the customs in order to prevent the exclusion from the overall successful club. Yet, considering the functioning of the community itself as the only unbiased testimony

[4]The group's general adherence to the belief may be beneficial to the individual, but it does not make personal adherence to the *belief* beneficial to the individual.

[5]Empirical evidence suggests costly rites increase the longevity of religious communes (Sosis 2000).

of the belief by individuals, we find indeed a truly costly, but indirect signal exhibited by the general member of the group, suggesting that specifically for *moralistic* religions, costly signaling can enhance the evolutionary stability.

An individual within a moralistic religious group would strictly have to judge other individuals' attendance to the rites as non-costly: They have direct costs, but exclusion plausibly following non-attendance could itself have higher real material costs, so that the relative costs of attending could naturally be seen as below zero. Only considering the religious belief as the ultimate source on which enforcement of the morality relies, and this morality being essential for the functioning of the community, an observer finds indeed a directly observable signal of costly commitment: If members generally were unbelieving, the community would be malfunctioning, so that the sheer functioning of the religious society itself constitutes the required signal. This mechanism does not generally support belief systems involving costly rites as membership criteria and, potentially, supernatural beings, without moralistic code that specifically relies on the belief itself (rather than on largely rational reciprocity and punishment), for the (relatively) vigilant learner would not find evidence for a truly costly signal confirming the belief.

In a straightforward framework, it could thus indeed suffice that individuals accept widespread, really (individually) costly action as a truthful signal, in order for learners to specifically accept moralistic religions, explaining the stability of such belief systems. Moralistic religious belief systems, which emerged in genetically very recent times, are thus *exceptional* self-enforcing equilibria in the way they may be largely explicable based on attendance to costly signaling of their models, even absent a fully rational foundation.

Are such systems easily exploited? The clue is that the believer has no incentive to exploit the community—supernatural wrath incurred by defection can be prohibitively large. Nevertheless, the systems seem exploitable, and indeed, the very formers and transformers of belief systems seem to have shaped religions in ways supportive of their own worldly aims throughout history (Wright 2010). Self-enhancement (Sect. 12.4) can partly explain such moves, even absent a conscious abuse of power by such leaders: Their convictions may themselves have been unconsciously biased (Wright 2010).

Self-enhancement offers a further explanation for the ease of adoption of many religious narratives: Promised heavenly justice seems often a potent relief from life circumstances and hence presents a naturally attractive option for a mind predisposed to believe what pleases.

12.4 Self-Enhancement

Mechanisms proposed as evolutionary explanations for self-enhancement act on an intrapersonal or an interpersonal level. Intrapersonal explanations, proposed notably by psychologists, identify benefits the self derives from misleading itself (Greenwald 1980; Alloy and Abramson 1979; Lewinsohn et al. 1980; Sedikides

and Skowronski 2000; Krebs and Denton 1997). For example, an overly positive self-perception is thought to prevent depression, which naturally would suggest a fitness benefit of biased perception by means of avoiding unhealthily low life activity and other effects characterizing depression. Proposed interpersonal benefits include notably reduction of the cognitive cost and increased credibility when deceiving others (von Hippel and Trivers 2011; Myers and Ridl 1979; Alexander 1987; Trivers 1985), high self-esteem increasing esteem by others (Sedikides and Skowronski 2000), as well as the beneficial treatment of happy persons in social interactions (von Hippel and Trivers 2011).

12.4.1 Perceptions and Their Translation into Action

The behavior of a self-aware individual can be described with two mapping functions, one translating real-world observations into perceptions of the self and the world, and a decision rule leading to actions based on perceptions. The intrapersonal explanations for self-enhancement have been criticized as evolutionarily inconsistent because it is more efficient to optimize decision rules given an accurate perception rather than to distort perceptions which must induce overhead costs (e.g., Pinker 2011). This critique can be extended to interpersonal explanations of self-enhancement. Similarly to our own decisions, our personal perceptions can affect others only via our personal, conscious, and unconscious decision rules. As biased perception has high extra costs (Leary 2007; Funder 2011), this means an ESS would be to adapt directly the decision rules and to leave own perceptions unbiased, in line with Pinker's critique (2011). Fully thinking through the Pinker's critique hence implies that proposed explanations of self-enhancement, in so far as they consider the trait as an ESS, seem not entirely satisfactory. Yet, most proposed explanations of self-enhancement, explicitly or implicitly, seem to advance the trait as an optimal long-run strategy.

12.4.2 Moderately Skilled Liars

There is an ongoing debate about the degree to which humans are good liars or, instead, good lie detectors. Evidence seems mixed; depending on circumstances, humans appear to get away with lies rather successfully or, instead, be caught rather easily (for an overview see von Hippel and Trivers 2011; Vrij 2011). It may thus be fair to call humans rather moderately skilled in lying. This view will ultimately square well with the reflections that follow.

12.4.3 Ad Hoc Evolution

The following suggests that delivering a genuinely 'optimal' self-perception (and decision rule) in terms of evolutionary fitness would be by far too demanding for the framework within which the genetic foundation of the human mind evolved. Instead, the evolutionary time since the beginning of the explosive development of conscious, abstract intelligence of humans or hominids, was likely far too short to support largely optimal adaptation. Strong evolutionary pressure raised support for the spread of features in line with (but not identical to) the evolutionary optimum, but due to fundamental genetic laws was unable to increase directly the pool of new genetic arrangements to draw from. This meant that even features only roughly in the direction of the evolutionary fitness pressure, and with potentially substantial overhead costs, were selected; *ad hoc* evolution rather than ideal evolution. On the order of 200,000–100,000 years may have been at disposal since the emergence of our species Homo sapiens, with enlarged brains (e.g., IHO 2008) presumably fundamental for the distinctly human cognitive capacities (e.g., Donald 1991). This is very brief on a timescale for genetic human evolution, especially considering the fine-tuning of novel machinery such as the enhanced human cognitive apparatus. In times of moderate evolutionary pressure, one might simply expect relatively moderate genetic changes, but the explosive change in cognitive capabilities, accompanied by demographic and social changes, meant an extraordinary evolutionary pressure (e.g., Haidt 2012) that likely was able to help spreading a range of beneficial though suboptimal adaptations, and the following explains how self-enhancing biases are plausibly among them. Conceptually, this means that while thinking in terms of ESS may be adequate for evolution in gradually changing environments (or when plenty of time follows a change), the rapid pace of changes in the case of recently emerged human cognitive capacities warrants a broader perspective, an 'everything that tends into the right direction goes' view. ESS, i.e., traits that would be long-run optimal, will in this case, rather than themselves emerge, figure as guidance for *ideal* features against which realistically selected traits are to be benchmarked.

Having evolved in intimate bands of a few handfuls of people, early hominids and ancestors interacted with their peers day in, day out. Individuals knew their peers' qualities almost as well as their own, having stood by or been close to most major happenings of everyone's life. In such an environment, little scope is to be expected for deceiving others about most of one's personal qualities. With others witnessing one's qualities directly, the brunt of the social costs of personal misperception would be borne individually, as others would not be fooled. The cost of misperception and the likely following misjudgment and suboptimal decisions means that self-enhancing traits would have a net fitness cost and would be evolutionarily unstable. But after several million years of life in bands and smaller tribes, what implication does emergence of living in larger groups, which over time lead to tribes with higher headcounts, then to chiefdoms of often hundreds, potentially up to many thousands of people, imply, regarding expressed, and felt,

self-esteem? A state with biased self-esteem and misled interactants appears unstable in two ways. First, one would ideally only *express* a better picture about oneself, rather than believe it personally. Second, others should develop a healthy skepticism that dismisses cheap talk about good characteristics lacking evidence beyond personal communication. In this sense, a substantial level of self-enhancement as an ESS seems unlikely. However, as we have seen, there was scant time for evolution to allow optimal adaptation. In a dynamic system, starting from a mind tuned to act honestly given its perceptions, the new situation with enlarged groups would clearly have supported an ad hoc adaptation by the individual to personally perceive one's own skills as better than they are: This addressed the shortcoming of the individuals' historically limited ability to actively deceive and allowed exploiting an (equally historically grounded) naivety of the audience. In parallel, the same evolutionary pressure that selected such misperception could have selected traits enhancing the ability to *consciously* deceive. Furthermore, conscious and unconscious deception, in turn, implied a pressure for the audience to develop more vigilance against deception. Time was scarce so none of the three adaptations were perfect; the evolutionary support for each of them remained intact.

In addition to the genetically very short time that was available to fine-tune the recently enhanced human cognitive apparatus, the high overhead cost of self-enhancement supports this theory of self-enhancement as ad hoc adaptation rather than an optimal trait (Leary 2007; Funder 2011). Moreover, the theory fits particularly well to experimental evidence of self-enhancement (Paulhus 1998): Self-enhancing individuals appear disadvantaged in intense (long) social interactions, but advantaged in more casual interactions, in line with the idea of self-enhancing as an adaptation to exploit an environment of more casual or anonymous interactions in enlarged social groups.

12.5 Bringing It All Together

The previous sections considered evolutionary explanations for altruistic, spiritual, and self-enhancing traits relatively independently. The three phenomena are strongly interlinked, occasionally even difficult to dissociate—famously, religiosity appears often so tightly linked to altruistic behavior that the former is sometimes considered indispensable for the latter, and the evidence for an increased popularity of narratives of heavenly redemption in dire times (Atran and Henrich 2010) fits the idea of an opportunistic mind whose subconsciousness makes believe whatever pleases. How do the traits, and the explanations for them, relate to each other?

The presented discussion about altruism maintains the possibility of differences in survival rates of groups based on their fitness to have supported a low level of positive other-regarding preferences. However, self-enhancement is proposed to have emerged as an unconscious strategy for the deception of peers, potentially at a substantial social cost. This paradox, of the simultaneous selection of an

individually costly trait for its benefit to the group and a trait costly to the group for its benefit to the individual, is readily solved. Whether the selective pressure at the group level suffices to outweigh individual fitness costs depends on trait-specific, quantitative particularities. Essential public goods, whose provision would have been difficult to maintain in groups of purely selfish reciprocators that psychopathically tried to exploit others wherever possible, may have implied a very strong group advantage of some altruism. Self-enhancement, in contrast, though not entirely petty, may not have had nearly as dramatic consequences for group survival; rather than primarily reducing the social surplus, it could have mainly led to a shift of resources from some to those most 'skilled' in self-enhancement. According to Sedikes and Skowronski (1997), self-enhancement could even have fostered altruistic behavior, as warm-glow effects related to the desire to see ourselves as nice increase our propensity to help. Furthermore, if self-enhancement selectively allows the self to justify certain egoistic positions, it may even have increased the cost-effectiveness of the altruistic provision of key public goods: Rather than behaving altruistically in all aspects of communal life, a properly shaped bias could lead the individual to justify contributing only in areas key for group strength rather than for peers personally.

A moralistic element has become a widespread component of spirituality only recently, along with the enlargement of human tribes, chiefdoms, or states, and its prior absence has been explained by the redundancy of a heavenly surveillance camera (Gervais and Norenzayan 2012) when members of small-sized bands are in intimate contact. Genes evolving more slowly, the independent component of innate altruism maintained above has presumably emerged much earlier. How can moral religious rules have been redundant in an earlier past, while genuine altruism was central for social life? On the one hand, originally smaller group sizes meant that an adequate level of public goods provision could be sustained even with a relatively limited degree of truly good intention, so that the limited level of sustainable genuine altruism (aided not least by reciprocity) could have sufficed. On the other hand, the increase in group size in recent times meant that public goods gained in importance, and their provision in the more anonymous societies became increasingly difficult, so that the altruism inherited from past life in smaller groups became relatively helpless.[6] Moreover, supernatural supervision and punishment or reward specifically addressed the problem of anonymity accruing as group size increased. The problem of anonymity may not have been so much at the centre in the small bands, where instead other issues (e.g., containment of tendencies by strong individuals to overtly exploit others) could have been dominant.

Self-enhancement seems naturally conducive to the stability of religion: A mind prone to believe what pleases it seems attracted to comforting promises that things eventually turn out well (Gorelik and Shackelford 2011; Krebs and Denton 1997).

[6]While the evolutionary sustainability of altruism beyond kin in *small* groups is a matter of ongoing debate (cf. Sect. 12.1), for large groups of hundreds or more members, it is widely considered implausible.

Assuming the benefit from comforting belief to increase in dire times, this suggests a positive relationship between material deprivation and the level of religiosity, a relationship that is confirmed in experiments and surveys (Atran and Henrich 2010). In a dynamic world with intermittent environmental pressure affecting the survival of group, this can imply a tandem between self-enhancement and religiosity leading to a flexible level of culturally underpinned altruism. The bias means that cultural altruism is strongest in dire periods, when it is most useful for the group's survival. Habermacher (2014) shows with numerical simulations in a multi-level selection framework with continuous degrees of self-enhancement, religiosity, and genuine altruism, how such a tandem of self-enhancement and religiosity could in the long run crowd out more stable genuine altruism despite overhead costs of bias and religious cult.

A further relationship plausibly links the evolutionary origin of altruism and self-enhancement. Self-enhancement leads to an overly positive view not only about the *self* but equally about groups one identifies with, and it can explain a negative view of out-group members (Krebs and Denton 1997). In this regard, self-enhancement may have played a role in helping to direct or 'contain' altruism toward in-group members, lowering the requirement to strictly restrict the underlying altruistic compassion itself to group members. It would allow for deep-rooted human altruism itself to have been 'imprecisely' designed as rather encompassing from the outset—we could again talk of an *ad hoc* adaptation. In the past, a separate, negative bias against foreign groups grounded in self-enhancement was able to contain within the group the effect of the otherwise more general altruism, so that a strict group benefit was obtained. Ultimately, this would imply that the today observed extension of altruism beyond clearly defined kin or group circles is, in some sense, more natural than otherwise implied by the big mistake hypothesis. In this case, the main challenge boils down to explaining how modern culture leads to the containment of negative out-group perceptions that are based on deep-rooted self-enhancement. Plausibly, the increased availability of information about the people of the world largely explains this; it may simply have become difficult to decry all others as inherently repugnant if the facts show that they are, after all, just quite like us.

12.6 Conclusions

Altruism, religiosity, and self-enhancement are complex and intertwined facets of human societies, and a multi-disciplinary literature studies their evolutionary origins. Biological altruism is often seen as mainly relevant for kin relationships, for which it is readily explained by kin selection, with remaining phenomena explained as strategic cooperation, attributable to reciprocity with repeated interaction. This ignores the deeper reality of human, genuine altruism, extending—arguably on a low level—to humanity rather generally. Such an extent seems an unlikely expansion for other-regarding preferences ultimately based purely on direct

kin-level selection. Selection on multiple levels with non-negligible force acting at the level of groups could offer an explanation, and rather than devaluating group-selection forces because of their quantitatively limited impact, this weakness fits well with the apparent low *level* of the general component of human altruism.

Literature has proposed explanations for unmoral spirituality, moral religiosity, and self-enhancement, often as ESS, implying them to be optimal adaptations to a given environment either in the past or today. Critics have pointed out caveats of such explanations; ultimately, prevalent theories may have to be revised.

Here, it is proposed that rather than as ESS, the traits should be viewed as what might best be called ad hoc adaptations. Time available since the development of the higher human cognitive capabilities, roughly the past 200,000–100,000 years, was scant for an optimal fine-tuning of all brain functions. It seems perfectly conceivable that in the longer run, our brains would have improved in distinguishing genuine causal relationships from random events, limiting superstition. It seems equally plausible that in the longer run, self-enhancement, widely considered an aid to avoid cognitive costs when deceiving others, would largely be crowded out by an increased ability to consciously deceive. The reason for the presence of the apparently imperfect traits is that this evolutionary 'long run' has never materialized, because on an evolutionary timescale, the modern brain is young; the limited time available for adaptation to have been 'complete' offers a natural explanation for the evolutionarily interesting state of the human mind.

References

Aiello LC, Dunbar RIM (1993) Neocortex size, group-size, and the evolution of language. Curr Anthropol 34:184–193

Akerlof GA (1970) The market for "lemons": quality uncertainty and the market mechanism. Q J Econ 84:488–500

Alexander RD (1987) The biology of moral systems. Aldine de Gruyter, New York

Alicke MD, Sedikides C (2011) Handbook of self-enhancement and self-protection. Guilford, New York

Alloy LB, Abramson LY (1979) Judgement of contingency in depressed and nondepressed students: sadder but wiser? J Exp Psychol Gen 108:441–485

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm glow giving. Econ J 100:464–477

Atran S, Henrich J (2010) The evolution of religion: How cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. J Theor Biol 5:18–30

Babcock L, Loewenstein SI et al (1995) Biased judgments of fairness in bargaining. Am Econ Rev 85:1337–1343

Beck J, Forstmeier W (2007) Superstition and belief as inevitable by-products of an adaptive learning strategy. Hum Nat 18:35–46

Bloom P (2012) Religion, morality, evolution. Ann Rev Psychol 63:179–199

Bowles S (2006) Group competition, reproductive level, and the evolution of human altruism. Science 314:1569–1572

Bowles S (2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? Science 324:1293–1298
Boyd R, Richerson PJ (2002a) Group beneficial norms can spread rapidly in a structured population. J Theor Biol 215:287–296
Boyd R, Richerson PJ (2002b) Solving the puzzle of human cooperation. In: Levinson S (ed) Evolution and culture. MIT, Cambridge MA
Bshary R, Bergmüller R (2008) Distinguishing four fundamental approaches to the evolution of helping. J Evol Biol 21:405–420
Codol J-P (1995) On the so-called 'superior conformity of the self' behavior: twenty experimental investigations. Eur J Soc Psychol 5:457–501
Crumpler H, Grossman PJ (2008) An experimental test of warm glow giving. J Pub Econ 92:1011–1021
Darwin Ch (1869) On the origins of species by means of natural selection, 5th edn. Murray, London
Darwin Ch (1871) The descent of man and evolution in relation to sex. Murray, London
Dawkins R (1989) The selfish gene, 2nd edn. Oxford University Press, Oxford
Donald M (1991) Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press, Cambridge MA
Dunbar RIM, Shultz S (2007) Evolution in the social brain. Science 317:1344–1347
Farrell J, Rabin M (1996) Cheap talk. J Econ Perspect 10:103–118
Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425:785–791
Fehr E, Fischbacher U (2004) Social norms and human cooperation. Trends Cogn Sci 8:185–190
Frazer JG (1922) The golden bough: a study in magic and religion. Macmillan, New York
Foster KR, Kokko H (2009) The evolution of superstitious and superstition-like behavior. P R Soc B 276:31–37
Frank SA (2012) Natural selection. III. Selection versus transmission and the levels of selection*. J Evol Biol 25:227–243
Funder D (2011) Directions and beliefs of self-presentational bias. Behav Brain Sci 34:23–24
Galen LW (2012) Does religious belief promote prosociality? a critical examination. Psychol Bull 138:876–906
Gervais WM, Norenzayan A (2012) A camera in the sky? thinking about god increases public self-awareness and socially desirable responding. J Exp Soc Psychol 48:298–302
Gorelik G, Shackelford TK (2011) Culture of deception. Behav Brain Sci 34:24–25
Grafen A (1984) Natural selection, kin selection and group selection. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 62–84
Greenwald AG (1980) The totalitarian ego: fabrication and revision of personal history. Am Psychol 35:603–618
Grove M (2013) Evolution of hominin social systems. In: eLS. John Wiley & Sons Ltd, Chirchester UK. http://onlinelibrary.wiley.com/book/10.1002/047001590X/homepage/ForAuthors.html
Habermacher F (2014) Altruism, religion and irrationality—emergence of cultural cooperation dominating innate altruism under intermittent evolutionary pressure. Paper presented at the 18th evolutionary biology meeting, Association pour l'Etude de l'Evolution Biologique, Marseille, 16–19 September 2014
Haidt J (2012) The righteous mind—why good people are divided by politics and religion. Pantheon, New York
Hamilton WD (1964) The genetical evolution of social behavior. J Theor Biol 7:1–16
Hamilton WD (1970) Selfish and spiteful behaviour in an evolutionary model. Nature 228:1218–1220
Heine SJ, Lehman DR et al (1999) Is there a universal need for positive self-regard? Psychol Rev 106:766–794
Henrich J (2004) Cultural group selection, coevolutionary processes and large-scale cooperation. J Econ Behav Organ 53:3–35

Henrich J (2009) The evolution of costly displays, cooperation and religion: credibility enhancing displays and their implications for cultural evolution. Evol Hum Behav 30:244–260
Von Hippel W, Trivers R (2011) The evolution and psychology of self-deception. Behav Brain Sci 34:1–16
Institute of Human Origins (IHO) (2008) Becoming human. School of Human Evolution and Social Change, Arizona State University. http://www.becominghuman.org. Accessed 18 Jan 2015
Johnson DP, Bering JM (2006) Hand of God, mind of man: punishment and cognition in the evolution of cooperation. Evol Psychol 4:219–233
Kirchgässner G (1992) Towards a theory of low-cost decisions. Eur J Polit Econ 8:305–320
Kirchgässner G (2008) Homo oeconomicus: the economic model of behaviour and its applications in economics and other social sciences. Springer, New York
Kirchgässner G (2010) On minimal morals. Eur J Polit Econ 26:330–339
Krebs JR, Davies NB (1993) An introduction to behavioural ecology, 3rd edn. Blackwell, Oxford
Krebs JR, Denton K (1997) Social illusions and self-deception: the evolution of biases in person perception. In: Simpson JA, Kenrick DT (eds) Evolutionary social psychology. Psychology Press, New York, pp 21–48
Lack D (1966) Population studies of birds. Clarendon, Oxford
Leary MR (2007) Motivational and emotional aspects of the self. Annu Rev Psychol 58:317–344
Levin BR, Kilmer WL (1974) Interdemic selection and the evolution of altruism: a computer simulation study. Evolution 28:527–545
Lewinsohn PM, Mischel W et al (1980) Social competence and depression: the role of illusory self-perceptions. J Abnorm Psychol 89:202–212
Markus HR, Kitayama S (1991) Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev 98:224–253
Myers DG, Ridl J (1979) Can we all be better than average? Psychol Today 13:89–98
Norenzayan A, Shariff AF (2008) The origin and evolution of religious prosociality. Science 322:58–62
Nowak MA (2006) Five rules for the evolution of cooperation. Science 314:1560–1563
Nowak MA, Sigmund (2007) How populations cohere: five rules for cooperation. In: May R, McLean AR (eds) Theoretical ecology: principles and applications. Saunders, Philadelphia PA, pp 7–16
Paulhus DL (1998) Interpersonal and intrapsychic adaptiveness of trait self-enhancement: a mixed blessing? J Personal Soc Psychol 74:1197–1208
Pinker S (2011) Representations and decision rules in the theory of self-deception. Behav Brain Sci 34:35–37
Okasha S (2013) Biological altruism. In: Zalta EN (ed) The Stanford encyclopedia of philosophy. http://plato.stanford.edu/archives/fall2013/entries/altruism-biological/. Accessed 30 Jan 2015
Saroglou V (2012) Is religion not prosocial at all? comment on Galen (2012). Psychol Bull 138:907–912
Saroglou V (2013) Religion, spirituality, and altruism. In: Pargament KI, Exline JJ et al (eds) The APA handbook of psychology, religion and spirituality. APA, Washington DC, pp 439–457
Sedikides C, Gregg AP (2008) Self-enhancement: food for thought. Perspect Psychol Sci 3: 102–116
Sedikides C, Skowronski JJ (1997) The symbolic self in evolutionary context. Pers Soc Psychol Rev 1:80–102
Sedikides C, Skowronski JJ (2000) On the evolutionary functions of the symbolic self: the emergence of self-evaluation motives. In: Tesser A, Felson RB et al (eds) Psychological perspectives on self and identity. APA, Washington DC, pp 91–117
Sedikides C, Gaertner L et al (2003) Pan-cultural self-enhancement. J Pers Soc Psychol 84:60–79
Sen AK (1977) Rational fools: a critique of the behavioral foundations of economic theory. Philos Public Aff 6:317–344

Shermer M (1998) Why people believe weird things: pseudoscience, superstition, and other confusions of our time. Freeman, New York
Shermer M (2008) Patternicity. Sci Am 299:48
Shermer M (2009) Agenticity. Sci Am 300:36
Skinner BF (1948) Superstition in the pigeon. J Exp Psychol 38:168–172
Smith JM (1964) Group selection and kin selection. Nature 201:1145–1147
Smith JM, Price GR (1973) The logic of animal conflict. Nature 246:15–18
Sober E, Wilson DS (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge MA
Spencer H (1864) The principles of biology. William and Norgate, London
Sosis R (2000) Religion and intragroup cooperation: preliminary results of a comparative analysis of utopian communities. Cross-Cult Res 34:70–87
Stark R (2001) Gods, rituals, and the moral order. J Sci Stud Relig 40:619–636
Traulsen A, Shoresh N et al (2008) Analytical results for individual and group selection of any intensity. B Math Biol 70:1410–1424
Trivers R (1985) Social evolution. Cummings, Menlo Park, CA
Tylor EB (1958) Religion in primitive culture. Harper and Brothers, New York
Vail KE, Rothschild ZK et al (2010) A terror management analysis of the psychological functions of religion. Pres Soc Psychol Rev 14:84–94
Vrij A (2011) Self-deception, lying, and the ability to deceive. Behav Brain Sci 34:40–41
Vyse SA (2013) Believing in magic: the psychology of superstition—Updated Edition. Oxford University Press, Oxford
Warneken F, Tomasello M (2009) The roots of human altruism. Brit J Psychol 100:455–471
Weibull W, Salomonsson M (2006) Natural selection and social preferences. J Theor Biol 239: 79–92
West SA, Griffin AS et al (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415–432
West SA, Griffin AS et al (2008) Social semantics: how useful has group selection been? J Evol Biol 21:374–385
West SA, Gardner A (2010) Altruism, spite, and greenbeards. Science 327:1341–1344
Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton
Wilson DS (1987) Altruism in Mendelian populations derived from sibling groups: the haystack model revisited. Evolution 41:1059–1070
Wilson DS (2012) Richard Dawkins, Edward O Wilson and the consensus of the many. This view of life 29
Wright R (2010) The evolution of God: the origins of our beliefs. Hachette, New York

Chapter 13
Investigating Trade-offs in Sexual Populations with Gene Flow

Zachary N. Ardern and Matthew R. Goddard

Abstract Understanding the processes underlying trade-offs between environments, where adaptation to one results in decreased fitness in another, is important in understanding evolutionary processes across a wide range of organisms. The molecular basis of this evolutionary phenomenon is a key question in biology generally. Unravelling the basis of trade-offs has application in understanding the maintenance of sexual reproduction in most eukaryotic lineages in spite of apparent costs. In this chapter, we discuss the evolutionary problem of sexual reproduction, and its relationship with trade-offs, working from August Weismann's suggestion that sex improves the efficiency of natural selection. We argue that microbial experimental evolution is an important way in which claims about trade-offs and sex can be tested and that these experiments need to be developed to better represent real world ecological and evolutionary problems. We review experiments, including from our laboratory, which bear on the question of the benefits of sex in complex environments. We also argue for the necessity of a genomic rather than merely genetic perspective on these questions.

13.1 Introducing Complexity into Experimental Evolution

In recent years, evolutionary biology has been equipped with new molecular resources to aid in its explanations of the natural world and corresponding opportunities to revise its theoretical foundations in the light of the molecular data. Two important challenges are explaining the genomic basis of adaptation to complex environments and the maintenance of sexual reproduction in spite of apparent costs. Progress on each of these will benefit from the flood of molecular data which is becoming available. While the large-scale patterns in life history

Z.N. Ardern (✉) · M.R. Goddard
School of Biological Sciences, University of Auckland, Auckland, New Zealand
e-mail: zard001@aucklanduni.ac.nz

M.R. Goddard
The School of Life Sciences, Lincoln University, Lincoln, UK

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_13

resulting from evolution via natural selection have been well established since the publication of Darwin and Wallace's seminal article, the molecular details of the processes undergirding the phenotypic changes inferred remain to be determined (Ball 2013). The emerging field of microbial experimental evolution (Adams and Rosenzweig 2014) is one source of new empirical insights into fundamental evolutionary questions. In this chapter, we discuss findings from experimental evolution which relate to genotype *x* environment interactions and the maintenance of sexual reproduction over the course of generations; two closely related concepts in the evolutionary literature. An ecological trade-off occurs when a population increases in adeptness in one environment but decreases in another. This concept is an important interface between genes and environments and has exciting applications in understanding the evolution of antibiotic resistance, cancer tumour cell genetics and the adaptation of populations to changing environments.

The majority of rigorous experiments in microbial evolution have involved asexual haploid organisms adapting to a single simple environment, often nutrient-limited or with another stress added such as a change in temperature. These experiments have become more prominent in recent years, but have been conducted ever since the Rev. William Dallinger set-up an incubator for microbes in the time of Darwin (Kassen 2014). Dallinger over a number of years grew populations of flagellated monads in an incubator with a slowly increasing temperature, changing from 16 °C to 70 °C over the course of the experiment. Starting populations could not grow at the higher temperature, and the ending populations could not grow at the starting temperature of 16 °C (Xu 2012)—a shift of ecological niche had occurred. There have been few tests involving key factors in more complex real world evolutionary scenarios such as divergent selection, variegated environments, diploidy or sexual reproduction, to name just a few variables which are believed to be important in evolutionary dynamics.

We begin this chapter by discussing the evolutionary challenge of explaining the maintenance of sexual reproduction in the face of apparent costs, in the light of a discussion of adaptation. We then address the molecular basis of genotype *x* environment interactions in complex environments, crucial for understanding the opportunities for natural selection to act following the generation of phenotypic variance through recombination between genomes. In Sects. 13.3 and 13.4, we discuss genetic factors contributing to trade-offs between environmental niches, and experimental evidence for different models of trade-off, including data from our research group. In the final section, we suggest consequences for the evolutionary problem of the maintenance of sexual reproduction and for ecological theory more generally.

13.2 Sex and Adaptation

When assessing evolutionary processes underlying the origin or maintenance of putatively adaptive features such as sexual reproduction, we must ask how adaptation generally proceeds. One well-studied account is Ronald Fisher's geometric

model, which represents the fitness landscape as a high-dimensional phenotypic space around a point of optimal fitness; these dimensions model the different phenotypic components of fitness. Mutations are represented as vectors in this space, such that a shift closer to the optimum is beneficial and a shift away is deleterious. In this model, a fitness landscape is locally smooth around the optimum (Orr 2006). As developed by Fisher, the model implies that the mutations fixed over the course of an adaptive walk will follow an exponential distribution when mutations of very small effect are disregarded (Orr 2006). As adaptation proceeds, the mean phenotypic effect of each subsequently fixed beneficial mutation decreases by a nearly constant proportion (Orr 2002). As a consequence, when taking the general approach of Fisher's model, mechanisms which assist in sifting through and combining mutations of small effect become important in the evolutionary process; adaptation does not require fixing mutations with large beneficial effects.

A key challenge for evolutionary theories is explaining the near-ubiquity of sexual reproduction amongst eukaryotic lineages (Burt 2000). Following Goddard (2007) we define sex as involving recombination, random assortment and syngamy. These terms refer to meiotic recombination (occurring between non-sister chromatids), random assortment (segregation) of chromosomes and the event of fertilisation, i.e. the fusing of gametes. Sexual reproduction is commonly associated with a 'twofold cost' when compared to asexually reproducing populations. While meiosis is a costly process, the cost has various components and is unlikely to be precisely twofold (Goddard 2007). Two aspects of the cost of sex are firstly those associated with anisogamy, i.e. when male and female gametes are different and one of each kind are required for reproduction, and secondly, the costs involved in undergoing the more complex process of meiosis rather than merely mitosis. The most prominent strands of theory concerning the evolutionary maintenance of sexual reproduction can be traced back to August Weismann's suggestion towards the end of the nineteenth century that the adaptive value of sex lies in the introduction of variation, which thereby increases the effectiveness of natural selection (Poulton et al. 1889). A number of theorists have developed a line of research in this vein (Bell 1982; Kondrashov 1988; Burt 2000; Hoekstra 2005). The extent to which these accounts have been successful is contentious (Otto 2009). As an example, one critic claims that the advantage has not been empirically demonstrated 'in real-case studies', 'despite huge efforts by scientists for more than 40 years on many organisms and therefore does not appear convincing' (Gouyon 2015). However, experimental evolution has directly shown a competitive advantage in sexual populations (Goddard et al. 2005). The nature of this advantage has also been explored. Earlier work from the Goddard laboratory suggests that sex may provide a benefit in heterogeneous environments by both increasing the speed of adaptation through combining beneficial mutations together but also by removing accumulated deleterious mutations (Gray and Goddard 2012b). The nature of 'variation' needs to be explicated in further depth. It is likely that sex actually reduces variation at the genomic and chromosomal levels while increasing it at the level of individual genes (Gorelick and Heng 2011). As Sarah Otto (2009) notes, increasing variation through sex can reduce a population's fitness, as a lot of

variation is not beneficial. Rearrangement of alleles through segregation can break associations favoured by selection, at a particular locus, while rearrangement through recombination can break associations between loci.

The Weismann hypothesis presumes that sex has an evolutionary advantage for individuals in some way connected to the action of natural selection on beneficial and deleterious mutations, i.e. that selection acts within rather than between species. In this chapter, we focus on this thesis, but two other possibilities should be noted. Firstly, processes of selection and neutral evolution operating at the species level can be considered. Selection between species is a controversial concept based on the observation that groups of related species can originate and become extinct at different rates, a process on which selection may act (Lieberman and Vrba 2005). It is plausibly claimed that this phenomenon cannot explain the origin of sex but may explain its maintenance for species exhibiting anisogamy, i.e. this evident cost may be offset by its effects on rates of speciation (Gouyon et al. 2015). Alternatively, but also at the species level, neutral models may be able to explain the loss of asexual species observed in the fossil record (Schwander and Crespi 2009). As there are also some asexual lineages which survive over long periods of time (Judson and Normark 1996; Butlin 2002), it may be that the supposed detriments associated with asexual reproduction have been overstated, but nonetheless, the widespread extent of the complex features of sexual reproduction seems in need of some explanation. A different approach, where sex is beneficial for reasons different than those commonly posited, is offered by analysis of DNA repair mechanisms, which have been proposed as providing the original adaptive advantage which gave rise to sex and facilitates its maintenance as an adaptation. This is argued with reference to the functions of the enzymatic machinery of meiosis (Bernstein and Bernstein 2013).

Mueller's ratchet occurs in asexual populations when they accumulate mutations through neutral processes. It has been posited as a major cost to asexuality (Felsenstein 1974). There are a number of related concepts which contribute to a limitation on evolution from linkage between genomic loci. Clonal interference occurs when separate lineages have different beneficial mutations; this can result in some beneficial variants being lost from a population when the lineage goes to extinction. Genetic hitchhiking is when loci increase in frequency in a population due to the action of selection on linked loci. When an adaptive variant sweeps through a population it may drag linked loci, likely to be neutral or slightly deleterious, with it. Finally, purifying selection against deleterious mutations can act to eliminate linked beneficial variants, termed 'background selection'. These processes are reviewed by Hartfield and Keightley (2012), and they can all reflect the association of deleterious alleles with linked beneficial alleles, which is known as Hill-Robertson interference (Lively and Moran 2014). Sexual reproduction, by the process of recombination, can act to reduce Hill-Robertson interference.

As noted, the theories reviewed are in need of empirical testing. A system appropriate for testing differences in adaptation between sexual and asexual populations was developed by Goddard et al. (2005). *Saccharomyces cerevisiae* is a unicellular yeast that asexually divides when nutrients are plentiful, but undergoes

meiosis when starved. The haploid recombined spores that are produced germinate when nutrients are encountered again, and may mate with spores of the opposite mating type. Previous work has compared the evolutionary trajectories of yeast population that were variously periodically starved or never starved creating sexual and asexual lineages. While these experiments were well conducted, it has been demonstrated that the process of starvation used to induce sporulation is mutagenic (Marini et al. 1999), and thus, it cannot be deciphered whether any difference in evolutionary trajectories is due to the process of sex or the mutagenic effect of the sporulation treatment. To circumvent this, Goddard et al. created diploid populations of *S. cerevisiae* which do not undergo meiosis (but still may sporulate) by removing two genes involved in recombination (Goddard et al. 2005). Sexual populations with the wild-type genes are otherwise isogenic, and there is no significant fitness effect resulting from the knockout (Goddard et al. 2005) meaning that the two strains are generally comparable. In line with Weismann's suggestion, the initial experiment simply compared the rate of adaptation between these sexual and asexual populations and demonstrated that sexual populations adapted more rapidly and to a greater extent than asexual populations when under directional selection. Subsequent experiments pursued by Jeremy Gray have used this system to test the effects of sex and gene flow on adaptation to more complex heterogeneous nutrient-limited environments, during 300 generations of evolution in chemostats. The two environments used were termed 'hot C' and 'osmotic N', on account of being, respectively, raised in temperature and limited in glucose and raised in salt concentration and limited in nitrogen. Results from this experimental system are discussed as part of the 'Experimental Evidence' below.

13.3 Heterogeneous Environments and Trade-offs

The expression and function of a genetic sequence always depends on environmental context, and genotypes often result in different functional effects in different environments with this phenomenon termed phenotypic plasticity. When the phenotypic plasticity of an organism differs according to genotype, this phenomenon is known as 'genotype by environment interaction', also 'gene-x-environment interaction' (El Soda et al. 2014), and similar. Genotype by environment interactions can be measured in terms of either a particular phenotype, or of overall fitness, contributed to by multiple phenotypes. Refer to Fig. 13.1. According to Stearns (1990), drawing on the work of Bell (1982) and following on in the tradition of Weismann, the question of why sex is maintained is able to be reduced to questions about the origination and maintenance of genotype by environment interactions. Concerning various environmental factors which have been discussed in the literature on sex, such as parasites, disease and other environmental variation, Stearns notes that 'each is effective only insofar as it is involved in a genotype by environment interaction for fitness'. We are therefore drawn into considering the genetic basis for adaptation across multiple environments.

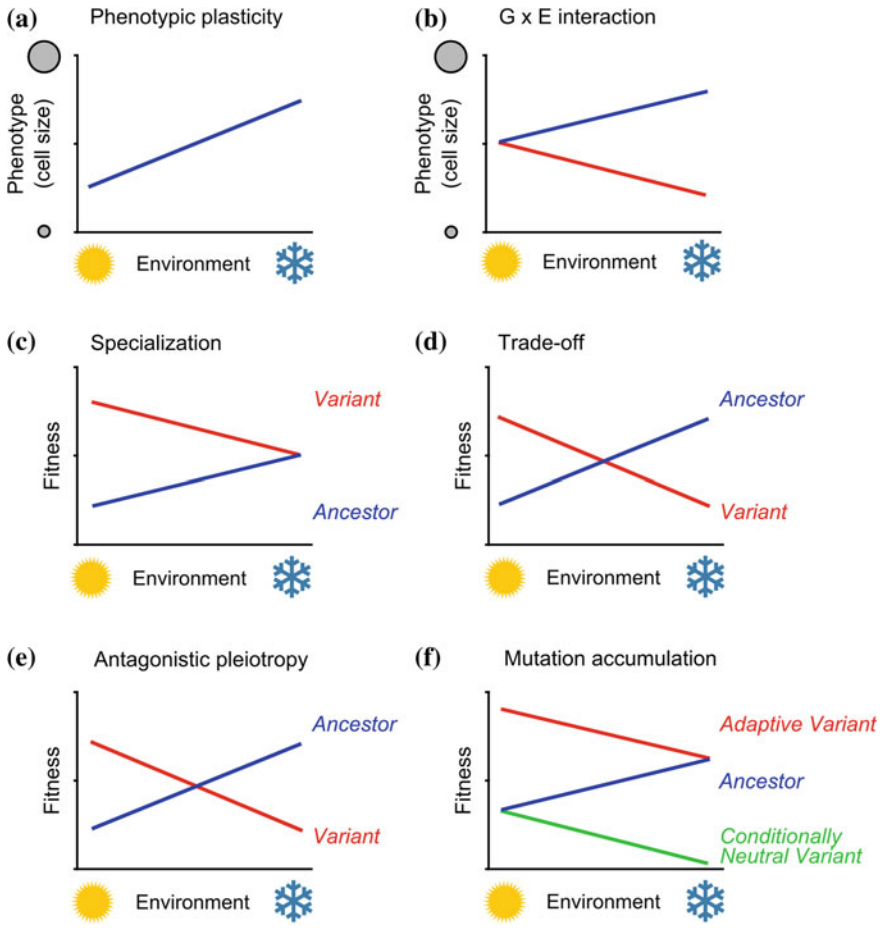


Fig. 13.1 Different forms of relationship between genotypes and environments. **a** Shows a situation where a genotype differs in phenotype across environments. **b** Shows two different genotypes which differ in their response to a change in environment—for instance from hot to cold. The remaining graphs show different kinds of genotype *x* environment interaction for evolutionary. **c** Specialisation occurs when a variant genotype has increased fitness in the (hot) environment of selection—in this example, no change has occurred in the alternate (cold) environment. **d** A trade-off—adaptation of the variant to the hot environment results in less fitness than the ancestral genotype has in the alternate environment. **e** When a trade-off is due to antagonistic pleiotropy, the mutations responsible are antagonistic in effect across the two environments. **f** When a trade-off is due to mutation accumulation, the mutations responsible for adaptation are distinct from those which conditionally neutral mutations which result in the cost in the alternate environment. Modified from Grishkevich and Yanai (2013)—Fig. 13.1

Ecological specialisation is one consequence of different genotypes differing in phenotypic responses to environmental variation. A demonstration of this phenomenon was likely offered by William Dallinger's experiment mentioned earlier

(Kassen 2014), although possible confounding factors such as contamination cannot be ruled out. Dependent on genotype by environment interactions then, adaptation of populations can result in either specialists with improved fitness in one particular niche or generalists with fitness improved across multiple niches. When considering specialisation, related terms include 'local adaptation', 'trade-off' and 'cost of adaptation', and we review them briefly here. Local adaptation occurs when a population becomes better adapted to an environmental niche than populations evolved in other contexts are, i.e. resident genotypes have higher fitness than genotypes evolved elsewhere (Kawecki and Ebert 2004). The concept of 'trade-off' is often found in discussions of specialisation, but has received more than one definition. Some have used the term for any case of specialisation (Hereford 2009; Kassen 2014), while others limit it to cases where the specialised population has decreased fitness in alternate niches relative to the founding population (Fry 1996; Bennett and Lenski 2007; Leiby 2014). In these latter cases, a trade-off involves attaining a 'cost of adaptation' relative to the ancestral genome. We will use trade-off in this second sense, but the processes involved are also relevant when considering genotype *x* environment interactions more broadly.

Contrary to what may be expected, relatively few trade-offs have been empirically demonstrated in reciprocal transplant experiments designed to test for them (Fry 1996; Remold 2012; Jasmin and Zeyl 2013; Leiby and Marx 2014). Hereford (2009) notes that in 43 % of the reciprocal transplant data from published studies analysed, at least one of the populations increased in fitness in both environments, while 48 % showed a trade-off. We can generalise that it is not the usual case for adaptive variants to exhibit 'antagonistic pleiotropy', where the same variant (in a particular genomic context) causes an increase in fitness in one environmental niche and regress in another. This mechanism of trade-off has also been referred to as 'functional interference' (Bell 2008). Another type of trade-off is termed 'mutation accumulation' (Cooper and Lenski 2000) or 'mutational degradation' (Bell 2008), where neutral variants accumulate through processes of genetic drift, but are deleterious in an alternate environment. Examples of these processes are discussed in the section on Experimental Evidence.

The ecological processes leading to specialisation to a particular niche have received a lot of discussion in the literature. A key driver of specialisation is divergent selection. Divergent selection operates when the 'force' exerted by natural selection acts to promote the maintenance of different alleles according to different environmental niches to which a population is exposed, i.e. 'selection that acts in contrasting directions in two populations' (Nosil et al. 2009). Under divergent selection, natural selection may favour different alleles at the same locus in different niches or else may act on different loci in each niche. The adaptive consequences of these two situations may differ, depending on a population's exposure to each environment and the extent to which recombination between lineages is achieved. Divergent selection is believed to be a contributor to speciation in many instances (Schluter 2009). When divergent selection leads to phenotypic divergence, the result can be an adaptive radiation, where populations diversify to fill differing ecological niches (Rainey and Travisano 1998).

Populations are unlikely to be equally well adapted to two different environments. While costs relative to the ancestor may be apparent less frequently, different environments are associated with different fitness optima, for any particular genotype. Adaptation to one environment may not be costly in alternate environments in cases where the population was not already well adapted to the alternate environment (Martin and Lenormand 2015). As an example, in the case of the Goddard system described earlier, there is good reason to believe that the isogenic starting population was reasonably well adapted to the nitrogen-limited high osmotic pressure environment, and less adapted to the warm carbon-limited environment (Gray and Goddard 2012a). As such, we would expect that the populations evolved solely in the C-limited niche would evolve less of a trade-off in the alternate, N-limited, niche than those evolved solely in the N-limited niche. This expectation appears to be borne out, with the average fitness in those populations evolved in the N-limited niche with no gene flow lower than the fitness of those evolved in the C-limited niche with no gene flow; however, the difference is not statistically significant (one-tailed t-test, $p = 0.09$). Fisher's geometric model discussed earlier can be extended to describe situations with more than one fitness optima, for instance when two environments are each associated with a different optima. This has been formalised by Martin (2015, in preparation). This extension of Fisher's model can be illustrated with a simple diagram, as in Fig. 13.2a. The model implies that the further away from the two optima that a population starts, the higher the proportion of conceivable mutations beneficial in one environment that are also beneficial in the other, moving the population closer to both optima, i.e. on this model, adaptive mutations can be expected to be generalist in effect when the starting point is distant from both optima. This trend is shown in the graph in Fig. 13.2b. This is a hypothesis that is open for falsification in microbes, and results for or against would be useful for understanding the mapping between phenotype and genotype beyond a single environment, but it has not been directly tested to our knowledge.

A key question to ask when classifying trade-offs is whether selection is able to 'see' the cost and therefore may act to mitigate it if appropriate genetic variation becomes available. Many trade-offs are merely hypothetical, if the relative alternate environment is not encountered or is able to be avoided by the organism. If a population does experience a fitness cost in an alternate environment, there are a number of mechanisms by which this can be reduced. The theoretical background for understanding the evolution of trade-offs is discussed in the next section.

13.4 The Contexts of Trade-offs

We discuss details of the genomic contexts of trade-offs in a forthcoming article and summarise our arguments here. Trade-offs are not always necessary consequences of a particular environment; different populations which have adapted to the same environment from the same genetic starting point can exhibit different trade-offs by following different routes through genotype space (Jasmin and Zeyl 2013).

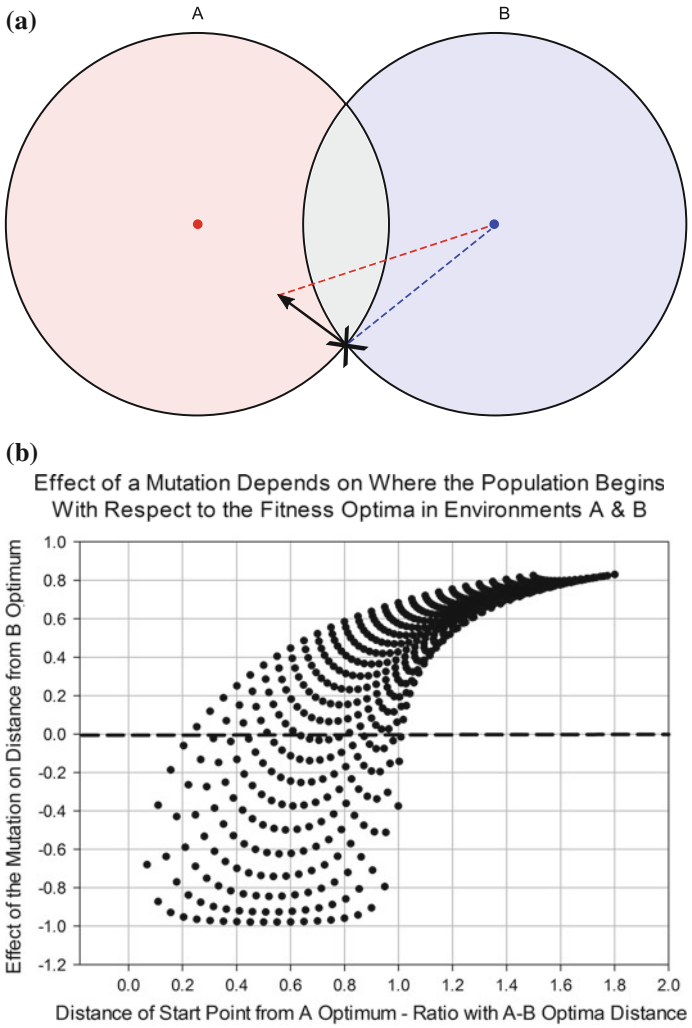


Fig. 13.2 **a** A simple extension of Fisher's Geometric Model to two environments. An adaptation towards the optimum in environment A (*centre of A circle*), represented as a vector, here increases the distance from the optimum genotype in B (*centre of B circle*)—a trade-off occurs. The effect in the alternate environment of an adaptive mutation will depend on the population's starting point in relation to each optimum. Adapted from Kassen (2014, p. 64). **b** Graph showing the relationship between a population's starting point and the fitness effect, in the alternate environment, of a mutation (change in distance to the optimum). The data plotted are for a particular 'distance' in phenotypic space between the optima and a particular mutation size, but the same trend holds across a wide range of values

Different genetic backgrounds can predispose a population to the later selection of particular adaptive variants. This phenomenon has been termed 'potentiation' (Kassen 2014). This is one way in which the wider genomic context is important in

the evolution of trade-offs. As with many genetic features, trade-offs can evolve, moderated by both selective pressures and stochastic processes. When trade-offs do occur they need not be gained in a 'symmetrical' way. Adaptation to environment A may cause a cost of adaptation in B, even though adaptation to B tends to cause no corresponding loss, and perhaps even a fitness increase, in A (Kassen 2014). This might occur due to differences in the pleiotropy of the affected genes in each environment; antagonistic pleiotropy would promote different adaptive routes being followed in each environment, as variants which are adapted in one environment would be maladaptive in the other and so unlikely to be fixed. Mutation accumulation can also occur differently in different environments when different genes are expressed in each; if fewer genes are expressed in a particular environment, the mutational target size (number of inactive genes) in which neutral processes can lead to the accrual of mutations is larger (Kassen 2014).

The biological 'level' at which trade-offs are assessed differs markedly between studies. The most obvious way to measure a trade-off is in terms of fitness, which in microbial populations, for instance, can be measured by comparing growth rates with samples from an ancestral population. Trade-offs can also be measured in terms of particular phenotypes such as metabolic abilities. An example of this was the use of Biolog plates, culture media plates used to compare metabolic abilities on different carbon sources, to measure metabolism of populations from the Lenski long-term *E. coli* evolution experiment (Cooper and Lenski 2000). Alternatively, at a more fundamental level of function, the function of a particular gene's product such as an enzyme could also be tested directly. Or, the genetic basis of a trade-off can be investigated in yet another way, in terms of quantitative trait loci (QTLs), or in more detail at the level of the gene or specific nucleotide. When considering the genetic basis for a trade-off, it is important to recognise that variants that may be causally efficacious in trade-off go beyond single nucleotide polymorphisms (SNPs) and need not be in regions of the genome which directly encode protein sequence. As an example of this, changes in the regulation of the gene RpoS have been shown to undergird differential fitness between environments in *E. coli* strains (King et al. 2004). Possible variants which can contribute to a trade-off include not just a range of small scale changes in sequence, but also changes in genome structure such as chromosomal rearrangements and duplications. Non-nuclear, e.g. mitochondrial, genomes may also contribute to trade-offs between environments (Ballard and Pichaud 2014).

A contributor to the possibility of compensatory mutations and the amelioration of trade-off effects, as well as the origin of trade-offs, is epistasis between genomic loci. Epistasis can occur between genes; for instance, if mutations in each of two genes are beneficial separately but detrimental when in the same genetic background this is known as reciprocal sign epistasis. This pattern was observed between a mutation in the yeast gene MTH-1 and amplification of HXT6/HXT7 (Kvitek and Sherlock 2011). It can also occur within a gene, as illustrated by a study of variants in hepatitis C virus NS3 protease (Parera and Martinez 2014), where a particular substitution was introduced into 56 different genetic backgrounds. Fitness effects of this single substitution were predominantly deleterious,

but ranged from lethal (8.9 % of backgrounds) to beneficial (7.1 %). Even when epistasis is observed between genes, it needs to be recognised that different mutations within a gene may have very different effects, so specific loci do need to be named. Additionally, epistatic interactions exist not only pair-wise, but between multiple mutations (Paaby and Rockman 2014). There are other higher order interactions which may also need to be explored in future, such as 'genotype *x* genotype *x* environment' interactions, or what could be called 'epistasis by environment interactions'. In other words, epistatic interactions may often differ according to the environment in which a genome is situated.

It can be argued that genome size and complexity as well as population size are likely to influence the dominant forms of trade-offs occurring in a population. We briefly review this argument here—some elements are well established—while others are more speculative. The adaptive trajectories of smaller populations tend to be influenced more by stochastic processes, allowing neutral and slightly deleterious mutations to accumulate, so we would expect more 'mutation accumulation' trade-offs in smaller populations, such as in higher eukaryotes. An important caveat to this however is that one of the effects of clonal interference is to reduce effective population sizes in asexual populations (Haag and Roze 2007); how this affects trade-offs is a question for future research. It is also possible that small populations fix mutations which compensate for antagonistically pleiotropic variants more slowly than large ones (Qian et al. 2012), and so small populations exhibit more antagonistic pleiotropy than large ones. We suggest that our hypothesised smaller initial incidence of antagonistic pleiotropy in multicellular eukaryotes compensates for this possible effect, but this is also a matter for further research. Greater genomic complexity, associated with greater redundancy (Sanjuan and Elana 2006) may allow for more neutral mutations. A related concept is the accumulation of 'cryptic' variation (Paaby and Rockman 2014), and we suggest that this is likely to be more common in higher eukaryotes, again increasing the potential for mutation accumulation, as a functional element that is redundant in one environment may have a fitness effect if interrupted in an alternate environment.

More speculatively, microbes may experience more 'loss of function' mutations than higher eukaryotes due to a more modular genome structure, where perhaps components can be lost with less effect on the overall system than in multicellular organisms. If most adaptive mutations in microbial systems can be classed as 'loss of function' (Behe 2010), then it is reasonable to expect trade-offs may result in alternate environments. Whether this claim holds true beyond the microbial systems focused on in Behe's review deserves further analysis. Another plausible contributor to a difference between the rate of 'loss of function' mutations in microbes and multicellular eukaryotes is that asexual microbes may experience more large-scale chromosomal rearrangements than sexual species, if sex does in fact act to reduce large-scale variations as suggested by Gorelick and Heng (2011). Genomic rearrangements can be a cause of antagonistically pleiotropic trade-offs (Wenger et al. 2011). Hence, on the basis of both genome structure and reproductive mode, asexual microbes may exhibit more trade-offs than multicellular eukaryotes due to antagonistic pleiotropy.

13.5 Some Experimental Evidence

There is experimental evidence for both antagonistic pleiotropy and mutation accumulation in microbial systems. This has been reviewed by Kassen (2002). Here, we lay out some evidence for each process in microbes. The most well-known project in experimental evolution is the previously mentioned long-term experiment in the laboratory of Richard Lenski at Michigan State University, involving daily serial transfer of *E. coli* populations through what now approaches 60,000 generations of cell division and evolution in a glucose-limited environment (Leiby 2014). The molecular details of some of the phenotypic changes observed have been probed in depth (Blount et al. 2012). The Lenski group initially inferred functional interference as the basis of a trade-off observed in glucose-limited environments due to the loss of unused catabolic functions in a glucose-rich medium, i.e. loss of catabolic function was selected for rather than increasing in the population through a process of genetic drift, and this resulted in decreased fitness in an alternate environment. It has since been shown (Leiby and Marx 2014) that most trade-offs in this system are due to mutation accumulation instead. The original paper in *Nature* (Cooper and Lenski 2000) was probably the most prominent evidence for antagonistic pleiotropy for fitness between environments; however, there is plenty of other evidence for this process, particularly in bacteria and viruses (Pedersen 2013; Carroll et al. 2013; Penterman et al. 2014; Garcia-Arenal and Fraile 2013), in yeast (Wenger et al. 2011; Qian et al. 2012), but also occasionally in multicellular eukaryotes such as the plant *Arabidopsis lyrata* (Leinonen et al. 2013).

A result that has not been emphasised much in discussion of the paper from Leiby and Marx (2014), for instance in the discussion by Cooper (2014), is that the development of trade-offs in that experiment was much less widespread than had initially been suggested (Cooper and Lenski 2000). The trade-offs that did occur were only in 'mutator' strains with particularly high mutation rates. Additionally, while the claim that trade-offs were widespread and due to antagonistic pleiotropy has been overturned, at least one trade-off that evolved in that system was due to antagonistic pleiotropy—the loss of ribose catabolism was due to a loss of function mutation which increased in frequency due to selection (Cooper et al. 2001). Maclean et al. (2004) demonstrated antagonistic pleiotropy to be the cause of a cost of adaptation in the bacteria *Pseudomonas fluorescens*; specialisation to the air–liquid interface of a structured environment resulted in decreased fitness on alternate carbon substrates when compared to the ancestor. However, this experiment used Biolog plates to measure metabolic abilities. The result from Leiby & Marx raises the question of whether other results using Biolog plate data as a proxy for fitness may also need to be reassessed. Another experiment with *P. fluorescens* (Maclean and Bell 2002) explored the dynamics of adaptive radiations across a wide range of carbon sources. During the course of adaptation to a particular niche, costs on alternate carbon sources arose in a stochastic manner, favouring mutation accumulation as the source of costly variants.

In the Goddard system, mutation accumulation was inferred as the basis of trade-offs in fitness following adaptation to osmotic N-limited and high-temperature glucose-limited environments (Gray and Goddard 2012a). The existence of superior generalists—i.e. populations experiencing full gene flow from the alternate niche which suffered no loss in either environment—in sexual but not asexual populations suggests that perhaps the sexual populations were able to remove the variants responsible for trade-off. This is compatible with the 'mutation accumulation' model for trade-off, where costs of adaptation are due to mutations which are neutral in the selective environment but deleterious in the alternate niche. Such mutations should be able to be removed through processes of recombination followed by selection. The superiority of meiosis in clearing detrimental mutations discussed above means that under the mutation accumulation model, sexual populations are able to remove variants causing trade-off, while in asexual populations, they remain in clonal lineages which remain in the population, selected due to linked sites which are adaptive in one or both environments. This understanding follows from the 'ruby in the rubbish' account of the evolutionary benefit of sexual reproduction (Peck 1994).

There are however reasons to reassess the inference to mutation accumulation as the cause of trade-offs. Sexual reproduction was not introduced into a system without sex, instead it was always present in the sexual populations. Sex was not demonstrated to have removed a trade-off per se, as there may have been no or few trade-off-causing variants present in the population. It may be that no trade-off developed in these populations effectively exposed to both environments as only variants which did not cause a trade-off were able to fix under this selective regime. The asexual populations which experienced full gene flow do raise the question of how the sexual populations avoided trade-offs and it is plausible that they did so by removing the variants responsible for trade-off in the asexual population. However, these variants need not have been neutrally accumulated in the asexual populations. The demonstrated superiority of sexual populations in removing deleterious mutations and fixing beneficial mutations may have assisted in sifting through beneficial variants for those which did not have antagonistically pleiotropic effects between these two environments. It may be that antagonistic pleiotropy is responsible for trade-offs in the asexual populations even if many adaptive variants are not antagonistically pleiotropic. In other words, perhaps 'generalist' variants are better able to be fixed in sexual populations. As Elena and Lenski (2003) suggest, it may be that beneficial variants can be divided into those which are antagonistically pleiotropic and those which are generalist in effect and that populations evolving in heterogeneous environments will be enriched for the generalist alleles.

13.6 Sex and Trade-offs

In situations where populations are exposed to heterogeneous environments, environmental trade-offs derived from mutation accumulation should be mitigated by recombination, which breaks apart linked variants, enabling natural selection to

work more efficiently in removing deleterious variants (Gray and Goddard 2012a). If our earlier suggestion that ecological specialisation in multicellular eukaryotes is predominantly due to mutation accumulation is true, this provides a testable hypothesis—there should be a general trend where populations with higher rates of recombination exhibit fewer trade-offs between environments. Similarly, in conditions where trade-offs are due to conditionally deleterious variants accumulated through neutral processes in an original selective niche, sexual populations should be able to expand into alternate niches with greater success than asexual populations. However, the test suggested is by no means definitive, as higher rates of recombination may help remove trade-offs due to antagonistic pleiotropy, as suggested in the earlier discussion. Even when trade-offs between environments are due to antagonistic pleiotropy, sex may facilitate the evolutionary processes of genome rearrangement that allow the cost of adaptation to be avoided. This suggests a broader hypothesis: that the relevant processes of regulatory evolution, chromosomal rearrangement and/or the fixation of compensatory mutations, which undergird the mitigation of trade-offs proceed more quickly in sexual populations exposed to multiple niches, all other things being equal. Sexual populations may be more efficient than asexual populations at reaching a region in the fitness landscape which is free of antagonistic clashes.

We assume for the sake of argument that mutation rates in sexual and asexually reproducing populations are similar, although the actual situation is more complex given that there is evidence of meiosis increasing the mutation rate in *Saccharomyces cerevisiae* (Rattray et al. 2015). On the assumption of mutation rate equality, it is clear that there should be less hitch-hiking of neutral and deleterious mutations in sexual than in asexual populations, because the recombination enabled by sexual reproduction decreases linkage between genetic loci. Hitch-hiking is also called 'genetic draft', in distinction to genetic drift which is due to non-representative sampling of alleles which is a larger contributor to allele frequencies in small populations. If hitch-hiking is a major cause of trade-offs, even those exposed to just one niche should exhibit less trade-offs if sexually reproducing than if asexually reproducing. The mutation load due to genetic drift is expected to be approximately the same in large sexual and asexual populations, assuming no population structure (Haag and Roze 2007)—but the rate of draft is expected to be much higher in asexual populations. The rate of 'neutral evolution' in the classic sense is therefore the same, but the overall rate of accumulation of neutral mutations may be very different, if many mutations spreading through draft are selectively neutral. Whether these are conditionally deleterious, contributing to trade-offs in alternate niches, is a separate question. If however, at least in some systems, many hitch-hiking mutations are conditionally neutral/deleterious across relevant ecological niches, then sexual reproduction should provide a particular advantage in reducing trade-offs between these environments.

To conclude, we note another area for future work—in bacteria, an important adaptive advantage of 'parasexual' (Colegrave 2012) processes of conjugation, transformation and transduction may also be the advantage offered in new niches through acquiring new genetic information. Sex may assist in niche expansion. We

therefore suggest that while the advantage offered by sexual reproduction is likely to be multifaceted, and the reasons for costs of adaptation differ according to genome and population structure, meiosis and recombination are of particular adaptive advantage in heterogeneous environments. Trade-offs between environmental niches, whether due to processes of mutation accumulation (particularly common, we have argued, in multicellular eukaryotes) or antagonistic pleiotropy, are likely to be better avoided—or ameliorated through the selection of secondary mutations—in sexual than in asexual populations.

Acknowledgments V. Ardern and J. Forrest for assistance with diagrams, and Prof. G. Bell for discussions.

References

Adams J, Rosenzweig F (2014) Experimental microbial evolution: history and conceptual underpinnings. Genomics 104:393–398. doi:10.1016/j.ygeno.2014.10.004

Ballard JWO, Pichaud N (2014) Mitochondrial DNA: more than an evolutionary bystander. Funct Ecol 28:218–231. doi:10.1111/1365-2435.12177

Ball P (2013) DNA: celebrate the unknowns. Nature 496:419–420. doi:10.1038/496419a

Behe MJ (2010) Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". Q Rev Biol 85:419–445. doi:10.1086/656902

Bell G (1982) The masterpiece of nature: the evolution and genetics of sexuality. Croom Helm, London

Bell G (2008) Selection: the mechanism of evolution, 2nd edn. Oxford University Press, Oxford, New York

Bennett AF, Lenski RE (2007) An experimental test of evolutionary trade-offs during temperature adaptation. Proc Natl Acad Sci 104:8649–8654. doi:10.1073/pnas.0702117104

Bernstein H, Bernstei C (2013) Evolutionary origin and adaptive function of meiosis. Meiosis

Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513–518. doi:10.1038/nature11514

Burt A (2000) Sex, recombination, and the efficacy of selection—was Weisman right? Evolution 54:337–351. doi:10.1111/j.0014-3820.2000.tb00038.x

Butlin R (2002) The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet 3:311–317. doi:10.1038/nrg749

Carroll SM, Lee M-C, Marx CJ (2013) Sign epistasis limits evolutionary trade-offs at the confluence of single- and multi-carbon metabolism in *Methylobacterium extorquens* AM1. Evolution n/a–n/a. doi: 10.1111/evo.12301

Colegrave N (2012) The evolutionary success of sex. EMBO Rep 13:774–778. doi:10.1038/embor.2012.109

Cooper VS (2014) The origins of specialization: insights from bacteria held 25 years in captivity. PLoS Biol 12:e1001790. doi:10.1371/journal.pbio.1001790

Cooper VS, Lenski R (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407

Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183:2834–2841. doi:10.1128/JB.183.9.2834-2841.2001

Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet 4:457–469. doi:10.1038/nrg1088

El-Soda M, Malosetti M, Zwaan BJ, et al (2014) Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends Plant Sci doi:10.1016/j.tplants.2014.01.001
Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78:737–756
Fry JD (1996) The evolution of host specialization: are trade-offs overrated? The American Naturalist
García-Arenal F, Fraile A (2013) Trade-offs in host range evolution of plant viruses. Plant Pathol 62:2–9. doi:10.1111/ppa.12104
Goddard MR (2007) Why bother with sex? answers from experiments with yeast and other organisms. In: Taylor JW, Kronstad JW, Heitman J, Casselton LA (eds) Sex in fungi. American Society of Microbiology, pp 489–506
Goddard MR, Godfray HCJ, Burt A (2005) Sex increases the efficacy of natural selection in experimental yeast populations. Nature 434:636–640. doi:10.1038/nature03405
Gorelick R, Heng HHQ (2011) Sex reduces genetic variation: a multidisciplinary review. Evolution 65:1088–1098. doi:10.1111/j.1558-5646.2010.01173.x
Gouyon P-H, de Vienne D, Giraud T (2015) Sex and evolution. In: Heams T, Huneman P, Lecointre G, Silberstein M (eds) Handbook of evolutionary thinking in the sciences. Springer, The Netherlands, pp 499–507
Gray JC, Goddard MR (2012a) Gene-flow between niches facilitates local adaptation in sexual populations. Ecol Lett 15:955–962. doi:10.1111/j.1461-0248.2012.01814.x
Gray JC, Goddard MR (2012b) Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. BMC Evol Biol 12:43. doi:10.1186/1471-2148-12-43
Grishkevich V, Yanai I (2013) The genomic determinants of genotype *x* environment interactions in gene expression. Trends Genet 29:479–487. doi:10.1016/j.tig.2013.05.006
Haag CR, Roze D (2007) Genetic load in sexual and asexual diploids: segregation, dominance and genetic drift. Genetics 176:1663–1678. doi:10.1534/genetics.107.073080
Hartfield M, Keightley PD (2012) Current hypotheses for the evolution of sex and recombination. Integr Zool 7:192–209. doi:10.1111/j.1749-4877.2012.00284.x
Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. Am Nat 173:579–588. doi:10.1086/597611
Hoekstra RF (2005) Evolutionary biology: why sex is good. Nature 434:571–573. doi:10.1038/434571a
Jasmin J-N, Zeyl C (2013) Evolution of pleiotropic costs in experimental populations. J Evol Biol 26:1363–1369. doi:10.1111/jeb.12144
Judson OP, Normark BB (1996) Ancient asexual scandals. Trends Ecol Evol 11:41–46. doi:10.1016/0169-5347(96)81040-8
Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity: Experimental evolution in variable environments. J Evol Biol 15:173–190. doi:10.1046/j.1420-9101.2002.00377.x
Kassen R (2014) Experimental evolution and the nature of biodiversity. Roberts and Company, Greenwood Village
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7:1225–1241. doi:10.1111/j.1461-0248.2004.00684.x
King T, Ishihama A, Kori A, Ferenci T (2004) A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. J Bacteriol 186:5614–5620. doi:10.1128/JB.186.17.5614-5620.2004
Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–440. doi:10.1038/336435a0
Kvitek DJ, Sherlock G (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet 7:e1002056. doi:10.1371/journal.pgen.1002056

Leiby N (2014) Adaptation and specialization in the evolution of bacterial metabolism. Harvard University, Cambridge
Leiby N, Marx CJ (2014) Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. PLoS Biol 12: e1001789. doi:10.1371/journal.pbio.1001789
Leinonen PH, Remington DL, LeppäLä J, Savolainen O (2013) Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. Mol Ecol 22:709–723. doi:10.1111/j.1365-294X.2012.05678.x
Lieberman BS, Vrba ES (2005) Stephen Jay Gould on species selection: 30 years of insight. Paleobiology 31:113–121. doi: 10.1666/0094-8373(2005)031[0113:SJGOSS]2.0.CO;2
Lively CM, Morran LT (2014) The ecology of sexual reproduction. J Evol Biol 27:1292–1303. doi:10.1111/jeb.12354
MacLean RC, Bell G (2002) Experimental adaptive radiation in *Pseudomonas*. Am Nat 160:569–581. doi:10.1086/342816
MacLean RC, Bell G, Rainey PB (2004) The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. Proc Natl Acad Sci 101:8072–8077. doi:10.1073/pnas.0307195101
Marini A, Matmati N, Morpurgo G (1999) Starvation in yeast increases non-adaptive mutation. Curr Genet 35:77–81. doi:10.1007/s002940050435
Martin G, Lenormand T (2015) The fitness effect of mutations across environments. Fisher's geometrical model with multiple optima. Evolution doi:10.1111/evo.12671
Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. Mol Ecol 18:375–402. doi:10.1111/j.1365-294X.2008.03946.x
Orr HA (2002) The population genetics of adaptation: the adaptation of DNA sequences. Evolution 56:1317–1330. doi:10.1111/j.0014-3820.2002.tb01446.x
Orr HA (2006) The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J Theor Biol 238:279–285. doi:10.1016/j.jtbi.2005.05.001
Otto SP (2009) The evolutionary enigma of sex. Am Nat 174:S1–S14. doi:10.1086/599084
Paaby AB, Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate. Nat Rev Genet 15:247–258. doi:10.1038/nrg3688
Parera M, Martinez MA (2014) Strong epistatic interactions within a single protein. Mol Biol Evol 31:1546–1553. doi:10.1093/molbev/msu113
Peck JR (1994) A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics 137:597–606
Pedersen SD (2013) Systems biology investigations of *Pseudomonas aeruginosa* evolution in association with human airway infections. Technical University of Denmark
Penterman J, Nguyen D, Anderson E et al (2014) Rapid evolution of culture-impaired bacteria during adaptation to biofilm growth. Cell Rep. doi:10.1016/j.celrep.2013.12.019
Poulton EB, Poulton EB, Schönland S et al (1889) Essays upon heredity and kindred biological problems. In: Poulton EB, Schönland S, Shipley AE (eds, Authorised translation). Clarendon Press, Oxford
Qian W, Ma D, Xiao C et al (2012) The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep 2:1399–1410. doi:10.1016/j.celrep.2012.09.017
Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 394:69–72. doi:10.1038/27900
Rattray A, Santoyo G, Shafer B, Strathern JN (2015) Elevated mutation rate during meiosis in *Saccharomyces cerevisiae*. PLoS Genet 11:e1004910. doi:10.1371/journal.pgen.1004910
Remold S (2012) Understanding specialism when the jack of all trades can be the master of all. Proc R Soc B Biol Sci 279:4861–4869. doi:10.1098/rspb.2012.1990
Sanjuan R, Elena SF (2006) Epistasis correlates to genomic complexity. Proc Natl Acad Sci 103:14402–14405. doi:10.1073/pnas.0604543103

Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323:737–741. doi:10.1126/science.1160006

Schwander T, Crespi BJ (2009) Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. Mol Ecol 18:28–42. doi:10.1111/j.1365-294X.2008.03992.x

Stearns SC (1990) The evolutionary maintenance of sexual reproduction: the solutions proposed for a longstanding problem. J Genet 69:1–10. doi:10.1007/BF02931662

Wenger JW, Piotrowski J, Nagarajan S et al (2011) Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. PLoS Genet 7:e1002202. doi:10.1371/journal.pgen.1002202

Xu J (2012) Rapid evolution in experimental populations of major life forms. In: Singh RS, Xu J, Kulathinal RJ (eds) Rapidly evolving genes and genetic systems. Oxford University Press, Oxford, pp 40–52

Chapter 14
Promoter Order Strategy and Bacterial PspF Regulon Evolution

Goran Jovanovic, Parul Mehta, Christopher McDonald and Martin Buck

Abstract Successful gene regulation that governs the information flow from DNA underpins programs of differentiation and adaptation across all kingdoms of life. Transcription in bacteria is controlled by RNA polymerase containing the housekeeping sigma factor(s) or the alternative, e.g. sigma54 factor, its function being regulated via bacterial enhancer binding proteins. Elaboration of the enhancer DNA sequence acting *in cis* with the RNA polymerase is believed to drive species diversification through enlargement and finessing of regulons that help cells survive stresses and become more complex. The cellular adaptation and associated gene expression can translate to heterogeneity across cell populations, potentially causing physiological diversity that leads to successful and stable colonisation of ecological niches under unstable changing environments typically faced by bacteria. Here, we present the Phage shock protein F (PspF) regulon as a model system to describe the actions and dynamics of *cis*- and *trans*-regulatory elements that govern the control of sigma54-dependent transcription of *psp* genes leading to adaptation of enterobacteria to the inner membrane stress. We discuss how the interplay of *psp* expression control elements may influence the evolution of bacterial regulon.

G. Jovanovic (✉) · P. Mehta · C. McDonald · M. Buck
Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK
e-mail: g.jovanovic@imperial.ac.uk

P. Mehta
e-mail: pm432@cam.ac.uk

C. McDonald
e-mail: cmcdonald11@imperial.ac.uk

M. Buck
e-mail: m.buck@imperial.ac.uk

P. Mehta
Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_14

14.1 Introduction

Gene transcription is often controlled by specialised activator protein complexes that stimulate key functions of the multi-subunit RNA polymerase (RNAp) needed to copy DNA into RNA. RNAps are structurally and functionally conserved between Bacteria, Archaea and Eukarya. In bacteria, different sigma factors that recognise the specific promoter sequences associate with the RNA polymerase to enable initiation of transcription from the target promoter(s). There are two general groups of sigma factors in bacteria. One group comprises the family of housekeeping sigma70 factors, while the alternative sigma54 (RpoN) factor belongs to the other group. The major difference between these two groups of sigma factors is that RNAp containing the housekeeping sigma factor that recognises the −35/−10 promoter DNA sequence (in respect to transcription start site) can spontaneously initiate transcription upon promoter binding, while the sigma54-RNAp binds different class promoter (−24/−12) and requires the action of the transcription activator complex. These transcription activator complexes often act from so-called enhancer DNA sites remote from the sigma54 target promoter and transcription start site, and can be large homomeric and heteromeric protein assemblies (Buck et al. 2006; Bush and Dixon 2012; Joly et al. 2012). Bacterial enhancer sequence is typically positioned ∼150 nucleotides from the transcription start site and comprises at least one upstream activating sequence (UAS) to which bacterial enhancer binding protein (bEBP), transcription activator, binds via its C-terminal helix-turn-helix (HTH) DNA-binding motif. bEBPs are the members of the AAA+ (ATPases Associated with diverse cellular Activities) family which in their active form impose their function as ATP-bound hexameric complexes. DNA bending enabled by either intrinsic DNA curvature or by binding of the DNA architectural element, integration host factor (IHF) protein, brings active bEBP bound to a UAS in close proximity to the closed promoter complex formed by promoter-associated sigma54-RNAp. Then, bEBP hexamer interacts with (mainly) sigma54 and following the ATP hydrolysis driven by central AAA domain of subunits opens the promoter complex and initiates transcription. Importantly, this mode of transcriptional regulation closely resembles control of eukaryotic RNA polymerase II driven transcription.

The sigma54 factor is used in bacteria to respond to different growth conditions (Shingler 2011; Bush and Dixon 2012) and is a key factor that controls the bacterial environment being involved in transport and synthesis of lipids, lipoproteins and peptidoglycan (Francke et al. 2011). Also the *rpoN* mutants have increased sensitivity to antibiotics (Liu et al. 2010). Most of the bEBPs that activate sigma54-dependent promoters contain an N-terminal regulatory domain which receives the environmental stress signal and controls activity of AAA domain or DNA-binding via HTH motif. However, the activity of some bEBPs (e.g. NifA) or constitutively active bEBPs with no obvious regulatory domain (e.g. PspF, HrpS) is controlled with cognitive negative regulator (e.g. NifL, PspA, HrpV) that acts as a primary signal receiver (Martinez-Argudo et al. 2004; Joly et al. 2010; Jovanovic et al. 2011). The complexity and modality of bacterial cell response to environmental

cues is not only increased *in trans* by bEBPs and their regulators signal recognition but also using combinations of *cis*-control elements, including the DNA enhancers which can contain different number of UASs. In turn, this can increase the heterogeneity of response to stress rendering the subpopulation(s) of cells with different capacity to adapt to given environmental condition. This may be a driving force responsible for rearrangements of DNA control elements or regulation rewiring. Therefore, the studies of sigma54 and bEBP controlled systems can reveal not only how these systems function but may also give us a clue how they change and evolve. One such model system extensively studied is the Phage shock protein (Psp) response system which expression is controlled by master bEBP transcription activator PspF and induced by environmental conditions and stress agents that impair the integrity of inner membrane (IM) in Gram-negative bacteria, including enterobacteria.

14.2 PspF Regulon

The Psp system is one classical and widely conserved extracytoplasmic stress system, recognised as of being of major importance in numerous stress conditions including protein translocation, biofilm formation, secretin production and virulence, bacterial infection of macrophages, antibiotic resistance and persistence in enterobacteria (Rowley et al. 2006; Joly et al. 2010; Darwin 2013). In *Escherichia coli*, the Psp system consists of *pspF* gene and divergently oriented *psp* operon (*pspABCDE*) positioned on 29.45 centisom and of physically separated *pspG* gene at 91.84 centisom (Fig. 14.1). The *psp* operon and *pspG* transcriptions are under control of sigma54 promoter (Joly et al. 2010). The regulatory region of *psp* operon (also *pspA*) contains an enhancer consisting of two UAS sequences, UAS I and UAS II, separated from sigma54 promoter by IHF-binding site (Fig. 14.1). It seems that UAS II ($\sim$130 nucleotides upstream from the sigma54 promoter) has a major impact on activation of *psp* operon (Jovanovic and Model 1997; GJ, PM, MB unpublished). Complexity of this regulatory region is increased by the presence of *pspF* sigma70 promoter which overlaps UAS I and partially UAS II (Fig. 14.1). In addition, in *Yersinia enterocolitica*, a low-level sigma70-dependent transcription of *psp* operon has been shown to exist (Maxson and Darwin 2006). The *pspG* regulatory region upstream from sigma54 promoter comprises IHF-binding site and enhancer composed of one UAS with very similar position and sequence to UAS II in *pspA* regulatory region (Fig. 14.1).

Molecular genetics and biochemical experiments revealed that transcriptions of *pspA* and *pspG* are co-activated by PspF (Lloyd et al. 2004), a bEBP that belongs to AAA+ family (Model et al. 1997; Joly et al. 2010). The active, ATP-bound form of the PspF hexamer binds to UASI and/or UASII via its C-terminal HTH motif. In a *pspA* regulatory region, the PspF effectively binds UAS II, but binding to UAS I is facilitated by the presence of UAS II. The binding of IHF to DNA co-operates binding of PspF (and vice versa) to UAS I/II and UAS II in both *pspA* and *pspG* regulatory regions, respectively (Jovanovic and Model 1997; Lloyd et al. 2004).

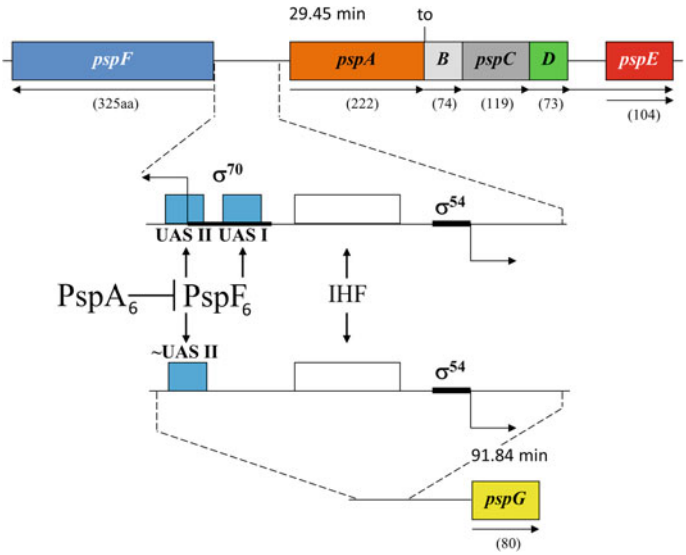


Fig. 14.1 Schematic representation of the PspF regulon and the gene expression control elements in *psp* operon and *pspG* regulatory regions (see text for details). Bar, presents the negative control of the PspF ATPase activity imposed by PspA to the weak *psp* operon internal transcription termination; number in parenthesis, the number of amino acid residues of the protein encoded by the corresponding gene above

Importantly, IHF upon binding bends DNA and enables the functional interactions of UAS-bound PspF and promoter-bound sigma54 factor associated with core RNAp leading to ATP hydrolysis by PspF AAA domain, opening of the sigma54 promoter complex and initiation of *pspA* and *pspG* transcription.

The PspF lacks N-terminal regulatory region, and so it is constitutively active bEBP (Jovanovic et al. 1996). However, the action of PspF is negatively controlled by PspA protein (Fig. 14.1). PspA is a dual-action peripheral IM coiled-coil protein, and besides being the negative regulator of PspF under non-stress conditions, PspA is a major Psp effector upon IM stress. As a negative regulator, PspA does not bind DNA but directly interacts with the PspF hexamer through hydrophobic cluster on PspF surface-exposed "W56 loop" and inhibits its ATPase activity (Zhang et al. 2013). The ratio of PspA–PspF binding in this inhibitory complex is mainly 1:1, but the minimal number of PspA subunits interacting with PspF is ~3. Under non-stress conditions, PspF-PspA inhibitory complex prevail enabling only basal-level transcription of *pspA* and *pspG*. Under stress conditions which damage the IM, PspA negative control is lifted (see Sect. 14.3) and PspF is free to induce the expression from *pspA* and *pspG* sigma54 promoters.

The binding of PspF to UASs in *pspA* regulatory region is a prerequisite for activation of basal-level expression or induction of *psp* operon transcription. This, however, collides with *pspF* transcription driven from its sigma70 promoter (see

Fig. 14.1) and so activation of PspF expression is negatively autogenously controlled by means of PspF acting as a classical repressor of its own transcription upon binding to UAS I and/or UAS II (Jovanovic et al. 1997). The PspF autogenous control is independent on interactions with PspA and imposed negative control, but IHF does influence PspF binding to target UASs and affects autogenous control of *pspF* transcription (Mehta et al. 2013). Even though the low-level transcription of *pspF* is partially compensated by high stability of *pspF* mRNA, the intracellular concentration of PspF is very limited (~20 hexamers per cell) (Jovanovic et al. 1997). Otherwise, transcription of *pspF* from its sigma70 promoter is constitutive, namely in *E. coli* cells encoding, the PspF variant that lacks HTH motif and so is unable to bind UASs and activates *pspA* and *pspG* transcription, the expression of *pspF* mutant gene is constitutive and at much higher level (>10-fold). Notably, this level of PspF expression is not sufficient for UAS-independent activation of *psp* sigma54 promoters. Only if overexpressed 100-fold over its native intracellular concentration, this PspF variant is able to bypass specific UAS binding and directly interact with sigma54 and so activates the *psp* transcription (Jovanovic et al. 1999). However, this high-level PspF expression also causes an activation of many other *psp* unrelated sigma54 promoters in cell (GJ, MB unpublished). This emphasises that keeping the PspF level of expression low through the autogenous control and low intrinsic *pspF* sigma70 promoter activity is important to limit PspF crosstalk with other sigma54 promoters to specifically target the *psp* UASs and PspF regulon genes expression.

The actions and control of PspF activator are of key importance for regulation of Psp system response under non-stress and stress conditions. Therefore, we consider *psp* operon, *pspG* and *pspF* genes to be constituents of the PspF regulon.

14.3 PspF Regulon Response to IM Stress

The Psp response in enterobacteria is triggered by extracytoplasmic stresses that damage the IM and consequently dissipate the proton motive force (pmf) and change the membrane potential, redox state and energy status of bacterial cell (Joly et al. 2010). There is a wide variety of stimuli such as extreme environmental conditions (extreme heat and osmotic shock), defected in protein translocation systems and production of secretions (e.g. pIV, YscC, PulD, OutD) that cause high-level production of Psp regulators and effectors proteins resulting in cell adaptation to stress (Model et al. 1997; Darwin 2005; Joly et al. 2010). The PspF regulon respond to IM damage at the level of transcriptional control by means of releasing the inhibitory PspF–PspA complex prevalently formed under non-stress conditions and so inducing the sigma54 transcription from *pspA* and *pspG* promoters. Only in the absence of PspA is the negative control fully abolished suggesting that even under stress conditions, some inhibitory complexes are still present. The differential transcription assays conducted in the absence of stress or upon IM stress reveal that production of pIV in *E. coli* and *Salmonella* or YscC in

Yersinia cause specific up-regulation in transcription of *psp* genes (Lloyd et al. 2004; Seo et al. 2007). In all bacteria tested, the transcription of *pspA* is ~100-fold increased under stress and then declines for the rest of the genes in corresponding *psp* operon. Within the *psp* operon, a possible transcription termination site between *pspA* and *pspB* has been identified (see Fig. 14.1) which could account for the large amount of PspA produced upon induction compared to the downstream proteins. The transcription of *pspG* is ~20-fold up-regulated compared to non-stress conditions. The similar qualitative with more pronounced quantitative outcome for *psp* operon and *pspG* transcriptions has been shown in *E. coli* lacking PspA negative control. This suggests that *pspA* and *pspG* sigma54 promoters have different power in activating transcription, probably due to difference in number of *cis*-control UAS elements. Importantly, the transcription of *pspF* is unchanged emphasising that autogenous control and limited amount of PspF is constant under non-stress and IM stress conditions. This is consistent with observation that PspF negative autogenous control and so *pspF* transcription do not depend on interactions with PspA acting as a negative regulator.

The origin of the *psp*-inducing signal(s) caused by IM damaging stimuli is debated for a long time. It is now believed that upon IM stress, the originated signal is complex and includes changes in chemical (e.g. anionic phospholipids charge) and physical (e.g. curvature, fluidity) properties of the lipids as well as changes in membrane potential and red-ox state of the cell (Joly et al. 2010; Engl et al. 2011; Jovanovic et al. 2014b). The strength of the signal and signal transduction depends on sternness of the stimuli. The PspB and PspC IM proteins are IM stress sensors and positive regulators of the Psp system. PspB and PspC can both recognise a diverse signals, associate and form the assembly able to interact with the PspF-PspA inhibitory complex. Notably, PspC is a polytopic protein which may change its topology in response to changes in membrane potential and/or lipids charge upon IM stress bringing its C-terminal leucine zipper motif in position favourable for the cytoplasmic interaction with PspA coiled-coil protein (Jovanovic et al. 2010; Flores-Kim and Darwin 2012). Under non-stress conditions, PspBC are mainly inactive and only transiently interact with inhibitory complex (Fig. 14.2a). On the onset of IM stress, it is likely that PspBC recognise the signals close to the threshold (e.g. upon pIV, YscC secretin production), integrate them and then transduce via protein–protein interactions by directly and stably interacting with PspA within the PspF-PspA inhibitory complex (Jovanovic et al. 2014b). This releases the inhibitory complex and negative control enabling PspF bound to UASs to greatly elevate the transcription of *psp* operon and *pspG*. In turn, the PspBC sequester PspA regulator at the IM (Fig. 14.2b). The well-characterised agent that induces Psp response in PspBC-dependent manner is mislocalisation of the outer membrane secretin pIV into the IM. The more severe treatments (e.g. extreme heat shock) and the corresponding signal(s) at the IM (e.g. changes in physical properties of lipids) can directly, via PspA–IM interactions, recruit and subsequently release the PspF-PspA inhibitory complex in PspBC-independent fashion (Model et al. 1997; Jovanovic et al. 2014a, b; McDonald et al. 2015). Possibly, in many instances, the IM stress

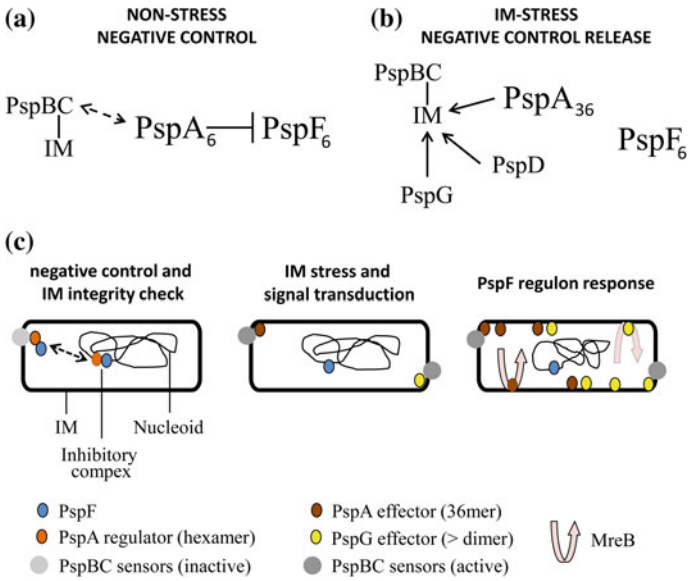


Fig. 14.2 Protein–protein and protein–IM interactions in PspF regulon response to IM stress. **a** PspF and PspA as hexamers interact and form the inhibitory complex which transiently interacts with the PspBC IM integrity sensors under non-stress conditions. **b** Under IM stress conditions, following stable interactions between the inhibitory complex and PspBC sensors, PspF a hexameric activator is released leading to induction of *psp* genes expression. PspA then switches to the high-order oligomeric effector and interacts with the IM to repair the damage. The additional Psp effectors, PspD and PspG, interact with the IM and help PspA to confer the adaptation to stress. **c** Spatio-temporal distribution of key Psp regulators and effectors under non-stress and IM stress growth conditions (see text for details)

signal transduction to the PspF-PspA inhibitory complex involves synergistic action of PspBC-dependent and PspBC-independent mechanisms.

Following the PspBC-assisted or direct interaction with the IM-promoted release of interaction with PspF, the PspA hexameric regulator switches to IM-bound high-order oligomeric (36mer) effector complex able to repair the membrane damage and conserve the pmf (Kobayashi et al. 2007; Jovanovic et al. 2014B) (Fig. 14.2b and see Sects. 14.6 and 14.7). In *E. coli*, the PspA acts in conjunction with the IM effector proteins PspD and PspG (Fig. 14.2b) and the periplasmic protein PspE to confirm the adaptation to IM stress (Joly et al. 2010) (see Sect. 14.6). The PspBC sensors in *Y. enterocolitica* (PspC has an extra amino acids at N-terminus compared to the *E. coli* PspC) are shown to have also the effector function in maintaining the IM stress due to secretin production (Horstman and Darwin 2012).

The molecular genetics and biochemical methods that produce an averaged data set across observations could not answer the questions concerning the spatio-temporal organisation of Psp response in vivo including the real-time gene expression and distributions, dynamics and stoichiometries of regulators, sensors,

effectors and their complexes. Therefore, we employed a single-cell single-molecule live cell imaging (SMI) to start addressing these questions.

14.4 Cell Biology of the PspF Regulon Response

The SMI was used to study quantitatively in vivo gene expression from *psp* promoters and to establish the spatio-temporal distribution of key PspF regulon proteins in *E. coli* model system. The SMI is a powerful methodology because it can be used to capture the dynamics of biological action and transient intermediates in live cells to provide insights into mechanisms of gene regulation avoiding the ensemble averaging. Additionally, SMI can reveal heterogeneous behaviour across cell populations and cellular adaptation-associated gene expression. The stable and functional PspF, PspA, their variants and PspG fluorescently tagged with eGFP and/or Venus expressed at physiological concentrations either from the plasmid or from the chromosome were used to employ SMI methodology. It was established that under non-stress conditions, PspF-PspA inhibitory complex mainly resides in nucleoid bound to *psp* UASs and only transiently, driven by PspA, commutes between nucleoid and membrane polar regions in PspBC-dependent manner (Mehta et al. 2013; Jovanovic et al. 2014a, b) (Fig. 14.2c). Accordingly, the PspF variant lacking its HTH UAS-binding domain or PspA variant with PspF interactions abolished does not associate with nucleoid. On the onset of IM stress, PspF-PspA inhibitory complex now stably interacts with the curved polar membrane regions leading to release of PspA and PspF interactions (Fig. 14.2c). This IM association depends on the presence of PspBC sensors and anionic lipid cardiolipin which is in principle distributed in curved membrane regions. The complementary studies in *Y. enterocolitica* confirmed that upon IM stress the PspF–PspA inhibitory complex indeed interacts with PspBC sensors in polar regions where PspA exchanges its PspF partner for the PspBC complex (Yamaguchi et al. 2010, 2013). In *Y. enterocolitica*, PspA complex is then found only in polar IM regions. However, in *E. coli*, following PspA switch from hexameric regulator to the IM-bound high-order oligomeric effector, PspBC assist PspA association with bacterial actin MreB. This enables dynamic distribution of PspA effector complexes to the distinct lateral IM regions, many of those being marked by RodZ, the MreB-interacting protein and part of the cell wall biosynthesis machinery (Jovanovic et al. 2014b) (Fig. 14.2c). Apparently, the PspA variant unable to bind the IM localises either in nucleoid or in polar membrane PspBC-dependent regulatory complexes and does not exhibit the effector complexes.

In majority of IM stressed cells, the PspF is rarely seen to commute with the IM but is rather static, stably nucleoid bound and engaged in elevated activation of *psp* transcription (Mehta et al. 2013) (Fig. 14.2c). This separates actions of PspF and PspA proteins in different bacterial cell compartments. The great-level production of Psp regulators and effectors imposes the positive feedback on Psp response since as long as the stress persists, the PspBC will be activated and PspA negative control

diminished. The more severe the stress, the number of cells with nucleoid-associated PspF will increase and more PspA lateral effector complexes will be formed. Notably, the PspF W56A variant with no PspA negative control imposed resembles PspF localisation and dynamics seen upon IM stress conditions. Under IM stress, similar to PspA, the PspG will interact with PspC, oligomerise and form the MreB-dependent dynamic lateral membrane complexes (Engl et al. 2009; Jovanovic et al. 2010) (Fig. 14.2c). It seems that for many stimuli, the PspBC complex represents the main IM integrity and stress signalling checkpoint used by Psp regulators and effectors. The PspBC-dependent induction of PspF regulon response is good example how the high-order assemblies in signal transduction maintain the threshold, integration and amplification of the signal leading to reduction of noise and oscillation in gene expression and temporal and spatial control of the stress response system (Toni et al. 2011; Jovanovic et al. 2014a, b).

14.5 *pspA* and *pspG* Regulatory Regions Interact Via PspF

The SMI methodology enabled a direct visualisation of the movements and localisations of bEBP PspF activator and its interactions with target regulatory *in cis* and negative regulator PspA, architectural element IHF and sigma54 *in trans* components. This is important, since these factors directly impact on making open sigma54 promoter complex and determine the rate of transcription initiation. The factors that alter the ATPase activity (e.g. PspA) of the PspF and/or the relative frequency of contact between enhancer/UAS-bound PspF and the promoter Promoter-bound sigma54-RNAp (e.g. IHF) can change the control of sigma54 promoters (Mehta et al. 2013). The PspF–PspA interactions seem to control the dynamics of PspF mostly at the level of inhibitory complex assemblies and communication with the IM. The diffusion and localisation of PspF are also influenced by the presence of sigma54 and IHF. The presence of sigma54, besides promoting the PspA expression, seems to facilitate the movement of DNA-bound PspF, most likely by the clearance of PspF from the open promoter complex. The IHF enhances PspF nucleoid association independently of PspA consistent with assisting binding of PspF to *pspA* and *pspG* UAS's DNA (see Sect. 14.2). This tightens the PspF autogenous control and facilitates PspF-closed promoter complex interactions through DNA-looping. The details of this intricate balance between PspF associations with *psp* regulatory regions and IM supported by layers of cooperativity between the Psp proteins, regulatory sequences and the local architectural element, IHF, escaped detection in classical in vivo and in vitro assays.

In the same manner, a detail analysis of SMI data regarding PspF localisation, dynamics and stoichiometry in live bacterial cells under different growth conditions strongly suggests that the consequence of PspF being at a very limiting intracellular concentration is that the PspF regulon control is not simply proportional to the association of the activator with the *pspA* and *pspG* transcription apparatus. The main nucleoid UAS-bound form of PspF in inhibitory complex or when the negative

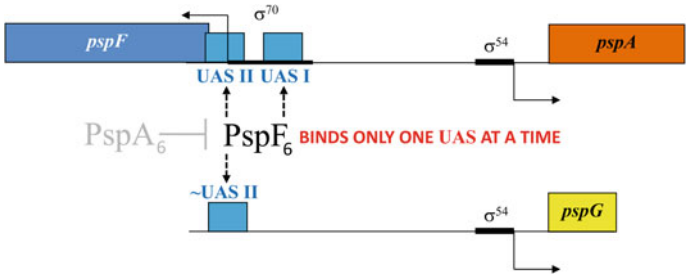


Fig. 14.3 PspF-UAS interactions. PspF hexamer as an activator being in inactive PspA-bound or active form binds only single UAS/enhancer in *pspA* or *pspG* regulatory regions. This suggests that certain promoter order strategy exists resulting in *psp* enhancers communication via PspF and balanced *psp* genes expression under non-stress or IM stress conditions

control is lifted is a hexamer (assemblies from dimer up to active hexamer has been observed) (Mehta et al. 2013). Accordingly, in nucleoid-associated inhibitory complex or in regulatory PspBC-PspA(-PspF) polar IM complexes, the PspA is found to be primarily hexameric, while as an IM-bound effector, the PspA is a high-order oligomer (up to 36mer) (Lenn et al. 2011; Jovanovic et al. 2014b) (see Sects. 14.6 and 14.7). Since the stoichiometry of the repressed and the activating ATP hydrolysing form of PspF is the same, the changes in the oligomeric state of this bEBP are not the mean of regulating its activity in vivo.

Significantly, as an active hexamer, PspF predominantly (>95 %) binds a single *psp* regulatory sequence at a time (Mehta et al. 2013) (see Fig. 14.2c) and so would need to bind *pspA* and *pspG* target UASs on average in turn for a balanced control activation of corresponding sigma54 promoter (Fig. 14.3). The binding dynamics of PspF does not depend on the absence or presence of IM stress or the presence of PspA and negative control in logarithmic growth phase (see Fig. 14.2c). Importantly, the *pspA* and *pspG* *cis*-regulatory sequences are different (see Sect. 14.2); *pspA* region comprises two PspF binding sites, while *pspG* region contains only one (see Fig. 14.3) which may account for regulatory flexibility and potentially to a particular PspF-binding order. Consequently, the complete lack of *pspG* regulatory region slows down diffusion and dynamics of PspF and leads to increased expression of the *pspA*. Hence, the limited intracellular concentration of PspF and the binding capacities of *psp* regulatory regions might be the key factors in determining intracellular dynamics of PspF and communication between *psp* regulatory regions and so activation of the *psp* promoters. This may as well be a destabilising factor in *psp* genes expression leading to heterogeneity in IM stress response within bacterial cells population.

The recent studies of PspF subcellular localisations and dynamics in early stationary phase *E. coli* cells confirmed this presumption. Compared to logarithmic growth phase, in stationary phase the dynamics of PspF has been found to be ~10-fold reduced under non-stress or stress conditions or in the presence or absence of PspA (Mehta et al. 2015; GJ, PM, MB unpublished). More importantly,

in stationary growth phase the number of nucleoid-associated UAS-bound PspF gradually increases from the single to two or three complexes per cell depending on the number of *psp* regulatory regions (*pspA* alone, *pspA* and *pspG*, *pspA*, *pspG* and P_{pspA}-*lacZ*) compared to mainly one nucleoid-associated PspF complex being observed in logarithmic phase. In addition, the relative intracellular protein density is increased by 33 % in stationary phase compared to logarithmic phase. Since in the stationary phase the cytoplasm crowding is substantially increased with the reduction in relative cell size and volume in concert with more pronounced glass-like properties of the cytoplasm (Parry et al. 2014; Mehta et al. 2015), and considering that the concentration of IHF is also elevated, it is very likely that the binding of PspF to its target *pspA* and *pspG* enhancers/UASs is more stable and prolonged under given growth conditions. In stationary growth phase, DNA supercoiling in *E. coli* is more relaxed in comparison with logarithmic phase and although impact of DNA supercoiling differ depending on the chromosomal region, in general it enables the more effective DNA binding of transcription factors (Reyes-Dominguez et al. 2003). This may as well facilitate both the PspF-UAS interactions and its target closed promoter complex formation in stationary phase. In addition, it might contribute to variation in *psp* promoter's activity. The experiments to determine precise and real-time levels of nascent *pspA-BC* and *pspG* mRNA's synthesis (smFISH methodology) in the absence or presence of PspA, P_{pspG} and IM stress are currently underway in our laboratory. They may answer the questions regarding the expression dynamics and nucleoid structure proximity of the *psp* operon and *pspG* genes/regulatory regions.

Therefore, different number of the enhancers/UASs required for PspF DNA-binding and chromosomal localisation of the *psp* regulatory regions can determine not only the strength of promoters but also PspF dynamics and order of *psp* enhancers binding, leading to overall different levels of two *psp* sigma54 promoter activities. The interaction between *psp* regulatory regions via PspF accounts for the non-simultaneous and balanced expression of *pspA* and *pspG* promoters as well as for the coordinated autogenous control of *pspF* sigma70-driven transcription (Fig. 14.3). The observed promoter order strategy may well reflect the *psp* operon expansion to a regulon (e.g. addition of *pspG*) as controlled by a limiting amount of the PspF activator.

14.6 Expansion of *psp* Operon into the PspF Regulon

The extensive systematic analysis based on the genes that constitute the Psp system in *E. coli* revealed the evolutionary history of the system (Fig. 14.4a) (Huvet et al. 2009, 2011). The comparative analyses of genome/operon organisation within ~700 Proteobacteria Gram-negative genomes, including alpha-, gamma- and delta-proteobacterium, suggest that the Psp system evolved around PspF and PspA as a small protein interaction network leading to system-level correlated evolution. The new genes are mainly recruited into a *psp* operon and their additions improve

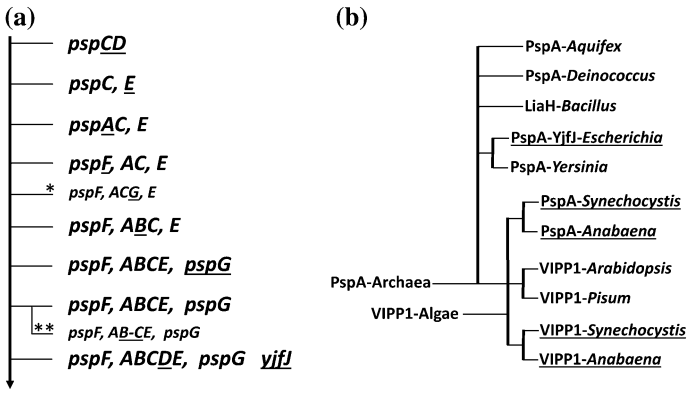


Fig. 14.4 Evolution of the PspF regulon genes and PspA homologues. **a** The Psp system orthologues appearances are organised based on their position in simplified phylogenetic tree inferred from the 23S rRNA sequences of Proteobacteria. Newly acquired genes are underlined. *Asterisks*, one of the examples of *pspG* being a part of the operon in *Aeromonas* species and PspBC being a chimeric protein in *Sodalis glossinidius*. The *yjfJ* gene is present in Proteobacteria (e.g. *Caulobacter*, *Mesorhizobium*, *Zymomonas*, *Pseudomonas*, *E. coli*, *Salmonella*) and was potentially part of the PspF regulon in enterobacteria (see text). **b** Distribution of PspA homologues/orthologues (PspA, LiaH, VIPP1) in Gram-negative (i.e. *E. coli*) and Gram-positive (i.e. *Bacillus*) bacteria, cyanobacteria (i.e. *Synechocystis*), archaea (i.e. *Haloferax volcanii*), green algae (i.e. *Chlamydomonas*) and higher plants (i.e. *Arabidopsis*). Two PspA homologues co-expressed in a same species are underlined

and modulate sensory, signalling and effector functionalities. The key regulatory proteins PspF and PspA, PspA being also the major effector of the system as well, constitute minimal system which respond to more than one environmental signal establishing the potential in regulon evolution from standpoint of regulatory signal reception (e.g. PspC and then PspB) (see also Sects. 14.3 and 14.4). In addition, a plethora of stimuli expands the variation of IM damage and potential in evolution of effector proteins (e.g. PspD, PspG). Apparently, the variety in effector protein functionality was a driving force for the expansion of *psp* operon into the PspF regulon.

In *E. coli*, the PspA and PspG complementary functions, signalling and spatio-temporal organisations underline their synergistic actions leading to adaptation to IM stress (Joly et al. 2010; Jovanovic et al. 2010) (see Sects. 14.3 and 14.4 and Fig. 14.2b, c). However, they have distinct features and functionalities that may explain why initially the *pspG* was part of the *psp* operon and later in evolution exists as a physically separated and co-regulated member of the Psp system (see Fig. 14.4a).

PspA acts as a peripheral IM high-order oligomeric effector (see Sects. 14.3 and 14.4) and in vitro binds anionic phospholipids phosphatidylglycerol and phosphatidylserine to repair the membrane damage and conserve the pmf (Kobayashi et al. 2007). The PspA lipids binding is facilitated by stored curvature elastic stress (McDonald et al. 2015). This membrane feature might as well be the main target for

the function of PspA effectors which upon binding might relax and stabilise the membrane conserving pmf and energy status of the cell. In addition, PspA effector actions decrease motility that in addition conserves pmf and greatly increase the spermidine uptake and synthesis found to be implicated in growth, cell wall synthesis, acid resistance, biofilm formation and virulence (Joly et al. 2010). Notably, the *psp* operon in enterobacteria is adjacent to gene cluster involved in spermidine synthesis and acid resistance.

The PspG is an integral IM protein which following the induction and interaction with PspC upon IM stress affects the metabolism to help conserve the pmf in vivo (Joly et al. 2010; Jovanovic et al. 2010). High level of PspG in concert with PspA modulates the metabolism by diminishing the glycerol shift and aerobic respiration, favouring glycerol-3-phosphate conversion into phospholipids (Jovanovic et al. 2006; Bury-Mone et al. 2009). Accordingly, the inhibition of phospholipids biosynthesis induces the Psp response. The actions of PspG specifically up-regulates the fast responding *narG*, *nirB*, *focA* and *yfiD* genes and so may fine-tune the respiration to microaerobic nitrate-formate mode to conserve energy and pmf under IM stress. Importantly, cardiolipin shown to be important for Psp signalling is also necessary for in vitro assembly of the respiratory super-complexes. Notably, the activity of cardiolipin-associated NarGHI complex is enhanced by anionic phospholipids that confer the interaction with the quinol (Arias-Cartin et al. 2011). The quinone/quinol ratio regulates the activity of IM red-ox sensor ArcB found to be implicated in signals integration by PspBC and is important for the ArcAB–PspBC-dependent activation of *psp* in microaerobiosis (Jovanovic et al. 2009; Joly et al. 2010). Notably, in enterobacteria the *pspG* is positioned on a chromosome in proximity of *qor* gene encoding for quinone oxidoreductase. Therefore, the PspG role may be to complement the actions of PspA by modulating the cells respiration (aerobic versus microaerobic) to provide the conditions favourable for conserving the pmf upon IM stress.

In *E. coli* cells, peripheral IM protein PspD exhibits to some extent the PspA-like effector function (Jovanovic et al. 2006) and might assist PspA. It is of importance to note that the PspD resumed presence as a *psp* operon member coincide with the PspG appearance separated from the *psp* operon cluster (see Fig. 14.4a). This suggests the specialisation of PspA function supported by PspD lead to overtaking the part of the function by PspG. Finally, the periplasmic PspE, a major *E. coli* bona fide rhodanese (Cheng et al. 2008), in addition to sigma54, has its own independent sigma70 control and stable conservation throughout the evolution (see Fig. 14.4a). PspE could be involved in the IM respiratory enzyme [Fe-S] clusters and the periplasmic S–S bond repairs upon IM stress.

The PspF regulon formation empowers bacteria to be more effective in encountering different environmental growth conditions. However, the *psp* operon has been defined as functionally coherent unit with shared transcriptional control and balanced mRNA/protein expression levels. The *pspF* and *pspABC* genes have been found to be highly conserved in many enterobacteria and are the core members of the Psp response (Darwin 2005; Huvet et al. 2011). There is a question of fitness of bacteria containing a simplified version of the Psp system under different

oxygen-related environments. Surprisingly, under IM stress conditions in anaerobiosis *pspG* and/or *pspBC pspG* mutants perform better than wild type. Intriguingly, this supports the notion of the "selfish operon" where co-localisation and co-expression is advantage to the operon genes but not to the organism, depending on growth conditions. This also suggests that under certain growth conditions, the potential low selective pressure on *pspG* together with regulatory flexibility due to combinatorial effects of UASs and limited amount of the activator might through promoter order strategy lose control over *pspG* (and/or subsequently lose *pspG* gene itself) and so reduce the already formed PspF regulon.

14.7 Distribution and Structure Function of PspA Homologues

PspA is a key protein of the Psp response regarding both control and the effector functions. The PspA contains 222 amino acids which constitute four alpha-helical domains, HD1–HD4, which accommodate coiled-coil structure (residues 30-187/204-222). HD1–HD3 are implicated in interactions with PspF and negative control, while HD1 and HD4 are involved in IM-binding and high-order oligomer formation (Joly et al. 2009). More specifically, the N-terminal region of HD1 comprises two amphipathic helixes (ah), ahA and ahB. The ahB is a determinant for PspA interaction with PspF "W56 loop" hydrophobic cluster and ahA is a major determinant for the PspA–IM binding (Jovanovic et al. 2014a). The interplay between ahB and ahA determines the oligomeric state and function of PspA. Upon engagement of ahB in interaction with PspF ahA is unable to stably bind the membrane and PspA is a low-order oligomer. Following release of PspF, ahA binds the IM and enables PspA high-order oligomer formation via HD4. The PspA ahA and ahB talk is possible only in the presence of highly conserved residue proline 25 (P25) positioned between these ahs.

The PspA homologues/orthologues in Gram-positive bacteria (PspA or LiaH), Gram-negative cyanobacteria (PspA) and archaea (PspA) (Fig. 14.4b) have very similar general HD1–HD4 structures and are also implicated in plasma membrane protection upon extracytoplasmic IM stress conditions including antibiotic resistance of *Bacillus* (Joly et al. 2010). These PspA orthologues retained membrane specific ancestral function, but the vesicle inducing protein in plastids (VIPP1), the PspA orthologues in cyanobacteria, green algae and higher plants (Fig. 14.4b); besides, the ancestral structure function has an additional domain HD5. This domain is probably responsible for specialisation of function of VIPP1 which is involved in thylakoid membranes protection/biogenesis and photosynthesis (Votknecht et al. 2012; Zhang and Sakamoto 2013). The expression of these proteins is regulated by means of difference from PspA in enterobacteria and there is no PspF homologue/orthologue present. Although these proteins have slightly different substructure within HD1–HD4(5), all of them possess the PspA ahA- and

ahB-like amino acids sequences and conserved residue P25. Similar to PspA, ahA is the VIPP1 IM-binding determinant and LiaH and VIPP1 have been shown to form the IM-associated high-order oligomeric effectors (Wolf et al. 2010; Otters et al. 2013; Dominguez-Escobar et al. 2014).

The archaeal PspA orthologue is probably the result of acquisition of bacterial *pspA* gene, and phylogenetic analyses show that VIPP1 very likely has a prokaryotic origin as well. The specialisation of VIPP1 may be connected to the evolution of PspA to oxygenic photosynthesis. The plants contain only VIPP1 but many cyanobacteria carry both *pspA* and *vipp1* genes (Fig. 14.4b) indicative of a requirement for both proteins. Indeed, PspA in cyanobacteria is implicated in maintenance of the protein translocation defects, while VIPP1 function is related to thylakoid membranes protection and biogenesis. The *vipp1* and *pspA* genes are adjacent to each other in some cyanobacterial genomes (separated by 162 base pairs in *Anabaena* sp. PCC 7120) and it has been suggested that Vipp1 has evolved in cyanobacteria from PspA by gene duplication (Vothnecht et al. 2012). Passage of *vipp1* to plants then occurred via the cyanobacterial endosymbiotic event, with *pspA* either absent in the original endosymbiont or lost in the plant lineage shortly after the event.

Similar to the situation with Gram-negative cyanobacteria, some enterobacteria carry both PspA and its homologue YjfJ (Fig. 14.4b). The *E. coli* YjfJ has 232 amino acids with significantly different structural features from PspA. YjfJ is predicted to contain a reduced coiled-coil structure (residues 28-83/90-117), different ahA and no P25 (true for all YjfJ homologues and unique for all PspA homologues), while the ahB is similar to PspA. Accordingly, purified YjfJ is an exclusive low-order oligomer and YjfJ exhibits the residual inhibition of *psp* transcription in vivo both of which depend on the presence of ahB (Jovanovic et al. 2014A; GJ, MB unpublished). *yjfJ* gene is part of *yjfIJ* cluster positioned at 95 centisom (*pspA* is at 29.45 and *pspG* at 91.84 centisom positions, see Sect. 14.2) and adjacent to the *yjfKLMC* operon. The YjfJ effect on *psp* expression and gene positioning on the chromosome might have significance for a functional relationship with the PspF regulon and YjfKM proteins, namely the YjfKM proteins coordinately act with the PspF regulon to handle the metabolic shutdown caused by membrane stress and pmf dissipation in *tolC* mutants (Dhamdhere and Zgurskaya 2010). This suggests that now the YjfJ (and YjfI) might be a part of the *yjfKLMC* operon stress response and in the past it may have been the member of the PspF regulon in enterobacteria (Fig. 14.4a).

Significantly, the effects of selection seem to extend well beyond the *psp* gene-coding sequences. The *psp* sigma54 promoter sequences bear GT (or GA) nucleotides at position-12 instead of GC consensus, and this is conserved signature of *psp* sigma54 promoter(s) (Huvet et al. 2011) (Fig. 14.5a). Apparently, in *E. coli* several of the *psp* regulatory sequence features exist in PspF and the sigma54-independent regulatory region driving the expression of the YjfI–YjfJ. The inspection of *yjfIJ* regulatory region nucleotide sequence revealed the presence of *psp* sigma54 signature promoter possibly inactivated by the addition of two nucleotides (see Fig. 14.5a). Moreover, a part of the PspF-binding enhancer UAS II is present at exact position ~130 nucleotides upstream from the inactivated

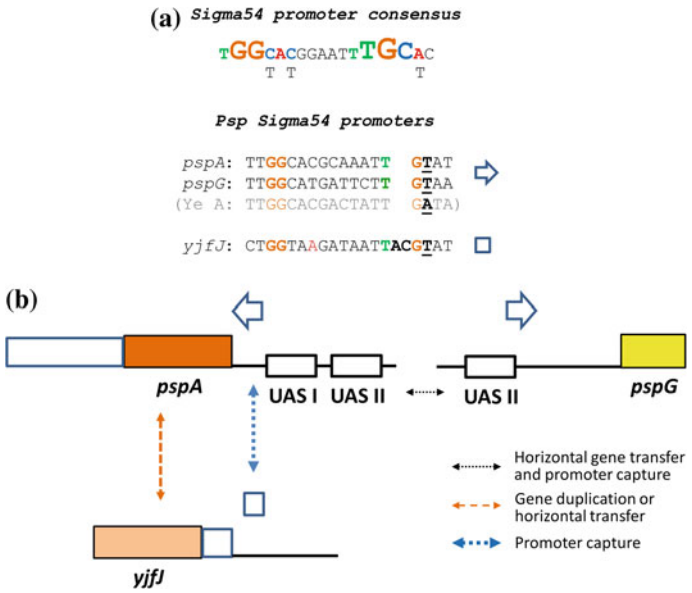


Fig. 14.5 Capture and loss of *psp* sigma54 promoter control. **a** The active (*open arrow*) *pspA* and *pspG* and inactive (square) *yjfIJ* sigma54 promoter sequences compared to the consensus (with most often substitutions) sigma54 promoter sequence, -12 GC and -24 GG being virtually 100 % conserved. Ye A represents the *pspA* sigma54 promoter in *Yersinia enterocolitica*. **b** The model of promoter order strategy potentially leading to reduction in control of the extended PspF regulon (see text for details)

sigma54 promoter as in *pspA* or *pspG* regulatory region. Therefore, YjfI–YjfJ expression was potentially regulated by sigma54 and PspF, and then, this control was switched off (see Fig. 14.5b). Consistently, the *yjfIJ* transcription is not affected upon IM stress, otherwise inducing the sigma54-dependent transcription of the PspF regulon (Lloyd et al. 2004).

14.8 Promoter Order Strategy and Loss of a PspF Regulatory Circuit

Based on the information presented above and considering the current concept of evolution of bacterial regulatory circuits (Perez and Groisman 2009), the potential model for the role of promoter order strategy in evolution of the PspF regulon in enterobacteria is presented in Fig. 14.5b. Horizontal transfer of genes (75 % of genes) can shape the evolution of bacteria. The *pspG* may as a newly acquired gene by horizontal gene transfer capture the *psp* operon sigma54 control elements and be incorporated into a Psp ancestral regulatory circuit leading to PspF regulon formation. The restructuring of the *psp* activator DNA-binding sites (e.g. deletion of

UAS I enhancer and retain of UASII and IHF-binding site) allows regulatory flexibility and balanced PspF regulon control (Fig. 14.5b). This can enable a modulation to IM stress response but may also cause the variable order of promoter use generating heterogeneity of gene expression and stress response across population and flexibility in regulon membership depending on growth conditions (e.g. *pspG*, see Sect. 14.6). This signal-dependent indeterminacy that can produce the phenotypic differences between cells is distinct from the stochastic noise in individual cell.

Further expansion of the PspF regulon might occur either by horizontal transfer of *yjfIJ* cluster or by *pspA* gene duplication any of which can be followed by *psp* sigma54 promoter/UAS II capture (Fig. 14.5b). Since YjfJ is a PspA homologue and YjfI is a conserved protein that interacts with glycerol-3-phosphate dehydrogenase and stationary phase-related regulators, including sigma54 activator RavA, it is more likely that enterobacteria acquired *yjfIJ* by horizontal gene transfer as a unit. Both *pspG* and *yjfIJ* horizontally acquired genes may be embedded under regulation of an ancestral transcription factor to associate with respiration control and strengthen up the adaptation to stress under simultaneous changes in metabolism developed upon IM stress and PspA effector actions. The IM stress besides membrane damage causes redox changes and increases reducing capacity of the cytoplasm and so PspG complements PspA effector functions and mainly acts by modulating the respiration upon IM stress (see Sects. 14.3 and 14.6). The *yjfJ* is associated with *yjfKM* genes involved in response to metabolic shutdown (increase in reducing capacity and high NADH/NAD^+ ratio) following the membrane stress and drop in pmf that simultaneously induce PspA (see Sect. 14.7).

Considering the potential overlap in PspA and YjfJ functions and the fact that a promoter order strategy might formally be in action through diverse PspF DNA-binding UAS/enhancer control elements in *pspA*, *pspG* and *yjfIJ* regulatory regions, we speculate that two different scenarios may determine the fate of the PspF regulon structure. In the first one, the *psp* promoter captured by *yjfIJ* first alters the differential expression pattern of PspF regulon control; then, due to differences *in cis*-regulatory elements and the consequent infrequent binding of PspF and promoter activation, the regulatory region sequence of *yjfIJ* is changing which is a prelude for YjfIJ losing the ancestral sigma54 control (see Fig. 14.5b). The predominant growth conditions and cell fitness will determine whether this can lead to loss of the *yjfIJ* cluster or to a subsequent change in YjfJ protein function finally resulting in transcriptional rewiring of *yjfIJ* to *yjfKM* regulatory circuit. Alternatively, the second scenario assumes that PspA paralog YjfJ which adopts the Psp *cis*-regulatory elements already carries the functional domain innovations more compatible with metabolic shutdown stress control than specifically with the IM stress. The accumulation of beneficial mutations can lead to protein function divergence and partial segregation in function from PspA shaping YjfJ function towards specialisation to support the actions of YjfKM in metabolic shutdown maintenance. These may facilitate the changes in *yjfIJ cis*-control elements and in concert with promoter order strategy cause the gene regulation divergence from PspF regulon control and rewiring. To discriminate between the impacts on changes

in regulatory region versus protein function of YjfJ, one possible approach would be to either swap the YjfJ and PspA regulatory region or proteins and test the outcome, i.e. strain fitness after prolonged growth under selective pressure of IM stress or metabolic shutdown in *tolC* mutants.

Therefore, the promoter order strategy may either contributes in variability of the *psp* expression and stress response or in redesigning the PspF regulatory circuit leading to regression of control and coincidental loss of PspF regulon members in some enterobacterial genomes.

14.9 Conclusions

It has been shown recently that the extensive *cis*-regulatory rewiring implicating evolutionary innovations and loss may be a critical driving force to separate the mouse and human genomes (Vierstra et al. 2014). The changes in nucleotide sequence or numbers of the *cis*-regulatory sequences could as well play a central role in establishing developmental differences between bacterial strains (Stern 2000; Carroll 2005; Loehlin and Werren 2012). The promoter order strategy dependent on number of *cis*-control elements and the intracellular concentration of a master regulator is described here as a potential factor in bacterial regulon evolution. This strategy might be a general feature of stress responding regulons controlled by a limiting amount of an activator and may occur when a regulon expands. The single regulatory region occupancies within the regulon may cause the variability in gene expression control between cells under stress leading to heterogeneity among population of cells and/or control regression within the regulon and loss of genes. Specifically, the non-simultaneous promoter use in IM stress responding PspF regulon can affect the bacterial cell fitness with direct implications in enterobacterial cell adaptation upon biofilm formation, virulence and the multi-drug resistance of persister cells.

It seems that multi-factorial control of gene expression in bacteria can be exploited as a selective pressure target in many different ways depending upon the particular dynamic environmental conditions. The changes in the *cis*-regulatory sequences and promoter order strategy may act synergistically to drive bacterial regulon evolution and contribute in gene regulation divergence and the subsequent change in protein functions and microevolution of chromosomes leading towards differentiation of species.

Acknowledgements The work conducted in our laboratory has been supported by the Wellcome Trust and Biotechnology and Biological Sciences Research Council (BBSRC) research grants and is currently funded by The Leverhulme Trust research project grant (RPG-2012-705).

References

Arias-Cartin R, Grimaldi S, Pommier J, Lanciano P, Schaefer C, Arnoux P, Giordano G, Guigliarelli B, Magalon A (2011) Cardiolipin-based respiratory complex activation in bacteria. Proc Natl Acad Sci USA 108:7781–7786

Buck M, Bose D, Burrows PC, Cannon W, Joly N, Pape T, Rappas M, Schumacher J, Wigneshweraraj S, Zhang X (2006) A second paradigm for gene activation in bacteria. Biochem Soc Trans 34:1067–1071

Bury-Mone S, Nomane Y, Reymond N, Barbet R, Jacquet E, Imbeaud S, Jacq A, Bouloc P (2009) Global analysis of extracytoplasmic stress signalling in *Escherichia coli*. PLoS Genet 5: e1000651

Bush M, Dixon R (2012) The role of bacterial enhancer binding proteins as specialized activators of σ^{54}-dependent transcription. Microbiol Mol Biol Rev 76:497–529

Carroll SB (2005) Evolution at two levels: on genes and form. PLoS Biol 3:e245

Cheng H, Donahue JL, Battle SE, Ray WK, Larson TJ (2008) Biochemical and genetic characterization of PspE and GlpE, two single-domain sulfurtransferases of *Escherichia coli*. Open Microbiol J 2:18–28

Darwin AJ (2005) The phage-shock-protein response. Mol Microbiol 57:621–628

Darwin AJ (2013) Stress relief during host infection: the phage shock protein response supports bacterial virulence in various ways. PLoS Pathog 9:e1003388

Dhamdhere G, Zgurskaya HI (2010) Metabolic shutdown in *Escherichia coli* cells lacking the outer membrane channel TolC. Mol Microbiol 77:743–754

Domínguez-Escobar J, Wolf D, Fritz G, Höfler C, Wedlich-Söldner R, Mascher T (2014) Subcellular localization, interactions and dynamics of the phage-shock protein-like Lia response in *Bacillus subtilis*. Mol Microbiol 92:716–732

Engl C, Jovanovic G, Lloyd LJ, Murray H, Ying L, Errington J, Buck M (2009) In vivo localizations of membrane stress controllers PspA and PspG in *Escherichia coli*. Mol Microbiol 73:382–396

Engl C, Beek AT, Bekker M, de Mattos JT, Jovanovic G, Buck M (2011) Dissipation of proton motive force is not sufficient to induce the phage shock protein response in *Escherichia coli*. Curr Microbiol 62:1374–1385

Flores-Kim J, Darwin AJ (2012) Phage shock protein C (PspC) of *Yersinia enterocolitica* is a polytopic membrane protein with implications for regulation of the Psp stress response. J Bacteriol 194:6548–6559

Francke C, Kormelink TG, Hagemeijer Y, Overmars L, Sluijter V, Moezelaar R, Siezen RJ (2011) Comparative analyses imply that the enigmatic sigma factor 54 is a central controller of the bacterial exterior. BMC Genom 12:385–405

Horstman NK, Darwin AJ (2012) Phage shock proteins B and C prevent lethal cytoplasmic membrane permeability in *Yersinia enterocolitica*. Mol Microbiol 85:445–460

Huvet M, Toni T, Tan H, Jovanovic G, Engl C, Buck M, Stumpf MPH (2009) Model-based evolutionary analysis: the natural history of phage-shock stress response. Biochem Soc Trans 37:762–767

Huvet M, Toni T, Sheng X, Thorne T, Jovanovic G, Engl C, Buck M, Pinney JW, Stumpf MPH (2011) The evolution of the phage shock protein response system: interplay between protein function, genomic organization, and system function. Mol Biol Evol 28:1141–1155

Joly N, Burrows PC, Engl C, Jovanovic G, Buck M (2009) A lower-order oligomer form of phage shock protein A (PspA) stably associates with the hexameric AAA^+ transcription activator protein PspF for negative regulation. J Mol Biol 34:764–775

Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MPH, Buck M (2010) Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol Rev 34:797–827

Joly N, Zhang N, Buck M, Zhang X (2012) Coupling AAA protein function to regulated gene expression. Biochim Biophys Acta 1823:108–116

Jovanovic G, Weiner L, Model P (1996) Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the Escherichia coli stress-induced psp operon. J Bacteriol 178:1936–1945

Jovanovic G, Model P (1997) PspF and IHF bind co-operatively in the *psp* promoter-regulatory region of *Escherichia coli*. Mol Microbiol 25:473–481

Jovanovic G, Dworkin J, Model P (1997) Autogenous control of PspF, a constitutively active enhancer-binding protein of *Escherichia coli*. J Bacteriol 179:5232–5237

Jovanovic G, Rakonjac J, Model P (1999) In vivo and in vitro activities of the *Escherichia coli* sigma54 transcription activator, PspF, and its DNA-binding mutant, PspFΔHTH. J Mol Biol 285:469–483

Jovanovic G, Lloyd LJ, Stumpf MPH, Mayhew AJ, Buck M (2006) Induction and function of the phage shock protein extracytoplasmic stress response in *Escherichia coli*. J Biol Chem 281:21147–21161

Jovanovic G, Engl C, Buck M (2009) Physical, functional and conditional interactions between ArcAB and phage shock proteins upon secretin-induced stress in *Escherichia coli*. Mol Microbiol 74:16–28

Jovanovic G, Engl C, Mayhew AJ, Burrows PC, Buck M (2010) Properties of the phage-shock-protein (Psp) regulatory complex that govern signal transduction and induction of the Psp response in *Escherichia coli*. Microbiol (SGM) 156:2920–2932

Jovanovic M, James EH, Burrows PC, Rego FGM, Buck M, Schumacher J (2011) Regulation of the co-evolved HrpR and HrpS AAA+ proteins required for *Pseudomonas syringae* pathogenicity. Nat Commun 2:177. doi:10.1038/ncomms1177

Jovanovic G, Mehta P, McDonald C, Davidson AC, Uzdavinys P, Ying L, Buck M (2014a) The N-terminal amphipathic helices determine regulatory and effector functions of phage shock protein A (PspA) in *Escherichia coli*. J Mol Biol 426:1498–1511

Jovanovic G, Mehta P, Ying L, Buck M (2014b) The anionic lipids and cytoskeletal proteins MreB and RodZ define the spatial distribution and functioning of membrane stress controller PspA in *Escherichia coli*. Microbiol (SGM) 160:2374–2386

Kobayashi R, Suzuki T, Yoshida M (2007) *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol Microbiol 66:100–109

Lenn T, Gkekas CN, Bernard L, Engl C, Jovanovic G, Buck M, Ying L (2011) Measuring the stoichiometry of functional PspA complexes in living bacterial cells by single molecule photobleaching. Chem Comm 47:400–402

Liu A, Tran L, Becket E, Lee K, Chinn L, Park E, Tran K, Miller JH (2010) Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. J Bacteriol 54:1393–1403

Lloyd LJ, Jones SE, Jovanovic G, Gyaneshwar P, Rolfe MD, Thompson A, Hinton JC, Buck M (2004) Identification of a new member of the phage shock protein response in *Escherichia coli*, the phage shock protein G (PspG). J Biol Chem 279:55707–55714

Loehlin DW, Werren JH (2012) Evolution of shape by multiple regulatory changes to a growth gene. Science 335:943–947

Martinez-Argudo I, Little R, Shearer N, Johnson P, Dixon R (2004) The NifL-NifA system: a multidomain transcriptional regulatory complex that integrates environmental signals. J Bacteriol 186:601–610

Maxson ME, Darwin AJ (2006) Multiple promoters control expression of the *Yersinia enterocolitica* phage-shock-protein A (*pspA*) operon. Microbiol (SGM) 152:1001–1010

McDonald CL, Jovanovic G, Ces O, Buck M (2015) Membrane stored curvature elastic stress modulates recruitment of maintenance proteins PspA and Vipp1. mBio (Under Review)

Mehta P, Jovanovic G, Lenn T, Bruckbauer A, Engl C, Ying L, Buck M (2013) Localisations, dynamics and stoichiometry of a regulated enhancer-binding protein in live *Escherichia coli* cells. Nat Commun 4:1997. doi:10.1038/ncomms2997

Mehta P, Jovanovic G, Ying L, Buck M (2015) Is the cellular and molecular machinery docile in the stationary phase of *Escherichia coli*? Biochem SocTrans (In press)

Model P, Jovanovic G, Dworkin J (1997) The *Escherichia coli* phage-shock-protein (*psp*) operon. Mol Microbiol 24:255–261

Otters S, Braun P, Hubner J, Wanner G, Vothknecht UC, Chigri F (2013) The first α-helical domain of the vesicle-inducing protein in plastids 1 promotes oligomerization and lipid binding. Planta 237:529–540

Parry BR, Surovtsev IV, Cabeen MT, O'Hern CS, Daufresne ER, Jacobs-Wagner C (2014) The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. Cell 156:183–194

Perez JC, Groisman EA (2009) Evolution of transcriptional regulatory circuits in bacteria. Cell 138:233–244

Reyes-Dominguez Y, Contreras-Ferrat G, Ramírez-Santos J, Membrillo-Hernández J, Gómez-Eichelmann MC (2003) Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. J Bacteriol 185:1097–1100

Rowley G, Spector M, Kormanec J, Roberts M (2006) Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. Nat Rev Microbiol 4:383–394

Seo J, Savitzky DC, Ford E, Darwin AJ (2007) Global analysis of tolerance to secretin-induced stress in *Yersinia enterocolitica* suggests that the phage-shock-protein system may be a remarkably self-contained stress response. Mol Microbiol 65:714–727

Shingler V (2011) Signal sensory systems that impact σ^{54}-dependent transcription. FEMS Microbiol Rev 35:425–440

Stern DL (2000) Perspective: evolutionary developmental biology and the problem of variation. Evolution 54:1079–1091

Toni T, Jovanovic G, Huvet M, Buck M, Stumpf MPH (2011) From qualitative data to quantitative models: analysis of the phage shock protein stress response in *Escherichia coli*. BMC Syst Biol 5:69. doi:10.1186/1752-0509-5-69

Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS et al (2014) Mouse regulatory DNA landscapes reveal global principles of *cis*-regulatory evolution. Science 346:1007–1012

Vothknecht UC, Otters S, Hennig R, Schneider D (2012) Vipp1: a very important protein in plastids?! J Exp Bot 63:1699–1712

Wolf D, Kalamorz F, Wecke T, Juszczak A, Mader U, Homuth G, Jordan S, Kirstein J, Hoppert M, Voigt B, Hecker M, Mascher T (2010) In-depth profiling of the LiaR response of bacillus subtilis. J Bacteriol 192:4680–4693

Yamaguchi S, Gueguen E, Horstman NK, Darwin AJ (2010) Membrane association of PspA depends on activation of the phage-shock-protein response in *Yersinia enterocolitica*. Mol Microbiol 78:429–443

Yamaguchi S, Dylan AR, Rothenberg E, Darwin AJ (2013) Changes in Psp protein binding partners, localization and behavior upon activation of the *Yersinia enterocolitica* phage shock protein response. Mol Microbiol 87:656–671

Zhang L, Sakamoto W (2013) Possible function of VIPP1 in thylakoids. Plant Signal Behav 8: e22860

Zhang N, Simpson T, Lawton E, Uzdavinys P, Joly N, Burrows PC, Buck M (2013) A key hydrophobic patch identified in an AAA+ protein essential for in trans inhibitory regulation. J Mol Biol 425:2656–2669

Part IV
Speciation and Biodiversity

Chapter 15
Genomic Admixture and Species Delimitation in Forest Trees

Amanda R. De La Torre

Abstract The study of natural hybrid zones can provide insights into the interplay of gene flow and divergent selection in the maintenance of species barriers between hybridizing species. In tree species, porous genomes and incomplete barriers to reproduction may allow neutral and selectively advantageous loci to freely cross species barriers, whereas loci under divergent selection or linked to those under selection will be retained, contributing to reproductive isolation and the maintenance of species barriers. Tree species are characterized by their long-generation times, outcrossing mating systems, and effective seed and pollen dispersal. Several of these features have certainly influenced the genetic structure and the patterns of hybridization we see nowadays. Hybrid zones of tree species may present some distinctive features in relation to hybrid zones in other taxa, such as the widespread occurrence of exogenous selection and environmental-dependent hybrid zones, their ancient nature, the asymmetry of introgression, and the permeability of tree genomes. This chapter aims to review the recent literature in hybridization and hybrid zones studies in forest tree species, and to discuss the implications of these studies for the study of evolution.

15.1 Introduction

The maintenance of species differences in the face of gene flow is of great biological interest (Minder and Widmer 2008). Gene flow between species may allow the spread of neutral and advantageous alleles across genomes unless the alleles are tightly linked to loci that contribute to reproductively isolation (Barton 1979). In hybridizing species showing differential adaptations, divergent selection may act on particular genes preventing introgression in surrounding genomic regions. Species

A.R. De La Torre (✉)
Umeå Plant Science Centre, Umeå, Sweden
e-mail: amandarodltc@gmail.com

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_15

boundaries are then maintained despite hybridization and introgression (Andrew and Rieseberg 2013).

Natural hybrid zones, which are areas in which two species meet, mate, and produce hybrids, can be seen as laboratories for evolutionary studies because they provide insights into the interplay between natural selection and gene flow among divergent populations (Barton and Hewitt 1989; Harrison 1986, 1993). In contrast to experimental hybridization, natural hybrid zones contain a wide variety of genotypes that may result from many generations of hybridization and recombination, offering a unique opportunity to study the evolutionary processes that influence interactions between hybridizing species.

Natural hybrid zones are particularly important in the study of hybridization in forest tree species. In contrast with annual plant species, in which tests of experimental hybridization are common, the formation of multiple generations of artificial crosses in tree species is largely impractical due to their long-generation times. Therefore, the use of natural hybrid zones provides a good opportunity to investigate the role of the opposing forces of gene flow and selection in the maintenance of species barriers in natural settings. In addition, the identification of genes involved in reproductive barriers between species, or conferring a selective advantage in some environments, will result in a better understanding of the genomic basis of local adaptation.

Tree species are characterized by their long-generation times, outcrossing mating systems, and effective seed and pollen dispersal. In addition, many temperate trees have wide distribution areas, occurring across a diverse range of climate and ecotypes. Several of these features have certainly influenced the genetic structure and the patterns of hybridization we see nowadays. Hybrid zones of tree species may present some distinctive features in relation to hybrid zones in other taxa. From these, the widespread occurrence of exogenous selection and environmental-dependent hybrid zones, the ancient nature of many hybrid zones, the asymmetry of introgression, and the permeability of tree genomes seem to be the most prominent (Table 15.1). This chapter aims to review the recent literature in hybridization and hybrid zones studies in forest tree species, and to discuss the implications of these studies for the study of evolution.

15.2 Gene Flow, Species Delimitation, and Species Concepts

The nature of species differences is of great biological interest (Minder and Widmer 2008). Identifying species as independent groups is a key step in conservation, genetics, taxonomy, and ecology. Under the biological species concept (Mayr 1942), species are "groups of interbreeding natural populations that are reproductively isolated from other groups." According to this concept, species and thus their genomes, behave as cohesive or coadaptive genetic units, suggesting that species

Table 15.1 Hybrid zones studies in tree species since year 2005

Species	Location	Individuals/markers	Hybrid zone distribution	Ecological differences	Hybrid classes	Hybrid zone maintenance model	Fitness/RI tests	Reproductive barriers	Asymmetric introgression	Reference
Picea										
P. engelmanniii × *P. glauca*	SW USA	336 individuals, 17 nSSR	Parapatric	Not stated	High introgression, several hybrid classes	Mosaic zone	No	Temporal isolation (not tested)	Yes	Haselhorst and Buerkle (2013)
P. engelmanniii × *P. glauca*	SW North America	15498 individuals (fitness analysis), 86 SNPs	Parapatric	Drought tolerance, length of growing seasons, snow precipitation	High introgression, several hybrid classes	Bounded hybrid Superiority	Yes	Temporal isolation, postzygotic barriers	Yes	De La Torre et al. (2014a)
P. engelmanniii × *P. glauca*	SW North America	745 individuals, 384 SNPs	Parapatric	Drought tolerance, length of growing seasons, snow precipitation	High introgression, several hybrid classes	Bounded hybrid Superiority	No	Temporal isolation, postzygotic barriers	Yes	De La Torre et al. (2014b)
P. engelmanniii × *P. glauca*	SW North America	805 individuals, 86 SNPs, 10 nSSR	Parapatric	Drought tolerance, length of growing seasons, snow precipitation	High introgression, several hybrid classes	Bounded hybrid Superiority	No	Temporal isolation, postzygotic barriers	Yes	De La Torre et al. (2015)
P. engelmannii, *P. glauca*, *P. sitchensis*	SW North America	1516 individuals, 71 SNPs	Parapatric and sympatric	Precipitation	High introgression, several hybrid classes	Not stated	No	Not stated	Yes	Hamilton, De la Torre and Aitken (2015)
P. sitchensis × *P. glauca*	SW North America	721 individuals, 268 SNPs	Sympatric	Precipitation	High introgression, several hybrid classes	Mosaic zone	No	Not stated	Yes	Hamilton et al. (2013)
P. likiangensis × *P. purpurea*	China	39 populations, 2 mtDNA, 3cpDNA	Sympatric	Not stated	Not stated	Not stated	No	Not stated	Yes (only mtDNA)	Du et al. (2011)
Pinus										
P. massoniana × *P. hwangshanensis*	China	270 individuals, 14 EST-SSR	Parapatric	Temperature, soil moisture, moisture	Mainly backcrosses	Not stated	No	Germination rates (not tested)	Yes	Zhang et al. (2014)
P. contorta × *P. banksiana*	Canada	536 individuals, 1536 SNP	Sympatric	Temperature, soil moisture, drought	High introgression, several hybrid classes	Mosaic zone	No	Not stated	Not stated	Cullingham et al. (2013)

(continued)

Table 15.1 (continued)

Species	Location	Individuals/markers	Hybrid zone distribution	Ecological differences	Hybrid classes	Hybrid zone maintenance model	Fitness/RI tests	Reproductive barriers	Asymmetric introgression	Reference
Picea										
P. contorta × *P. banksiana*	Canada	1998 individuals, 11 nSSR	Sympatric	Temperature, soil moisture, drought	High introgression, several hybrid classes	Mosaic zone	No	Not stated	Unequal ancestry despite equal gene flow	Cullingham et al. (2012)
Juniper										
J. tibetica, *J. saltuaria*, *J. convallium*, *J. przewalskii*	China	53 populations, 13 nuclear loci	Parapatric and sympatric	Not stated	Not stated	Not stated	No	Not stated	Yes	Li et al. (2010)
Populus										
P. alba × *P. tremula*	Europe and Northern Africa	414 individuals, PCR-RFLPs in 6 cpDNA	Parapatric	Not stated	Variable introgression	Not stated	No	Not stated	Not stated	Fussi et al. (2010)
P. alba × *P. tremula*	Italy, Austria and Hungary	693 individuals, 77 nSSR	Sympatric	Habitat type (lowland vs upland)	Mainly backcrosses and advanced-generation hybrids	Mosaic zone	No	Postzygotic	Not stated	Lindtke et al. (2012)
P. alba × *P. tremula*	Italy, Austria and Hungary	692 individuals, 93 SSR	Sympatric	Habitat type (lowland vs upland)	Mainly backcrosses and advanced-generation hybrids	Mosaic zone	No	Premating: temporal isolation and assortative mating, postzygotic		Lexer et al. (2010)
P. alba × *P. tremula*	Italy	621 individuals, 11976 SNPs	Sympatric	Habitat type (lowland vs upland)	Mainly backcrosses and advanced-generation hybrids	Mosaic zone	Yes (germination and survival)	Postzygotic: reduced hybrid viability	Yes	Lindtke et al. (2014)
P. balsamifera × *P. deltoides*	Canada	142 individuals, 36 SNPs, 2 cpDNA	Parapatric	Habitat type (lowland vs upland)	Low introgression, mostly F1s	Not stated	Yes	Pre and postzygotic: hybrid viability, germination rates	Yes	Roe et al. (2014)
P. balsamifera, *P. deltoides*, *P. nigra*	Canada	635 individuals, 35 SNPs	Parapatric	Habitat type (lowland vs upland)	Low introgression, mostly F1s	Not stated	No	Pre and postzygotic	Yes	Thompson et al. (2010)
P. deltoides, *P. fremontii*, *P. angustifolia*	Colorado, USA	270 individuals, 26 nSSR, 5 cpDNA	Parapatric and sympatric	Habitat type (lowland vs upland)	Low introgression, mostly F1s	Tension zone	No	Postzygotic	No	Hersch-Green et al. (2014)

(continued)

Table 15.1 (continued)

Species	Location	Individuals/markers	Hybrid zone distribution	Ecological differences	Hybrid classes	Hybrid zone maintenance model	Fitness/RI tests	Reproductive barriers	Asymmetric introgression	Reference
Picea										
P. euphratica × *P. pruinosa*	Asia–Africa	290 individuals, 2cpDNA, 2nrITS sequence data and 8 nSSR	Sympatric	Drought tolerance	Low hybridization	Not stated	No	Premating: temporal isolation (not tested)	Yes	Wang et al. (2011)
Juglans										
J. nigra × *J. regia*	Italy	483 individuals, 9 nSSR	Sympatric	Not stated	Low hybridization	Not stated	Yes	Premating: temporal isolation, differences in floral size	Yes	Pollegioni et al. (2013)
Eucalyptus										
E. aggregata × *E. rubida*	Southern Australia	412 individuals, 6 nSSR markers	Sympatric	Soil and moisture	F1, F2 and backcrosses	Not stated	No	Premating:style lengths	Yes	Field et al. (2010)
E. aggregata × *E. rubida*	Southern Australia	731 individuals, 6 nSSR markers	Sympatric	Soil and moisture	Not stated	Not stated	No	Premating: temporal isolation and assortative mating	Yes	Field et al. (2010)
E. globulus × *E. cordata*	Tasmania	388 individuals, 542 AFLP, cpDNA	Sympatric	No	F1 × cordata	Not stated	No	Not stated	Not stated	McKinnon et al. (2010)
Fraxinus										
F. excelsior × *F. angustifolia*	Ireland	96 individuals, 6 nSSR	Sympatric	Not stated	Mainly backcrosses	Not stated	No	Not stated	Not stated	Thomasset et al. (2013)
F. excelsior × *F. angustifolia*	Europe	456 individuals, 19 nSSR	Sympatric	Frost days and summer temperature	Not stated	Environment-dependent	No	Premating: temporal isolation (not tested)	Yes	Gerard et al. (2013)
Quercus										
Q. robur × *Q. petraea*	France	855 individuals, 262 SNPs	Sympatric	Late-successional versus pioneer species	Low introgression, several hybrid classes	Mosaic zone	No	Not stated	Yes	Guichoux et al. (2013)
Q. robur × *Q. petraea*	France	256 parents, 3213 offspring, 12-plex EST-SSRs, 384-plex SNPs	Sympatric	Drought tolerance	Low introgression	Not stated	No	Pollen limitation and pollen competition	Yes	Lagache et al. (2013)

(continued)

Table 15.1 (continued)

Species	Location	Individuals/markers	Hybrid zone distribution	Ecological differences	Hybrid classes	Hybrid zone maintenance model	Fitness/RI tests	Reproductive barriers	Asymmetric introgression	Reference
Picea										
Q. robur × *Q. petraea*	France	155 individuals, 6 nSSR	Sympatric	Drought and shade tolerance	Not stated	Not stated	Yes	Prezygotic: pollen germination, pollen tube progress;postzygotic: germination rate and seed production ratio	Yes	Abadie et al. (2011)
Q. robur, Q. petraea, Q. pubescens, Q. pyrenaica	France	208 individuals, 10 nSSR	Sympatric	Drought tolerance, differences on soil type	Mainly backcrosses, with low levels of F1s and 3-way hybrids	Not stated	Yes	Assortative mating, (possibly) germination rate	Yes	Lepais and Gerber (2010)
Q. engelmannii, Q. cornelius-mulleri, Q. berberidifolia, Q. durata, Q. dumosa	California, U.S.A	343 individuals, 9 nSSR	Sympatric	Hybrids and parents occupy different environmental niches	Low introgression, several hybrid classes	Mosaic zone	No	Pre and postzygotic barriers (not tested)	Yes	Ortego et al. (2014)
Q. pyrenaica × *Q. petraea*	Spain	176 adults, 96 offspring, 5 nSSR	Sympatric	Drought tolerance, differences on soil type	Low introgression, several hybrid classes	Not stated	No	Not stated	No	
Q. mongolica × *Q. liaotungensis*	China	1166 individuals, 19 nSSRs, AFLP, 6 cpSSR	Sympatric	Not stated	Variable introgression	Not stated	No	Postzygotic (not tested)	Not stated	Zeng et al. (2011)
Q. robur, Q. petraea, Q. pubescens, Q. frainetto	West-central Romania	269 individuals, 6 nSSR	Sympatric	Not stated	Mainly backcrosses, with low levels of F1s hybrids	Not stated	No	Premating: temporal isolation and assortative mating	Yes	Curtu et al. (2009)

Notes
nSSR—Nuclear microsatellite markers
SNPs—Single nucleotide polymorphism markers
mtDNA—Mitochondrial DNA markers
cpDNA—Chloroplast DNA markers

divergence only occurs through whole-genome isolation. One of the most common criticisms of the biological species concept (BSC) has been that related species rarely show complete reproductive isolation and that natural hybridization is more common than previously thought (Coyne and Orr 2004). Recent studies in hybridizing species showing differential adaptations have shown that divergent selection may act on particular genes and thus prevent introgression in surrounding genomic regions, challenging the view that species differentiation is a genome-wide phenomenon and supporting Wu (2001)'s genetic view of speciation that states that the gene and not the genome is the unit of differentiation (Minder and Widmer 2008; Sambatti et al. 2012; DeFaveri et al. 2013; Martin et al. 2013).

15.2.1 Species Delimitation and Molecular Markers

When selecting molecular markers for delimiting species, researchers have mostly focused on their variability; however, recent studies have shown that focusing in the amount of gene flow and on sex-biased dispersal may result in more accurate species delimitation (Petit and Excoffier 2009). Caution is advised against the use of uniparentally inherited markers for species delimitation when they are inherited only from the least-dispersing sex. Instead, multiple unlinked high gene flow markers are suggested to be the best option to delimitate species (Petit and Excoffier 2009).

15.2.1.1 Mitochondrial Versus Chloroplast DNA Markers in Tree Species Delimitation

In conifer trees, mitochondrial (mtDNA) and chloroplast DNA (cpDNA) experience very different levels of gene flow due to their contrasting modes of inheritance (maternal for mtDNA, gene flow mediated by seeds; paternal for cpDNA, gene flow mediated by pollen) (Du et al. 2011). Their genomes experience very different levels of gene flow because gene flow by seeds is much more restricted than gene flow by pollen (Petit et al. 2005). Recent studies in the Pinaceae taxa (*Picea*, *Pinus* and *Abies)* suggest that mtDNA haplotypes are often shared among related species, matching geographic variation rather than taxonomic status, whereas cpDNA haplotypes are found to be more species-specific, differentiating species or groups of related species (Du et al. 2009, 2011; Bouille et al. 2011). These results are coincident with the hypothesis that the low-dispersed genome of the local species (mtDNA) could be captured by an expanding species through successive backcrossings with first-generation hybrids (Du et al. 2009; Petit and Excoffier 2009; Bouille et al. 2011).

In angiosperm tree species with maternal inheritance of cpDNA such as *Quercus* (Petit et al. 2002, 2003), *Betula* (Palme et al. 2004), and *Eucalyptus* (McKinnon et al. 2001, 2010), extensive, geographically structured sharing of cpDNA have been observed among species. Cases of cpDNA haplotype sharing were also

reported in *Populus* species (Fussi et al. 2010; Wang et al. 2011). The extensive interspecific hybridization found in *Quercus* and *Eucalyptus* was proposed to act as a mechanism of species dispersal, in which an established pioneer species' population could be invaded by a later successional species by pollen swamping (resulting from recurrent and asymmetric hybridization and introgression) (Petit et al. 2003). This mechanism of species dispersal, in which plant species could disperse by pollen in an area previously colonized by a sister species, may also play a significant role in species range shift nowadays, where migration is a way to cope with changing climates and environmental disturbances (Lepais and Gerber 2010).

15.3 Natural Hybridization and Genomic Admixture

Despite the different viewpoints about the significance of hybridization and its role in evolution [from "evolutionary dead end" (Mayr 1942) to "potent evolutionary force" (Arnold 1997)], there is consensus about the frequency and widespread occurrence of hybridization in plant species. The movement of genes between species resulting from hybridization, and hybridization followed by backcrossing (introgression) may have several different outcomes (Rieseberg and Carney 1998). In porous genomes with low reproductive barriers, there may be a balance between hybridization and selection, with only some parts of the genome introgressing between hybridizing populations. As the strength of reproductive barriers increases, there will be larger areas of the genome protected from introgression (Wu et al. 2001; Via et al. 2009). In this scenario, hybrid zones can lead to speciation via "reinforcement" (evolution of premating barriers to gene exchange in response to selection against hybrids) (Servedio and Noor 2003). The opposite scenario occurs when barriers to reproduction break down leading to the fusion of the parental types (Harrison 1993). In the case of expanding species with higher fitness, one species can invade other species' range resulting in the extinction of the other parental form. Finally, hybridization may generate adaptive divergence and novel phenotypes through transgressive segregation or adaptive introgression, creating populations of mixed ancestry that remain distinct from both parental populations and may eventually evolve into new hybrid species (Abbott et al. 2010, 2013). In all cases, the patterns of hybridization currently observed only represent a single snapshot of a complex and continuously changing process, in which complete reproductive isolation may take hundreds or millions of generations to evolve (Abbott et al. 2013).

15.3.1 Natural Hybridization in Tree Species

Natural hybridization is pervasive in tree species, and in contrast with annual plant species, hybridization in trees may be facilitated by their mating systems, which are

predominantly outcrossing because of self-incompatibility mechanisms and severe inbreeding depression; their usually large natural distribution areas; their efficient pollen and seed flow (including long-distance gene flow); and the absence of strong reproductive barriers between closely related species. Recent studies show the prevalence of hybridization in several different taxa such as *Picea* (De La Torre et al. 2014a, b, 2015; Hamilton et al. 2015; Sun et al. 2014; Hasselhorst and Buerkle, 2013), *Pinus* (Zhou et al. 2010; Li et al. 2010; Gao et al. 2012; Cullingham et al. 2013; Zhang et al. 2014), *Populus* (Lexer et al. 2007; Thompson et al. 2010; Lindtke et al. 2014), *Juglans* (Pollegioni et al. 2013), *Fraxinus* (Gerard et al. 2013; Thomasset et al. 2013), *Quercus* (Lepais and Gerber 2010; Zeng et al. 2011; Ortego et al. 2014), *Eucalyptus* (McKinnon et al. 2010), and *Aesculus* (Thomas et al. 2008).

15.3.2 Secondary Contact or Primary Intergradation?

Hybrid zones can be formed as a result of secondary contact between populations that have differentiated in allopatry (Mayr 1942), or as a consequence of environmental gradients and consequent varying selection pressures in sympatry or parapatry (Endler 1977). Differentiating between patterns of primary intergradation or secondary contact may be very difficult because both can produce identical patterns of variation; therefore, we should be cautious when inferring process from pattern (Harrison 1993). In closely related tree species, hybrids may occur at geographical or ecological margins, at intermediate environments between parentals' habitats (parapatric distribution), or may share the range with the parental species (sympatric distribution). Hybrid zones with parapatric distribution extending along elevational gradients are particularly interesting for the study of the genetic basis of local adaptation and the maintenance of species differences in the face of gene flow because adaptive differentiation in temperate tree species usually evolves as a response to environmental gradients along elevation, producing morphological and genetic clines (Abbott and Brennan 2014). Some examples can be found in *Picea* (De La Torre et al. 2014a, b), *Pinus* (Watano et al. 2004; Zhang et al. 2014), *Populus* (Roe et al. 2014; Fussi et al. 2010), and *Quercus* (Albarran-Lara et al. 2010). In all of these studies, parental species occupied different ecological niches characterized by differences mostly associated with temperature and precipitation, but differences in the length of growing seasons, drought tolerance, and soil moisture were also observed. In *Quercus* and *Pinus*, there was a smooth clinal change in morphology associated with environmental gradients along elevation (Watano et al. 2004; Albarran-Lara et al. 2010), making difficult to classify the hybrid zones. In another study of *Pinus* (*Pinus massoniana x P. hwangshanensis*) in China, although parental species can be distinguished on the basis of morphological, cytological, and timber anatomic characteristics and hybrids occupy a distinct ecological and altitudinal niche, increased introgression (probably due to pollen swamping) from *P. hwangshanensis* toward the hybrids has resulted in the genetic fusion of the two groups (Zhang et al. 2014).

Recent studies in North American conifer species suggest there is enough evidence to confirm successive episodes of secondary contact as a consequence of range expansion and contractions in relation with glacial periods (De La Torre et al. 2014b). *Picea glauca* and *P. engelmannii* are closely related species that diverged in allopatry during the Pliocene (5 MYA). During the Last Glacial Maximum, most of Canada and northwestern USA were covered by ice (Laurentide and Cordilleran ice sheets), which led to the displacement of the species south of their present distribution. Both *Picea* species survived the Last Glacial Maximum in two refugia, where they stayed until the temperatures warmed up. Ecological niche modeling combined with fossil and pollen records suggest these species most recently had the potential to come into secondary contact by 21,000 YBP, in the southern Rocky Mountains in Colorado and Wyoming, considerably south of its present distribution. Due to species ranges' expansions and contractions, the species came into contact several times before they reached their current distribution in northwest Canada and USA. Similar patterns of secondary contact were also suggested in *Pinus contorta x P. banksiana* (Cullingham et al. 2012) and in *Picea sitchensis x P. glauca* hybrid zones (Hamilton et al. 2013).

15.3.3 Extension and Direction of Introgression

In permeable genomes exchanging genes, barriers to gene exchange might accumulate when spatial isolation or physical obstacles are present; however, the progress toward complete reproductive isolation will only be possible when associations among the loci that influence isolation build up (Smadja and Butlin 2011; Abbott et al. 2013). This implies that divergent selection may act on particular genes that when associated with other loci prevent introgression in surrounding genomic regions (often called islands, continents, or signatures of divergence). Genes or alleles not affected by selection will freely cross species barriers. Recent studies suggest that genes contributing to reproductive isolation might be widely distributed in plant genomes (Sambatti et al. 2012; De La Torre et al. 2014b; Lindkte et al. 2012). However, information about the size and distribution of these isolating regions in tree species will only come from fine-scale linkage mapping and experimental crossings.

The study of the extent and direction of hybridization and the frequency of hybrid classes can provide insights about the processes that generate and maintain species differences. The direction of introgression is usually asymmetric in tree species hybrid zones due to several pre- and postzygotic barriers. The extent of hybridization varies widely among and within tree species' populations. A group of hybrid zones is composed by mainly F1 hybrids (*Quercus*: Nason et al. 1992; *Populus*: Hersch-Green et al. 2014), suggesting the presence of reproductive barriers and restricted gene flow between parental species. The distribution of this hybrid zone will tend toward a bimodal distribution of hybrid genotypes associated with strong assortative mating or fertilization (prezygotic isolation) but only weakly

with postzygotic isolation (Jiggings and Mallet 2000). Other hybrid zones may contain a very small proportion of F1 hybrids and many backcrosses (*Populus*: Lexer et al. 2005; Lindkte et al. 2014; *Pinus*: Watano et al. 2004; *Eucalyptus*: Field et al. 2010; and *Quercus*: Lepais and Gerber 2010). Finally, in other cases, mostly advanced-generation hybrids are present (*Picea*: De La Torre 2014a, b; Hamilton et al. 2015; *Pinus*: Cullingham et al. 2012; and *Fraxinus:* Thomasset et al. 2013), resulting in broad hybrid swarms with a wide variety of recombinants.

15.4 Environment Role in the Maintenance of Hybrid Zones

Investigating how hybrid zones are maintained is crucial to understand the contrasting effects of gene flow and natural selection in the evolution of the species. Two models explaining hybrid zone maintenance have been suggested. In the environment-independent model, hybrid zones are maintained by a balance between dispersal and selection against hybrids, with selection being independent of the environment (endogenous selection) (Mayr 1942; Barton 1979). Hybrid inferiority occurs as a result of genetic incompatibilities between the two parental species or due to the breakup of coadapted gene complexes that affect fitness traits (Barton and Hewitt 1989). In these models, hybrids are less fit than their parents regardless of location. The tension zone (Barton and Hewitt 1989) is an example of an environment-independent model.

In the environment-dependent model, hybrid zones are maintained through selection gradients due to environmental heterogeneity and hybrid fitness varies with the environment (exogenous selection) (Endler 1973, 1977; Moore 1977; Harrison 1986). The bounded hybrid superiority model (Moore 1977), in which hybrid individuals are fitter than either parental species in environments that are intermediate to the parental habitats, but are less fit than parental species in their respective native habitats, is an example of an environment-dependent model. Another example is the mosaic model (Harrison 1986), in which parental species are distributed in a spatial mosaic rather than a discrete separation of two environments or along a unidirectional environmental gradient. A combination of exogenous and endogenous selection may also occur in both environment-dependent and environment-independent models (Arnold 1997).

An interesting feature about hybrid zones in tree species is the prevalence of environment-dependent hybrid zones. Whereas environment-independent models (tension zones) and hybrid inferiority are common in animals and some plant species, this type of hybrid zones seems to be scarce in tree species. Even when tension zones are suggested, as in the case of *Populus deltoides x P. angustifolia*, exogenous selection and genotype-by-environment interactions are known to influence the structure of the hybrid zone (Hersch-Green et al. 2014). Some reasons underlying this widespread occurrence may include the wide distribution range of some tree

species that comprises an array of different climates and ecological conditions, and the porous nature of their genomes. Usually, parental species occupy different environmental niches and are adapted to different conditions (such as temperature, soil moisture, and precipitation, between others) that occur in sympatry or in parapatry along elevation gradients (De La Torre et al. 2014a; Cullingham et al. 2012; Lindkte et al. 2012). In some cases, hybrids occupy a different environmental niche than those of parental species (De La Torre et al. 2014b; Zhang et al. 2014).

Although several studies have suggested the presence of environmental-dependent hybrid zones such as the mosaic zone (*Quercus*: Guichoux et al. 2013; Ortego et al. 2014; *Populus*: Lidkte et al. 2014; Lexer et al. 2010; *Pinus:* Cullingham et al. 2012; and *Picea*: Hamilton et al. 2013) and the bounded hybrid superiority (De La Torre et al. 2014a), to explain the maintenance of hybrid zones in tree species, very few have performed experimental tests on fitness differentials between hybrids and pure species (common-garden studies), or tests of reproductive barriers. The estimation of lifetime fitness is extremely complicated in tree species due to their long-generation times; therefore, fitness proxies (such as germination, survival, growth, and disease susceptibility) have been used to test the fitness of pure species and hybrids (De La Torre et al. 2014a; Roe et al. 2014; Abadie et al. 2011).

For many tree species, variation in ecologically important traits has long been studied in provenance trials, which are common-garden experiments similar to reciprocal transplantation experiments (Savolainen and Pyhäjarvi 2007). These tests have proven to be useful in the study of local adaptation of tree species; however, very few have been used in hybrid zones studies. A recent study about the parapatric species *P. glauca x P. engelmanniii* in North America has used an extensive data set (15,498 individuals) of growth measurements, phenology, and survival obtained from long-term (25 years) provenance trials. Fitness differentials suggested both pure species and hybrids occupy different environmental niches and are adapted to different climatic conditions, in which key factors for survival are adaptation to the length of the growing seasons and the depth of the snowpack. In this hybrid zone, hybrids are fitter than parental species in intermediate environments but less fit than pure species in their respective habitats, suggesting the bounded hybrid superiority model may be maintaining the hybrid zone (De La Torre et al. 2014a). Considering the extension of many trees hybrid zones, the amount of advanced-generation hybrids and backcrosses, and the low barriers to reproduction, it is reasonable to think that hybrids are usually viable and able to produce offspring, even when postzygotic barriers are present (Lindkte et al. 2014). However, more studies involving common-garden or reciprocal transplantation experiments and experimental crossings are required to understand the factors that have limited or allowed the survival and establishment of hybrid genotypes in hybrid zones.

15.4.1 Studies of Reproductive Barriers

Barriers to gene exchange can be divided in pre- and postzygotic barriers. Prezygotic isolation can arise as a consequence of habitat and temporal isolation, immigrant inviability, and sexual (pollinator) isolation, whereas postzygotic isolation can be a result of intrinsic postzygotic isolation (hybrid inviability and sterility), ecologically dependent postzygotic isolation (ecological inviability and behavioral sterility), and sexual selection against hybrids. Divergent selection on genes affecting ecological traits can be transmitted directly or indirectly to genes causing reproductive isolation. In the former case, speciation is facilitated when reproductive isolation arises as the pleiotropic effect of divergent selection. In the latter, the genetic association between the genes under selection and those causing reproductive isolation is caused by linkage disequilibrium (Rundle and Nosil 2005).

A recent review on reproductive barriers indicates that prezygotic isolation tends to be greater than postzygotic isolation in plant species. Very strong reproductive isolation often requires a combination of barriers, which are found to be very diverse in plant species. The number of loci involved in prezygotic and postzygotic isolating barriers varied among systems and traits, being the loci of large effect (more than 20 % of the phenotypic variance) the more frequent (Lowry et al. 2008). In tree species, prezygotic barriers also seem to be more frequent; however, the limited number of studies testing postzygotic barriers may be just a result of the difficulties in detecting postzygotic isolation once prezygotic barriers are complete (Lindkte et al. 2014). The difficulty of duplicating divergent natural selection under experimental conditions makes the study of ecologically dependent postzygotic isolation almost impossible in long-generation tree species, but feasible in other organisms such as yeast (Dettman et al. 2007).

Various intrinsic and extrinsic isolating barriers have been suggested to explain the widespread occurrence of asymmetric introgression in tree species studies. Premating barriers such as temporal isolation (differences in flowering phenology) can influence the direction of pollen-mediated gene flow in species such as *Quercus* (Curtu et al. 2009) and *Picea* (De La Torre et al. 2014b; Hasselhorst and Buerkle 2013). After pollination, factors such as pollen competition or pollen tube and ovary interactions may favor the formation of hybrids resembling one of the parental species in *Quercus* (Abadie et al. 2011; Lagache et al. 2013). In addition, differences in style length, which can prevent species with the smaller flowers to fertilize the species with the larger flowers, have been reported in *Eucalyptus* (Field et al. 2010). Finally, after the zygote is formed, postzygotic isolation barriers may act upon the viability and fertility of the hybrids. Epistatic interactions may cause segregation distortion in advanced-generation hybrids in *Populus*, causing sterility or lethality of some hybrid classes, which in turn results in asymmetric patterns of introgression (Thomson et al. 2010). Postzygotic barriers may also reduce hybrids' ability to backcross with parental species, process that could also induce asymmetric introgression (Lepais et al. 2013).

A renewed interest in the study of natural hybrid zones has resulted in a rich recent literature on tree species hybridization. In the upcoming years, the wealth of genomic data combined with the recent sequencing and mapping of some tree species genomes will translate in a better understanding on how species barriers are maintained in the face of gene flow and on the role of divergent selection in the evolution of reproductive barriers between closely related species. The study of hybrid zones in tree species will certainly lead to the identification of ecologically important genes conferring a selective advantage in some environments, therefore contributing to dissect the genomic basis of local adaptation. In the face of climate change, information about the genomic composition of hybridizing species would be extremely valuable to define future management and conservation strategies.

References

Abadie P, Roussel G, Dencausse B et al (2011) Strength, diversity and plasticity of postmating reproductive barriers between two hybridizing oak species (*Quercus robur* L. and *Quercus petraea* (Matt) Liebl). J Evol Biol 25:157–173

Abbott RJ, Brennan AC (2014) Altitudinal gradients, plant hybrid zones and evolutionary novelty. Phil Trans R Soc B 369:20130346

Abbott RJ, Hegarty MJ, Hiscock SJ et al (2010) Homoploid hybrid speciation in action. Taxon 59:1375–1386

Abbott R, Albach D, Ansell S et al (2013) Hybridization and speciation. J Evol Biol 26:229–246

Albarran-Lara AL, Mendoza-Cuenca L, Valencia-Avalos S et al (2010) Leaf fluctuating asymmetry increases with hybridization and introgression between *Quercus magnoliifolia* and *Quercus resinosa* (Fagaceae) through an altitudinal gradient in Mexico. Int J Plant Sci 171:310–322

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution 67–9:2468–2482

Arnold ML (1997) Natural hybridization and evolution. Oxford University Press, Oxford

Barton NH (1979) The dynamics of hybrid zones. Heredity 43:341–359

Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. Nature 341:497–503

Bouille M, Senneville S, Bousquet J (2011) Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. Tree Genet Genomes 7:469–484

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates Inc, Massachusetts

Cullingham CI, James PMA, Cooke JEK et al (2012) Characterizing the physical and genetic structure of the lodgepole pine x jack pine hybrid zone: mosaic structure and differential introgression. Evol Appl 5(8):879–891

Cullingham CI, Cooke JEK, Coltman DW (2013) Effects of introgression on the genetic population structure of two ecologically and economically important conifer species: lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*Pinus banksiana*). Genome 56:577–585

Curtu AL, Gailing O, Finkeldey R (2009) Patterns of contemporary hybridization inferred from paternity analysis in a four-oak-species forest. BMC Evol Biol 9:284

De La Torre AR, Wang T, Jaquish B et al (2014a) Adaptation and exogenous selection in the *Picea glauca* x *P. engelmannii* hybrid zone and its implications for forest management under cimate change. New Phytol 201(2):687–699

De la Torre AR, Roberts D, Aitken SN (2014b) Genome-wide admixture and ecological niche modeling reveal the maintenance of species boundaries despite long history of interspecific gene flow. Mol Ecol 23:2046–2059

De La Torre AR, Ingvarsson PK, Aitken SN (2015) Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. Heredity 00:1–12, doi:10.1038/hdy.2015.19

DeFaveri J, Jonsson PR, Merila J (2013) Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. Evolution 67(9):2530–2546

Dettman JR, Sirjusingh C, Kohn LM et al (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature 447:585

Du FK, Petit RJ, Liu JQ (2009) More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. Mol Ecol 18:1396–1407

Du FK, Peng XL, Liu JQ, Lascoux M (2011) Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. New Phytol 192:1024–1033

Endler JA (1973) Gene flow and population differentiation. Science 179:243–250

Endler JA (1977) Geographic variation, speciation and clines. Princeton University Press, Princeton

Field DL, Ayre DJ, Whelan RJ, Young AG (2010) Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare *Eucalyptus aggregate* and common *E. rubida*. Heredity 1–13

Fussi B, Lexer C, Heinze B (2010) Phylogeography of *Populus alba* (L.) and *Populus tremula* (L.) in Central Europe: secondary contact and hybridization during recolonization from disconnected refugia. Tree Genet Genomes 6:439–450

Gao J, Wang B, JF M, Ingvarsson P et al (2012) Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. Mol Ecol 21(19):4811–4827

Gerard PR, Temunovic M, Sannier J et al (2013) Chilled but not frosty: understanding the role of climate in the hybridization between the Mediterranean *Fraxinus angustifolia* Vahl and the temperate *Fraxinus excelsior* L. (Oleaceae) ash trees. J Biogeogr 40:835–846

Guichoux E, Garnier-Gere P, Lagache L et al (2013) Outlier loci highlight the direction of introgression in oaks. Mol Ecol 22:450–462

Hamilton JA, Lexer C, Aitken SN (2013) Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis x P. glauca*) hybrid zone. New Phytol 197(3):927–938

Hamilton JA, De La Torre AR, Aitken SN (2015) Fine-scale environmental variation contributes to introgression in a three-species spruce hybrid zone. Tree Genet Genomes 11:817

Harrison RG (1986) Pattern and process in a narrow hybrid zone. Heredity 56:337–349

Harrison RG (1993) Hybrids and hybrid zones: historical perspective. In: Harrison RG (ed) Hybrid zones and the evolutionary process. Oxford University Press, Oxford, pp 3–12

Haselhorst MSH, Buerkle CA (2013) Population genetic structure of *Picea engelmannii, P. glauca* and their previously unrecognized hybrids in the central rocky mountains. Tree Genet Genomes 9(3):669–681

Hersch-Green EI, Allan GJ, Whitman TG (2014) Genetic analysis of admixture and patterns of introgression in foundation cottonwood trees (Salicaceae) in southwestern Colorado, USA. Tree Genet Genomes 10:527–539

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. TREE 15:250–256

Lagache L, Klein EK, Guichoux E, Petit RJ (2013) Fine-scale environmental control of hybridization in oaks. Mol Ecol 22:423–436

Lepais O, Gerber S (2010) Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. Evolution 65–1:156–170

Lepais O, Roussel G, Hubert F, Kremer A, Gerber S (2013) Strength and variability of postmating reproductive isolating barriers between four European white oak species. Tree Genet Genomes 9:841–853

Lexer C, Fay MF, Joseph JA, Nica MS, Heinze B (2005) Barrier to gene flow between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Mol Ecol 14:1045–1057

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay M (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. Heredity 98:74–84

Lexer C, Joseph JA, van Loo M, Barbara T, Heinze B, Bartha D, Castiglione S, Fay MF, Buerkle CA (2010) Genomic admixture analysis in European Populus spp. reveals unexpected patterns of reproductive isolation and mating. Genet 186:699–712

Li Y, Stocks M, Hemmilä S, Källman T et al (2010) Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. Mol Biol Evol 27(5):1001–1014

Lindkte D, Buerkle CA, Barbara T, Heinze B et al (2012) Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. Mol Ecol 21(20):5042–5058

Lindkte D, Gompert Z, Lexer C, Buerkle CA (2014) Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. Mol Ecol 23:4316–4330

Lowry DB, Modliszewski JL, Wright KM et al (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. Phil Trans R Soc B 363:3009–3021

Martin SH, Dasmahapatra KK, Nadeau NJ et al (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res 23:1817–1828

Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York

McKinnon GE, Vaillancourt RE, Jackson HD, Potts BM (2001) Chloroplast sharing in the Tasmanian eucalypts. Evolution 55:703–711

McKinnon GE, Smith JJ, Potts BM (2010) Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange between two eucalypt species. Mol Ecol 19:1367–1380

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Mol Ecol 17:1552–1563

Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. Q Rev Biol 52:263–277

Nason JD, Ellstrand NC, Arnold ML (1992) Patterns of hybridization and introgression in populations of oaks, manzanitas and Irises. Am J Bot 79(1):101–111

Ortego J, Gugger PF, Riordan EC, Sork VL (2014) Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. J Biogeogr 41:1895–1908

Palme´ AE, Su Q, Palsson S, Lascoux M (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. Mol Ecol 13:167–178

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24(7):386–393

Petit RJ, Latouche-Halle´ C, Pemonge MH, Kremer A (2002) Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. For Ecol Manage 156:115–129

Petit RJ, Bodenes C, Ducousso A, Roussel G et al (2003) Hybridization as a mechanism of invasion in oaks. New Phytol 161:151–164

Petit RJ, Duminil J, Fineschi S, Hampe A et al (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689–701

Pollegioni P, Olimpieri I, Woeste KE et al (2013) Barriers to interspecific hybridization between *Juglans nigra* L. and *J. regia* L species. Tree Genet Genomes 9:291–305

Rieseberg LH, Carney SE (1998) Tansley review N.102: plant hybridization. New Phytol 140:599–624

Roe AD, MacQuarrie CJK, Gros-Louis MC et al (2014) Fitness dynamics within a poplar hybrid zone: I. Prezygotic and postzygotic barriers impacting a native poplar hybrid stand. Ecol Evol 4 (9):1629–1647

Rundle HD, Nosil P (2005) Ecological speciation. Ecol Lett 8:336–352
Sambatti JBM, Strasburg JL, Ortiz-Barrientos D et al (2012) Reconciling extremely strong barriers with high levels of gen exchange in annual sunflowers. Evolution 66:1459–1473
Savolainen O, Pyhäjärvi T (2007) Genomic diversity in forest trees. Curr Opin Plant Biol 10:162–167
Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. Annu Rev Ecol Syst 34:339–364
Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. Mol Ecol 20:5123–5140
Sun Y, Abbott RJ, Li L, Li L, Zou J (2014) Evolutionary history of purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. Mol Ecol 23:343–359
Thomas DT, Ahedor AR, Williams CF, DePamphilis C (2008) Genetic analysis of a broad hybrid zone in *Aesculus* (Sapindaceae): is there evidence of long-distance pollen dispersal? Int J Plant Sci 169:647–657
Thomasset M, Fernandez-Manjarres JF, Douglas GC et al (2013) Assignment testing reveals multiple introduced source populations including potential ash hybrids (*Fraxinus excelsior x F. angustifolia*) in Ireland. Eur J For Res 132:195–209
Thompson SL, Lamothe M, Meirmans PG et al (2010) Repeated unidirectional introgression towards *Populus balsamifera* in contact zones of exotic and native poplars. Mol Ecol 19 (1):132–145
Via S (2009) Natural selection in action during speciation. Proc Natl Acad Sci USA 106:9939–9946
Wang B, Mao JF, Gao J, Zhao W, Wang XR (2011) Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. Mol Ecol 20:3796–3811
Watano Y, Kanai A, Tani N (2004) Genetic structure of hybrid zones between *Pinus pumila and P. parviflora var. pentaphylla* (Pinaceae) revealed by molecular hybrid index analysis. Am J Bot 91(1):65–72
Wu CI (2001) The genic view of the process of speciation. J Evol Biol 14:851–865
Zeng YF, Liao WJ, Petit RJ, Zhang DY (2011) Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. Mol Ecol 20:4995–5011
Zhang D, Xia T, Yan M, Dai X et al (2014) Genetic introgression and species boundary of two geographically overlapping pine species revealed by molecular markers. PLoS ONE 9(6): e101106
Zhou YF, Abbott RJ, Jiang ZY, Du FK et al (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. Evolution 64 (8):2342–2352

Chapter 16
Apomixis as a Facilitator of Range Expansion and Diversification in Plants

Diego Hojsgaard and Elvira Hörandl

Abstract Apomixis, the asexual reproduction via seed, often occurs in huge plant polyploid complexes with large geographical distributions. However, the long-term evolutionary fate of asexuals traditionally was regarded as doomed by extinction. A seven-step evolutionary model is presented to explain the role of sex → apomixis shifts on geographical cytotype distributions, and the potential consequences of reversals apomixis → sex on plant diversity. Accordingly, apomictic polyploid genotypes act as facilitators for range expansions of asexual taxa in agamic complexes by functioning as pioneer explorers of new niches. High intragenomic (allelic) diversity and epigenetic variability may help for rapid adaptation. Therefore, they could rapidly expand the distribution areas of their progenitor sexual populations by occupying new ecological niches and geographical areas. Hence, apomixis would result in divergent patterns of geographic distribution between sexual and asexuals, a pattern described as "geographical parthenogenesis," in which apomicts occupy extensive geographical areas and higher latitudinal zones while sexual relatives are restricted to small refugees. Later on, reversals to complete sexuality would allow for the establishment of new sexual populations in different habitats without the long-term disadvantages of asexuality. The new sexual recombinants will be genetically isolated from the original sexual populations and consequently predisposed to a divergent evolution, and potentially enabled to evolve into new sexual species. The present model stresses a previously unidentified evolutionary significance of the geographical parthenogenesis as a motor for plant diversification.

D. Hojsgaard (✉) · E. Hörandl
Department of Systematics, Biodiversity and Evolution of Plants, Albrecht-Von-Haller Institute for Plant Sciences, University of Goettingen, Goettingen, Germany
e-mail: diego.hojsgaard@biologie.uni-goettingen.de

E. Hörandl
e-mail: elvira.hoerandl@biologie.uni-goettingen.de

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_16

16.1 Introduction

Angiosperms are the most successful group of the plant kingdom. Features such as hybridisation and polyploidy are major contributors to evolution and diversification in angiospermous lineages (Soltis and Soltis 2009; Jiao et al. 2011), often connected to shifts between sexuality and asexuality (Asker and Jerling 1992). On the other hand, asexuality in angiosperms has long been seen as a reproductive strategy that rather restricts evolution by limiting genetic reshuffling (Stebbins 1950).

Asexual taxa are characterized by the lack or malfunctioning of normal mechanisms enhancing variability in natural populations (i.e., meiotic, chromosomal, and gametic recombination). Traditional views regarded apomixis as a blind alley of evolution (Darlington 1939) because of the lack of variation and hence reduced adaptive potentials.(Stebbins 1950) noted that apomictic complexes harbour just variation on the same theme, and that apomicts failed to evolve new genera.

Here, we present an alternative view of apomixis (i.e., the asexual reproduction via seeds = agamospermy) as promoter of diversification and evolution. Present geographical distributions of apomictic complexes rather suggest broad range expansions and altogether broad ecological niches compared to sexual relatives (Kearney 2005, Hörandl 2006). This phenomenon termed "geographical parthenogenesis" was so far regarded just a short-term success of sexuality (Van Dijk 2003), and most explanations relate the causality to short-term ecological differentiation. This view probably neglects the dynamics that is inherent in agamic polyploid complexes. Research of the last decades suggests that the dynamics in origin, genetic, and developmental features within agamic complexes has been strongly underestimated. Here, we will discuss (1) developmental variability enhancing genetic variation, (2) natural origins agamic complexes and dynamics among lineages as sources for genetic variation, (3) intragenomic allelic diversity and epigenetic variation as potential factors for adaptation to environmental heterogeneity, (4) recent findings of geographical patterns and niche differentiation in various model systems, and (5) hypothetical scenarios of reversals from asexuality to sexuality which may open further evolutionary potentials for speciation.

16.2 Asexual Plants as Uniform Clones: A View with Many Scarcities

In this section, we will introduce a simple definition of clone at the organismal level, considering single plants as individuals rather than the "evolutionary individual" in the sense first raised by Janzen (1977), to place our discussion and pinpoint snags surrounding the view of clonality.

A clone is a number of ramets which belong to the same genet (Richards 1997), i.e., a number of physiologically independent individuals with the same genetic constitution (or genotype). Hence, a cloning process must exactly reproduce the

genetic composition of the individual in the absence of genetic recombination. In angiosperms, plants can reproduce asexually and generate "clones" following dissimilar processes. The two most important categories are vegetative propagation and apomixis.

Vegetative propagation involves the formation of new individuals through specialized structures without production of seeds or spores, and hence, meiosis and syngamy are avoided. There are many types of structures to propagate vegetative daughter plants, including plantlets and bulbils (vivipary), stem offshoots or adventitious shoots, rhizomes and stolons, runners, bulbs, and tubers. Dispersal capacities of each of these propagation strategies are variable and can be limited to a small area usually close to the mother plant or big extensions, occupied by one or few clones (e.g., the invasive Canadian pondweed, *Elodea canadensis* Michx.; Gustafsson 1946).

Apomixis includes many different developmental patterns, clustered in three general groups, i.e., diplospory,apospory, and adventitious embryony (Asker and Jerling 1992). Diplospory starts the development from an unreduced megaspore, while in apospory, a somatic cell takes over a megaspore-like cell fate; in both cases, an unreduced egg cell is formed which develops parthenogenetically. Adventitious embryony involves the direct development of embryos out of somatic cells. All three general patterns involve the formation of new individuals through seeds, although meiosis and syngamy might be partially or completely avoided. Consequently, apomictic plants take the advantage of producing dispersal units (diaspores) with higher dispersal capacity compared to vegetative propagation strategies. Most important, apomictic development of embryos starts from a single-cell stage, either an unfertilized egg cell or a somatic cell, while vegetative development starts from multicellular tissues (Mogie 1992).

Although clones are expected to be genetically uniform, because of several reasons, 100 % pure clones rarely exist in nature (Avise 2008). Individual plants are formed by thousands, even millions of mitotically derived cells. Even though DNA repair mechanisms inside cells are efficient, they are not perfect. A few *de novo* mutations will arise in a genome during DNA replication and will be retained during the plant development. While in vegetative propagated plants, meristems are multicellular and hence mutations produce chimeric tissues, apomictic plants goes through a single-cell stage that increases on the one hand the chance of the elimination of deleterious mutations as all daughter cells will inherit this mutation (Grosberg and Strathman 2007) and on the other hand the chances of establishment of distinct genotypes carrying particular adaptive mutations (Van Dijk et al.2009). Moreover, another three processes can slightly reduce the 100 % genetic identity expected among clonemates (Avise 2008). The first one is gene conversion, a process in which recombination between members of the same gene family (e.g., duplicated genes) occurs at ectopic places (nonallelic or paralogous recombination) involving unidirectional exchange of DNA sequences (e.g., Datta et al. 1997). In angiosperm genomes, there is an appreciable rate of gene conversion between genes duplicated during ancient polyploidization events (i.e., paleologous genes) (Wang and Paterson 2011).

The second process is mitotic recombination, an atypical crossover event between homologous chromosome segments that can occur during repair of spontaneous DNA damage (Andersen and Sekelsky 2010). However, as pointed out by Gorelick (2014), although mitotic recombination in the sense of crossing over events per cell division occurs at much lower rates than normal meiotic recombination (Pontecorvo and Käfer 1958; Andersen and Sekelsky 2010), a single individual plant undergoes a lot more mitotic than meiotic cell divisions, especially in large clonal plants. The third process showing increasing relevance on studies of clonality is what Martin (2005) described as "epigenetic drift." Epigenetic drift refers to the differential acquisition of epigenetic marks in organisms with the same genetic constitution and, naturally, affected by different or even slightly different environmental sceneries and ecological interactions. Thus, epigenetic variation acquired in long-lived asexual organisms may critically influence genotype development and plasticity by differential gene regulation, creating genetic differences among clonemates (e.g., Verhoeven and Preite 2013; Douhovnikoff and Dodd 2015). The age and growth rates of each particular genet/genotype will influence on the relative impact of those gene-based processes on clones' genetic composition. Depending on particular cases, the longevity of clonal plants can vary between few and several thousand years (de Witte and Stöcklin 2010). Henceforth, *de novo* mutations, gene conversion, mitotic recombination, and epigenetic drift processes may certainly work as an important source of genetic and genomic variation which would make the assumption of genetic uniformity of large and long-standing clones in nature something hardly ever valid.

An extra decisive point unconnected to the last-mentioned gene-based processes that occur only in apomictic plants is that they usually are not obligate apomicts in the sense that the normal meiotic pathway still can be functional (Asker and Jerling 1992). Therefore, most apomicts produce low levels of sexually derived individuals together with the clonally derived ones. Consequently, natural populations of apomicts show greater genotype variability compared to that expected from a clone. Although levels of functional sexuality are generally low, they show a large range of variation depending upon genotypic, environmental, and ecological conditions (e.g., Aliyu et al. 2010; Bicknell et al. 2003; Cosendai et al. 2011; Hojsgaard et al. 2008, 2013, 2014b; Majeský 2013; Noyes and Givens 2013; Quarin 1986; Rebozzio et al. 2011, and have clear consequences on asexual genome evolution, e.g. buffering Müller's ratchet effect of mutation accumulation expected in strictly asexual organisms (Müller 1964; Hojsgaard and Hörandl 2015). Hence, the proportion of recombinant seeds and individuals will be different for each particular genetic system and probably for each area and time of the season considered. Anyhow, if meiosis can still be functional, there will be high probabilities of injecting new recombinant genotypes in any apomitic population. Whether such genotypes will be successful and establish a new lineage in a particular situation would depend upon different factors. Such factors may include internal causes (developmental and genetical or genomic related; e.g., Hojsgaard et al. 2013; Pellino et al. 2013; Hojsgaard and Hörandl (2015) or external ones (competition against other genotypes, niche shifts, etc.; e.g., Pearman et al. 2007).

16.3 Apomixis and the Formation of Agamic Complexes

Diploid plants are the primary source of new polyploids (Ramsey and Schemske 1998). Multiple origins of polyploids from different hybridization events between diploid parental species results in various polyploid derivatives, which can also interact with each other (Koch et al. 2003; Guo et al. 2004; Hörandl et al. 2009; Lo et al. 2010, Sochor et al. 2015 among others). Polyploids arisen in a diploid population can also backcross with their parental species in different ways and will have diverse outcomes. The interaction between such derivative polyploids and their ancestral diploids form polyploid complexes that reshape many aspects of the species variability (morphologically, physiologically, cytogenetically, ecologically, genetically, etc.). Babcock and Stebbins (1938, pp. 55–56) defined a polyploid complex as:

> *a group of species, containing forms with different chromosome numbers, of which those with the lowest number (i.e., the diploids) are more or less distinct from one another morphologically, and* ***are usually isolated from one another by sterility barriers****, but in which some of the* (*aneuploid or*) *polyploid type*(*s*) (*are*) is *intermediate between the diploids or show different recombinations of their characteristics.*

Since polyploidy in plants may or may not be associated with reproductive shifts, two types of complexes can be found: (1) sexual polyploid complexes, in which only sexual diploid–polyploid associations are observed; and (2) agamic polyploid complexes, in which the normal sexual reproduction in the new polyploids is partially or completely replaced by some kind of asexual propagation for dispersal, e.g., bulbils, nucellar embryony, and gametophytic apomixis (Babcock and Stebbins 1938; Grant 1981). While the first type is the most frequent in plants, the second is less recurrent, connected to the inferior frequency of apomixis in angiosperms (Hojsgaard et al. 2014a). Thus, agamic complexes are shaped by polyploidy, hybridization, and apomixis, three major processes responsible of evolutionary change and the formation of an array or "complex" of microspecies carrying different character combinations (morphological, cytological, and genetic) from parental species and higher ploidy taxa (usually derivatives). The establishment of "microspecies" is the primary result of apomixis, and consequently microspecies connect morphological gaps between the diploid sexual species blurring discontinuities and causing taxonomic difficulties (Babcock and Stebbins 1938; Gustafsson1947; Stebbins 1950; Grant 1981; Hörandl 1998; Haveman 2013). From an evolutionary point of view, those microspecies represent relatively successful cases of new character combinations, new genotypes, and/or cytotypes produced by introgression of sexual parentals and/or apomicts derivatives in which residual levels of sexuality are still functional. The level of taxonomic complexity is thought to be contingent with the evolutionary stage of development of the agamic complex (see examples in Babcock and Stebbins1938; Gustafsson 1947; Grant 1981). Thus, a "young" complex will be considered that one in which parental sexual species together with few microspecies occur in nature and show restricted geographical distribution. In the same direction, a "mature" agamic complex would be that one presenting sexual parentals together with a relatively high number of agamospermous microspecies and widely

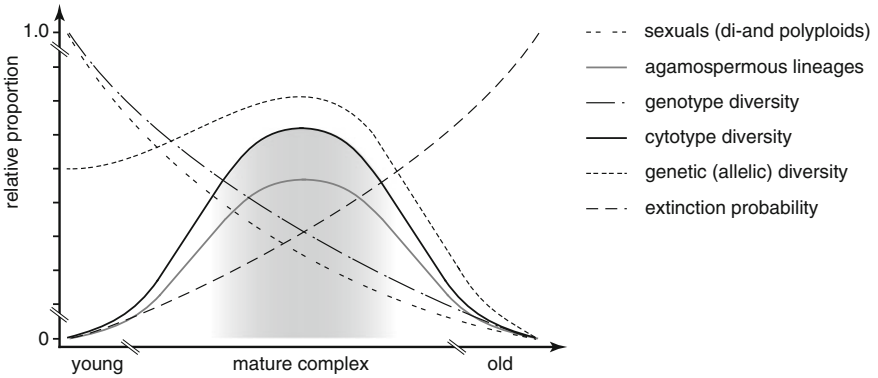
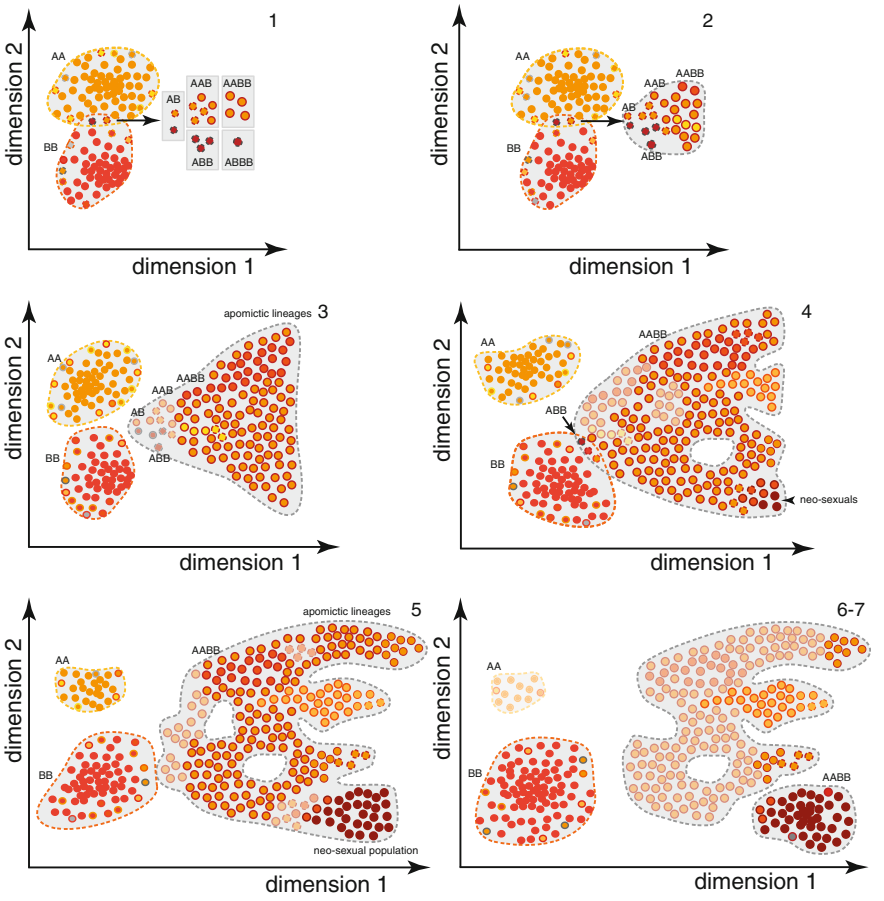


Fig. 16.1 Properties of agamic complexes expected and observed from theoretical models and empirical data after analyses of different successful natural agamic complexes (without considering reversals to sexuality). The abscissa represents a time gradient starting with the origin of the first apomictic lineage. Distribution of values in both axes can vary according to genomic properties and evolutionary flexibility of each particular genetic system. Sexual individuals and/or cytotypes will start to decrease while a proportional increase of asexual lineages. Sexual populations and/or species will decrease following normal extinction rates plus competition with asexual polyploid lineages. Agamospermous lineages will increase by recurrent formation and mutation/recombination processes acting on previously established lineages. Genotype diversity will decrease with elimination of sexual individuals/populations, but will be partially slowed down by the formation of new asexual lineages. Cytotype diversity will increase with the formation of new polyploid and aneuploid lineages mainly via hybridization and polyploidization, but also via partial apomixis (i.e., fertilization of unreduced female gametes), and formation of chromosomally unbalanced gametes. Genetic diversity will be partially maintained or increased during first stages of establishment of the agamic complex by formation of new polyploids that would serve as reservoirs of genetic variability. Finally, the probability of extinction of the agamic complex will increase with the concurrent decrease of sexual individuals/populations, which will drastically influence recurrent formation of new lineages as well as the surviving capacities of established agamospermous lineages along with the other diversity parameters

distributed, while an "old" one would have no extant sexual parentals and few surviving agamospermous microspecies (Grant 1981). In natural populations, the sympatric occurrence of sexual (frequently, but not exclusively) diploids associated with apomictic (almost always) polyploids is not uncommon, but in each case, the evolutionary developmental stage should be considered according to particular features of individual complexes. Figure 16.1 represents a summary of theoretical and empirical features expected and observed in different natural agamic complexes (Babcock and Stebbins 1938; Stebbins 1950; Grant 1981). Although initially agamic complexes were seen more as a "closed system" whose final doom depends upon its sexual members (Babcock and Stebbins 1938, p. 61; Stebbins 1950, p. 417), today's view of agamic complexes has changed. Following Carman´s hypothesis (Carman 1997), apomixis in angiosperms is a result of asynchronous expression of duplicate genes. This hypothesis had found support on different research areas (i.e., embryology, cytology, and gene expression) (e.g., Grimanelli et al. 2003; Polegri et al. 2010; Sharbel et al. 2010; Hojsgaard et al. 2013, 2014b). Under this view, in natural

agamic complexes, apomixis would function as a reproductively stable transitional stage that promotes the evolution of asexual neo-polyploids into regular and developmentally novel sexual paleopolyploid species by functional elimination of gene duplications responsible for apomixis. Therefore, cycles of polyploid agamospermous microspecies formation evolving into new sexually stable species would serve as a facilitator for diversification. The transition-phase hypothesis of Carman (1997) was recently expanded by Hörandl and Hojsgaard (2012) who considered a more dynamic role for new apomictic polyploids within agamic complexes. According to these authors, facultative apomicts with restricted genetic recombination and genotype segregation works as perfect forerunners to explore new eco-geographical areas and niches available only to certain specific genotypes. While sexuality would break down such particular genotypes hampering the access to those areas or niches, facultative apomixis can easily multiply a single genotype without altering gene combinations, whereas it keeps certain level of variation (sexual recombinants). Moreover, although sexual populations theoretically present a cohesive gene pool through a uniform gene flow network among individuals, facultative apomixis allows for the creation of a relatively restricted gene flow network that cracks such uniform gene reservoir within the lineage and isolates populations. With sufficient time, in some areas of widely distributed microspecies, cases of reversal to sexuality could result in a new, and genetically isolated from the parents, sexual population with ample potential to divergent evolution and therefore to maximize species diversification capacities. Following previous features, a model and general outline for the evolution of agamic complexes in nature is drafted (Fig. 16.2).

Even though obligate apomixis was considered as "an escape into a blind alley" with a certain extinction fate (Darlington 1939, p. 113, but also Babcock and Stebbins 1938; Stebbins 1950; Grant 1981), whenever a strict asexual spread into new geographic and ecological areas, it will be subject to new environmental conditions that would promote possibilities for a cytological and genomic re-stabilization of the sexual (meiotic) reproductive program and a complete reversal to sex. The end of the alley may not be strictly a blind one. Based on evolutionary and biogeographical information on apomixis occurrences, Hojsgaard et al. (2014a) found support for the expanded version of Carman's original hypothesis, as apomixis in angiosperms is highly correlated to those clades with higher levels of diversity (Hojsgaard and Hörandl 2012; Hojsgaard et al. 2014a). Below, we will focus on most important microevolutionary aspects playing a role in different steps of development and evolution of agamic and (when correspond) clonal complexes, as well as their potential role on enhancing plant diversification.



16.4 Apomicts Temporarily Restrict Genetic Variation and Potentiate Plant Ecological Capacities

Apomictic species can found populations via single individuals and are therefore better colonizers than related outcrossers; this colonization ability is most efficient after long distance-dispersal of seeds (Baker's law; Baker 1965). Vegetative propagation, in contrast, remains in terrestrial biota usually spatially restricted. Hence, apomixis combines the benefits of seed dispersal, increasing the potential for long distance-dispersal, and reproductive assurance in the introduced area without the need of pollinators (Hörandl 2006). Even in the case of pseudogamy, self-compatibility enables apomictic plants to use self-pollen for the required endosperm fertilization (Hörandl 2010).

◀ **Fig. 16.2** Hypothetical model for the origin, development, and dynamic of agamic polyploid complexes based on most important microevolutionary processes along a chronological gradient of time (steps). Here, we focus on evolutionary destiny of one single asexual lineage, but many lineages can arise by recurrent formation from sexual parental species through polyploidization and/or hybridization. In the graphics, sexual individuals are represented by uniform colored *circles* (*orange and red*) and apomictic individuals are shown with different color combinations. Species and lineage distributional areas are included within *dotted lines*. Developmentally and genetically unstable individuals are represented by *dotted circles*. Circle sizes follow individuals´ ploidy, and opaqueness identifies ecologically unfitted individuals. *1*, a first step would involve a reproductive shift from sexuality to facultative apomixis associated with polyploidization and/or hybridization processes; *2*, diversification of apomictic lineages by facultative sexuality, mutation, chromosome re-arrangements, hybridization; *3*, range expansions of eco-physiologically fitted genotypes and additional (limited) agamic complex differentiation trough mutation and residual sexuality into "cracked" gene pools by gene flow restrictions; *4*, occasional re-stabilization of the meiotic reproductive pathway in some lineages (*arrowhead*); *5*, complete reversal to sexuality, promoted by niche shift and/or environmental conditions and establishment of a new sexual population (*dark red circles*); *6–7*, allopatric speciation of newly formed, biologically and geographically isolated sexual populations (*dark red circles*), and further diversification of genetically cohesive biological species (with the eventual evolution of new genera along time). As expected, each agamic complex is subject to extinction, but when persists, different complexes can be found at different stages of this evolutionary frame. Stages *3–5* may represent mature agamic complexes, while at *6–7* there is a drastic decay of complexity mainly due to ecological changes not alleviated by sexual populations nor asexual lineages which drive them to extinction. Contact zones and introgression with parental species could appear along any of these stages (here is represented in *4*). AA, BB: diploid sexual populations/species; AB: diploid hybrids; AAB, ABB: triploid hybrids; AABB: tetraploid hybrids. Dimension 1 and 2 represent any biological, geographical, or ecological gradient outlining the hypothetical distribution, niche and dynamics of populations represented here

However, after colonization, populations have to adapt to novel habitats in the introduced area. Although a "very narrow population niche width" is one of main disadvantages of one agamospermous lineage (Richards 1997), it may not be such in a uniformly patched environment. In a sexual panmictic population, a huge array of genetic variants is expected, each occurring at a certain frequency in space and time. However, a fitted genotype is hardly ever re-created, and such genotype will be lost even when their genes will be uploaded to the population gene pool. Only the population as a whole can adapt to a certain niche. On the contrary, in an agamospermous polyploid complex, broad arrays of individuals derived from few genotypes are being produced, each fitting to a particular niche. Altogether arrays of differentiated clones can fill the total niche space more efficiently than sexual species can do, especially by occupying the extreme niches (Vrijenhoek 1984). This so-called frozen niche variation model was empirically confirmed in many studies (review in Vrijenhoek and Parker 2009; Lo et al. 2013; Mau et al. 2015).

Here, we illustrate how the apomictic genotype can be superior to a sexual population in the same niche space. The same apomictic genotype will be multiplied promoting its spread into the uniformly patched environment. To exemplify this idea, we depict a simple situation where the genotype *Aa* is best adapted to, and hence can disperse on the long term in a relatively uniform ecological niche space (Fig. 16.3). In both sexual and asexual reproductive situations, self-compatibility is

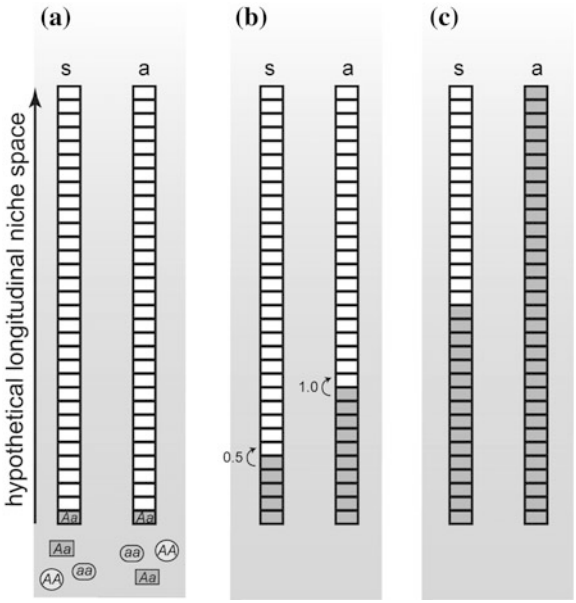


Fig. 16.3 Scheme representing species ecological capacity to occupy a hypothetical patched niche offered (vacant) under two different reproductive strategies and a single dispersal unit (seeds). After first arrival of a single seed from a (distant) population into a free niche patch (**a**), the fitted genotype will be able to produce self-pollinated seeds and conquer one adjacent patch per generation. For the plant species with a sexual strategy, Mendelian chromosome segregation would produce 0.25*AA*, 0.5*Aa,* and 0.25*aa* individuals. Since mainly genotypes *Aa* can adapt well to specific niche requirements, through sexuality the species can occupy one patch every second generation. For the species with the asexual strategy, apomixis will multiply exclusively the fitted genotype *Aa*, and hence, one free patch can be conquer every generation. Thus, apomicts will widespread here twice as fast as sexuals (**b**), but when considering genotype complexity and adaptation to a particular heterogeneous niche, the relative occupancy speed gained by producing clonal individuals would be even higher as proportions of fitted genotypes when considering many gene combinations in segregating sexuals will certainly be lower than 0.5. After *n* generations (**c**), only the species with the asexual strategy occupy the whole niche space. By restricting genetic segregation, apomicts can potentiate ecological capacities of fittest genotypes within particular niches. Individuals fitting dissimilar niche requirements are represented with different shapes and greyscale colors

assumed, seeds are the dispersal units and only one seed can move to the next accessible site or niche patch per generation. From the graphic is possible to visualize the expected consequences the temporal restriction of genetic variation in asexuals can have potentiating and/or stabilizing ecological plant capacities to occupy faster than sexuals and vastly uniform niches independently of the niche size.

The consequences of vegetative propagation on niche space occupancy would be essentially the same only when consecutive niche patches are close enough to allow dispersal of clonal units. Otherwise, sexuality would have an advantage over vegetative propagation as seeds are the most efficient dispersal units.

16.5 Genomic Background of Adaptation to Environmental Heterogeneity

Although asexuality can be advantageous in uniform ecological conditions, natural landscapes are patchy and environmental heterogeneity, both spatial and temporal, is rather the rule (Stewart et al. 2000). Yet, pioneer plants can use different strategies to adapt rapidly to novel and patchy environments. Several studies have reported shifts of ecological niches of pioneer plants from the native to the new conquered area (Early and Sax 2014; Tingley et al. 2014). In general, rapid adaptation can be facilitated by polyploidy, hybridization, and stress-induced modification via epigenetic change (Prentis et al. 2008). Vegetative propagation may take advantage of clonal integration to exploit environmental heterogeneity, while apomixis does not restrict the formation of new biotypes which could adapt to heterogeneous ecological conditions.

Clonal integration, i.e., connection among ramets that share water, nutrients, and other substances can improve the plant´s utilization of heterogeneous resources thus affecting growth, biomass allocation, photosynthetic performance, etc. and facilitating spatial occupation of new habitats at a local scale (You et al. 2014).

Apomixis, instead, copes more regional scales by using the long distance-dispersal capacities of seeds. Apomictic plants can deal with regional environmental heterogeneity because they are not strictly clonal, they can create new genotypes primarily through residual sexuality and secondly through mutations (e.g., Paun et al. 2006; Hörandl and Paun 2007; but see further details under the section *Apomixis and the formation of agamic complexes*). Because apomixis is mostly facultative, even low levels of functional sexuality are enough to create new genotypic variants and thus facilitate the occupation of heterogeneous environments. In this sense, genotypic variability is higher surrounding the center of distribution of diploid sexual forms or in areas of co-occurrence of diploid sexuals-polyploid apomictics (e.g., Babcock and Stebbins 1938; Daurelio et al. 2004; Cosendai et al. 2013). Moreover, since apomixis is usually connected to polyploidy, and often apomicts are hybrids (Carman 1997), apomictic plants sustain highly heterozygous, vigorous genotypes, which can colonize different environments with, for example, general purpose genotypes (Baker 1965). Therefore, colonization and niche shifts into novel abiotic conditions and environmental heterogeneity could be conducted by genotypes with a high standing intraindividual genetic variation, i.e., intragenomic (allelic) diversity and heterozygosity (Prentis et al. 2008). Additionally, multiple introductions can increase genotypic variability in a newly occupied area (Barrett et al. 2008; Molins et al. 2014) and thus enhance niche space occupancy.

Detection of small-scale genotype–environment interactions is strongly indicative of local adaptation at microsite of a few squares meters and that small-scale environmental heterogeneity is maintaining genetic variation in asexual species (Stratton 1994; McLeod et al. 2012). Polyploidy is another factor that helps asexuals to afford the requirements of diverse ecological niches by, for example,

preserving levels of genetic variability despite the absence of meiotic recombination and genetic segregation.

While even low percentages of genotypic variation could help asexual plants to fulfill ecological demands of heterogeneous environments, the capacity to acclimatize to new conquered microhabitats after exposition to novel abiotic and biotic conditions and to environmentally induced stress seems to be related to epigenetic variation (e.g., DeWalt and Hamrick 2004; Poulin et al. 2005; Hardesty et al. 2012; Roiloa et al. 2014). Such epigenetic fine-tuning would favor the selection of fittest genotypes possible adapted to a variety of microenvironments and the entire occupancy of niche spaces. Experimental work on apomictic clonal dandelions demonstrated immediate response of polyploid apomictic plants to abiotic stress conditions and heritability of epigenetically controlled traits (Verhoeven et al. 2010a, b; Verhoeven and Preite 2013). Asexual plants further change their morphological phenotypes significantly in introduced areas with the same speed as sexual plants (Dalrymple et al. 2015) and show a high ecophysiological plasticity (Molina-Montenegro et al. 2013). Morphological variation is in apomictic plants correlated to epigenetic rather than to genetic variation (Rois et al. 2013). Epigenetic markers are in plants at least partially heritable (Verhoeven et al. 2010b) and can be directly influenced by abiotic environments (Bossdorf et al. 2008; Zhang et al. 2013). Other than in animals, the plant germline cells separate late during development in the adult sporophyte, and re-setting of epigenetic methylation does not occur during plant meiosis (Jacobsen and Grossniklaus 2011). Hence, transgenerational inheritance of environmentally induced epigenetic change is present and potentially a powerful mechanism for rapid local adaptation of plants.

Thus, not only genotypic but also epigenetic variations are factors enhancing genotype plasticity and adaptation (e.g., Verhoeven and Preite 2013; Douhovnikoff and Dodd 2015), and hence the capacity of asexual plants to cope better with environmental heterogeneity and niche gradients, promoting the dispersal of fitted genotypes and contributing to the success observed in clonal plants.

16.6 Geographical Distribution and the Dual Effect of Being Asexual: Benefits May also be Constraints

Asexual plants suffer from several restrictions connected mainly to genetic limitations raised by the absence of sex. Because of polyploidy and/or hybridization, levels of genetic variability and heterozygosity may remain similar to those observed in sexual parental species, despite reductions in genotypic diversity (Hörandl and Paun 2007; Beck et al. 2011; Cosendai et al. 2013). Although evolutionary capacities of asexual plants are certainly influenced by such restrictions, still it is not completely clear whether the consequences of mode of reproduction are negative or positive. Natural distribution of sexual–asexual systems offers opportunities to evaluate probably consequences of asexuality in species range

expansion. In some cases, ecological capacities of asexual plants seem to be broader than that of sexual parents as supported by geographical parthenogenesis patterns (i.e., a pattern where sexual diploid and asexual polyploid populations have different distribution areas; Vandel 1928; Bierzychudek 1985; Asker and Jerling 1992; Hörandl et al. 2008), but such assertion has yet not been tested with specific studies. Asexuality had certainly played a crucial role in many cases (e.g., Kearney 2005; Hörandl 2006), whether it is a temporary or lasting role is still uncertain. Range expansion is usually associated with ecological and/or genomic circumstances such as landscape changes, hybridization, or polyploidy, which opens new opportunities for species expansion by releasing new niches and/or creating new combinations of characters that allow the invasion of new habitats (e.g., Babcock and Stebbins 1938; Levin 2000). In some cases morphological and/or genetic differentiation of ecologically divergent lineages may obscure species boundaries and with this researcher's ability to find signs of geographical parthenogenesis patterns since differentiation between sexual and asexual morphotypes will not be possible without a reproductive characterization. Nevertheless, the known cases where asexuality and apomixis were morphologically, embryologically and/or genetically tracked show contrasting patterns of geographic distributions between sexuals and asexuals. Well-known examples such as the *Crepis* complex (Babcock and Stebbins 1938), *Ranunculus kuepferi* (Cosendai and Hörandl 2010), the *Taraxacum* sect. *Ruderalia* complex (Van Dijk 2003), *Paspalum simplex* (Urbani et al. 2002), *Townsendia hookeri* (Thompson and Whitton 2006), and *Crataegus* (Lo et al. 2013) display the classic pattern of geographical parthenogenesis, where sexuals occupy geographically restricted areas usually related to ancient glacial refuges while apomicts exhibit a more extensive occurrence into ecologically more diversified or severe environments. Several non-exclusive hypothesis had been claimed to play a role in establishment of geographical parthenogenesis patterns (reviewed in Hörandl 2009; Vrijenhoek and Parker 2009), some related with reproductive or genetic advantages such as uniparental reproduction (Baker's law, Baker 1955), polyploidy (Comai 2005; Mau et al. 2015), hybridization (Stebbins 1959), and others related to ecological genetic models such as general purpose genotypes (Baker and Stebbins 1965; Lynch 1984) or frozen niche variation (Vrijenhoek 1979, 1984). However, although no single factor might explain geographical parthenogenesis patterns in each particular case, there are many other cases where apomicts do not show a geographically clear asymmetric distribution compared to sexuals, or even they display inverse patterns with narrow distributions and/or occupy less diversified ecological areas. A clear example could be that of the *Boechera holboellii* complex, diploids (sexual or apomictic) prevailed rather in the northern USA, whereas polyploid apomicts predominate in the southern USA (Dobeš et al. 2004b), with a diffuse geographical pattern between sexuals and apomicts (Sharbel et al. 2005), probably connected to evidence suggesting that diploid apomicts in *Boechera* are trapped into ecological niches of their parental diploid sexuals (Mau et al. 2015). Another possible case that does not show the typical pattern of geographical parthenogenesis would be that of *Bouteloua curtipendula* Michx., where apomicts are restricted to semi-arid regions from

southwestern USA and northeastern Mexico while sexual forms occurs from central and eastern USA till central Mexico (Gould 1959).

Patterns of geographical distribution depend also upon specific biogeographical history of an apomictic group (Hörandl 2006), whenever a complex biogeographical scenario of, for example, past fragmentation, recolonization, isolation-by-distance, and glacial refuges have certainly shaped geographical range of sexual and asexual biotypes (e.g., Dobeš et al. 2004a, b). As mentioned by Hojsgaard et al. (2014b), it became clear that geographical parthenogenesis is an opportunistic strategy that is not necessarily realized by every agamic complex.

The overall observed geographic patterns of sexual–asexual plant systems suggest that either (1) it is an effect of the evolutionary stage of the agamic complex and then smaller distributions of asexuals is related to the age (too young or too old) of the agamic complex rather than to a disadvantage of asexuality; or (2) the consequences of apomixis, or features associated with, are sensed differentially by diverse plant systems and hence have a dual effect, in some cases potentiating while in others restraining species ecological capacities to foster range expansion.

Accordingly, the impact of parthenogenesis on plant ecological capacities may depend upon other particularities related more to specific features of the taxon, such as physiological (phenotypic) plasticity, intragenomic (allelic) diversity and heterozygosity, little tolerance to genome doubling, among others that could constrict plant ecological capacity to recognize and access open or free niches. Thus, if a plant cannot fulfill requirements for niche occupancy, the mentioned theoretical advantages of apomixis will turn into ecologically trivial, fallow features. A graph representing plant ecological capacity or plasticity to fulfill requirements (or not) and hence colonize an open niche and take advantage of apomixis to expand species distribution is presented (Fig. 16.4). The degree of stabilization and penetrance of apomixis after emergence would also influence the plant's capacity to explore new habitats and increase range distributions when clonal genotypes with narrow niche breadth are tightly fitted to, and occupy a new patch, since high levels of sexuality may break up the fittest genotype and create new "sub-adapted" gene combinations no coping with niche requirements of consecutive patches. In spite of this, certain level of sex would be crucial to facilitate the conquest of new niches exposed to the facultative apomictic lineages during range expansion.

16.7 A Returning to Sex Can Fuel Ecological Genetic Divergence and Diversification

Clonal propagation fosters the expansion of certain genotypes at the micro-geographic scale. Depending on the type of clonal propagation, range distributions can vary, as runners expand easily and faster than rhizomes. However, clonal propagation at local geographical scales has the disadvantage that once a flowering period is activated (whether by ecological and/or environmental factors), gene flow between diverse clonal ramets would create new recombinant genets, uniformize the

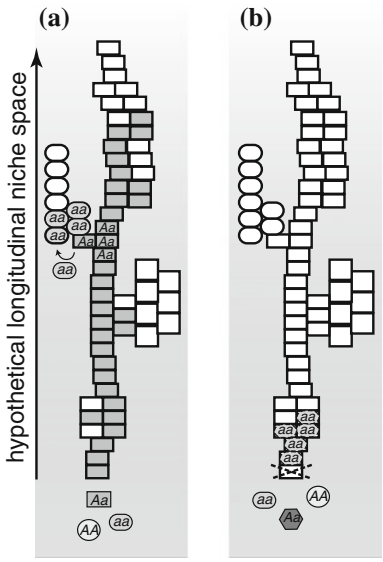


Fig. 16.4 Scheme representing a hypothetical niche offered to two species with different ecological plasticities and capacities to fulfill niche requirements. Because of particular genetic, epigenetic and ecophysiological features, species in (**a**), can fulfill the niche requirements (*rectangles*) and occupy the niche space (*colored rectangles*) by multiplying apomictically the same genotype while propagating into consecutive patches and thus accessing moderately different niches (*squares* or *ovals*). Facultative sexuality or other processes generating variation in clonal progenies (e.g., mutations or epigenetic changes; see details in the text) could facilitate the formation of genotypes with novel ecological capacities to access new open niches and occupy them by taking advantages of asexual reproduction. Contrary, species (**b**), has a genetic background and ecophysiological features that not allow filling niche requirements (no "rectangle" genotypes), although sub-adapted genotypes ("oval" genotypes) may access some patches. In such cases, only a fine-tunning toward a fitted genotype might allow the exploitation of apomixis advantages to expand species distribution range and/or access other connected niches.

gene pool and generate a net of closely connected individuals with restricted chances of further diversification and differentiation. Even if single clonally propagated genets are locally restricted in subpopulations, selection and divergence of genets adapted to particular microenvironments will be limited by regular flowering (i.e., relative rate of sexuality), reducing pair-wise allelic divergence and hence coalescence times (Austerlitz et al. 1997; Bengtsson 2003). The extent of such process would depend on the extent of the range expansion (i.e., geographic sub-divisions) and the time to the last flowering period (sexuality rate), and nonetheless will certainly shape patterns of genetic and genotypic variation within populations.

Apomictic plants, on the contrary, disperse clonal seeds and reach more regional geographic scales. In this case, not only genotypic but also epigenetic variation are factors enhancing the capacity of asexual plants to cope better with environmental heterogeneity and niche gradients, promoting the spreading of fitted genotypes. Such range expansion occurs first at local and then at regional geographical scales

and generate a mesh of relatively restricted gene flow among apomictic individuals within and among populations that will work as a dense sieve, slowing down or even interrupting genetic cohesion and isolating *ipso facto* populations located at distances that, in a sexual species, rates of gene flow are allowed and expected. The absence of the genetic cohesiveness granted by sexuality is a major consequence of apomixis in plants, which creates independent lineages, genetically isolated, and promotes faster genetic divergence among them (e.g., Bengtsson 2003) by changes other than recombination and genomic resetting of meiosis. Therefore, changes which usually had limited influence on sexual reproducing populations (e.g., mutations and epigenetic signals) had magnified effects among ameiotic non-recombinant asexual individuals. Consequently, sexuals show regional gene pools, whereas apomicts had several unique gene pools (e.g., Paun et al. 2006; Cosendai et al. 2013; Molins et al. 2013).

This is in agreement with observations that indicate that the higher is the distance to the sexual parental populations, the higher the degree of clonality and the less the genotypic diversity, so that remote areas are dominated by genotypes selected by the harshness of the environment (e.g., Babcock and Stebbins 1938; Daurelio et al. 2004). On the other hand, the farther away the distance of the founder event from the diversity center (sexual parental populations), the less probable to be reached by any migrants, and hence the less genotypically variable will be the population (depending too of the degree of residual sexuality present). Since reproductive shifts can be expected to occur in both directions (Hörandl and Hojsgaard 2012), a return to sexuality in any of the apomictic widespread lineages would found a new, geographically and genetically isolated population with a definite divergent fate. In remote areas, the pressure for a reversal to sexuality—whether realized or not—would probably be higher whenever the clonal genotype cannot adapt to or fulfill environmental requirements. Yet, reversals to sex are so far rough of examination. A few cases have been documented as likely reversions from apomictic to sexually (near) obligated reproduction in plants (see Hörandl and Hojsgaard 2012). Reasons can be diverse. Primarily, once a new sexual population is established, it can be hard to distinguish whether its evolutionary source is from an agamospermous lineage or from a sexual putative parental species. Secondly, because of constraints to the establishment of a sexual population (see next section), new sexual individuals may have a short-lived existence. Nevertheless, once established and mainly due to the facts mentioned above, the new sexual population will undergo genetic isolation that will drive it into an independent evolutionary path. Even though a complete block of the gene flow is not apparent in most of apomictic complexes studied (e.g., Hörandl et al. 2009; Cosendai and Hörandl 2010), the present levels of sexuality at the micro-geographic scale crack gene pools and can work as a very efficient "isolation net" at the macro-geographic scale. Whether limiting or interrupting gene flow among geographically related or distant apomictic clones, the only chance of genetic interchange between the neo-sexual and any other sexual population would be subject to migrants and hence, to the distance to the former sexual parental populations/species and the specific dispersal abilities of the species. Thus, the farther away the neo-sexual population is found, the less

probable to be reached by any migrants from sexual parental species, and the most likely to have divergent, distantly located, sexual populations.

The lifespan and destiny of both the new seeded and the orginal sexual populations would depend upon diverse factors, but if lasting, will certainly have a robust evolutionary potential for speciation and further diversification. This model is supported by the observation that occurrence of apomixis in angiosperms is highly correlated to clades with higher levels of diversity (i.e., higher genera and/or species numbers) (Hojsgaard et al. 2014a), reinforcing the hypothesis that apomictic pioneer plants often promote range expansion and enhances plant diversification by founding neo-sexual isolated populations.

16.8 Challenges of the Neo-Sexual Populations

Most striking challenges of reverted sexual individuals that will have a strong influence on plant population dynamics are related to Allee-effects, i.e. negative effects of low population density or small population size on mean individual fitness (Courchamp et al. 2008). Many mechanisms are known to cause an Allee-effect. The best-documented mechanism for plant species is pollination limitation (e.g. Groom 1998; Davis et al. 2004; Xia et al. 2013a), but also predation, breding system, reproductive facilitation, interspecific reproductive interference, inbreeding depression can lead to an Allee-effect (e.g. Fischer et al. 2000; Wagenius et al. 2007; Xia et al. 2013b). Since (polyploid) apomictic taxa do not undergo recombination during the reproduction process, they are not strongly affected by inbreeding consequences and maintain high levels of heterozygosity (Hörandl 2006). Conversely, self-compatible sexual species may display high levels of homozygosis due to founder effects following introduction into a new niche and therefore experience severe losses of genetic diversity and heterozygosity (e.g., Lynch 1984). Thus, a few reverted sexual individuals will suffer the same genetic consequences that will constrain the establishing of reproductively stabilized sexual individuals

Additionally, because apomictic plants produce seed embryos without fertilization, and consequently, sexual individuals cannot fertilize apomictic ones, a reversal to sex may have an ephemeral existence if unidirectional introgression of apomixis is not avoided before the establishment of a statistically representative number of reverted individuals. Usually, although few examples of interploid gene flow are known in the wild (e.g., Chapman and Abbott 2010; Pinheiro et al. 2010), differences in ploidy sets represent a relatively efficient barrier to character introgression, including complex features such as apomixis into sexual species (e.g., Hörandl and Temsch 2009). However, in the case of a new sexual population derived from an agamospermous lineage, ploidies will be equal and hence without a niche shift or any disadvantage discouraging the coexistence of the neo-sexual and the putative parental apomictic lineages, the establishment phase for a new sexual population will be highly vulnerable.

Another major disadvantage to the new sexual individuals could be related to pollinator limitation, as they would be surrounded by apomictic plants and hence not available for pollen transfer between sexual (self- or outcrosser) plants. Self-compatibility or shifts in phenology could overcome this minority (reproductive) cytotype disadvantage.

Finally, the re-establishment of sexuality could stimulate the access to new niches, mainly at the borders of apomictic distributions where genetic and epigenetic limitations would not allow further expansion. Thus, the success of the neo-sexual individuals would be mainly related to their relative capacity to became ecologically differentiated and conquer new niches for which the apomicts are not adapted and/or some genetic restrictions limit them to access such habitats.

Once a reversal is successfully established, a genetically cohesive gene pool canalized through high levels of gene flow would strengthen the isolation from surrounding asexual plants and facilitate the divergence of the new population from the original one(s).

Acknowledgements This work was funded under by the German Research foundation (DFG projects HO 5462/1-1 to D.H. and HO 4395/1-1 to E.H.) and funds of the Universitätsbund Göttingen e.V. (to D.H.). The comments of anonymous referees are gratefully acknowledged.

References

Aliyu OM, Schranz ME, Sharbel TF (2010) Quantitative variation for apomictic reproduction in the genus *Boechera* (Brassicaceae). Am J Bot 97:1719–1731

Andersen SL, Sekelsky J (2010) Meiotic versus mitotic recombination: two different routes for double-strand break repair. The different functions of meiotic versus mitotic DSB repair are reflected in different pathway usage and different outcomes. BioEssays 32:1058–1066

Asker SE, Jerling L (1992) Apomixis in plants. CRC Press, Boca Raton

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997) Evolution of coalescence times, genetic diversity and structure during colonization. Theor Pop Biol 51:148–164

Avise JC (2008) Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals. Oxford University Press, Inc, New York

Babcock EB, Stebbins GL (1938) The American species of *Crepis*: their interrelationships and distribution as affected by polyploidy and apomixis. Carnegie Institution of Washington, Publication number 504, Washington DC

Baker HG (1955) Self-compatibility and establishment after long distance dispersal. Evolution 9:347–349

Baker HG (1965) Characteristics and modes of origin of weeds. In: Baker HG, Stebbins GL (eds) The genetics of colonizing species. Academic Press, New York, pp 147–168

Baker HG, Stebbins GL (1965) The genetics of colonizing species. Academic Press, New York

Barrett SCH, Colautti RI, Eckert CG (2008) Plant reproductive systems and evolution during biological invasion. Mol Ecol 17:373–383

Beck JB, Alexander PJ, Allphin L, Al-Shehbaz IA, Rushworth C, Bailey CD, Windham MD (2011) Does hybridization drive the transition to asexuality in diploid *Boechera*? Evolution 66:985–995

Bengtsson BO (2003) Genetic variation in organisms with sexual and asexual reproduction. J Evol Biol 16:189–199

Bicknell RA, Lambie SC, Butler RC (2003) Quantification of progeny classes in two facultatively apomictic accessions of Hieracium. Hereditas 138: 11–20. ISSN 0018-0661
Bierzychudek P (1985) Patterns in plant parthenogenesis. Experientia 41:1255–1264
Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. Ecol Lett 11:106–115
Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Lin Soc 61:51–94
Chapman MA, Abbott RJ (2010) Introgression of fitness genes across a ploidy barrier. New Phytol 186:63–71
Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6:836–846
Cosendai A-C, Hörandl E (2010) Cytotype stability, facultative apomixis and geographical parthenogenesis in *Ranunculus kuepferi* (Ranunculaceae). Ann Bot 105:457–470
Cosendai A-C, Wagner J, Ladinig U, Rosche C, Hörandl E (2013) Geographical parthenogenesis and population genetic structure in the alpine species Ranunculus kuepferi (Ranunculaceae). Heredity 110: 560–569
Courchamp F, Berec J, Gascoigne J (2008) Allee effects in ecology and conservation. Oxford, New York, USA: Oxford University Press
Dalrymple RL, Buswell JM, Moles AT (2015) Asexual plants change just as often and just as fast as do sexual plants when introduced to a new range. Oikos 124:196–205
Darlington CD (1939) The evolution of genetic systems. Cambridge University Press, Cambridge
Datta A, Miyono H, Lipsitch M, Jinks-Robertson S (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc Natl Acad Sci USA 94:9757–9762
Daurelio LD, Espinoza F, Quarin CL, Pessino SC (2004) Genetic diversity in sexual diploid and apomictic tetraploid populations of *Paspalum notatum* situated in sympatry or allopatry. Plant Syst Evol 244:189–199
Davis HG, Taylor CM, Lambrinos JG, Strong DR (2004) Pollen limitation causes an Allee effect in a wind-pollinated invasive grass (Spartina alterniflora). Proc Natl Acad Sci 101: 13804–13807. doi:10.1073/pnas.0405230101
de Wide LC, Stöcklin J (2010) Longevity of clonal plants: why it matters and how to measure it. Ann Bot 106:859–870
DeWalt SJ, Hamrick JL (2004) Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, *Clidemia hirta* (Melastomataceae). Amer J Bot 91:1155–1163
Dobeš C, Mitchell-Olds T, Koch MA (2004a) Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A.* × *divaricarpa*, and *A. holboellii* (Brassicaceae). Mol Ecol 13:349–370
Dobeš C, Mitchell-Olds T, Koch MA (2004b) Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera* × *divaricarpa*, and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrative approach. Am J Bot 91:2087–2101
Douhovnikoff V, Dodd RS (2015) Epigenetics: a potential mechanism for clonal plant success. Plant Ecol 216:227–233
Early R, Sax DF (2014) Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. Global Ecol Biog 23:1356–1365
Fischer M, van Kleunen M, Schmid B (2000) Genetic Allee effects on performance, plasticity and developmental stability in a clonal plant. Ecol Lett 3: 530–539
Gorelick R (2014) Defining clonality and individuals in plant evolution. Ideas Ecol Evol 7:84–88
Gould FW (1959) Notes on apomixis in sideoats grama. J Range Manage 12:25–28
Grant V (1981) Plant speciation. Columbia University Press, New York. 2nd edition
Grimanelli D, García M, Kaszas E, Perotti E, Leblanc O (2003) Heterochronic expression of sexual reproductive programs during apomictic development in *Tripsacum*. Genetics 165:1521–1531
Groom MJ (1998) Allee effects limit population viability of an annual plant. Am Nat 151: 487–496

Grosberg RK, Strathmann RR (2007) The Evolution of multicellularity: a minor major transition? Annu. Rev. Ecol. Evol. Syst. 38:621–54
Gustafsson Å (1946) Apomixis in higher plants. I. The mechanism of apomixis. Lunds Univ. Årsskrift 42: 1–67
Gustafsson Å (1947) Apomixis in higher plants. III. Biotype and species formation. Lunds Univ. Årsskrift 43: 181–370
Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004) Allelic variation of gene expression in maize hybrids. The Plant Cell 16: 1707–1716
Hardesty BD, Le Roux JJ, Rocha OJ, Meyer JY, Westcott D, Wieczorek AM (2012) Getting here from there: testing the genetic paradigm underpinning introduction histories and invasion success. Divers Distrib 18:147–157
Haveman R (2013) Freakish patterns—species and species concepts in apomicts. Nordic J Bot 31:257–269
Hojsgaard D, Hörandl E (2015) A little bit of sex matters for genome evolution in asexual plants. Front Plant Sci 6:82. doi:10.3389/fpls.2015.00082
Hojsgaard DH, Schegg E, Valls JFM, Martinez EJ, Quarin CL (2008) Sexuality, apomixis, ploidy levels, and genomic relationships among four *Paspalum* species of the subgenus *Anachyris* (Poaceae). Flora 203:535–547
Hojsgaard DH, Martinez EJ, Quarin CL (2013) Competition between meiotic and apomictic pathways during ovule and seed development results in clonality. New Phytol 197:336–347
Hojsgaard DH, Klatt S, Baier R, Carman JG, Hörandl E (2014a) Taxonomy and biogeography of apomixis in Angiosperms and associated biodiversity characteristics. Crit Rev Pl Sci 33:1–14
Hojsgaard D, Greilhuber J, Pellino M, Paun O, Sharbel TF, Hörandl E (2014b) Emergence of apospory and bypass of meiosis via apomixis after sexual hybridisation and polyploidisation. New Phytol 204:1000–1012
Hörandl E (1998) Species concepts in agamic complexes: applications in the *Ranunculus auricomus* complex and general perspectives. Folia Geobot 33:335–348
Hörandl E (2006) The complex causality of geographical parthenogenesis. New Phytol 171:525–538
Hörandl E (2009) A combinational theory for maintenance of sex. Heredity 103:445–457
Hörandl E (2010) The evolution of self-fertility in apomictic plants. Sex Pl Repr 23:73–86
Hörandl E, Hojsgaard D (2012) The evolution of apomixis in angiosperms: a reappraisal. Plant Biosyst 146:681–693
Hörandl E, Paun O (2007) Patterns and sources of genetic diversity in apomictic plants: implications for evolutionary potentials. In: Hörandl E, Grossniklaus U, van Dijk P, Sharbel T (eds) Apomixis: evolution, mechanisms and perspectives. Gantner, Liechtenstein, pp 169–194
Hörandl E, Temsch E (2009) Introgression of apomixis into sexual species is in the *Ranunculus auricomus* complex inhibited by mentor effects and ploidy barriers. Ann Bot 104:81–89
Hörandl E, Cosendai A-C, Temsch EM (2008) Understanding the geographic distributions of apomictic plants: a case for a pluralistic approach. Plant Ecol Divers 1:309–320
Hörandl E, Greilhuber J, Klímova K, Paun O, Temsch E, Emadzade K, Hodálová I (2009) Reticulate evolution and taxonomic concepts in the *Ranunculus auricomus* complex (Ranunculaceae): insights from analysis of morphological, karyological and molecular data. Taxon 58:1194–1215
Jacobson J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Curr Opin Plant Biol 14:195–203
Janzen D (1977) What are dandelions and aphids? Am Nat 111:586–589
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97–100
Kearney M (2005) Hybridization, glaciation and geographical parthenogenesis. TREE 20:495–502

Koch MA, Dobeš C, Mitchell-Olds T (2003) Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American Arabis divaricarpa (Brassicaceae). Mol Biol Evol 20: 338–350

Levin DA (2000) The origin, expansion, and demise of plant species. Oxford University Press, New York

Lo EYY, Stefanović S, Dickinson TA (2010) Reconstructing reticulation history in a phylogenetic framework and the potential of allopatric speciation driven by polyploidy in an agamic complex in *Crataegus* (Rosaceae). Evolution 64:3593–3608

Lo EYY, Stefanović S, Dickinson TA (2013) Geographical parthenogenesis in Pacific Northwest hawthorns (*Crataegus*; Rosaceae). Botany 91:107–116

Lynch M (1984) Destabilizing hybridization, general-purpose genotypes and geographical parthenogenesis. Quart Rev Biol 59:257–290

Majeský Ľ (2013) Microevolutionary processes in apomictic genus *Taraxacum*. Dissertation, Palacký University, Olomouc

Martin GM (2005) Epigenetic drift in aging identical twins. Proc Natl Acad Sci USA 102:10413–10414

Mau M, Lovell JT, Corral JM, Kiefer C, Koch MA, Aliyu OM, Sharbel TF (2015) Hybrid apomicts trapped in the ecological niches of their sexual ancestors. Proc Natl Acad Sci. doi:10.1073/pnas.1423447112

McLeod KA, Scascitelli M, Vellend M (2012) Detecting small-scale genotype–environment interactions in apomictic dandelion (*Taraxacum officinale*) populations. J Evol Biol 25:1667–1675

Mogie M (1992) The evolution of asexual reproduction in plants. Chapman and Hall, London

Molina-Montenegro MA, Plama-Rojas C, Alcayaga-Oliveras Y, Oses R, Corcuera LJ, Cavieres LA, Gianoli E (2013) Ecophysiological plasticity and local differentiation help explain the invasion success of *Taraxacum officinale* (dandelion) in South America. Ecography 36:718–730

Molins MP, Corral JM, Aliyu OM, Koch MA, Betzin A, Maron JL, Sharbel TF (2014) Biogeographic variation in genetic variability, apomixis expression and ploidy of St. John's wort (*Hypericum perforatum*) across its native and introduced range. Ann Bot 113:417–427

Müller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1: 2–9

Noyes RD, Givens AD (2013) Quantitative assessment of megasporogenesis for the facultative apomicts *Erigeron annuus* and *Erigeron strigosus* (Asteraceae). Int J Plant Sci 174:1239–1250

Paun O, Greilhuber J, Temsch E, Hörandl E (2006) Patterns, sources and ecological implications of clonal diversity in apomictic *Ranunculus carpaticola* (*Ranunculus auricomus*) complex, Ranunculaceae. Mol Ecol 15:897–930

Pearman PB, Guisan A, Broennimann O, Randin CF (2007) Niche dynamics in space and time. TREE 23:149–158

Pellino M, Hojsgaard D, Schmutzer T, Scholz U, Hörandl E, Vogel H, Sharbel TF (2013) Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. Mol Ecol 22:5908–5921

Pinheiro F, de Barros F, Palma-Silva C, Meyer D, Fay MF, Suzuki RM, Lexer C, Cozzolino S (2010) Hybridization and introgression across different ploidy levels in the Neotropical orchids *Epidendrum fulgens* and *E. puniceoluteum* (Orchidaceae). Mol Ecol 19:3981–3994

Polegri L, Calderini O, Arcioni S, Pupilli F (2010) Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic Paspalum simplex Morong flowers. J Exp Bot 61:1869–1883

Pontecorvo G, Käfer E (1958) Genetic analysis based on mitotic recombination. Adv Genet 9:71–104

Poulin J, Weller SG, Sakai AK (2005) Genetic diversity does not affect the invasiveness of fountain grass (Pennisetum setaceum) in Arizona, California and Hawaii. Divers Distrib 11 (3):241–247

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. Trends Plant Sci 13(6):288–294

Quarin CL (1986) Seasonal changes in the incidence of apomixis of diploid, triploid, and tetraploid plants of Paspalum cromyorrhizon. Euphytica 35:515–522
Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29: 467–501
Rebozzio RN, Sartor ME, Quarin CL, Espinoza F (2011) Residual sexuality and its seasonal variation in natural apomictic Paspalum notatum accessions. Biol Plant 55:391–395
Richards AJ (1997) Plant breeding systems, 2nd edn. Chapman & Hall, London
Roiloa SR, Antelo B, Retuerto R (2014) Physiological integration modifies d15 N in the clonal plant Fragaria vesca, suggesting preferential transport of nitrogen to water-stressed offspring. Ann Bot 114:399–411
Rois AS, Rodríguez López CM, Cortinhas A, Erben M, Espírito-Santo D, Wilkinson MJ, Caperta AD (2013) Epigenetic rather than genetic factors may explain phenotypic divergence between coastal populations of diploid and tetraploid Limonium spp. (Plumbaginaceae) in Portugal. BMC Plant Biol 13:205
Sharbel TF, Mitchell-Olds T, Dobeš C, Kantama L, de Jong H (2005) Biogeographic distribution of polyploidy and B chromosomes in the apomictic Boechera holboellii complex. Cytogenet Genome Res 109:283–292
Sharbel TF, Voigt M-L, Corral JM, Galla G, Kumlehn J, Klukas C, Schreiber F, Vogel H, Rotter B (2010) Apomictic and sexual ovules of Boechera display heterochronic global gene expression patterns. Plant Cell 22:655–671
Sochor M, Vašut RJ, Sharbel TF, Trávníček B (2015) How just a few makes a lot: Speciation via reticulation and apomixis on example of European brambles (Rubus subgen. Rubus, Rosaceae) Mol Phyl Evol 89: 13–27. http://dx.doi.org/10.1016/j.ympev.2015.04.007
Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60:561–588
Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York
Stebbins GL (1959) The role of hybridization in evolution. Proc Am Philos Soc 103:231–251
Steward AJA, John EA, Hutchings MJ (2000) The world is heterogeneous: ecological consequences of living in a patchy environment. In: Hutchings MJ, John EA, Steward AJA (eds) The ecological consequences of environmental heterogeneity. Cambridge University Press, England, pp 1–8
Stratton DA (1994) Genotype-by-environment interactions for fitness of erigeron annuus show fine-scale selective heterogeneity. Evolution 48:1607–1618
Tingley R, Vallinoto M, Sequeira F, Kearney MR (2014) Realized niche shift during a global biological invasion. Proc Natl Acad Sci 111:10233–10238
Thompson SL, Whitton J (2006) Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (Townsendia hookeri). Molecular Ecology 15: 3389–3400 doi: 10.1111/j.1365-294X.2006.03020.x
Urbani MH, Quarin CL, Espinoza F, Penteado MIO, Rodrigues IF (2002) Cytogeography and reproduction of the Paspalum simplex polyploid complex. Plant Syst Evol 236:99–105
Van Dijk PJ (2003) Ecological and evolutionary opportunities of apomixis: Insights from Taraxacum and Chondrilla. Philos Trans R Soc Lond Ser B-Biol Sci 358:1113–1120
Van Dijk P, de Jong H, Vijverberg K, Biere A (2009) An apomixis-gene's view on dandelions. In: Schön I, Martens K, van Dijk P (eds.), Lost Sex: The evolutionary Biology of Parthenogenesis. Springer: Germany, pp 475–493
Vandel A (1928) La parthénogénèse géographique: contribution à l'édude biologique et cytologique de la parthénogénèse naturelle. Bulletin Biologique de France et Belgique 62:164–281
Verhoeven KJ, Preite V (2013) Epigenetic variation in asexually reproducing organisms. Evolution 68:644–655
Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A (2010a) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol 185:1108–1118
Verhoeven KJF, Van Dijk PJ, Biere A (2010b) Changes in genomic methylation patterns during the formation of triploid asexual dandelion lineages. Mol Ecol 19:315–324

Vrijenhoek RC (1979) Factors affecting clonal diversity and coexistence. Am Zool 19:787–797
Vrijenhoek RC (1984) Ecological differentiation among clones the frozen niche variation model. In: Woehrmann K, Loeschcke V (eds) Population biology and evolution. Springer, Berlin, pp 217–232
Vrijenhoek RC, Parker ED Jr. (2009) Geographical parthenogenesis: general purpose genotypes and frozen niche variation. In: Schön I, Martens K, van Dijk P (eds.), Lost Sex, The evolutionary biology of parthenogenesis Springer: Germany, pp 99–131
Wagenius S, Lonsdorf E, Neuhauser C (2007) Patch aging and the S-Allee effect: breeding system effects on the demographic response of plants to habitat fragmentation Am. Nat. 169: 383–397
Wang X-Y, Paterson AH (2011) Gene conversion in Angiosperm genomes with an emphasis on genes duplicated by polyploidization. Genes 2: 1–20; doi:10.3390/genes2010001
You W, Fan S, Yu D, Xie D, Liu C (2014) An invasive clonal plant benefits from clonal integration more than a co-occurring native plant in nutrient-patchy and competitive environments. PLoS ONE 9(5):e97246. doi:10.1371/journal.pone.0097246
Xia J, Lu J, Wang ZX, Hao BB, Wang HB, Liu GH (2013a) Pollen limitation and Allee effect related to population size and sex ratio in the endangered Ottelia acuminata (Hydrocharitaceae): implications for conservation and reintroduction. Plant Biol. 15: 376–383. doi:10.1111/j.1438-8677.2012.00653.x
Xia J, Sun SG, Liu GH (2013b) Evidence of a component Allee effect driven by predispersal seed predation in a plant (Pedicularis rex, Orobanchaceae). Biol Lett 9: 20130387. http://dx.doi.org/10.1098/rsbl.2013.0387
Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol 197:314–322

Chapter 17
Elopomorpha (Teleostei) as a New Model Fish Group for Evolutionary Biology and Comparative Genomics

Jhen-Nien Chen, Sarah Samadi and Wei-Jen Chen

Abstract Genome sequencing and related advances in research on model systems have provided many new insights into major questions of comparative genomics and evolutionary biology. However, without a comprehensive sampling of organisms that reflects the phylogenetic diversity of the studied taxa, the obtained conclusion may be distorted. Indeed, the existing model systems in Teleostei belong exclusively to one of the three main groups: the Clupeocephala (the others being Osteoglossomorpha and Elopomorpha). In this review, we propose the Elopomorpha, the basal-most lineage, as a new model group. General background on the taxa as well as research progress on their evolution including the state of the art of phylogenetic and genomic hypotheses is presented. We also discuss further directions for the research focusing on the Elopomorpha.

17.1 Introduction

The Elopomorpha (tarpons, bonefishes, eels, and relatives) is one of the three major extant teleost lineages, which includes 996 species (Eschmeyer and Fong 2015). The two other teleost lineages are Osteoglossomorpha (bony tongues) and Clupeocephala (all remaining teleosts) (Fig. 17.1). An indeterminate species of the genus *Elopsomolos* (*Elopsomolos* sp. 1) from the Kimmeridgian Fossil-Lagerstätte of Schamhaupten, Germany, documents the existence of the elopomorph fishes as early as in Jurassic (Arratia 2000). Available molecular results date the origin of elopomorphs to 150–200 million years ago, which is consistent with the fossil record (Alfaro et al. 2009; Near et al. 2012). After their appearance, the

J.-N. Chen · W.-J. Chen (✉)
Institute of Oceanography, National Taiwan University, 10617 Taipei, Taiwan
e-mail: wjchen.actinops@gmail.com

S. Samadi
Institut de Systématique, Évolution, Biodiversité—UMR 7205—CNRS, MNHN, UPMC, EPHE—Muséum National d'Histoire Naturelle, Sorbonne Universités, 75005 Paris, France

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_17

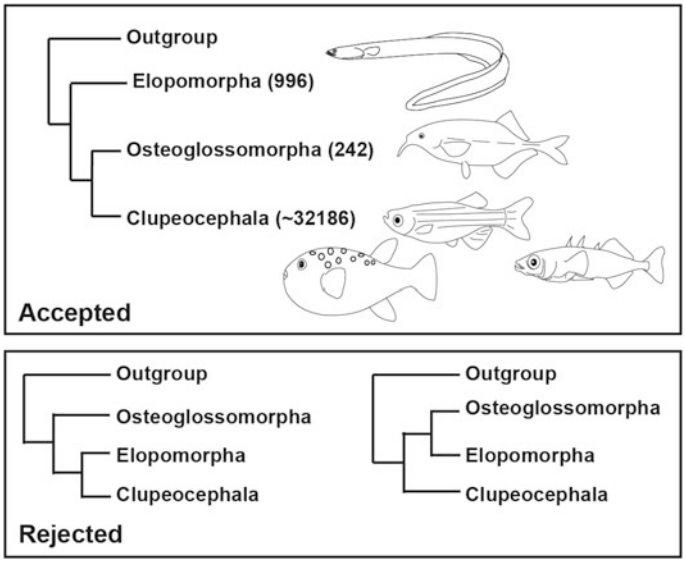


Fig. 17.1 Alternative hypotheses of interrelationship of three main teleost groups. Species numbers for each group referencing the Catalog of Fishes (Eschmeyer and Fong 2015) are indicated within *parentheses*. Illustration on *top* presents the currently accepted hypothesis

Elopomorpha experienced several episodes of radiation, leading to diverse descendant lineages living in various habitats from tropical to temperate and from shallow water to deep sea (Fig. 17.2). They are among the most successful vertebrate groups on the Earth and are important components of modern marine fauna.

The elopomorph fishes vary widely in their morphology, behavior, and life history (Fig. 17.2). Some of them, ladyfishes (Elopiformes: Elopidae), tarpons (Elopiformes: Megalopidae), and bonefishes (Albuliformes: Albulidae), look like primitive teleosts in retaining several ancestral features. Spiny eels and halosaurs (Notacanthiformes) have somewhat elongate body and may live in depths up to 2000 m. The true eels (Anguilliformes) are the most species-rich elopomorph group and are characterized by long cylindrical bodies with several reduced characters (e.g., no pelvic fins, and in some species, lack of pectoral fins). Most anguilliforms are marine fishes. The Anguillidae (18 species) is the only anguilliform lineage that has secondarily adapted to freshwater thanks to a catadromous life cycle (Aoyama 2009). Interestingly, a recent molecular study based on whole mitochondrial genomes showed evidence for a deep oceanic origin of these freshwater eels (Inoue et al. 2010). Gulper eels ("Saccopharyngiformes") are of the most extraordinary deep-sea living vertebrates (living as deep as 3000 m) with a unique appearance characterized by extremely large jaws and V-shaped myomeres instead of W-shaped ones as in all other fishes (Bertelsen et al. 1989). They are likely the most closely related taxa to these freshwater eels (Chen et al. 2014a; Santini et al. 2013). However, unlike the

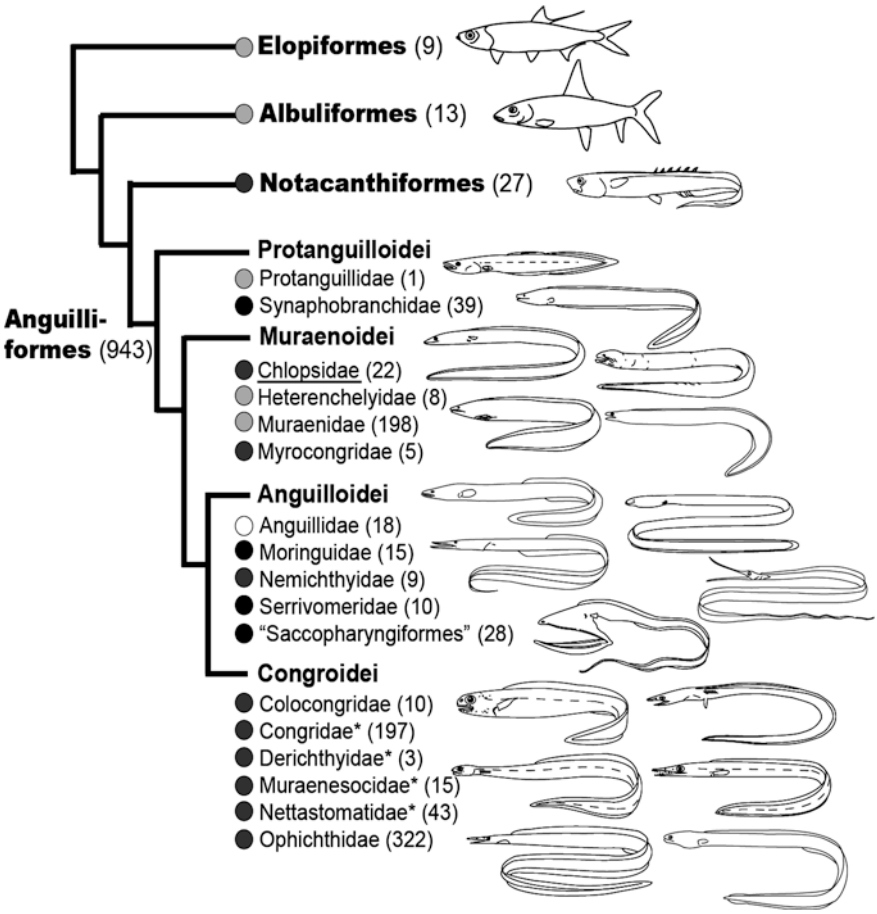


Fig. 17.2 Summary of evolutionary relationships among major elopomorph groups according to Chen et al. (2014a) and Santini et al. (2013). The order "Saccopharyngiformes" (including four families) should be discarded because it is nested within the anguilliform suborder, Anguilloidei. Taxonomic status of the Chlopsidae (underlined) is uncertain. *Asterisk* indicates the families that do not appear to be monophyletic from the newly obtained phylogenetic results. *Filled circles* indicate the different habitat preference of taxon (Froese and Pauly 2014). *Light-gray*, *dark-gray*, and *black circles* represent depth of 0–200 m, 200–1000 m, and deeper than 1000 m, respectively; the taxon with *open circle* proceeds a catadromous life cycle (i.e., their adults live in freshwater). Species numbers for each group referencing the Catalog of Fishes (Eschmeyer and Fong 2015) are indicated within *parentheses*

species from the Anguillidae on which extensive research has been conducted because of their commercial importance (e.g., European and Japanese eels) (Righton and Walker 2013), fewer research efforts have been made for other elopomorphs such as the gulper eels due mainly to availability of examinable specimens.

Specimen collection may be very challenging, especially from the deep sea. For instance, the specimens from gulper eels are scant. They are mainly accessible from deep-sea fishermen who find them casually in their nets.

17.2 Phylogeny and Classification: State of the Art

A comprehensive and resolved phylogeny is needed for testing some hypotheses in comparative biology such as the genomic evolution of the Teleostei or the Elopomorpha. However, the phylogenetic relationships of the Elopomorpha (at both intra-group and infra-group levels) have been a subject of debates in the past due primarily to the extensive morphological diversity of elopomorph adults that renders difficult securing adequate or homologous morphological characters used for phylogenetic inference. For instance, although their monophyly is currently accepted based on both molecular and morphological evidences (Chen et al. 2014a, b; Forey et al. 1996; Inoue et al. 2004; Nelson 2006; Wiley and Johnson 2010), it has sometimes been questioned (Filleul and Lavoué 2001; Gosline 1971; Hulet and Robins 1989). Wiley and Johnson (2010) listed five morphological synapomorphies, which support the monophyly of the Elopomorpha. Among these synapomorphies, the most remarkable character is the leptocephalus larval morphology (Greenwood et al. 1966; Hulet and Robins 1989). The leptocephalus is a pelagic larval form characterized by extreme dorsoventral compression, and willow leaf-shaped and highly translucent bodies (Hulet and Robins 1989). Leptocephali shrink during metamorphosis to the juvenile and adult forms (Fig. 17.3). Actually, the Elopomorpha is the only fish group whose development after hatching involves a conspicuous and relatively abrupt change in the body structure. How such a change is genetically determined through cell growth and differentiation is still unclear (Miller 2009; Pfeiler 1999).

Among the three main teleost groups (Elopomorpha, Osteoglossomorpha, and Clupeocephala; Fig. 17.1), all three possible alternative hypotheses of their interrelationship have been proposed (Arratia 1997; Greenwood et al. 1966; Lauder and

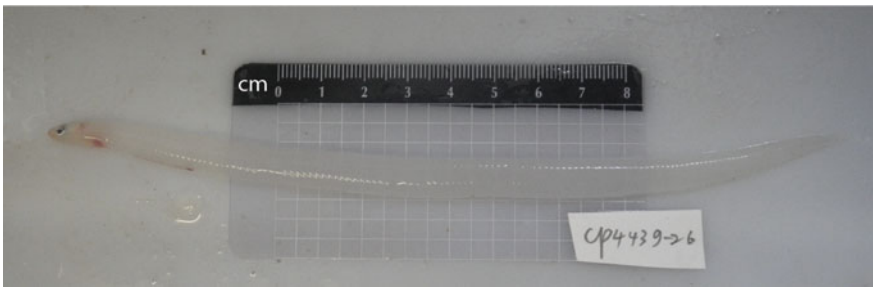


Fig. 17.3 An anguilliform species in its intermediate stage of metamorphosis between leptocephalus and juvenile caught at the depth of 1400 m

Liem 1983; Lê et al. 1993; Patterson and Rosen 1977). The consensus view reached by the recent multiple nuclear gene phylogenetic analyses (or phylogenomic analyses) shows that Elopomorpha is sister to the rest of teleosts (Fig. 17.1) (Alfaro et al. 2009; Chen et al. 2014a; Near et al. 2012; Faircloth et al. 2013).

Within the Elopomorpha, four to five orders were proposed to classify the around 1000 extant species in 25 families (Chen et al. 2014a; Eschmeyer and Fong 2015; Forey et al. 1996; Inoue et al. 2004; Nelson 2006; Obermiller and Pfeiler 2003; Wiley and Johnson 2010). The most up-to-date phylogenetic relationships among main elopomorph lineages based on molecular systematic results are summarized in Fig. 17.2. Briefly, gulper eels ("Saccopharyngiformes") are one of the lineages nested within the true eels or Anguilliformes (Chen et al. 2014a, b; Inoue et al. 2004, 2010; Johnson et al. 2012; Tang and Fielitz 2012). The monophyly of Albuliformes sensu lato [see Nelson (Nelson 2006)'s classification] is not recovered (Chen et al. 2014a, b; Santini et al. 2013). Indeed, the family Albulidae (i.e., the Albuliformes sensu Forey et al., 1996) is not the sister group to the Notacanthiformes. Thus, the elopomorph fishes should be redefined in four orders, namely Elopiformes, Albuliformes sensu Forey et al. 1996, Notacanthiformes, and Anguilliformes (inclusive of "Saccopharyngiformes"), and the "Saccopharyngiformes" should be discarded (Fig. 17.2) (Chen et al. 2014a; Santini et al. 2013; Tang and Fielitz 2012). Finally, within the Anguilliformes, four monophyletic groups (Protanguilloidei, Muraenoidei, Anguilloidei, and Congroidei) should be included (Chen et al. 2014a, b). It should also be noted that the monophyly of some anguilliform families from traditional point of views, such as, Congridae, Derichthyidae, Nettastomatidae, and Muraenesocidae, should be rejected (Chen et al. 2014a; Santini et al. 2013; Tang and Fielitz 2012). Protanguilloidei includes *Protanguilla palau*, the only species in its family, Protanguillidae. The individuals are presently known from a single fringing reef cave at 35 m depth in Palau discovered recently (Johnson et al. 2012). Based on morphology and mitochondrial genomic evidences, it was hypothesized that this family diverged earlier than other true eels (Johnson et al. 2012). However, the recent phylogenomic analyses show robust evidence of its sister group relationship to another member of the Protanguilloidei, the Synaphobranchidae (Chen et al. 2014a; Santini et al. 2013). Despite the recent efforts, the sister group relationships among $\sim$1000 species and their accurate taxonomy remain largely unresolved (Chen et al. 2014a; Dornburg et al. 2015; Santini et al. 2013; Tang and Fielitz 2012). Future work with denser sampling concomitant with the morphological evidences is still required to elucidate the evolutionary history and classification of this diverse group of fishes.

17.3 Model Fishes

Whole- or partial-genome sequencing of model systems has been helping us to resolve many questions in comparative genomics and evolutionary biology. Such data allow robustly testing some hypotheses such as whole-genome or gene

duplications, organismal phylogeny, and mechanisms underlying organismal speciation and adaptation. (Bernardi 2013; Chen and Mayden 2010; Nakamura et al. 2013; Spaink et al. 2014; Volff 2005). So far, the whole-genome sequences are available from the following fish models: zebrafish (*Danio rerio*) (Broughton et al. 2001), medaka (*Oryzias latipes*) (Kasahara et al. 2007), puffer fishes (*Takifugu rubripes* and *Tetraodon nigroviridis*) (Aparicio et al. 2002; Jaillon et al. 2004) and stickleback (*Gasterosteus aculeatus*) (Jones et al. 2012). The whole genomes of few cichlids have also been obtained (Brawand et al. 2014).

These model fishes include some laboratory models for vertebrates (such as zebra fish and medaka) (Howe et al. 2013; Porazinski et al. 2011). They have also been the subject of abundant researches in evolutionary developmental biology, speciation, and behavior (e.g., cichlids and sticklebacks) (Bernardi 2013; Chen and Mayden 2010; Huntingford and Ruiz-Gomez 2009; Irschick et al. 2013; Jones et al. 2012; Muschick et al., 2012). All these model fishes belong to the Clupeocephala (Fig. 17.1), but none to Osteglossomorpha. Although the Elopomorpha represents an astonishing diversity, the available information for comparative biology from the Elopomorpha is still scarce. Only two initiatives have selected representatives from the Elopomorpha for whole-genome sequencing and related analyses, but these studies are restricted to anguillid species (Japanese and European eels) (Coppe et al. 2010; Henkel et al. 2012a, b). Coppe et al. (2010) published the first European eel transcriptome data obtained by next-generation sequencing technique (NGS) with 454 pyrosequencing methodology of a normalized cDNA library, produced from a pool of juvenile eels. Over 310,000 reads were assembled in a total of about twenty thousand transcribed contigs. Overall 36 % of the contigs were annotated to known protein/nucleotide sequences and 35 putative miRNAs were identified. A database (EeelBase) of annotated transcriptome sequences of the European eel was created and available at http://compgen.bio.unipd.it/eeelbase. The first draft genome sequence of the Japanese eel assembled using a whole-genome shotgun sequencing strategy based on NGS with Illumina technology was published by Henkel et al. (2012b). The total assembled genome has a size of 1.15 Gbp (haploid), which is divided over thirty thousand scaffolds. Finally, draft sequences of the European eel genome obtained from a female juvenile are available from the same research team (Henkel et al. 2012a). The sequence data from these draft genomes can be accessed from the EELGENOME Web site (www.eelgenome.org) or from NCBI GenBank (see: Henkel et al. 2012a, b).

17.4 Elopomorpha, an Attractive Model Fish Group for Studying Genomic Evolution of the Teleostei

It has been shown that during the early evolution of the vertebrates, two rounds of whole-genome duplication (1R/2R WGD) have occurred (Dehal and Boore 2005). It was also inferred that an additional WGD (3R) event, known as fish-specific

genome duplication (FSGD), occurred in the common ancestor of teleost fishes (Van de Peer et al. 2009). The fact that teleost fishes often contain more copies of many genes than other vertebrates (e.g., opsin genes) could be the result of this extra run of genome-wide duplication event during their evolution (Chen and Mayden 2010; Christoffels et al. 2004; Van de Peer et al. 2009). According to the famous theory of evolution by gene duplication proposed by Susumu Ohno (1970), most duplicated genes are secondarily lost, yet it is assumed that these duplicates (when being maintained) may enable the evolution of sub- and/or novel gene functions, which might explain the great diversity of vertebrate/teleost species (Arnegard et al. 2010; Chen and Mayden 2010; Glasauer and Neuhauss 2014; Hoegg et al. 2004; Santini et al. 2009). Consequently, teleosts are seen as the best model systems to investigate biodiversity in relation to gene duplication or large-scale genomic change thanks to this specific event of ancient genome duplication or FSGD (Chen and Mayden 2010; Volff 2005). Since elopomorph fishes are the basal-most teleosts (Fig. 17.1) and that the origin of their common ancestor is estimated close to the time of the FSGD, several questions are interesting to be addressed. For instance, we may wonder whether or not the elopomorph fishes may have more opportunities to maintain the duplicated genes in their genomes? If yes, how these duplicated genes can be maintained from evolutionary perspective? Do the gene duplications (if any) correspond to the genome-wide duplication events (e.g., FSGD) or simply to some random gene duplications during their evolution? Is the diversification of elopomorph species a consequence of gene duplications? Actually, a single-copy nuclear gene (early growth response 3 gene, *EGR3*) in vertebrate genomes that has successfully been used as a phylogenetically informative marker on many studies investigating evolutionary relationships of diverse teleost groups (e.g., Chen et al. 2008; Santini et al. 2013; Chen et al. 2014b) was found to have an extra copy in the genomes of the Elopomorpha by a previous study (Chen et al. 2014a). This result and others (see the cases shown below) may support our supposition mentioned above.

17.5 *Hox* Gene Evolution Within the Teleostei and Elopomorpha

Hox genes (homeotic genes) encode transcription factors and contain highly conserved DNA sequences known as homeobox. *Hox* genes are associated with specification of body axial patterning, appendages, and organ system developments of animals (Duboule 2007; McGinnis and Krumlauf 1992). Duplicated or paralogous *hox* genes are tightly organized with conservative gene order into cluster(s) in animals (McGinnis and Krumlauf 1992). While earlier evolved animal lineages have only a single cluster in their genomes, teleosts should have retained eight duplicated clusters theoretically because of the three rounds of the vertebrate ancient genome duplications mentioned above. However, in some teleost genomes

(e.g., zebrafish, puffer fish, and medaka), *hox* genes were found to retain at most seven clusters (Amores et al. 2004; Kurosawa et al. 2006). The current comparative genomic studies revealed that the Elopomorpha, the earliest evolved teleost lineage, contains eight *hox* clusters.

Through a PCR-based survey in the Japanese eel and draft genome assembling in the European eel, at least eight duplicated *hox* gene clusters were identified (Guo et al. 2010; Henkel et al. 2012a). Further investigation with mRNA sequencing and in situ hybridizations indicated that all copies of *hox* genes are functional and expressed in early embryos (Henkel et al. 2012a). Compared to the other teleost genomes (zebrafish and puffer fish) for which *hox* clusters have been identified (Amores et al. 2004; Woltering and Durston 2006), the eels retain fully populated and duplicate *hox* clusters that might be the result of the FSGD (Guo et al. 2010; Henkel et al. 2012a). It has been hypothesized that the retention of *hox* gene clusters might provide additional opportunities of developmental complexity (Guo et al. 2010; Henkel et al. 2012a). In fact, the anguillid eels have evolved a catadromous life cycle, i.e., after the stage of leptocephalus larvae in open sea, metamorphosed juveniles (Fig. 17.3) and adults live in estuarine and freshwater habitats and then migrate to open sea for spawning (Aoyama 2009). A leptocephalus larva provides considerable survival benefits with fully transparent body and slowly metabolic rate (Bishop and Torres 1999; Pfeiler 1999). Henkel et al (2012a) demonstrated that nearly all *hox* genes were expressed in eels' early embryos and presumably functionally involved in determining cell fate during the embryonic and larval development. Indeed, as mentioned, leptocephalus larvae are an evolutionary innovation in the Elopomorpha since no other ray-finned fishes are known to have such the larval form. This fundamental developmental innovation may be the key for the explanation of retention of a maximum set of eight *hox* clusters in anguillid or all elopomorph genomes (Henkel et al. 2012a). Advanced research with other elopomorphs will be needed to elucidate the *hox* genes evolution and their role on metamorphosis within the Teleostei and the Elopomorpha.

17.6 Elopomorpha as a Model Group to Study Species Diversification in Relation to Gene Duplication

In addition to morphological and species diversity, elopomorph fishes exhibit an exceptional ecological diversity. Their habitats include sandy shore, coral reef, pelagic ocean, and deep-sea benthos. Tarpon and ladyfish (Elopiformes) are coastal fishes that can survive in estuarine environment where salinity is lower than the open sea. Bonefishes (Albuliformes) occur primarily in shallow, coastal waters on sand, or mud bottoms, in areas of relatively high salinity. Notacanthiform fishes such as halosaurs and spiny eels (Notacanthidae) live on or near the bottom in moderate to deep waters (about 500 to 3000 m). Although most of the anguilliform fishes are bathypelagic or bathydemersal, an idiosyncratic one, *P. palau*, was discovered in a shallow water reef (Johnson et al. 2012). Another reef-associated eels are moray eels

(Muraenidae). Besides, as mentioned earlier, anguillid eels and gulper eels are closely related, but they occupied two extreme sides of aquatic environments, freshwater and deep sea, respectively. This high ecological diversity and genomic complexity mentioned above make elopomorphs an ideal system to investigate the gene evolution in relation to species diversification and evolutionary adaptation, especially in new environmental niches (e.g., shallow water to deep sea).

17.7 Do Gene Duplication Play an Important Role in Adaptation of Visual Systems in Deep-Sea Teleost (Elopomorph) Fishes?

For the organisms living in the deep ocean, the physiological and/or evolutionary adaptation needed to overcome severe physical and chemical stresses (e.g., limited light source) is essential. The evolution and functional divergence of retinal opsin genes in fishes is a great example for this aspect especially on functional evolution accompanying gene duplication (Chen and Mayden 2010; Collin and Darwin 2005).

In vertebrates, the opsin genes or visual pigment genes are often expressed in either retinal cone cells (for cone opsin genes) or rod cells (for the rhodopsin gene, *RH1*) that mediate the essential step of color discrimination and image capture of organisms. The cone opsin genes include four classes (the short-wavelength sensitive class 1 and class 2, *SWS1* and *SWS2*; the medium- to long-wavelength-sensitive class, M/LWS; the mid-wavelength (green)-sensitive gene class, *RH2*), and their maximum wavelength absorption (λ_{max}) of light spectrum is between 355 and 570 nm which is located at the wavelength ranges of color and ultraviolet light spectrum (Fig. 17.4a) (Collin and Darwin 2005). Hence, for organisms living in a sufficient light environment (e.g., reef and freshwater lake), cone opsins play an important role in their color vision and subsequently in their evolution. Indeed, variations in organismal spectral sensitivity may have arisen through molecular variations of the cone opsin genes, evolution of gene regulation, and gene duplications (Chen and Mayden 2010; Sabbah et al. 2010). It is believed that such a visual adaptation has played a key role in the explosive radiation of cichlids from African Great Lakes (Bernardi 2013; Kocher 2004; Sabbah et al. 2010; Seehausen et al. 2008). However, in marine habitats, light source is often limited. With increasing depth, both the intensity and the spectral quality of light reduce. Actually, the spectrum in water becomes restricted to a narrow waveband of blue light (470–490 nm) and eventually faded to complete darkness in deep sea (below 200 m) (Fig. 17.4b) (Beebe 1935). Downwelling daylight and bioluminescence remain major light sources in deep-sea environment. Contrary to cone opsins, the rod opsin or rhodopsin (*RH1*) is responsible for the perception of light. Thus, the rhodopsin is essential for aquatic organisms, especially the deep-sea fishes, since it is sensitive to the restricted dim-light environment. The marine fishes rely on this sense of image caption to find food and mates and to maintain various interspecific and intraspecific associations that have a selective effect on their fitness.

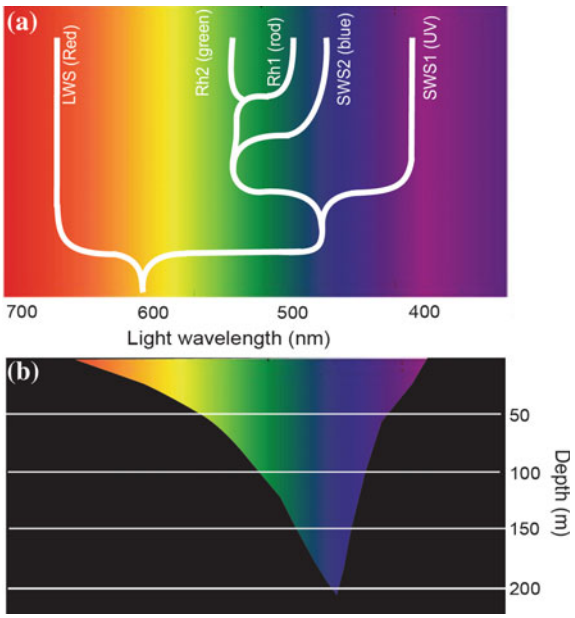


Fig. 17.4 **a** Phylogeny of the five major vertebrate opsin genes. The position of each branch on the background spectrum approximates the light spectral sensitivity of each opsin (figure modified from Collin and Darwin 2005). **b** General diagram of light spectrum penetration pattern in the open ocean at different depths

The adaptation of the visual systems of deep-sea fishes has been deduced from molecular studies. It has been shown that the peak of maximal absorbance (λ_{max}) in water agrees with the λ_{max} of rhodopsin carried by its fish host. A short-wavelength shift (from the normal value of rhodopsin of 500 nm), observed in the rhodopsin of deep-sea fishes, is believed to be produced by mutation of some critical amino acids sites during evolutionary time (Yokoyama 1997, 2008). Comparison of amino acid sequences of rhodopsin from few deep-sea fishes and others living in shallow waters confirms this hypothesis (Hope et al., 1997; Hunt et al. 2001; Yokoyama 2008), but there are few exceptions. Actually, the taxonomic sampling from these early papers is small and limited and does not cover the diverse spectra found for other fishes. A PhD work, using a sampling from a diverse set of 86 acanthomorphs (spiny-rayed teleost fishes from Clupeocephala), showed that there is no simple relationship between mutations at these critical positions and the spectral fit of the visual system of a fish to the light in which it lives (Chen 2001). Thus, a further examination with more rhodopsin gene data will be required to address this question. As elopomorph fishes exhibit a high diversity of habitats, research conducted using these fishes will provide potential opportunities for testing if or not the sampling bias may be the primary explanation for the disagreement of the results among the previous studies.

Nonetheless, if having rhodopsin with a lower λ_{max} is essential for the deep-sea fishes because of environmental constraint, this discordance [as shown in Chen (2001)] could have some other explanations. The molecular adaptation mechanism indicated above seems less parsimonious. “Waiting” for a specific amino acid replacement seems like a long and unpredictable way to achieve adaptation for particular light conditions. By contrast, other physical or developmental mechanisms

seem more feasible (Carleton et al. 2008; Yokoyama 1997). That is, a fish can adjust its levels of rhodopsin expression and thus to achieve concordance of the λ_{max} for rhodopsin and water. By using these mechanisms, the fish can adapt to its environment even with uncoordinated rhodopsin (i.e., without the expected mutation at the critical amino acid sites). In fact, it is known that fishes living in deep sea have evolved diverse visual system features allowing them to deal with the low dim-light environments, such as enlarged eyes or rod-cell-only retina to catch and utilize the light photons as many as possible. Some deep-sea fishes have well-developed luminescent system (producing and receiving the luminescence) to communicate or defend predators (Kenaley et al. 2014; Turner et al. 2009). In addition, mesopelagic fishes have great ability and can undergo diurnal vertical migrations to adjust their light needs.

Another hypothesis that may provide an explanation for this discordance is that fishes might have more than one rhodopsin gene. One of the other rhodopsins (or rhodopsin-like) might have concordant λ_{max} with water (resulting from amino acid mutations). Yet our standard protocols (directly sequencing from genomic DNA or from cDNA obtained from retinal tissue using standard primers) may fail to detect these paralogous copies, so that the observed existing rhodopsin sequences of deep-sea fishes do not have the expected mutations. Interestingly, two paralogous rhodopsin genes with different λ_{max} values (resulting from the amino acid mutation) have been shown in the Anguilliformes (from congers, Japanese and European eels) (Archer et al. 1995; Zhang et al. 2000; Zhang 2002). Expression of these two genes in Japanese eels in different sexual maturation stages helps them adapting to different environments during their life cycle (freshwater and deep sea) (Zhang et al. 2000). This kind of molecular compensatory mechanism might happen in other fishes such as deep-sea pearleyes (Aulopiformes from Clupeocephala) (Pointer et al. 2007). Although rhodopsin was historically considered as a single-copy gene in vertebrate genomes, with rare exceptions in few fishes mentioned above and in few cyprinids (Lim et al. 1997; Morrow et al. 2011), there are more and more evidences challenging this view (Chen and Mayden 2010; Mano et al. 1999; Morrow et al. 2011). For instance, a gene duplication event that might have occurred in common ancestor of all ray-finned fishes except the basal-most lineage, bichirs (Polypteriformes), gave rise to a novel rhodopsin (*ex-RH1*) of ray-finned fishes that is expressed in the pineal gland of the fish brain instead of retinal cells (Mano et al. 1999).

Through our current survey with gene-specific primers and PCR-based methods, we found that the Elopiformes, Albuliformes, and Notacanthiformes have single copy of rhodopsin gene, while most of the representatives of the Anguilliformes (except reef-associated Muraenidae) have two copies of rhodopsin gene (unpublished data). The question about whether these teleost-specific duplicated rhodopsin genes identified in elopomorph and other teleost fish genomes originated from the FSGD and then lost in some lineages is still uncertain. Nevertheless, based on the current results, it is likely that elopomorph fishes tend to maintain the duplicated rhodopsin gene during their evolution. Does the extra copy of gene express in the retinal cells in fish eyes? Are all the copies subject to simultaneous expression? Could these additional genes potentially provide a palette on which natural selection can act to tune fish eyes to fit into their visual environments under optimal

conditions? May different species use a different developmental pattern of gene expression? These questions when being answered will increase our understanding of opsin evolution in elopomorph fishes, the precise mechanisms of molecular adaptation described above for teleost fishes, and eventually highlight the potential consequence of gene duplications in diversification of the elopomorphs.

17.8 Conclusion and Perspectives

With the coalescence of advancements in molecular biology, informatics, and computational biology, a new era of phylogenomics in systematic biology has emerged, wherein large numbers of genomic data or DNA sequences are available for inferences of the sound Tree of Life of existing fish species (Chen and Mayden 2010). The established phylogenetic framework can be and should be used to examine a vast array of questions in comparative and evolutionary biology as well as biodiversity (Mayden 1992; Wiley 1981). Among the main teleost lineages, the Elopomorpha is resolved as the first-evolved one. Elopomorphs display a relatively high diversity (in morphology, ecology, and behaviors) and genomic complexity (a tendency to retain duplicated genes resulted perhaps from FSGD). The draft sequences of genomes and transcriptome data have recently become available for two anguillid species. All these make the Elopomorpha an ideal model for evolutionary and genomic studies in fishes. However, the recent studies were based only on part of the elopomorph diversity and generally restricted to a few commercially important species of the same family (e.g., Japanese and European eels). Such a bias in the taxonomic sampling may strongly distort our understanding of the causes of the observed pattern of diversity, thus of the evolution of the Elopomorpha, and more generally of the Teleostei. Increasing the sampling and on-site observations through more biodiversity exploration programs with a particular emphasis on deep-sea fauna are of the emergent needs for the future research.

Acknowledgments We appreciate the bilateral cooperation research funding from the Ministry of Science and Technology, Taiwan (MOST 102-2923-B-002 -001 -MY3), and the French National Research Agency (ANR 12-ISV7-0005-01) supporting this research.

References

Arratia G (1997) Basal teleosts and teleostean phylogeny. Palaeo Ichthyologica 7:5–168

Arratia G (2000) Remarkable teleostean fishes from the Late Jurassic of southern Germany and their phylogenetic relationships. Foss Rec 3:137–179

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Natl Acad Sci 106:13410–13414

Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH (2004) Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. Genome Res 14:1–10

Aoyama J (2009) Life history and evolution of migration in catadromous Eels (Genus *Anguilla*). Aqua-Bio 2:1–42

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J-M, Dehal P, Christoffels A, Rash S, Hoon S, Smit A et al (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310

Archer S, Hope A, Partridge JC (1995) The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. Proc Biol Sci 262:289–295

Arnegard ME, Zwickl DJ, Lu Y, Zakon HH (2010) Old gene duplication facilitates origin and diversification of an innovative communication system-twice. Proc Natl Acad Sci 107:22172–22177

Bernardi G (2013) Speciation in fishes. Mol Ecol 22:5487–5502

Beebe W (1935) Half Mile Down. The Bodley Head, London

Bertelsen E, Nielsen JG, Smith DG (1989) Suborder Saccopharyngoidei: families Saccopharyngidae, Eurypharyngidae, and Monognathidae. In: Bhlke EB (ed) Fishes of the western North Atlantic, Part 9, vol 1. Sears Foundation for Marine Research, New Haven, pp 636–655

Bishop R, Torres J (1999) Leptocephalus energetics: metabolism and excretion. J Exp Biol 202:2485–2493

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E et al (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381

Broughton RE, Broughton RE, Milam JE, Milam JE, Roe BA, Roe Ba (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Res 11:1958–1967

Carleton KL, Spady TC, Streelman JT, Kidd MR, McFarland WN, Loew ER (2008) Visual sensitivities tuned by heterochronic shifts in opsin gene expression. BMC Biol 6:22

Chen W-J (2001) La répétitivité des clades comme critère de fiabilité: application à la phylogénie des Acanthomorpha (Teleostei) et des Notothenioidei (acanthomorphes antarctiques). Ph.D., University of Paris VI

Chen W-J, Mayden RL (2010) A phylogenomic perspective on the new era of Ichthyology. Bioscience 60:421–432

Chen W-J, Miya M, Saitoh K, Mayden RL (2008) Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes (Ostariophysi) as a case study. Gene 423:125–134

Chen J-N, López JA, Lavoué S, Miya M, Chen W-J (2014a) Phylogeny of the Elopomorpha (Teleostei): evidence from six nuclear and mitochondrial markers. Mol Phylogenet Evol 70:152–161

Chen W-J, Santini F, Carnevale G, Chen J-N, Liu S-H, Lavoué S, Mayden RL (2014b) New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). Front Mar Sci 1:53

Christoffels A, Koh EGL, Chia JM, Brenner S, Aparicio S, Venkatesh B (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol 21:1146–1151

Collin SP, Darwin C (2005) Opsins : evolution in waiting. Curr Biol 15:794–796

Coppe A, Pujolar JM, Maes GE, Larsen PF, Hansen MM, Bernatchez L, Zane L, Bortoluzzi S (2010) Sequencing, *de novo* annotation and analysis of the first *Anguilla anguilla* transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel. BMC Genom 11:635

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3(10):e314

Dornburg A, Friedman M, Near TJ (2015) Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). Mol Phylogenet Evol 89:205−218

Duboule D (2007) The rise and fall of *Hox* gene clusters. Development 134:2549–2560

Eschmeyer WN, Fong JD (2015) Catalog of Fishes. Accessed May 2015
Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013) A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS ONE 8:e65923
Filleul A, Lavoué S (2001) Basal teleosts and the question of elopomorph monophyly. Morphological and molecular approaches. C R Acad Sci III 324:393–399
Forey PL, Littlewood DTJ, Ritchie P, Meyer A (1996) Interrelationships of elopomorph fishes. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Acadamic Press, San Diego, pp 175–191
Froese R, Pauly D (2014) FishBase. www.fishbase.org, version (11/2014)
Glasauer SMK, Neuhauss SCF (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics 289:1045–1060
Gosline WA (1971) Functional Morphology and Classification of Teleostean Fishes. The University Press of Hawaii, Honolulu
Greenwood PH, Rosen DE, Weitzman SH, Myers GS (1966) Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull Am Mus Nat Hist 131(4):339–456
Guo B, Gan X, He S (2010) Hox genes of the Japanese eel *Anguilla japonica* and *Hox* cluster evolution in teleosts. J Exp Zool B Mol Dev Evol 314:135–147
Henkel CV, Burgerhout E, de Wijze DL, Dirks RP, Minegishi Y, Jansen HJ, Spaink HP, Dufour S, Weltzien F-A, Tsukamoto K et al (2012a) Primitive duplicate *Hox* clusters in the European eel's genome. PLoS ONE 7:e32231
Henkel CV, Dirks RP, De Wijze DL, Minegishi Y, Aoyama J, Jansen HJ, Turner B, Knudsen B, Bundgaard M, Hvam KL et al (2012b) First draft genome sequence of the Japanese eel, *Anguilla japonica*. Gene 511:195–201
Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59:190–203
Hope AJ, Partridge JC, Dulai KS, Hunt DM (1997) Mechanisms of wavelength tuning in the rod opsins of deep-sea fishes. Proc Biol Sci 264:155–163
Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L et al (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503
Hulet WH, Robins CR (1989) The evolutionary significance of the leptocephalus larva. In Böhlke EB (ed) Fishes of the Western North Atlantic. Leptocephali. Part 9, vol 2. Sears Foundation for Marine Research, New Haven, pp 669–677
Hunt DM, Dulai KS, Partridge JC, Cottrill P, Bowmaker JK (2001) The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. J Exp Biol 204:3333–3344
Huntingford FA, Ruiz-Gomez ML (2009) Three-spined sticklebacks *Gasterosteus aculeatus* as a model for exploring behavioural biology. J Fish Biol 75:1943–1976
Inoue JG, Miya M, Miller MJ, Sado T, Hanel R, Hatooka K, Aoyama J, Minegishi Y, Nishida M, Tsukamoto K (2010) Deep-ocean origin of the freshwater eels. Biol Lett 6:363–366
Inoue JG, Miya M, Tsukamoto K, Nishida M (2004) Mitogenomic evidence for the monophyly of elopomorph fishes (Teleostei) and the evolutionary origin of the leptocephalus larva. Mol Phylogenet Evol 32:274–286
Irschick DJ, Albertson RC, Brennan P, Podos J, Johnson NA, Patek S, Dumont E (2013) Evo-devo beyond morphology: from genes to resource use. Trends Ecol Evol 28:267–273
Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A et al (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957
Johnson GD, Ida H, Sakaue J, Sado T, Asahida T, Miya M (2012) A "living fossil" eel (Anguilliformes: Protanguillidae, fam. nov.) from an undersea cave in Palau. Proc Biol Sci 279:934–943
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S et al (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y et al (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447(7145):714–719

Kenaley CP, Devaney SC, Fjeran TT (2014) The complex evolutionary history of seeing red: molecular phylogeny and the evolution of an adaptive visual system in deep-sea dragonfishes (Stomiiformes: Stomiidae). Evolution 68:996–1013

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288–298

Kurosawa G, Takamatsu N, Takahashi M, Sumitomo M, Sanaka E, Yamada K, Nishii K, Matsuda M, Asakawa S, Ishiguro H et al (2006) Organization and structure of *hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. Gene 370:75–82

Lauder GV, Liem KF (1983) Patterns of diversity and evolution in ray-finned fishes. In: Norrhcutt RG, Davis RE (eds) Fish Neurobiology. University of Muchigan Press, Ann Arbor, pp 1–24

Lê HLV, Lecointre G, Perasso R (1993) A 28S rRNA-based phylogeny of the gnathostomes: first steps in the analysis of conflict and congruence with morphologically based cladograms. Mol Phylogenet Evol 2:31–51

Lim J, Chang JL, Tsai HJ (1997) A second type of rod opsin cDNA from the common carp (*Cyprinus carpio*). Biochim Biophys Acta 1352:8–12

Mano H, Kojima D, Fukada Y (1999) Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland. Mol Brain Res 73:110–118

Mayden RL (ed) (1992) Systematics, historical ecology, and North American freshwater fishes. Stanford University Press, Stanford

McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. Cell 68:283–302

Miller M (2009) Ecology of Anguilliform leptocephali: remarkable transparent fish larvae of the ocean surface layer. Aqua-Bio Sci Monogr 2:1–94

Morrow JM, Lazic S, Chang BSW (2011) A novel rhodopsin-like gene expressed in zebrafish retina. Vis Neurosci 28:325–335

Muschick M, Indermaur A, Salzburger W (2012) Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol 22:2362–2368

Nakamura Y, Mori K, Saitoh K, Oshima K, Mekuchi M, Sugaya T, Shigenobu Y, Ojima N, Muta S, Fujiwara A et al (2013) Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proc Natl Acad Sci 110:11061–11066

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL (2012) Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci 109:13698–13703

Nelson JS (2006) Fishes of the World, 4th edn. John Wiley & Sons Inc, New York

Obermiller LE, Pfeiler E (2003) Phylogenetic relationships of elopomorph fishes inferred from mitochondrial ribosomal DNA sequences. Mol Phylogenet Evol 26:202–214

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, New York

Patterson C, Rosen DE (1977) Review of ichthyodectiformes and other Mesozoic teleost fishes and the theory and practice of classifying fossils. Bull Am Mus Nat Hist 158:81–172

Pfeiler E (1999) Developmental physiology of elopomorph leptocephali. Comp Biochem Physiol Part A 123:113–128

Pointer MA, Carvalho LS, Cowing JA, Bowmaker JK, Hunt DM (2007) The visual pigments of a deep-sea teleost, the pearl eye *Scopelarchus analis*. J Exp Biol 210:2829–2835

Porazinski S, Wang H, Furutani-Seiki M (2011) Essential techniques for introducing medaka to a zebrafish laboratory—towards the combined use of medaka and zebrafish for further genetic dissection of the function of the Vertebrate genome. In: Pelegri FJ (ed) Vertebrate embryogenesis SE- 8, methods in molecular biology. Humana Press, New York City, pp 211–241

Righton D, Walker AM (2013) Anguillids: conserving a global fishery. J Fish Biol 83:754–765

Sabbah S, Laria RL, Gray SM, Hawryshyn CW (2010) Functional diversity in the color vision of cichlid fishes. BMC Biol 8:133

Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin of teleosts? a comparative study of diversification in ray-finned fishes. BMC Evol Biol 9:194

Santini F, Kong X, Sorenson L, Carnevale G, Mehta RS, Alfaro ME (2013) A multi-locus molecular timescale for the origin and diversification of eels (Order: Anguilliformes). Mol Phylogenet Evol 69:884–894

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi Y, van der Sluijs Y, Schneider MV, Maan ME, Tachida H, Imai H, Okada N (2008) Speciation through sensory drive in cichlid fish. Nature 455:620–626

Spaink HP, Jansen HJ, Dirks RP (2014) Advances in genomics of bony fish. Brief Funct Genomics 13:144–156

Tang KL, Fielitz C (2012) Phylogeny of moray eels (Anguilliformes: Muraenidae), with a revised classification of true eels (Teleostei: Elopomorpha: Anguilliformes). Mitochondrial DNA 24:55–66

Turner JR, White EM, Collins MA, Partridge JC, Douglas RH (2009) Vision in lanternfish (Myctophidae): adaptations for viewing bioluminescence in the deep-sea. Deep Res Part I Oceanogr Res Pap 56:1003–1017

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nature Rev Genet 10:725–732

Volff JN (2005) Genome evolution and biodiversity in teleost fish. Heredity 94:280–294

Wiley EO, Johnson GD (2010) A teleost classification based on monophyletic groups. In: Nelson JS, Schultze H-P, Wilson MVH (eds) Oringin and Phylogenetic Interrelationships of Teleosts. Verlag Dr. Friedrich Pfeil, Germany, pp 123–182

Wiley RH (1981) Social structure and individual ontogenies: problems of description, mechanism, and evolution. In: Bateson PPG, Klopfer PH (eds) Perspecitves in Ethology, vol 4 Plenum Press, New York, pp 105–133

Woltering JM, Durston AJ (2006) The zebrafish *hoxDb* cluster has been reduced to a single microRNA. Nat Genet 38:601–602

Yokoyama S (1997) Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. Annu Rev Genet 31:315–336

Yokoyama S (2008) Evolution of dim-light and color vision pigments. Annu Rev Genomics Hum Genet 9:259–282

Zhang H (2002) Isolation of freshwater and deep-sea type opsin genes from the common Japanese conger. J Fish Biol 61:313–324

Zhang H, Futami K, Horie N, Okamura A, Utoh T, Mikawa N, Yamada Y, Tanaka S, Okamoto N (2000) Molecular cloning of fresh water and deep-sea rod opsin genes from Japanese eel *Anguilla japonica* and expressional analyses during sexual maturation. FEBS Lett 469:39–43

Chapter 18
Looks Can be Deceiving: Cryptic Species and Phenotypic Variation in *Rhodnius* spp., Chagas Disease Vectors

Márcio G. Pavan, Gustavo B.S. Rivas, Fernando B.S. Dias and Rodrigo Gurgel-Gonçalves

Abstract The aim of this chapter was to highlight the importance of applying morphological, ecological, behavioral, and molecular methods to analyze taxonomic problems among Chagas disease vector species. We structured the chapter as follows: an introductory section about the disease and the reason why studies on cryptic species, phenotypic variation, and ecological niches in *Rhodnius* spp. are relevant for the interruption of disease transmission and two sections containing general aspects of Chagas disease in three Latin American biomes (Amazon, Cerrado, and Caatinga), and taxonomic problem-solving examples. Finally, we present a section containing future trends in molecular systematics and behavior studies that might be useful for developing new vector control and surveillance strategies. Although this chapter is focused on insect vector species, any reader interested in ecology and molecular systematics will find valuable guidance on how to design a study that aims to answer taxonomic questions involving closely related species.

M.G. Pavan (✉)
Laboratório de Epidemiologia e Sistemática Molecular, Instituto Oswaldo Cruz, Av. Brasil, 4365, Pav. Leonidas Deane, sl. 308, Fiocruz, Manguinhos, Rio de Janeiro, RJ 21040-360, Brazil
e-mail: marciopavan@gmail.com

M.G. Pavan · G.B.S. Rivas
Laboratório de Biologia Molecular de Insetos, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil

F.B.S. Dias
Grupo de Pesquisa Triatomíneos, Centro de Pesquisas René Rachou, Fiocruz, Belo Horizonte, Minas Gerais, Brazil

F.B.S. Dias
Grupo de Pesquisa Mosquitos Vetores: Endossimbiontes e Interação Patógeno-Vetor, Centro de Pesquisa René Rachou, Fiocruz, Belo Horizonte, Minas Gerais, Brazil

R. Gurgel-Gonçalves
Laboratório de Parasitologia Médica e Biologia de Vetores, Área de Patologia, Faculdade de Medicina, Universidade de Brasília, Brasília, Distrito Federal, Brazil

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_18

18.1 Introduction

18.1.1 Neglected Tropical Diseases—Chagas Disease

Neglected tropical diseases are present in financially deprived regions of the Americas, Africa, Asia, and Oceania, affecting more than 1.4 billion people (Kline et al. 2013; WHO 2015). Four of the seventeen diseases inserted in this context are caused by pathogens which vectors are hematophagous insects and are present in Latin America—Chagas disease, dengue, filariasis, and leishmaniasis. The major goal aiming the interruption of disease transmissions is the development of new strategies for vector control, which should be preceded by a better understanding about the biology of these insects (WHO 2007).

The strategy used in the last 25 years to control Chagas disease vectors served as a lesson for future development of more effective monitoring methods. Although most of these hemimetabolous insects with obligatory hematophagy (Hemiptera: Reduviidae) are potentially capable of becoming naturally infected by the etiological agent *Trypanosoma cruzi*, control programs focused their efforts to eliminate the most epidemiological relevant vectors—those capable of colonizing human dwellings (*e.g.*, *Triatoma infestans* and *Rhodnius prolixus*). The main strategy was based on spraying insecticides in triatomine-infested houses (Dias 2006). Despite the elimination of *T. infestans* in Uruguay, Chile, Brazil, eastern Paraguay, southern Peru, and parts of Argentina, and of *R. prolixus* in southern Mexico, Guatemala, Honduras, Nicaragua, El Salvador, and Costa Rica (summarized in Abad-Franch et al. 2013a), Chagas disease is still recorded in Latin America with 50,000 new cases and 12,500 deaths per year (Rassi and Marin-Neto 2010; Senior 2007).

One-hundred and forty eight triatomine species have already been formally described (Galvão and Gurgel-Gonçalves 2015) and the majority is competent vectors of *T. cruzi*. These species can occupy different natural ecotopes, as palm tree crowns, bird and small mammal nests, rock piles, hollow trees, and bat caves. Generally, triatomines tend to remain in their ecotopes without getting into contact with humans. However, habitat deforestation and human occupation can cause extinctions of sylvatic birds and mammals, diminishing blood sources for hematophagous insects, and, therefore, opening new niches for triatomines. Moreover, anthropogenic landscape disturbances can increase triatomine abundance and their infection rates in forest remnants (Gottdenker et al. 2011, 2012), which might facilitate *T. cruzi* transmission to humans and development of domestic tendencies in sylvatic populations of these insects.

The main difficulty to interrupt the transmission of Chagas disease using traditional methods is reinfestation events of insecticide-treated households by native vectors (Cecere et al. 2006; Fitzpatrick et al. 2008). Sylvatic and peridomestic populations of *T. infestans* commonly reinfest houses in the Bolivian Chaco, Paraguay, and Argentina (Abad-Franch et al. 2010b; Ceballos et al. 2011; Quisberth et al. 2011), as well as *R. prolixus* sylvatic populations do in Venezuela and Colombia (Fitzpatrick et al. 2008; Guhl et al. 2009). Triatomine species, such as

Rhodnius pallescens, *Rhodnius stali*, *Rhodnius neglectus*, *Rhodnius nasutus*, *Rhodnius ecuadoriensis*, *Triatoma brasiliensis*, *Triatoma sordida, Triatoma pseudomaculata, Triatoma maculata*, *Triatoma carrioni*, *Panstrongylus megistus*, *Panstrongylus geniculatus,* and *Panstrongylus herreri,* also have the ability to colonize artificial environments in some regions of their occurrences (Abad-Franch and Monteiro 2007; Aguilar et al. 1999; Coura et al. 2002; Costa et al. 2003; Gurgel-Gonçalves et al. 2012; Vinhaes et al. 2014). Therefore, vector control programs in such regions require continuous monitoring and insecticide retreatment of any new domestic focus that might be detected.

Facing this complex scenario of disease transmission, political authorities and scientific researchers realized that more studies about taxonomy, population dynamics, ecology, and behavior of primary and secondary vectors are needed to map disease transmission risk and create intervention programs (Dumonteil and Gourbiere 2004; Abad-Franch and Monteiro 2005; Abad-Franch et al. 2005, 2010b; Gurgel-Gonçalves et al. 2008; Peterson 2014).

Taxonomic identification underpins the development of new vector control and surveillance strategies. However, species identification solely based on morphology traits can be problematic, since looks can be deceiving; populations with different chromatic patterns can genetically represent the same taxon, and more than one species can be morphologically very similar (or identical), but genetically different. Taxonomic identification is paramount in both cases, but especially in the latter; even though morphologically similar, cryptic species can present different vector capabilities.

18.1.2 The Genus Rhodnius

The genus *Rhodnius* comprises 18 taxonomic recognized species. Its natural populations occupy arboreal ecotopes (preferentially palm trees) in approximately 28 biogeographical provinces from Central America to southern Brazil (Abad-Franch and Monteiro 2007). Morphological traits of *Rhodnius* allow easy genus-level identification, but many species lack clear-cut diagnostic characters and thus remain difficult to identify (Abad-Franch et al. 2013b). Some species are near-sibling and others constitute species complexes with several cryptic taxa (Lent and Wygodzinsky 1979; Monteiro et al. 2003; Abad-Franch et al. 2009).

Biogeographic and genetic data suggest a northern Palaeo-Amazonian origin of *Rhodnius* spp., which comprises two main lineages: the "pictipes lineage" and the "robustus lineage" (Abad-Franch and Monteiro 2007; Abad-Franch et al. 2009). *R. pictipes* lineage is composed by Andean and Amazonian species (such as *R. pallescens*, *R. pictipes*, *R. colombiensis*, and *R. ecuadoriensis*). *Rhodnius robustus* lineage is composed by five recognized species from four different biomes: Amazon (*R. robustus sensu lato* and *R. prolixus*), Caatinga (*R. nasutus*), Cerrado (*R. neglectus*), and Atlantic forest (*R. domesticus*).

Cryptic speciation seems to have occurred in both *R. pictipes* and *R. robustus* lineages (Monteiro et al. 2003; Pavan and Monteiro 2007; Abad-Franch et al. 2009; Gómez-Palacio et al. 2012). The next section (see Sect. 18.2) gives detailed information about morphological, ecological, and molecular tools used to unravel *R. prolixus* and *R. robustus s.l.* species complex in the Amazon and Orinoco regions. Moreover, a preliminary study concerning the behavior of strictly sylvatic and domiciled cryptic species is presented and discussed in light of their vectorial capacities.

Natural selection has also played an important role favoring phenotypic variation in triatomines. Morphological differentiation without significant genetic divergence has already been recorded in several triatomine species (Schofield et al. 1999). Here, we provide a concise revision about the two sibling species *R. neglectus* and *R. nasutus* from Cerrado and Caatinga biomes. Furthermore, we present preliminary findings of natural selection acting in favor of better adapted phenotypes to explain different coloration observed in same species from different ecotopes (Sect. 18.3).

18.2 How Many *Rhodnius* Species Are There in the Amazon? *Rhodnius prolixus* and *Rhodnius robustus* Species Complex—A Case Study

18.2.1 Chagas Disease Vectors in the Amazon and Orinoco Regions

The Amazon Basin is the largest drainage basin on the planet, occupying approximately 8 million km^2 of total area, more than one-third of the South America continent (UNEP 2005). This region (together with the Orinoco basin) forms the Pan-Amazonian biogeographic area and encompasses the most species-rich terrestrial ecosystem in the world, including more than 40,000 plants, 2000 birds and mammals, and 2.5 million insect species (Hoorn et al. 2010). Although Amazon rainforests provide crucial ecosystem goods and services to humanity (Foley 2007), anthropogenic disturbance events, such as deforestation, are frequent and can lead to extinction of primary hosts; it also plays an important role in increasing disease risk to humans.

Amazon rainforests hosts more than 25 species of Chagas disease vectors that belong to nine genera and, with the exception of some isolated populations of *T. maculata*, *P. geniculatus*, *P. herreri*, and *R. stali*, these insects are entirely sylvatic (Abad-Franch and Monteiro 2007). Thus, Chagas disease can be considered hypoendemic in this region, since the prevalence of *T. cruzi* infection is between 1 and 4 % (Aguilar et al. 2007). Although discoveries of localized oral contamination outbreaks have been arising and alarming native populations, most transmission events depend upon direct sporadic contact between sylvatic triatomines infected with *T. cruzi* and susceptible humans (Abad-Franch et al. 2009). Therefore,

disordered occupation of natural environments can be cited as the main reason for maintenance of disease transmission to humans in this region (Abad-Franch et al. 2010a).

The Orinoco basin harbors at least 20 species of Triatominae that belong to nine genera and, together with southeastern Amazonia, seems to correspond to the center of endemism where triatomine parental lineages diversified (Abad-Franch et al. 2009). One of the most epidemiological important species of Chagas disease, *R. prolixus*, is endemic to this area.

R. prolixus is the primary vector in Venezuela (Feliciangeli et al. 2002), Colombia (Guhl 2007), and in Central America (Hashimoto and Schofield 2012), since it is capable of invading and colonizing human habitations. Although successful strategies against *R. prolixus* interrupted transmission of *T. cruzi* in some areas of Central America, taxonomic uncertainties have been hindering its control in South America. This species is morphologically very similar to a strictly sylvatic species that occurs in the Pan-Amazonian region, *R. robustus*. Thus, taxonomic validity of these species was questioned during the last decades.

18.2.2 Taxonomy of *Rhodnius prolixus* and *R. robustus* Species Complex

The taxonomic status of *R. prolixus* and *R. robustus* has long been a matter of controversy, caused by morphologic similarity, incorrect diagnosis, and poor sampling. Almost 70 years after the discovery of *R. prolixus* Stål 1859, a new species, called *R. robustus* Larrousse 1927, was described based primarily on two female specimens collected in French Guyana and Brazil. This species was considered to be larger and darker than a reference series of *R. prolixus*. These observations led scientists to routinely use insect body size as a diagnostic character to separate domestic and sylvatic populations (and thus identify these small-size and large-size specimens as *R. prolixus* and *R. robustus*, respectively; Pavan and Monteiro 2007). If true, this scenario (one exclusively sylvatic species and another exclusively domestic) would not denote a big challenge for vector control programs, since the application of insecticides would eliminate and eventually eradicate domestic populations. Lent and Jurberg (1969) and Lent and Wygodzinsky (1979) provided further additional observations about diagnostic characters, such as differences in male genitalia structures and color on the hind tibia of fourth- and fifth-stage nymphs. These traits, however, are either too variable to be used for diagnostic purposes (Harry 1993) or restricted to few stages of life.

The real epidemiological scenario in Venezuela started to be uncovered only after 1960, when many researchers confirmed the presence of *R. prolixus* in both human houses and sylvatic ecotopes. This species was found in bird and mammal nests (Gamboa 1961, 1970; Goméz-Nuñez 1963; Pifano 1973). A new controversy arose when Lent and Valderrama (1973) and Tonn et al. (1976) recorded *R.*

robustus in palm trees in Venezuela. These findings led to the discussion (which is still ongoing) of whether sylvatic populations (previously described as *R. prolixus*) might be in fact *R. robustus*.

Conflicted outcomes of isoenzymatic, morphometrical, biochemical, and molecular investigations during 1990's increased the discussion around the taxonomic status of *R. robustus*. The lack of reproductive barriers (tested with cross-mating experiments in laboratory conditions) and polymorphic loci of allozymes between *R. prolixus* and *R. robustus* (Harry et al. 1992; Harry 1993; Barrett 1995) raised the possibility of these species being the same taxon. This idea would implicate *R. prolixus* transmitting *T. cruzi* in the entire Pan-amazonian region and gene flow between sylvatic and domestic populations—the worst scenario for vector control.

Several studies were further conducted to resolve this taxonomic impasse. The results obtained from salivary protein analysis (Soares et al. 1998), RAPD (Feliciangeli et al. 2002), and wing morphometry (Villegas et al. 2002) led to the conclusion that *R. robustus* is a *bona fide* species and different from *R. prolixus*. Although very informative, these studies did not conclusively determined if sylvatic and domestic *R. prolixus* populations were freely interchanging genetic material.

Monteiro et al. (2003) provided the first in-depth study, which analyzed both field-collected and colony-reared specimens from the entire known distribution of *R. prolixus* and *R. robustus*. This study was driven by previous results of Lyman et al. (1999), in which DNA sequencing method was used for the first time in triatomines to solve taxonomic problems. Using a 399-bp fragment of the mitochondrial cytochrome b gene (mtCytb) and a 400-bp of the mitochondrial large subunit ribosomal RNA, Lyman et al. (1999) showed a clear separation between *R. prolixus* and *R. robustus*. Monteiro et al. (2003) extended this analysis and sequenced a larger fragment of the mtCytb gene (663-bp) and also 630-bp of the D2 variable region of the 28S RNA (D2) of 26 populations from seven Latin American countries. The authors concluded that *R. robustus* is different from *R. prolixus*, and represents a paraphyletic assemblage of, at least, four different cryptic lineages (which were provisionally referred to as *R. robustus* I, II, III, and IV; Fig. 18.1). Surprisingly, *R. robustus* I, which occurs in the Orinoco basin, is more closely related to *R. prolixus* than to the other Amazonian *R. robustus* lineages (II–IV).

The "Orinoco clade" comprises *R. prolixus* and *R. robustus* I. Although *R. prolixus* can be found well adaptedadapted to human domiciles in many Latin American countries, wild populations of these species are present only in Venezuela and Colombia, associated with native palms and agro-industrial palm plantations (Guhl 2007; Fitzpatrick et al. 2008). *R. robustus* I seems to be rare and shows restricted geographic distribution in sylvatic ecotopes. To date, it has only been found in Trujillo, Venezuela (Pavan et al. 2013). It inhabits 10 ecoregions [for more details about the ecoregions locations, *cf.* Morrone (2006)] from the Orinoco biogeographical area (Abad-Franch et al. 2009), occurring in sympatry with *R. prolixus*.

The "Amazonian clade" comprises the peripatric lineages *R. robustus* II, III, and IV. *R. robustus* II was already found in northern Rondônia and southern Amazonas

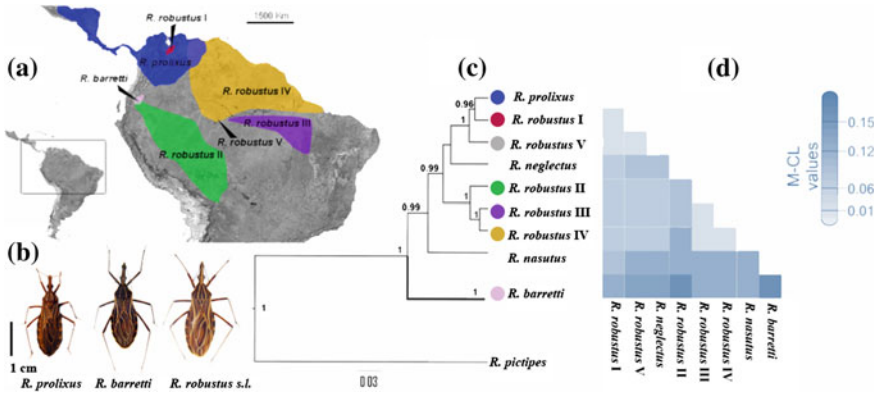


Fig. 18.1 The *R. robustus* lineage. **a** Geographic distribution of *R. prolixus*, *R. barretti*, and *R. robustus s.l.* based on collected specimens which were already genotyped. **b** *R. prolixus*, *R. barretti*, and *R. robustus s.l.* chromatic patterns. **c** Bayesian consensus tree based on 13 *R. robustus* lineage sequences of a 663-bp mtCytb fragment (*numbers above key nodes* represent the posterior probabilities). **d** Pairwise maximum-composite likelihood distances between sequences, with *darker colors* representing larger genetic distances (see scale bar). Modified from Abad-Franch et al. (2013b)

(Brazil), northern Ecuador (Monteiro et al. 2003; Pavan and Monteiro 2007), and in northeastern Bolivia (Justi et al. 2010). It inhabits 13 ecoregions from biogeographical areas of Napo, Inambari, and Rondônia (Abad-Franch et al. 2009). *R. robustus* III was already collected at northeastern Amazonas, northern Tocantins, and eastern Pará, Brazil (Monteiro et al. 2003; Pavan and Monteiro 2007), occupying six ecoregions from geographical areas of Rondônia, Pará, and Belém (Abad-Franch et al. 2009). *R. robustus* IV has been found in northern Pará, northern Amazonas, and Roraima, in Brazil, northern Venezuela (in sympatry with *R. prolixus*), and in French Guyana. This species is present in eight ecoregions from the biogeographical areas of Orinoco, Guyanan, Imerí, and Belém.

The investigation conducted by Monteiro et al. (2003), in addition to the great contribution in the systematics field, has changed the way genetic variability was measured in triatomines. Objective and variable characters—deoxyribonucleotides —arose in the taxonomy of these insects to complement morphological traits. Therefore, DNA sequencing analyses came as a natural choice for alpha systematics (Calleros et al. 2010; Monteiro et al. 2004, 2013; Mendonça et al. 2009) due to their efficacy in the guidance of control strategies (Miles et al. 2003).

Four years after the discovery of the four lineages of *R. robustus*, Abad-Franch and Monteiro (2007) suggested (based on field observations and preliminary molecular analysis) that this complex of morphological similar species could be even greater than Monteiro et al. (2003) previously observed. They reported the presence of a new taxon found in Amazonas, Brazil (referred to as *R. robustus* V), in sympatry with *R. robustus* IV, and another taxon that was collected in sylvatic

areas of the Ecuadorian and Colombian Amazon in sympatry with *R. robustus* II, and afterward it was referred to as *Rhodnius barretti* (Fig. 18.1).

A major step was taken toward a more objective and complete assessment of sibling species identification when Abad-Franch et al. (2013b) described *R. barretti* using an integrated analyses approach, which included mtDNA sequencing, morphometrics, and qualitative phenotype traits. Although slight differences in morphology and wing shape were observed, phylogenetic analysis provided consistent results to conclude that *R. barretti* is different from its sibling species *R. prolixus* and *R. robustus* I–V. Mitochondrial DNA sequences of *R. barretti* diverge from sequences of its closer relatives in more than 8.7 %, while *R. robustus* I–V diverges from each other in 2–4 % (Fig. 18.1). Therefore, *R. barretti* forms a basal clade to the rest of the "robustus lineage" species, which also comprises (besides *R. prolixus* and *R. neglectus*) *R. neglectus* and *R. nasutus*.

The foregoing molecular studies of *R. prolixus* and *R. robustus s.l.* species complex provided the background information required before taking the development of new strategies for vector control one step further. The correct identification of *R. prolixus* and *R. robustus* where they could occur in sympatry (such as Colombia and Venezuela) is a matter of paramount importance for vector surveillance programs to delineate the areas with major risks of (re-)infestation by primary vectors.

The creation of simple and cost-effective methods based on DNA information can be useful for species diagnosis purposes. Although DNA sequencing methods become more popular over years, many research groups involved in medical and epidemiological-related areas only need simple and cost-effective tools for a fast and reliable taxonomic identification.

18.2.3 Simple Taxonomic Diagnosis for *R. prolixus* and *R. robustus* Specimens

Pavan and Monteiro (2007) have designed an objective and cost-effective mtDNA multiplex-PCR assay to allow for accurate taxonomic identification of *R. prolixus* and *R. robustus s.l.* Using species-specific primers, mtDNA fragments of three different sizes are amplified by PCR. When submitted to an electrophoresis, these fragments have different migrations in agarose gel, providing species identification. This method amplifies a 239-bp fragment of the mtCytb gene from *R. robustus* II–IV, 285-bp from *R. prolixus*, and a 349-bp from *R. robustus* I DNA samples.

The limitation of this method is the mitochondrial-based nature, which is just maternally inherited. Therefore, it could generate misleading results in areas where *R. prolixus* and members of the *R. robustus* complex are sympatric and might hybridize, causing mitochondrial DNA introgression (Fitzpatrick et al. 2008). To overcome such limitation, Pavan et al. (2013) described a new non-coding, single-copy nuclear DNA fragment (TPS165, or AmpG) containing a single-nucleotide

polymorphism (SNP) that separates *R. prolixus* from members of the *R. robustus* cryptic species complex. This SNP is located on the 280th site of the 364-bp region of the fourth intron of the Transmembrane protein 165 gene (TP165), whereas all *R. prolixus* have an adenine, instead of a guanine, on that particular site, and thus, it is diagnostic for *R. prolixus*. Although this methodology requires DNA sequencing, as opposed to multiplex-PCR methods (Pavan and Monteiro 2007), no cloning is required, due to its single-copy nature. Since it needs just the identification of a single-nucleotide site, this method is also simple and straightforward.

18.2.4 Ecological Trends of *Rhodnius prolixus* and *R. robustus*—Implications for Epidemiology of Chagas Disease

The study of the potential geographic distribution of vector species is crucial for understanding geographic dimensions of risk transmission of the disease. In this context, ecologic niche modeling (ENM) is a tool that permits exploration of geographic and ecologic phenomena based on known occurrences of species (Peterson et al. 2011). ENM has been applied broadly to understand aspects of Chagas disease transmission in the last decade, including characterization of vector species niches, and analysis of relationships between vector and reservoir distributions (Costa and Peterson 2012; Gurgel-Gonçalves et al. 2012).

Much more care is required to prevent (re-)infestations of houses by *R. prolixus* populations due to its primary epidemiological relevance. Fitzpatrick et al. (2008) analyzed 10 microsatellite loci of 551 specimens (representing 34 populations) from six states of Venezuela and detected a high variability when populations from houses and peridomiciliary palms were compared. These results indicated that house colonization by insects from nearby palms is not a frequent event. Therefore, vector control inside houses in Venezuela must be optimized to avoid misapplication of insecticides that might result in recrudescent populations. Caution is also needed in the Brazilian Amazon, where *R. prolixus* can potentially occur in the southern part (Fig. 18.2a).

Although it is widely known that *R. prolixus* and *R. robustus s.l.* have distinct epidemiological relevance, neither any difference in vectorial capacity and competence between *R. robustus* lineages, nor the reason why *R. prolixus* (and not *R. robustus s.l.*) can colonize houses is clear. Insect behavior studies and molecular markers that might be involved in insect speciation processes might help clarify those issues.

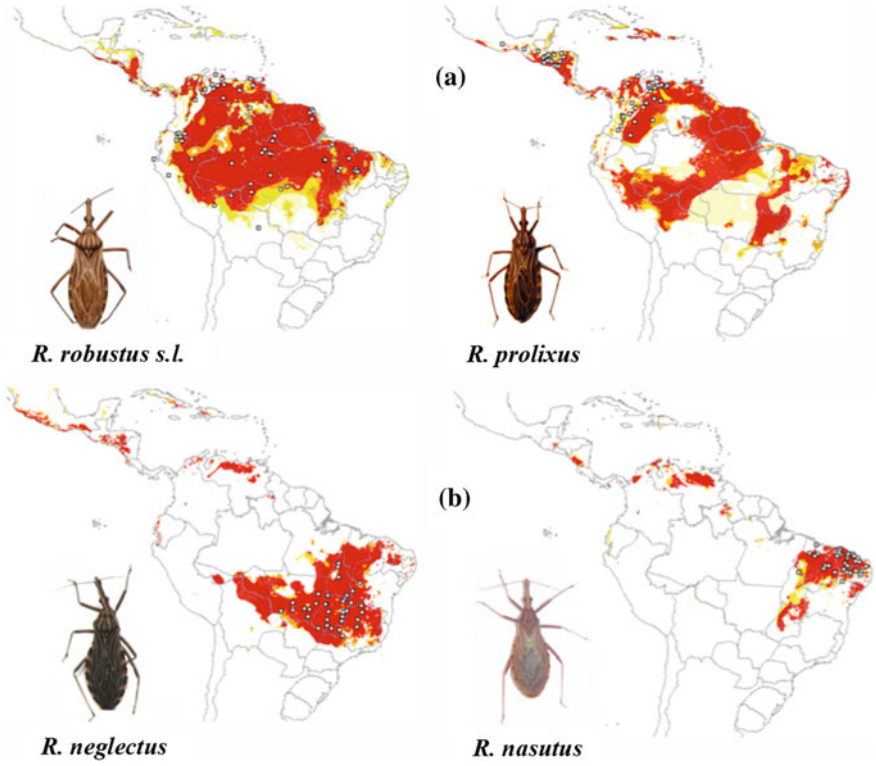


Fig. 18.2 Potential geographical distribution of some *Rhodnius* species in the Americas. *White squares* represent the occurrence points and the *yellow-red* gradient map indicates the environmental suitability in accordance with the ecological niche models generated by the GARP algorithm (*light yellow* low suitability; *dark red* high suitability). **a** *R. robustus s.l.* (*left*) and *R. prolixus* (*right*) predictions; **b** *R. neglectus* and *R. nasutus* predictions. Adapted from Batista and Gurgel-Gonçalves (2009)

18.2.5 Rhodnius prolixus and *R. robustus* Daily Activity Rhythms and Possible Correlation with Their Vector Capacities

Behavioral studies related to circadian control of insect activity rhythms (Clements 1999; Saunders 2002) can be particularly useful for understanding epidemiological-related questions. These rhythms affect time and degree of contact between vector and host and also vector and insecticide-sprayed surfaces (Lazzari and Lorenzo 2009).

An endogenous clock controls daily changes in physiology, metabolism, and behavior of insects in response to periodic changes in the environment, caused by the Earth's rotation period (Abruzzi et al. 2011; Ceriani et al. 2002; Keegan et al.

2007). Even in absence of external variations, this clock is able to maintain these rhythms for a period close to 24 h and is therefore called circadian rhythms (Saunders 2002).

The genetics of circadian rhythms has been extensively studied in *Drosophila melanogaster* (Hall 2003). The molecular pacemaker of this insect clock includes interconnected regulatory loops, which are responsible for controlling rhythms related to behavior, physiology, and other biological processes (reviewed in Allada and Chung 2010, Hardin 2011). The expression of most genes encoding proteins in these regulatory loops is rhythmic (Hall 2003; Hardin 2011).

One of the biological processes modulated by circadian rhythms is related to sexual behavior and reproductive isolation (Kyriacou and Hall 1986). For this reason, circadian genes have been considered as being "speciation genes" (Araki et al. 2009). One of these genes, called *period* (*per*), is involved in circadian control rhythms of pupae hatching, locomotor activity, and also in ultradian rhythm of cohort sounds (Kyriacou and Hall 1986; Wheeler et al. 1991). These sounds are involved in reproductive isolation between *D. melanogaster* and other phylogenetically related species (Ritchie et al. 1999). In fact, the analyses of *per* sequences have been very useful in the identification of cryptic species, such as the vectors of leishmaniasis *Lutzomyia longipalpis s.l.* (Bauzer et al. 2002; Araki et al. 2009), and also in population genetics of *D. melanogaster* (Peixoto 2002).

Circadian-related rhythms observed in locomotor activity are very useful to undermine different patterns related to vectorial capacity and to the formation of natural reproductive isolation barriers to gene exchange. Lima-Camara et al. (2011) have recently demonstrated that infection with dengue virus serotype 2 (DENV2) causes an increase in locomotor activity of *Aedes aegypti*. Since the nervous system of the mosquito is affected by dengue virus infection (Salazar et al. 2007). neurons responsible for circadian rhythm control probably altered the expression of clock genes, causing changes in the locomotor activity.

Sympatric species can show different temporal activity patterns. Rivas et al. (2008) reported that two sympatric *L. longipalpis* cryptic species showed a one-hour difference between their crepuscular activity phases, which in turn might be considered as an important adaptive trait to reinforce the pre-zygotic reproductive isolation. Moreover, different levels of activity observed in laboratory conditions can evidence reproductive isolation of sympatric species. This pattern was recently observed in *Anopheles gambiae* M and S molecular forms (Rund et al. 2012) and could lead to differences in migrations related to matting or host-seeking (Rund et al. 2012).

Those three examples raise a simple but fundamental question: How could these behavioral differences influence vectorial capacity of these species? Circadian rhythms have been extensively studied on the two most epidemiological relevant triatomine species, *R. prolixus* and *T. infestans*. In both species, these rhythms seem to control many biological processes related to reproduction and foraging, such as breeding and oviposition (Constantinou 1979; Lazzari 1991), dispersion (McEwen and Lehane 1993), and host-seeking (Barrozo et al. 2004; Bodin et al. 2008). Adult triatomines spend the day in refuges hiding from predators; during the night, they

remain active, searching for blood meals (Lorenzo and Lazzari 1998). However, it is imperative to understand if this pattern of behavior is fixed on triatomines or if they could vary depending on the species or ontogeny. Moreover, circadian rhythms were never explored under an evolutionary perspective in triatomines. *R. prolixus* and *R. robustus s.l.* can serve as a good model, since they constitute an extensive studied group of sister species with well-known phylogeny.

In a recent work, Pavan et al. (unpublished data) reported that *R. prolixus* and *R. robustus* nymphs share similar daily rhythms of activity and circadian period length. However, *R. prolixus* has higher levels of overall daily activity and a more preeminent activity rhythm under constant darkness.

R. prolixus is more active than *R. robustus*, what could increase its dispersion. These higher levels of activity of *R. prolixus* could explain its greater vector capacity. Recent mathematical models propose that the increase of locomotor activity in insect is directly proportional to biting rate (Luz et al. 2011). Thus, the activity of triatomines, blood feeding efficiency, and parasite spread are probably related.

18.3 One Theory to Rule Them All: Possible Role of Natural selection in Shaping Phenotypic Variation Among Sibling Species *R. neglectus* and *R. nasutus* Populations

18.3.1 Cerrado and Caatinga Biomes

Brazilian Amazon is bordered by two different dry biomes—Cerrado and Caatinga. Cerrado is the second largest Brazilian biome, occupying 21 % of country's land area. Climate is seasonal—wet from October to March and dry from April to September, and mild year around, with temperatures ranging from 22 to 27 °C. Diversity of vascular plants exceed that observed in most other biomes—about 7000 species, including known herbs, shrubs, trees, and lianas, 44 % of them being endemic. Moreover, there is a high diversity of vertebrate fauna (approximately 200 mammal, 840 bird, 180 reptile, 150 amphibian, and 1200 fish species). Thus, Cerrado is the richest tropical savanna in the world (Klink and Machado 2005), and one of the world's biodiversity hotspots (Myers et al. 2000). Unfortunately, high deforestation rates have been observed in the Cerrado, and conservation efforts have been modest, with only 2.2 % of its area under legal protection. Numerous animal and plant species are threatened with extinction, and approximately 20 % of these organisms are in unprotected areas.

The Caatinga biome is a semiarid scrub forest situated in the northeast of Brazil. It occupies 11 % of Brazilian territory, stretching across 300,000 square miles of subequatorial zone. Climate is hot, with long dry seasons; rainfall usually totals less than 750 mm/year, and it is concentrated in three consecutive months, between

November and June (summer/autumn). Temperature changes little, with an annual average of approximately 26 °C. Caatinga comprises a mosaic of vegetation types varying from dry thorn forest to open shrubby vegetation (Carvalho da Costa et al. 2007). It is extremely rich in natural resources but, when compared to rainforests, there is little information available on its biodiversity. Moreover, only 1 % of Caatinga habitats is protected (Silva et al. 2004).

Cerrado and Caatinga host approximately 25 and 15 species of Chagas disease vectors, respectively, which comprise almost 65 % of triatomine fauna registered in Brazil (Gurgel-Gonçalves et al. 2012). *T. cruzi* transmission by domiclliated triatomines is well documented in these biomes. Systematic use of chemical controls in these regions has reduced synanthropic triatomines (Silveira and Dias 2011); however, sylvatic vectors maintain widespread foci in natural habitats from which they regularly invade and sometimes colonize human dwellings.

In the Cerrado, the most broadly distributed species are *T. sordida* and *R. neglectus*. *T. sordida* is the most frequently captured species by entomological surveillance in Brazil (Pereira et al. 2013). However, the risk of *T. cruzi* transmission by *T. sordida* is relatively low, since it inhabits mainly chicken coops (Rossi et al. 2015). *R. neglectus* is frequently found in bird and mammal nests and plays an important role in the maintenance of *T. cruzi* transmission in the wild (Gurgel-Gonçalves et al. 2004b). Moreover, adult specimens have been invading houses in central Brazil (Gurgel-Gonçalves et al. 2008). Household infestation events (with adventitious bugs occasionally establishing breeding colonies) have been reported in several Brazilian states (Oliveira and Silva 2007; Gurgel-Gonçalves et al. 2012), maintaining a conspicuous risk of Chagas disease transmission in central Brazil.

T. brasiliensis, *T. pseudomaculata*, and *R. nasutus* are broadly distributed in the Caatinga. *T. brasiliensis* is the most important vector species in northeastern Brazil; on the other hand, *T. pseudomaculata* occurs mainly in peridomiciles, presenting low percentages of natural infection by *T. cruzi* (Dias et al. 2000). Moreover, the occurrence of adult *R. nasutus* specimens infected by *T. cruzi* inside houses in some Brazilian states has been observed (Lima et al. 2012a).

18.3.2 Ecological Niches and Geographic Distribution of *Rhodnius neglectus* and *R. nasutus*

R. neglectus is the main *Rhodnius* species found in the Cerrado. It inhabits various types of palm trees including *Attalea*, *Acrocomia*, *Mauritia*, *Oenocarpus*, and *Syagrus* (Gurgel-Gonçalves et al. 2004a; Abad-Franch et al. 2009). *R. neglectus* can be misidentified as other species, *R. nasutus* (Dias et al. 2008; Lima and Sarquis 2008), particularly in Caatinga and Caatinga–Cerrado transitional areas.

R. nasutus is found predominantly in Caatinga inhabiting *Copernicia prunifera* palm trees (Sarquis et al. 2004). Nonetheless, *R. nasutus* may also inhabit other

palms and trees in this region (Dias et al. 2008; Lima et al. 2012b). Although *R. nasutus* seems to be endemic to Caatinga, this species was also registered in transitional areas with Amazon forest (e.g., in Maranhão babassu forests) and also with Cerrado.

The high similarity between *R. nasutus* and *R. neglectus* leads to an uncertain definition of boundaries of their geographical distribution. Moreover, chromatic variation in both species (Barrett 1995) precludes the definition of possible co-occurrences in the same habitat. Novel analytical tools are needed to improve the identification of their geographical distributions, as well as to identify possible zones where these species are sympatric.

Recent ENM results show a high environmental suitability for *R. neglectus* occurrence in central Brazil. Moreover, these models indicate areas with potential occurrence of this species in northeastern Bolivia, Venezuela, and other countries of South and Central America, where it has not been registered yet (Fig. 18.2b). Also, the presence of *R. neglectus* in Caatinga has been recently registered (Abad-Franch et al. 2009; Gurgel-Gonçalves and Cuba 2009).

The potential distribution of *R. nasutus* covered the northeastern region of Brazil in the semiarid Caatinga and the Maranhão babassu forests (Fig. 18.2b). Although never found in central Brazil, the prediction maps of *R. nasutus* indicate areas of its potential occurrence in sympatry with *R. neglectus*. The absence of *R. nasutus* in central Brazil may have various causes, including historical restrictions (e.g., geographic barriers and/or lack of sufficient dispersal opportunities) and biotic interactions (such as competition with related species, such as *R. neglectus*) (Batista and Gurgel-Gonçalves 2009). Analysis of either congruence or discordance between predicted and actual distributions can be used in future investigations to evaluate ecological and historical factors that shaped the geographical distribution of species (Anderson et al. 2002). Nevertheless, dubious predictions must be confirmed by other methods, such as morphometrics and molecular data, to confirm species identification.

18.3.3 The Use of Geometric Morphometrics for Diagnosis Between *Rhodnius neglectus* and *R. nasutus*

Some morphological traits used for *R. neglectus* and *R. nasutus* diagnosis are not reliable. The identification of these species is mainly based on comparing chromatic patterns (of the body and antennae), and their overall body size (Lent and Wygodzinsky 1979). Chromatic variation is, however, present in both species (Barrett 1995) and size-related traits are influenced by environmental changes in triatomines (Dujardin et al. 1999; Abad-Franch et al. 2003). Therefore, these traits have to be cautiously interpreted in the context of species-level diagnosis (Gurgel-Gonçalves et al. 2008). Moreover, Harry (1993) detected no clear-cut differences in the male genitalia structures between *R. neglectus* and *R. nasutus* and

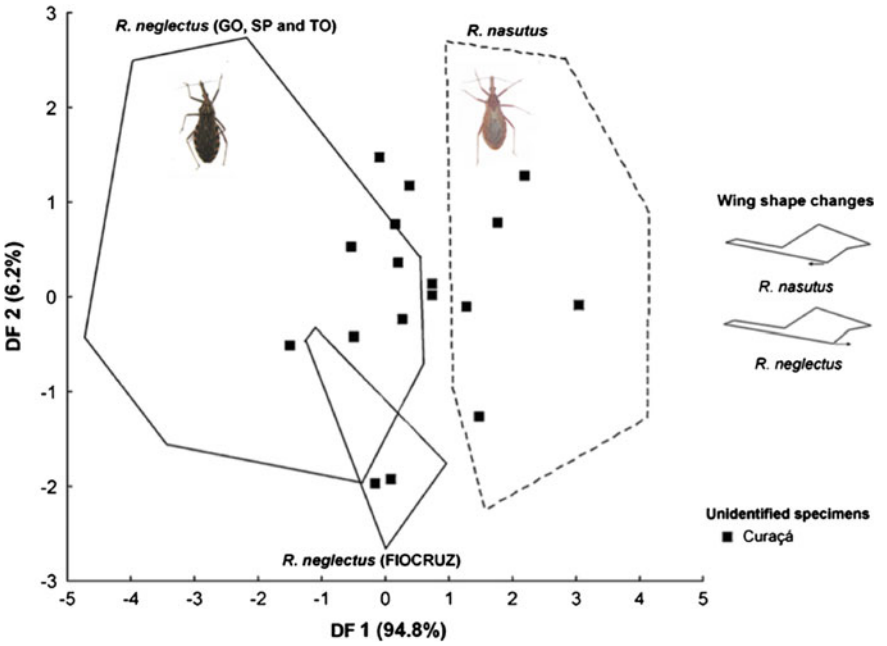


Fig. 18.3 Geometric morphometrics of near-sibling triatomine species. Factorial map on the plane defined by the two discriminant factors of wing shape variation (*DF1* and *DF2*). *Rhodnius neglectus* specimens used in the analysis were from Goiás (*GO*), São Paulo (*SP*), Tocantins (*TO*), and a laboratory-reared colony from Instituto Oswaldo Cruz (*FIOCRUZ*). *R. nasutus* specimens were collected in the Ceará state. In addition, unidentified *Rhodnius* specimens from Curacá, Bahia, (*black squares*) were added to the analysis. The percent contribution of each DF to total shape variation is shown on the axes (in *parentheses*). Polygons (convex hulls) enclose individual points (which were deleted for the sake of clarity) corresponding to specimens of known specific status. The least squares consensus wing configuration for each species is shown on the right; *arrows* indicate the main difference in wing shape (end of the postcubital vein) between *R. neglectus* and *R. nasutus*. Modified from Abad-Franch et al. (2009)

Monteiro et al. (2002) showed that these species differed only by one locus using allozyme analysis.

Several studies show that geometric morphometric approaches are valuable tools for triatomine alpha systematics and can be used to differentiate closely related species (e.g., Matías et al. 2001; Feliciangeli et al. 2007; Gurgel-Gonçalves et al. 2008, 2011; Dujardin et al. 2009; Márquez et al. 2011). Abad-Franch et al. (2009) applied a geometric morphometrics approach to evaluate if size and shape patterns can be used as taxonomic markers that differentiate *R. neglectus* from *R. nasutus*. Their results showed wing and head shape differences between these species (Fig. 18.3). Some specimens from Curaçá, Bahia (Brazil), were clustered within the *R. neglectus* group, while others were more similar to *R. nasutus*. Few specimens exhibited an intermediate shape pattern (Fig. 18.3). These results evidenced that both species are sympatric (and infest *C. prunifera* palms) in Curaçá, and display

different wing and head shapes, and therefore drew attention to researches on between-species ecological interactions.

18.3.4 *Phenotypic Variation Within* Rhodnius neglectus *and* R. neglectus

Triatomines are able to develop rapid morphological changes in response to adaptation to new habitats. Bugs from artificial habitats show smaller sizes than sylvatic ones (Dujardin et al. 1999; Catalá and Dujardin 2001; Dujardin et al. 2009; Batista et al. 2013; Rodríguez et al. 2013). Phenotypic variation is also associated with diet and microgeography (Abrahan et al. 2008).

Triatomines present great phenotypic heterogeneity, such as previously observed in *T. brasiliensis* (Costa 2000), *T. infestans* (Noireau et al. 1997), and *T. rubrovaria* (Almeida et al. 2002). Dujardin et al. (1999) suggested that morphological differentiation could be faster than development of reproductive and genetic barriers. In the *Rhodnius* genus, specifically in *R. nasutus* species, phenotypic variation was recently observed in Ceará Brazil, where it was collected in five different palm tree species (Dias et al. 2008). The holotype of *R. nasutus* has a pale brownish-yellow coloring, with a red-like appearance and dark brown dots in certain regions of the body and appendices (Lent and Wygodzinsky 1979). However, triatomines that were inhabiting *Copernicia prunifera* palms presented a reddish color, according to the original species description, while *R. nasutus* from *Acrocomia intumescens*, *Attalea speciosa*, *Mauritia flexuosa,* and *Syagrus oleracea* were chestnut-colored (Dias et al. 2008) (Fig. 18.4). These findings suggest that variations in triatomine color patterns, particularly those of certain *Rhodnius* species, could be influenced by habitat. Body colouration of *R. nasutus* specimens corresponded exactly to the fibers and base of fronds, providing camouflage for the insects and protecting them against predators.

A melanic *R. nasutus* form was recently discovered in *Attalea speciosa* palm trees (Dias et al. 2014) Reference populations of *R. nasutus* and its melanic form were reared in laboratory and were analyzed using geometric morphometric approaches. These results showed size and shape differences between different chromatic patterns (Fig. 18.5). Moreover, crossbreeding results between these *R. nasutus* chromatic forms indicated that the "dark-morph" coloration is a recessive trait inherited in a Mendelian fashion. Thus, the phenotypic variation observed in *R. nasutus* further reinforces intraspecific heterogeneity in the Triatominae subfamily, demonstrating the importance of using rigorous criteria for describing a new species.

Gurgel-Gonçalves and others (unpublished data) have recently observed *Rhodnius* specimens with dubious chromatic patterns at Caatinga and Caatinga/Cerrado transitional areas. Individuals collected in *Mauritia flexuosa* palms had a dark phenotype, with the same coloration and diagnostic traits of *R. neglectus*.

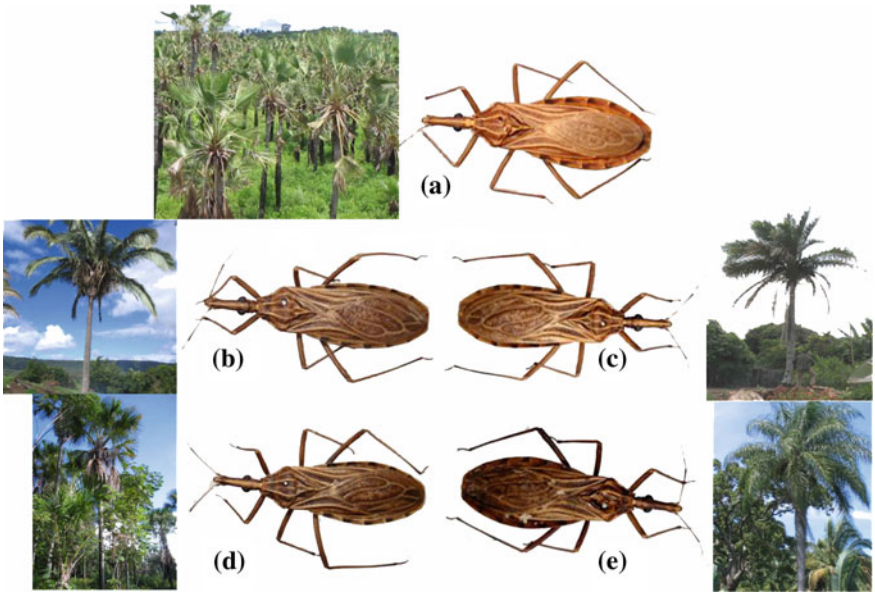


Fig. 18.4 Females of *Rhodnius neglectus* collected on different palm trees in the state of Ceará, Brazil, showing different chromatic forms. The different forms were collected in the following palm tree species: **a** *Copernicia prunifera* (where it was found the same chromatic pattern of the *R. nasutus* holotype); **b** *Attalea speciosa*; **c** *Syagrus oleracea*; **d** *Mauritia flexuosa*; **e** *Acrocomia intrumescens*. Adapted from Dias et al. (2008)

However, those collected in *C. prunifera* palms displayed a lighter chromatic pattern more similar to that of *R. nasutus* (Fig. 18.6). Previous analysis indicated that *R. nasutus* and *R. neglectus* populations co-occurred in those transitional areas (Abad-Franch et al. 2009), thus raising the possibility of natural hybridization.

An alternative explanation for this observation is that *R. neglectus* would exhibit one chromatic phenotype similar to *R. nasutus* that is different from the pattern described by Lent and Wygodzinsky (1979), which would therefore confound species identification. The lighter color of *R. neglectus* from *C. prunifera* may have improved its chances of survival and reproduction, since they would be camouflaged with the clear substrate of *C. prunifera* fibers. Therefore, populations with light phenotype increased in frequency in these palm trees (as a result of natural selection), since they would be less predated than the dark ones. The identification of these species in areas of co-occurrence should be improved with the application of molecular methods (Pavan and Monteiro 2007; Abad-Franch et al. 2013b; Pavan et al. 2013).

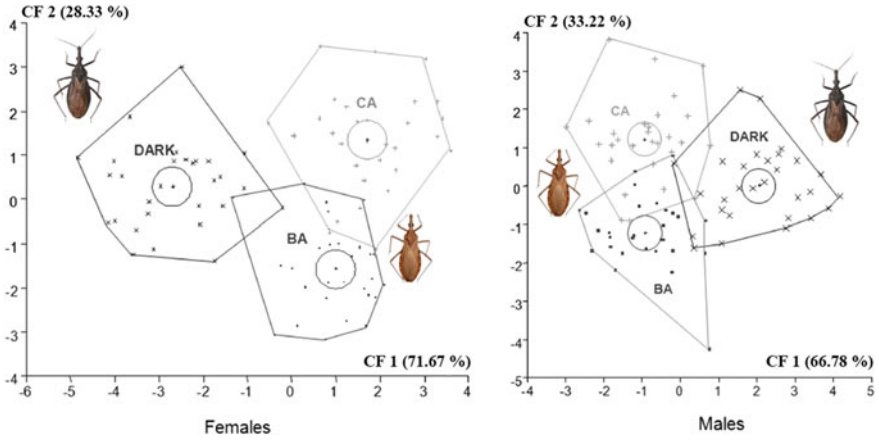


Fig. 18.5 Geometric morphometrics of *R. nasutus* and melanic *R. nasutus* males and females: factorial map on the plane defined by the two-conformational factors of wing shape variation (*CF1* and *CF2*). Each individual point corresponds to a specimen analyzed. *Cross marks*, referred as the "DARK" group, show melanic *R. nasutus* collected in *Attalea speciosa* (babassu) palms in state of Ceará, Brazil. *Plus mark* ("BA" group) and *square symbols* ("CA" group) correspond to the reference *R. nasutus* collected in *A. speciosa* and *Copernicia prunifera* (carnaúba) palm trees in Ceará, Brazil, respectively. The percent contribution of each CF to total shape variation is shown on the axes (in *parentheses*). Polygons enclose individual points that correspond to specimens from the same location and same chromatic pattern

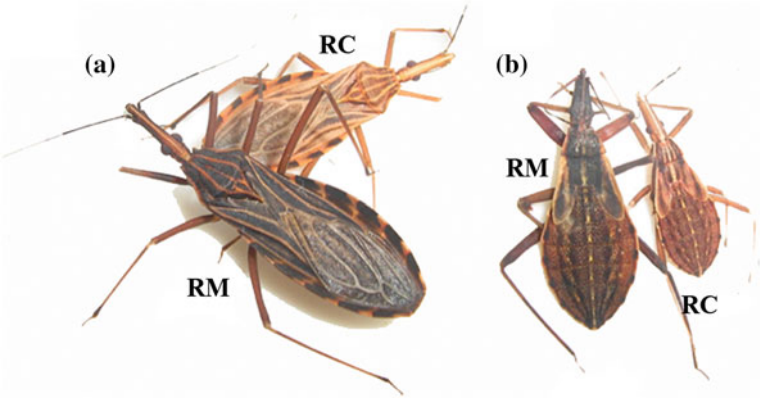


Fig. 18.6 Phenotypic variability of *Rhodnius neglectus/R. nasutus* from different palm tree species. **a** Chromatic forms of adult specimens captured in *Mauritia flexuosa* (RM) and *Copernicia prunifera* (RC) palm trees; **b** chromatic pattern of nymphs captured in these palm tree species

18.4 What Is Next? Perspectives in Behavior, Ecology, and Molecular Systematics of Triatomines

Here, we highlighted the importance of integrating morphological, ecological, behavioral, and molecular tools to disentangle some epidemiological and taxonomic unresolved questions in triatomines. We used as examples recent findings in a cryptic species complex that comprises, at least, seven lineages (*R. prolixus*, *R. robustus* I–V, and *R. barretti*), and a species that exhibits phenotypic variation, as a result of natural selection. These are just few examples of evolutionary issues that are waiting to be better explored with proper analytical tools.

The accomplishment of an integrative approach to validate *R. barretti* (morphologic, morphometrics, and molecular approaches) as a *bona fide* species exemplifies one of the trends in modern taxonomy. Moreover, knowledge in molecular systematics is likely to advance quickly after the release of *R. prolixus* genome (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/RevisedRhodniusSeq.pdf), since it will be possible to isolate and sequence new nuclear markers. Furthermore, the existence of this sequenced and annotated genome will open doors to new genome projects of other triatomine species, since the assembly of sequence reads should be easier using *R. prolixus* as a reference genome and thus will not require long read lengths, or high depth coverage, which in turn will decrease total costs.

During the past 15 years, numerous molecular markers were sequenced to help analyzing taxonomic problems in Triatominae (e.g., Gaunt and Miles 2002; Hypsa et al. 2002; Marcilla et al. 2002; Hwang and Weirauch 2012). However, only few markers were already tested and successfully separate closely related triatomine species, such as those maternally inherited (mtCytb, mtCOI, and mtCOII) (Abad-Franch et al. 2013b; Monteiro et al. 2003; Calleros et al. 2010; Justi et al. 2014), and also nuclear ribosomal internal transcribed spacers (ITS-1 and ITS-2) (Marcilla et al. 2001).

The exclusive use of mitochondrial genes as markers in molecular systematic studies had some limitations, such as (1) mitochondrial pseudogenes (NUMTs) in the nuclear genome, which can lead to misidentification of species; and (2) introgression of mitochondrial genetic material from one species into another, due to interspecific crosses (revised Mas-Coma and Bargues 2009). On the other hand, lack of nuclear markers useful for the separation of closely related species of triatomines hinders the detection of cryptic species yet unknown or natural hybrids.

The ITS-1 and ITS-2 markers seem to be in concerted evolution in some organisms, a homogenization phenomenon between different loci in gene families, as a consequence of gene conversion and unequal recombination (Liao 1999). This homogenization can be completed, or reveal intragenomic and intraspecific variation (of paralog sequences). As *Rhodnius* species seem to fit the latter case (Pavan, Lazoski, and Monteiro, unpublished data), analyses of these copies can be time-consuming and difficult to interpret. In this case, single-copy nuclear genes that are directly involved in sexual behavior and reproductive isolation, such as

those that control circadian and ultradian rhythm, seem to be good candidates to be tested in phylogenetic and populational data involving cryptic species complexes.

An alternative approach for generating genome data at a lower cost is becoming popular with population genetics studies and with phylogenetic analyses of closely related species, and probably will be the foundation for studies in the near future (mainly in non-model organisms)—the double-digested restriction-site-associated DNA sequencing (ddRAD-seq). ddRAD-seq are fragments of the genome that are located adjacent of sites that are digested by restriction enzymes. Since only these regions of the genome are sequenced, there has been an increase in coverage of the regions analyzed (i.e., a greater number of times a same base is sequenced) in comparison with sequencing a complete genome. Therefore, a better reliability in detection of possible SNPs is obtained (Davey and Blaxter 2011; Kai et al. 2014).

The choice of using different restriction enzymes and selecting different fragment sizes are useful to increase the number of targets and proportion of the genome to be analyzed (Rasic et al. 2014). Thousands of yeast, plant and animal SNPs, as well as microsatellites have been isolated and used to solve different questions in the molecular systematics field (Davey and Blaxter 2011; Etter et al. 2011).

On the concern of circadian clock and behavior of triatomines, it is important to clarify the role of some clock genes on the adjustment of daily activity and blood feeding rhythms, which are crucial to vectorial capacity. The recent achievement of RNAi experiments in triatomines (Paim et al. 2012) is an important milestone to further studies in their Chronobiology. In addition, the GAL4/UAS system, a biochemical method used to study gene expression in *Drosophila* (Brand and Perrimon 1993) has been very useful to understand the molecular circadian function. Recently, the GAL4/UAS system was established in *Ae. aegypti* (Kokoza and Raikhel 2011) and also should be adopted to study circadian rhythms of triatomines in the near future.

In the ecology field, analyses of site-occupancy are necessary for triatomine studies. One overlooked problem is that detecting insect infestation foci can be difficult when colonies are small and occupy structurally complex ecotopes. Traditional infestation indices (WHO 1991) are prone to negative bias, confounding ecological inferences. Analytical frameworks based on repeated ecotope-sampling data that explicitly incorporate detection failures (MacKenzie et al. 2006) should be used to understand ecological aspects of triatomine bugs (Abad-Franch et al. 2010a, 2014; Valença-Barbosa et al. 2014).

Applying this approach, it was found that (i) estimates of palm infestation by *Rhodnius* spp. in Amazonia are high (40–60 %), and well above the observed infestation rate (24 %); (ii) individual palm attributes are key drivers of infestation, suggesting that peridomestic palm tree management might help reduce triatomine infestation (Abad-Franch et al. 2010a); (iii) infestation estimates in man-made ecotopes in northeastern Brazil (44.5 ± 6.4 %) were ~2.4–3.9 times higher than naive indices computed (assuming perfect detection after single vector searches; Abad-Franch et al. 2014); and (iv) there is a major role of key-host availability in habitat selection by triatomines in Caatinga, such as *T. b. brasiliensis* (Valença-Barbosa et al. 2014). These results open new research venues for studying *T. cruzi* transmission risk under realistic field conditions.

References

Abad-Franch F, Monteiro FA, Patterson JS, Miles MA (2003) Phylogenetic relationships among members of the Pacific *Rhodnius* lineage (Hemiptera: Reduviidae: Triatominae). Infect Genet Evol 2:244–245

Abad-Franch F, Monteiro FA (2005) Molecular research and the control of Chagas disease vectors. An Acad Bras Cienc 77(3):437–454

Abad-Franch F, Monteiro FA (2007) Biogeography and evolution of Amazonian triatomines (heteroptera: reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. Mem Inst Oswaldo Cruz 102(Suppl 1):57–70

Abad-Franch F, Palomeque FS, Aguilar HM, Miles MA (2005) Field ecology of sylvatic *Rhodnius* populations (Heteroptera, Triatominae): risk factors for palm tree infestation in western ecuador. Trop Med Int Health 10(12):1258–1266

Abad-Franch F, Monteiro FA, Jaramillo ON, Gurgel-Gonçalves R, Dias FB et al (2009) Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: a multi-scale appraisal of the tribe Rhodniini (Triatominae). Acta Trop 110(2–3):159–177

Abad-Franch F, Ferraz G, Campos C, Palomeque FS, Grijalva MJ et al (2010a) Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by triatomine bugs at three spatial scales. PLoS Negl Trop Dis 4(3):e620

Abad-Franch F, Santos WS, Schofield CJ (2010b) Research needs for Chagas disease prevention. Acta Trop 115(1–2):44–54

Abad-Franch F, Diotaiuti L, Gurgel-Gonçalves R, Gürtler RE (2013a) Certifying the interruption of Chagas disease transmission by native vectors: cui bono? Mem Inst Oswaldo Cruz 108 (2):251–254

Abad-Franch F, Pavan MG, Jaramillo-O N, Palomeque FS, Dale C et al (2013b) *Rhodnius barretti*, a new species of Triatominae (Hemiptera: Reduviidae) from western Amazonia. Mem Inst Oswaldo Cruz 108(Suppl 1):92–99

Abad-Franch F, Diotaiuti L, Gurgel-Gonçalves R, Gürtler RE (2014) On bugs and bias: improving Chagas disease control assessment. Mem Inst Oswaldo Cruz 109(1):125–130

Abrahan L, Hernández L, Gorla D, Catalá S (2008) Phenotypic diversity of *Triatoma infestans* at the microgeographic level in the Gran Chaco of Argentina and the Andean valleys of Bolivia. J Med Entomol 45(4):660–666

Abruzzi KC, Rodriguez J, Menet JS, Desrochers J, Zadina A et al (2011) *Drosophila* CLOCK target gene characterization: implications for circadian tissue-specific gene expression. Genes Dev 25(22):2374–2386

Aguilar VHM, Abad-Franch F, Racines VJ, Paucar CA (1999) Epidemiology of Chagas disease in ecuador: a brief review. Mem Inst Oswaldo Cruz 94(Suppl 1):387–393

Aguilar HM, Abad-Franch F, Dias JCP, Junqueira ACV, Coura JR (2007) Chagas disease in the Amazon region. Mem Inst Oswaldo Cruz 102(Suppl. 1):47–55

Allada R, Chung BY (2010) Circadian organization of behavior and physiology in *Drosophila*. Annu Rev Physiol 72:605–624

Almeida CE, Pacheco RS, Noireau F, Costa J (2002) *Triatoma rubrovaria* (Blanchard, 1843) (Hemiptera: Reduviidae) I: isoenzymatic and chromatic patterns of five populations from the state of Rio Grande do Sul, Brazil. Mem Inst Oswaldo Cruz 97(6):829–834

Anderson JM, Lai JE, Dotson EM, Cordon-Rosales C, Ponce C et al (2002) Identification and characterization of microsatellite markers in the Chagas disease vector *Triatoma dimidiata*. Infect Genet Evol 1(3):243–248

Araki AS, Vigoder FM, Bauzer LG, Ferreira GE, Souza NA et al (2009) Molecular and behavioral differentiation among Brazilian populations of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). PLoS Negl Trop Dis 3(1):e365

Barrett TV (1995) Species interfertility and crossing experiments in Triatomine systematics. In: Schofield CJ, Dujardin JP, Jurberg J (eds) Proceedings of the international workshop on population genetics and control of Triatominae. INDRE Press, Mexico, pp 72–77

Barrozo RB, Minoli SA, Lazzari CR (2004) Circadian rhythm of behavioural responsiveness to carbon dioxide in the blood-sucking bug *Triatoma infestans* (Heteroptera: Reduviidae). J Insect Physiol 50(2–3):249–254

Batista TA, Gurgel-Gonçalves R (2009) Ecological niche modelling and differentiation between *Rhodnius neglectus* Lent, 1954 and *Rhodnius nasutus* Stål, 1859 (Hemiptera: Reduviidae: Triatominae) in Brazil. Mem Inst Oswaldo Cruz 104(8):1165–1170

Batista VS, Fernandes FA, Cordeiro-Estrela P, Sarquis O, Lima MM (2013) Ecotope effect in *Triatoma brasiliensis* (Hemiptera: Reduviidae) suggests phenotypic plasticity rather than adaptation. Med Vet Entomol 27(3):247–254

Bauzer LG, Souza NA, Ward RD, Kyriacou CP, Peixoto AA (2002) The period gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. Insect Mol Biol 11(4):315–323

Bodin A, Barrozo RB, Couton L, Lazzari CR (2008) Temporal modulation and adaptive control of the behavioural response to odours in *Rhodnius prolixus*. J Insect Physiol 54(9):1343–1348

Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118(2):401–415

Calleros L, Panzera F, Bargues MD, Monteiro FA, Klisiowicz DR et al (2010) Systematics of *Mepraia* (Hemiptera-Reduviidae): cytogenetic and molecular variation. Infect Genet Evol 10 (2):221–228

Carvalho da Costa R, Soares de Araújo F, Lima-Verde WL (2007) Flora and life-form spectrum in an area of deciduous thorn woodland (caatinga) in northeastern, Brazil. J Arid Environ 68:237–247

Catalá S, Dujardin JP (2001) Antennal sensilla patterns indicate geographic and ecotopic variability among *Triatoma infestans* (Hemiptera: Reduviidae) populations. J Med Entomol 38 (3):423–428

Ceballos LA, Piccinali RV, Marcet PL, Vazquez-Prokopec GM, Cardinal MV et al (2011) Hidden sylvatic foci of the main vector of Chagas disease *Triatoma infestans*: threats to the vector elimination campaign? PLoS Negl Trop Dis 5(10):e1365

Cecere MC, Vázquez-Prokopec GM, Ceballos LA, Gurevitz JM, Zárate JE et al (2006) Comparative trial of effectiveness of pyrethroid insecticides against peridomestic populations of *Triatoma infestans* in northwestern Argentina. J Med Entomol 43(5):902–909

Ceriani MF, Hogenesch JB, Yanovsky M, Panda S, Straume M et al (2002) Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. J Neurosci 22 (21):9305–9319

Clements AN (1999) The biology of mosquitoes. CABI, Wallingford

Constantinou C (1979) Circadian rhythms in insects. PhD. thesis, University of London, London

Costa J (2000) Distribution and characterization of different populations of *Triatoma brasiliensis* Neiva, 1911 (Hemiptera, Reduviidae, Tritominae). Cad Saude Publica 16(Suppl 2):93–95

Costa J, Peterson AT (2012) Ecological niche modeling as a tool for understanding distributions and interactions of vectors, hosts, and etiologic agents of Chagas disease. Adv Exp Med Biol 710:59–70

Costa J, Almeida CE, Dotson EM, Lins A, Vinhaes M et al (2003) The epidemiologic importance of *Triatoma brasiliensis* as a chagas disease vector in Brazil: a revision of domiciliary captures during 1993–1999. Mem Inst Oswaldo Cruz 98(4):443–449

Coura JR, Junqueira AC, Fernandes O, Valente SA, Miles MA (2002) Emerging Chagas disease in Amazonian Brazil. Trends Parasitol 18(4):171–176

Davey J, Blaxter M (2011) RADSeq: next-generation population genetics (vol 9, p 416, 2010). Briefs Funct Genomics 10(2):108

de Oliveira AW, da Silva IG (2007) Geographical distribution and indicators entomologic of synantropic Triatomines captured in the state of Goiás. Rev Soc Bras Med Trop 40(2):204–208

Dias JC (2006) Chagas disease: successes and challenges. Cad Saude Publica 22(10):2020–2021

Dias JC, Machado EM, Fernandes AL, Vinhaes MC (2000) General situation and perspectives of chagas disease in northeastern region, Brazil. Cad Saude Publica 16(Suppl 2):13–34

Dias FBS, Bezerra CM, Machado EM, Casanova C, Diotaiuti L (2008) Ecological aspects of *Rhodnius nasutus* Stål 1859 (Hemiptera: Reduviidae: Triatominae) in palms of the Chapada do Araripe in Ceará, Brazil. Mem Inst Oswaldo Cruz 103(8):824–830

Dias FBS, Jaramillo-O N, Diotaiuti L (2014) Description and characterization of the melanic morphotype of *Rhodnius nasutus* Stål 1859 (Hemiptera: Reduviidae: Triatominae). Rev Soc Bras Med Trop 47(5):637–641

Dujardin JP, Panzera P, Schofield CJ (1999) Triatominae as a model of morphological plasticity under ecological pressure. Mem Inst Oswaldo Cruz 94(Suppl 1):223–228

Dujardin JP, Costa J, Bustamante D, Jaramillo N, Catalá S (2009) Deciphering morphology in Triatominae: the evolutionary signals. Acta Trop 110(2–3):101–111

Dumonteil E, Gourbiére S (2004) Predicting *Triatoma dimidiata* abundance and infection rate: a risk map for natural transmission of chagas disease in the yucatan peninsula of Mexico. Am J Trop Med Hyg 70(5):514–519

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol 772:157–178

Feliciangeli MD, Dujardin JP, Bastrenta B, Mazzarri M, Villegas J et al (2002) Is *Rhodnius robustus* (Hemiptera: Reduviidae) responsible for Chagas disease transmission in western Venezuela? Trop Med Int Health 7(3):280–287

Feliciangeli MD, Sanchez-Martin M, Marrero R, Davies C, Dujardin JP (2007) Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the state of Barinas (Venezuela). Acta Trop 101(2):169–177

Fitzpatrick S, Feliciangeli MD, Sanchez-Martin MJ, Monteiro FA, Miles MA (2008) Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. PLoS Negl Trop Dis 2 (4):e210

Foley HM (2007) Amazonian revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. Front Ecol Environ 5(1):25–32

Galvão C, Gurgel-Gonçalves R (2015) Vetores conhecidos no brasil. In: Galvão C. (ed) Vetores da doença de chagas no Brasil [online]. Curitiba: Sociedade Brasileira de Zoologia, 2014, 289 p. Zoologia: guias e manuais de identificação series. ISBN 978-85-98203-09-6. Available from SciELO Books http://books.scielo.org

Gamboa JC (1961) Comprobación de *Rhodnius prolixus* extradomesticos. Boletín de la Direccion de Malariologia y Saneamiento Ambiental

Gamboa JC (1970). La Población Silvestre de *Rhodnius prolixus* en Venezuela. Boletin Informativo de la Direccion Malariologia y Saneamiento Ambiental

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol Biol Evol 19(5):748–761

Gomez-Nunez JC (1963) Notas sobre la ecologia del *Rhodnius prolixus*. Boletin Informativo de la Direction de Malariologia y Sanemiento Ambiental

Gómez-Palacio A, Jaramillo-O N, Caro-Riaño H, Diaz S, Monteiro FA et al (2012) Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (Hemiptera: Reduviidae: Rhodniini) from Colombia and Panama. Infect Genet Evol 12 (8):1975–1983

Gottdenker NL, Calzada JE, Saldaña A, Carroll CR (2011) Association of anthropogenic land use change and increased abundance of the Chagas disease vector *Rhodnius pallescens* in a rural landscape of Panama. Am J Trop Med Hyg 84(1):70–77

Gottdenker NL, Chaves LF, Calzada JE, Saldaña A, Carroll CR (2012) Host life history strategy, species diversity, and habitat influence *Trypanosoma cruzi* vector infection in changing landscapes. PLoS Negl Trop Dis 6(11):e1884

Guhl F (2007) Chagas disease in Andean countries. Mem Inst Oswaldo Cruz 102(Suppl 1):29–38

Guhl F, Pinto N, Aguilera G (2009) Sylvatic triatominae: a new challenge in vector control transmission. Mem Inst Oswaldo Cruz 104(Suppl 1):71–75

Gurgel-Goncalves R, Cuba CA (2009) Predicting the potential geographical distribution of *Rhodnius neglectus* (Hemiptera, Reduviidae) based on ecological niche modeling. J Med Entomol 46(4):952–960

Gurgel-Gonçalves R, Duarte MA, Ramalho ED, Palma AR, Romaña CA et al (2004a) Spatial distribution of Triatominae populations (Hemiptera: Reduviidae) in *Mauritia flexuosa* palm trees in Federal District of Brazil. Rev Soc Bras Med Trop 37(3):241–247

Gurgel-Gonçalves R, Ramalho ED, Duarte MA, Palma AR, Abad-Franch F et al (2004b) Enzootic transmission of *Trypanosoma cruzi* and *T. rangeli* in the Federal District of Brazil. Rev Inst Med Trop Sao Paulo 46(6):323–330

Gurgel-Gonçalves R, Abad-Franch F, Ferreira JB, Santana DB, Cuba CA (2008) Is *Rhodnius prolixus* (Triatominae) invading houses in central Brazil? Acta Trop 107(2):90–98

Gurgel-Gonçalves R, Ferreira JB, Rosa AF, Bar ME, Galvão C (2011) Geometric morphometrics and ecological niche modelling for delimitation of near-sibling Triatomine species. Med Vet Entomol 25(1):84–93

Gurgel-Gonçalves R, Cura C, Schijman AG, Cuba CA (2012) Infestation of *Mauritia flexuosa* palms by Triatomines (Hemiptera: Reduviidae), vectors of *Trypanosoma cruzi* and *Trypanosoma rangeli* in the Brazilian savanna. Acta Trop 121(2):105–111

Hall JC (2003) Genetics and molecular biology of rhythms in *Drosophila* and other insects. Adv Genet 48:1–280

Hardin PE (2011) Molecular genetic analysis of circadian timekeeping in *Drosophila*. Adv Genet 74:141–173

Harry M (1993) Use of the median process of the pygophore in the identification of *Rhodnius nasutus*, *R. neglectus*, *R. prolixus* and *R. robustus* (Hemiptera: Reduviidae). Ann Trop Med Parasitol 87(3):277–282

Harry M, Galindez I, Cariou ML (1992) Isozyme variability and differentiation between *Rhodnius prolixus*, *R.robustus* and *R.pictipes*, vectors of Chagas disease in Venezuela. Med Vet Entomol 6(1):37–43

Hashimoto K, Schofield CJ (2012) Elimination of *Rhodnius prolixus* in central America. Parasit Vectors 5:45

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A et al (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330 (6006):927–931

Hwang WS, Weirauch C (2012) Evolutionary history of assassin bugs (Insecta: Hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. PLoS ONE 7 (9):e45523

Hypsa V, Tietz DF, Zrzavý J, Rego RO, Galvao C et al (2002) Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a new world origin of the Asiatic clade. Mol Phylogenet Evol 23(3):447–457

Justi SA, Noireau F, Cortez MR, Monteiro FA (2010) Infestation of peridomestic *Attalea phalerata* palms by *Rhodnius stali*, a vector of *Trypanosoma cruzi* in the Alto Beni, Bolivia. Trop Med Int Health 15(6):727–732

Justi SA, Russo CA, Mallet JR, Obara MT, Galvão C (2014) Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). Parasit Vectors 7:149

Kai W, Nomura K, Fujiwara A, Nakamura Y, Yasuike M et al (2014) A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. BMC Genom 15:233

Keegan KP, Pradhan S, Wang JP, Allada R (2007) Meta-analysis of *Drosophila* circadian microarray studies identifies a novel set of rhythmically expressed genes. PLoS Comput Biol 3 (11):e208

Kline K, McCarthy JS, Pearson M, Loukas A, Hotez PJ (2013) Neglected tropical diseases of Oceania: review of their prevalence, distribution, and opportunities for control. PLoS Negl Trop Dis 7(1):e1755

Klink CA, Machado RB (2005) Conservation of the Brazilian Cerrado. Conserv Biol 19:707–713

Kokoza VA, Raikhel AS (2011) Targeted gene expression in the transgenic *Aedes aegypti* using the binary Gal4-UAS system. Insect Biochem Mol Biol 41(8):637–644

Kyriacou CP, Hall JC (1986) Interspecific genetic control of courtship song production and reception in *Drosophila*. Science 232(4749):494–497

Lazzari CR (1991) Circadian rhythm of egg hatching in *Triatoma infestans* (Hemiptera: Reduviidae). J Med Entomol 28(5):740–741

Lazzari CR, Lorenzo MG (2009) Exploiting Triatomine behaviour: alternative perspectives for their control. Mem Inst Oswaldo Cruz 104(Suppl 1):65–70

Lent H, Jurberg J (1969) Observações sobre o ciclo evolutivo, em laboratório, de *Pastrongylus geniculatus* (Latreille 1811) (Hemiptera, Reduviidae, Triatominae). An Acad Bras Ciências 41:125–131

Lent H, Valderrama A (1973) Hallazgo en Venezuela del Triatomino *Rhodnius robustus* Larrousse, 1927 en la Palma *Attalea maracaibensis* Martius (Hemiptera, Reduviidae). Boletin de la Direction de Maloriologia y Saneamiento Ambiental

Lent H, Wygodzinsky P (1979) Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. Bull Am Mus Nat Hist 163:125–520

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64(1):24–30

Lima MM, Sarquis O (2008) Is *Rhodnius nasutus* (Hemiptera; Reduviidae) changing its habitat as a consequence of human activity? Parasitol Res 102(4):797–800

Lima AF, Jeraldo VeL, Silveira MS, Madi RR, Santana TB et al (2012a) Triatomines in dwellings and outbuildings in an endemic area of Chagas disease in northeastern Brazil. Rev Soc Bras Med Trop 45(6):701–706

Lima MM, Sarquis O, de Oliveira TG, Gomes TF, Coutinho C et al (2012b) Investigation of Chagas disease in four periurban areas in northeastern Brazil: epidemiologic survey in man, vectors, non-human hosts and reservoirs. Trans R Soc Trop Med Hyg 106(3):143–149

Lima-Camara TN, Bruno RV, Luz PM, Castro MG, Lourenço-de-Oliveira R et al (2011) Dengue infection increases the locomotor activity of *Aedes aegypti* females. PLoS ONE 6(3):e17690

Lorenzo MG, Lazzari CR (1998) Activity pattern in relation to refuge exploitation and feeding in *Triatoma infestans* (Hemiptera: Reduviidae). Acta Trop 70(2):163–170

Luz PM, Lima-Camara TN, Bruno RV, Castro MG, Sorgine MH et al (2011) Potential impact of a presumed increase in the biting activity of dengue-virus-infected *Aedes aegypti* (Diptera: Culicidae) females on virus transmission dynamics. Mem Inst Oswaldo Cruz 106(6):755–758

Lyman DF, Monteiro FA, Escalante AA, Cordon-Rosales C, Wesson DM et al (1999) Mitochondrial DNA sequence variation among Triatomine vectors of Chagas' disease. Am J Trop Med Hyg 60(3):377–386

Pereira JM, Almeida PS, Sousa AV, Paula AM, Machado, RB, Gurgel-Gonçalves R (2013) Climatic factors influencing Triatomine occurrence in Central-West Brazil. Mem Inst Oswaldo Cruz 108(3)

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL et al (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier Academic Press, San Diego

Marcilla A, Bargues MD, Ramsey JM, Magallon-Gastelum E, Salazar-Schettino et al (2001) The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. Mol Phylogenet Evol 18(1):136–142

Marcilla A, Bargues MD, Abad-Franch F, Panzera F, Carcavallo RU et al (2002) Nuclear rDNA ITS-2 sequences reveal polyphyly of *Panstrongylus* species (Hemiptera: Reduviidae: Triatominae), vectors of *Trypanosoma cruzi*. Infect Genet Evol 1(3):225–235

Márquez E, Jaramillo-O N, Gómez-Palacio A, Dujardin JP (2011) Morphometric and molecular differentiation of a *Rhodnius robustus*-like form from *R. robustus* Larousse 1927 and *R. prolixus* Stal 1859 (Hemiptera, Reduviidae). Acta Trop 120(1–2):103–109

Mas-Coma S, Bargues MD (2009) Populations, hybrids and the systematic concepts of species and subspecies in Chagas disease Triatomine vectors inferred from nuclear ribosomal and mitochondrial DNA. Acta Trop 110(2–3):112–136

Matias A, de la Rive JX, Torrez M, Dujardin JP (2001) *Rhodnius robustus* in Bolivia identified by its wings. Mem Inst Oswaldo Cruz 96(7):947–950

McEwen PK, Lehane MJ (1993) Factors influencing flight initiation in the Triatomine bug *Triatoma infestans* (Hemiptera: Reduviidae). Insect Sci Appl 14:461–464

Mendonça VJ, da Silva MT, de Araújo RF, Júnior JM, Júnior MB et al (2009) Phylogeny of *Triatoma sherlocki* (Hemiptera: Reduviidae: Triatominae) inferred from two mitochondrial genes suggests its location within the *Triatoma brasiliensis* complex. Am J Trop Med Hyg 81 (5):858–864

Miles MA, Feliciangeli MD, de Arias AR (2003) American trypanosomiasis (Chagas' disease) and the role of molecular epidemiology in guiding control strategies. BMJ 326(7404):1444–1448

Monteiro FA, Lazoski C, Noireau F, Solé-Cava AM (2002) Allozyme relationships among ten species of *Rhodniini*, showing paraphyly of *Rhodnius* including *Psammolestes*. Med Vet Entomol 16(1):83–90

Monteiro FA, Barrett TV, Fitzpatrick S, Cordon-Rosales C, Feliciangeli D et al (2003) Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. Mol Ecol 12(4):997–1006

Monteiro FA, Donnelly MJ, Beard CB, Costa J (2004) Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in northeast Brazil. Mol Phylogenet Evol 32 (1):46–56

Monteiro FA, Peretolchina T, Lazoski C, Harris K, Dotson EM et al (2013) Phylogeographic pattern and extensive mitochondrial DNA divergence disclose a species complex within the Chagas disease vector *Triatoma dimidiata*. PLoS ONE 8(8):e70974

Morrone JJ (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. Ann Rev Entomol 51:467–494

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853–858

Noireau F, Flores R, Gutierrez T, Dujardin JP (1997) Detection of sylvatic dark morphs of *Triatoma infestans* in the Bolivian Chaco. Mem Inst Oswaldo Cruz 92(5):583–584

Paim RM, Pereira MH, Di Ponzio R, Rodrigues JO, Guarneri AA et al (2012) Validation of reference genes for expression analysis in the salivary gland and the intestine of *Rhodnius prolixus* (Hemiptera, Reduviidae) under different experimental conditions by quantitative real-time PCR. BMC Res Notes 5:128

Pavan MG, Monteiro FA (2007) A multiplex PCR assay that separates *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). Trop Med Int Health 12(6):751–758

Pavan MG, Mesquita RD, Lawrence GG, Lazoski C, Dotson EM et al (2013) A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). Infect Genet Evol 14:426–433

Peixoto AA (2002) Evolutionary behavioral genetics in *Drosophila*. Adv Genet 47:117–150

Peterson AT (2014) Mapping disease transmission risk: enriching models using biogeography and ecology. University Press, Baltimore

Peterson AT, Soberón J, Pearson RG, Anderson R, Martínez-Meyer E, Nakamura M, Araújo M (2011) Ecological niches and geographic distributions. Princeton University Press, Princetown

Pifano FC (1973) La dinámica epidemiológica de la enfermedad de Chagas en el Valle de los Naranjos, Estado de Carabobo, Venezuela. I. Contribución al estudio de los focos naturales silvestres del *Schizotripanum cruzi* Chagas 1909. Arch Venez Med Trop Parasitol Med 5:1–29

Quisberth S, Waleckx E, Monje M, Chang B, Noireau F et al (2011) Andean and non-Andean ITS-2 and mtCytB haplotypes of *Triatoma infestans* are observed in the Gran Chaco (Bolivia): population genetics and the origin of reinfestation. Infect Genet Evol 11(5):1006–1014

Rasic G, Filipovic I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. BMC Genomics 15:275

Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375(9723):1388–1402

Ritchie MG, Halsey EJ, Gleason JM (1999) *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. Anim Behav 58(3):649–657

Rivas GB, Souza NA, Peixoto AA (2008) Analysis of the activity patterns of two sympatric sandfly siblings of the *Lutzomyia longipalpis* species complex from Brazil. Med Vet Entomol 22(3):288–290

Rodríguez CS, Crocco L, Altamirano A, Catalá S (2013) Changes related to gender, geographic population and habitat in the antennal phenotype of *Triatoma patagonica* Del Ponte 1929 (Hemiptera: Reduviidae). Acta Trop 125(2):143–149

Rossi JCN, Duarte EC, Gurgel-Gonçalves R (2015) Factors associated with the occurrence of *Triatoma sordida* (Hemiptera: Reduviidae) in rural localities of Central-West Brazil. Mem Inst Oswaldo Cruz 1–9

Rund SS, Lee SJ, Bush BR, Duffield GE (2012) Strain- and sex-specific differences in daily flight activity and the circadian clock of *Anopheles gambiae* mosquitoes. J Insect Physiol 58 (12):1609–1619

Salazar MI, Richardson JH, Sánchez-Vargas I, Olson KE, Beaty BJ (2007) Dengue virus type 2: replication and tropisms in orally infected *Aedes aegypti* mosquitoes. BMC Microbiol 7:9

Sarquis O, Borges-Pereira J, Mac Cord JR, Gomes TF, Cabello PH et al (2004) Epidemiology of Chagas disease in Jaguaruana, Ceará, Brazil I: presence of triatomines and index of *Trypanosoma cruzi* infection in four localities of a rural area. Mem Inst Oswaldo Cruz 99 (3):263–270

Saunders DS (2002) Insect clocks, 3rd edn. Elsevier, Amsterdam, p 576

Schofield CJ, Diotaiuti L, Dujardin JP (1999) The process of domestication in Triatominae. Mem Inst Oswaldo Cruz 94(Suppl 1):375–378

Senior K (2007) Chagas disease: moving towards global elimination. Lancet Infect Dis 7(9):572

Silva JMC, Tabarelli M, Fonseca MT, Lins LV (2004) Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação. Ministério do Meio Ambiente, Brasília

Silveira AC, Dias JC (2011) The control of vectorial transmission. Rev Soc Bras Med Trop 44 (Suppl 2):52–63

Soares RP, Gontijo NF, Romanha AJ, Diotaiuti L, Pereira MH (1998) Salivary heme proteins distinguish *Rhodnius prolixus* from *Rhodnius robustus* (Hemiptera: Reduviidae: Triatominae). Acta Trop 71(3):285–291

Tonn RL, Carcavallo RU, Ortega R (1976) Notas sobre la biologfa, ecologfa y distribuci6n geograficade *Rhodnius robustus* Larrousse, 1927 (Hemiptera: Reduviidae). Bol Dir Malariol Saneam Ambien 16:146–152

UNEP-United Nations Environment Programme (2005) Geo Year Book 2004/5: An overview of our changing environment. (http://www.unep.org/yearbook/2004/pdf/geo_yearbook_2004.pdf)

Valença-Barbosa C, Lima MM, Sarquis O, Bezerra CM, Abad-Franch F (2014) A common Caatinga cactus, *Pilosocereus gounellei*, is an important ecotope of wild *Triatoma brasiliensis* populations in the Jaguaribe valley of northeastern Brazil. Am J Trop Med Hyg 90(6):1059–1062

Villegas J, Feliciangeli MD, Dujardin JP (2002) Wing shape divergence between *Rhodnius prolixus* from Cojedes (Venezuela) and *Rhodnius robustus* from Mérida (Venezuela). Infect Genet Evol 2(2):121–128

Vinhaes MC, de Oliveira SV, Reis PO, de Lacerda Sousa AC, Silva RA et al (2014) Assessing the vulnerability of Brazilian municipalities to the vectorial transmission of *Trypanosoma cruzi* using multi-criteria decision analysis. Acta Trop 137:105–110

Wheeler DA, Kyriacou CP, Greenacre ML, Yu Q, Rutila JE et al (1991) Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. Science 251 (4997):1082–1085

WHO-World Health Organisation (1991) Control of Chagas Disease. Report of a WHO Expert Committee, Genevanumber 811

WHO-World Health Organisation (2007) New global effort to eliminate Chagas disease

WHO-World Health Organisation (2015) Equitable and affordable health coverage against neglected tropical diseases is achievable

Chapter 19
Forest Tree Species Traced with a DNA-Based Proof for Illegal Logging Case in Poland

Justyna A. Nowakowska, Tomasz Oszako, Anna Tereba and Agata Konecka

Abstract Precise identification of biological samples remains the most important proof in the forensic science. Illegal logging has become the urgent issue in Poland during the last decades, and conventional methods of investigation turn out to be often insufficient. Recently, the DNA-based markers (SSR and cytoplasmic genes) can remarkably help in the forensic botany performed by the Forest Service Guards and the Police investigation in illegal logging of timber. The identification method relies on comparison of the piece of evidence (i.e., stolen wood fragments) with the piece of reference (e.g., tree parts remained in the forest). We present the usefulness of the DNA neutral markers (i.e., microsatellite loci) and cytoplasmic genes in forensic botany based on several case studies of illegal wood identification in Poland, concerning the most economically important coniferous tree species such as *Pinus sylvestris* L., *Picea abies* (L.) Karst., *Abies alba* Mill., and *Larix decidua* (L.). Thanks to the DNA profiles established on the basis of minimum 4 microsatellite nuclear DNA loci, and at least one cytoplasmic organelle (mitochondrial or chloroplast) DNA marker, the determination of the DNA profiles provided fast and reliable comparison between material of evidence (also wood and needles) and material of reference (first of all tree stumps) in the forest. These data strongly supported the decision taken by several District Courts in Poland, as far as the

J.A. Nowakowska (✉) · A. Tereba · A. Konecka
Laboratory of Molecular Biology, Forest Research Institute, 3 Braci Leśnej Street, 05-090 Sękocin Stary, Poland
e-mail: J.Nowakowska@ibles.waw.pl

T. Oszako
Department of Forest Protection, Forest Research Institute, 3 Braci Leśnej Street, 05-090 Sękocin Stary, Poland

T. Oszako
Faulty of Forestry, Technical University of Bialystok, 8 Piłsudskiego Street, 17-200 Hajnówka, Poland

T. Oszako
European Food Safety Authority (EFSA), Via Carlo Magno 1, 43100 Parma, Italy

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_19

identification of wood samples was proved with a high probability (approximately 98–99 %). The aim of the below publication is to present Polish case study on DNA use to fight illegal logging which became very successful among foresters.

19.1 Introduction

Production of quicker and larger ships as well as globalization phenomenon stimulates a rapid increase of worldwide trade. The wood as a raw material was recently recognized as renewable source of energy. The pattern of traditional division among wood constructing and pulp and paper industry has changed because the new customers appeared on the market dealing with biofuels and biorefineries. They strongly compete with wood panel industry using even wood residues left after tree harvesting, e.g., to produce particle boards. At the same time, societies in developed countries prefer to conserve high forests rather than mobilize their exploitation. New values such as tourism and recreation, and picking berries or mushrooms are going to become fully marketable goods, soon. The international agreements concerning sequestration of CO_2 also limit the use of forests. The fast increment of human population which will reach 9 billion in 2050 will only increase demand on wood and its products. The price of wood timber significantly increased during last years and became precious product being willingly harvested illegally and smuggled to Europe. It causes steady diminishing of the world forest area. Forest area covers c.a. 30 % of the world's land surface, providing renewable fuel, wood, timber, and bark (www.fao.org). Moreover, it offers many forest externalities such as a source of medicinal products, carbon sequestration, water protection, and habitat for many terrestrial species. In this sense, the illegal logging is much more than only a value of wood timber. It causes a major problem for many timber-producing developing countries, including the environmental damage; it promotes pathogen and pest transport and generates costs of billions of dollars in government's lost revenue. The international trade of illegally logged wood is a major problem for the countries legally producing timber in many Southeast Asian and South American countries. It is estimated that around 20 % of illegally harvested wood timber is reaching European Union. Without stopping the deforestation phenomenon, it is not possible to meet Europe's 2020 goal on the reduction of greenhouse gases by 20 %.

The European Commission decided to tackle with this topic as its priority and developed proper legislation. Firstly, the action was based on voluntary agreements (FLEGT) and recently thanks to the new EU Timber Regulation. The European Union Regulation No 995/2010 of the European Parliament and of the Council[1]

[1]Regulation (EU) No 995/2010 of the European Parliament and of the Council of 20 October 2010 laying down the obligations of operators who place timber and timber products on the market. Official Journal of the European Union L 295/23-34.

also known as the (illegal) Timber Regulation (EUTR) laid down the obligations of operators who place timber and timber products on the market. The Regulation entered into application on March 3, 2013. It counters the trade in illegally harvested timber and timber products. The wood stolen from the forest is equally treated as those wood timbers from which the VAT is not paid or the wood was harvested from the stand managed in non-sustainable way (e.g., too much cuttings). The EU Timber Regulation covers a wide range of timber products listed in its Annex using EU Custom's code nomenclature. If this initiative has to slow down or stop deforestation process in endanger parts of the world, the scientists should support policy makers with the proper tools. Such a chance is given by the quickly developing genetics and DNA-analyzing technologies (e.g., New-Generation Sequencing). Recent advances in forensic genetics provide unique opportunity for such an investigation, which is the focus of this chapter.

19.2 Scientific Bases for the Genetic Tool to Combat the Illegal Harvesting and Trade

It is estimated that timber theft generates annually the greatest losses (among other forms of crimes) incurred by the Polish State Forests, and thieves stay unpunished for such misconduct (Nowakowska and Pasternak 2014). Since far, the dendrochronological methods are used to trace illegal tree cuttings. The advantage of this technique basically resides in approximate data indication when the illegal cutting occurred (Yaman and Akkemik 2009). But, the insufficient number or lack of clear annual rings due to wood decay severely limits the use of dendrochronological identification of wood samples.

Recently, the forensic botany has become a useful tool, which can help to save forest resources. The method based on the developed set of DNA markers to identify timber genetic profiles could then provide new evidence in a litigation cases faced by the Forest Guard Service or the Police. It is based on the rule that each organism genome contains a large amount of DNA that represents a huge dataset for genetic profiling. For the first time the DNA-based proof was used to determine the identity of a crime perpetrator in United Kingdom (Seton 1988), nearly 20 years after the discovery of the DNA molecules in human cells. Molecular markers can solve the problem of illegal logging as far as DNA molecules are well conserved in wood during harvesting and processing, and the nucleic acid isolation follows standardized methods. The noncoding DNA fragments are a source of variation (polymorphism) of an organism and constitute unique characteristics (pattern) of an individual. Many changes occurring at the DNA molecule level result from a single base-pair mutation, deletions, or substitutions of nucleic acids during replication errors. If those mistakes in gene structure are transmitted to the progeny, the phylogenetic study confirming the relationship of people with animals in the parental line is possible.

Particularly, the microsatellite sequences, known as short sequence repeats (SSR) or short tandem repeats (STR), are prone to the genetic identification at the individual level. The conifer's microsatellite sequences are randomly distributed in the whole genome. The repeated base-pair motifs mostly occur around intergenic regions, in other repeat elements such as satellites or retro-elements. Uneven distribution of repeat motifs may be connected with their functional role in gene expression of some of them, and this kind of distribution has been reported for eukaryotic genomes (Echt and May-Marquardt 1997). For example, motifs $(AC)_n$ and $(AG)_n$ occur in different parts of conifer genomes, and this distribution is conserved in gymnosperms (Smith and Devey 1994; Schmidt et al. 2000). The precise identification of biological samples based on microsatellite loci remains a crucial proof in forensic science, including animal and human probes (Schmidt et al. 2000). In many genomes, the microsatellites in general arise by interaction between DNA motifs or can be carried by the transposons. They mutate by replication slippage and unequal crossing over during the recombination. After the mutation process, the SSRs can reach expansion equilibrium and random mutations, which can lead to break repetitive patterns or the mutation accumulation in some parts of genome (Buschiazzo and Gemmell 2006). Such a phenomenon can be connected with the replacing of the interrupt repetitive motives in particular hot spots (Buschiazzo and Gemmell 2006). So far, a few of STR-rich regions were studied in plants, mostly in nuclear genome (Karhu et al. 2000), but they were also observed in mitochondrial DNA (Jaramillo-Correa et al. 2013).

19.3 Genetics to Provide Evidence for the Courts

Forensic genetics aims to provide evidence in criminal cases basing on the characteristics of the genetic code of the seized material (collected at the crime scene) and the reference one (taken from the alleged perpetrator). The analysis of the genetic profiles for forensic case work includes all kind of organisms ranging from humans and animals, to insects, microorganisms, and plants. The plant material can generate very discriminating DNA profiles as far as their genomes are the biggest among all living organisms. This allows the match of the biological evidence from the scene of crime to an individual with a high degree of confidence.

The forensic botany compares the DNA profiles between the tree tissue (wood, needles as material of evidence) and tree stumps (material of reference) in the forest (Fig. 19.1). The phenomenon of illegal logging has become an urgent issue to be solved in Polish forestry during the last decades, and conventional methods of investigation often turned out to be insufficient. The DNA-based proof becomes of great assistance in many investigations performed by the Polish Forest Service Guards, District Courts, and the Police. The basic identification method relies on the comparison of the piece of evidence (i.e., wood fragments originating from the stolen timber) with the piece of reference (e.g., stumps remained in the forest stand). A small amount of material (100 mg of wood, leaves, or needles) collected in the

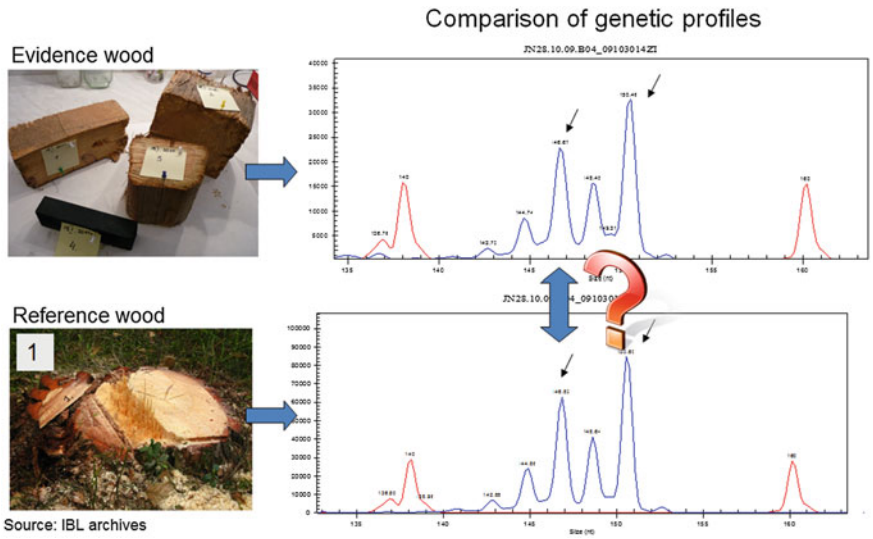


Fig. 19.1 General scheme of genetic profile comparison in wood sample identification procedure

field is required for biochemical analyses. The most frequent cases of illegal conifer wood identification in Poland concern Scots pine, Norway spruce, European silver fir, and European larch. Basic methodology for the DNA study relies on the following stages, i.e., DNA extraction (based on the lysis of cell walls), DNA amplification via polymerase chain reaction (PCR) method with specific primers for nuclear microsatellite or organelle alleles, genotyping of the PCR products in automated sequencer, and finally comparison of the DNA profiles obtained for all samples. This method avoids timely collection of data such as tree age, diameter, height, and thickness, although such a piece of information may be advantageous in wood identification process.

19.4 Difficulties Occurring in DNA-Based Analyses of Wood

The main difficulty in DNA-based analyses remains in proper DNA extraction method from wood tissues, because of the high amount of polysaccharides and polyphenolic compound residuals which inhibit the Taq polymerase during the PCR (Tibbits et al. 2006). The removal of those contaminants guarantees the success of further amplification and accurateness of DNA fragment (allele or gene) detection after the capillary electrophoresis performed in automated sequencer. Many methods of DNA extraction from plant tissue (including wood) have been proposed, e.g., CTAB-based isolation described by Doyle and Doyle (1990),

DNeasy Plant Mini Kit (Qiagen®) used by Dumolin-Lapègue et al. (1999), and MagAttract 96 DNA Plant Core Kit (Qiagen®) (Rachmayanti et al. 2006). Those methods yielded c.a. 2 μg of DNA per 50–100 mg of dried wood sample suitable for nuclear and organelle DNA amplification. The tissue type, i.e., cambium, sapwood, or hardwood, can generate different yield of the DNA molecules, in favor of cambial cells in *Pinus radiata* (Tibbits et al. 2006) and *Quercus robur* (Nowakowska 2011). Good-quality DNA was also extracted by Asif and Cannon (2005) and Tibbits et al. (2006), who supplemented the classical CTAB method with buffer containing NaCl and BSA effectively removing coextracted contaminants. In our case study, the DNeasy Plant Mini Kit (Qiagen®) used for Scots pine, NucleoSpin Plant II (Macherey-Nagel®) applied for Norway spruce, and Phire® Plant Direct PCR kit (Finnzymes®) method used in case of European silver fir and European larch wood resulted in correct amplification patterns of the studied forensic samples.

Generally, the DNA molecules are well preserved in the recently seized wood material. The best DNA isolation is achieved with an actively growing plant tissue (i.e., leaves, needles, buds), while the wood material seized in the field is often subjected to unfavorable climatic conditions during weeks or months. A potential problem occurs with ancient and/or potentially damaged DNA extracted from low amounts of tissues, which leads to the artefacts due to polymerase errors occurring during the amplification process. In many cases, chloroplast or mitochondrial DNA molecules represented by numerous copies per cell even in 100-year-old wooden material provide better templates for the PCRs than the nuclear genome being present in only two copies in each cell and being susceptible to all types of mutations (Dumolin-Lapègue et al. 1999). During archaeological investigations, cytoplasmic markers helped in the study of the geographical origin or taxonomic status of wood used in buildings, furniture, handcrafts, or barrels (Marco et al. 1994; Asif and Cannon 2005).

19.5 Cases of Different Forest Tree Species Recognition by DNA Method

In 2013, **five Scots pine samples** were seized by the Forest Service Guard in the Forest District of G. in Poland (Table 19.1) and sent to the Laboratory of Molecular Biology (LMB) in FRI (IBL, Poland) for relevant genetic analyses. The genomic DNA was extracted with DNeasy Plant 250 Kit (Qiagen®) according to manufacturer's instructions. Six nuclear microsatellite loci were amplified, i.e., SPAG 7.14, SPAC 11.4, SPAC 11.6, and SPAC 12.5 (Soranzo et al. 1998), and SsrPt-ctg4363, Rptest11 (Chagné et al. 2004) according to the LMB procedures (Nowakowska 2011). Eight chloroplast microsatellite DNA markers, i.e., PCP26106, PCP30277, PCP36567, PCP450712, PCP719872, PCP873142, Pt302042, and Pt71936 (Provan et al. 1998, Vendramin et al. 1996), were analyzed in order to generate the DNA profile of the samples.

Table 19.1 Genetic profiles of evidence and reference wood samples of Scots pine with 6 nuclear[1)] and 8 chloroplast[2)] microsatellite DNA markers

No.	Material[a]	SPAG 7.14[1)]		SPAC 12.5[1)]		SPAC 11.4[1)]		SPAC 11.6[1)]		SsrPt-ctg 4363[1)]		Rptest11[1)]	
1	Evidence wood	199	225	131	197	114	146	109	141	96	98	211	214
2	**Evidence wood**	**199**	**223**	**131**	**197**	**114**	**146**	**107**	**141**	**101**	**103**	**211**	**214**
3	Evidence wood	199	223	133	199	116	144	107	141	134	138	202	205
1	**Reference wood**	**199**	**223**	**131**	**197**	**114**	**146**	**107**	**141**	**101**	**103**	**211**	**214**
2	Reference wood	189	191	163	165	112	118	105	109	94	96	193	199

No.	Material[a]	PCP 26106[2)]	PCP 30277[2)]	PCP 36567[2)]	PCP 45071[2)]	PCP 71987[2)]	PCP 87314[2)]	Pt 30204[2)]	Pt 71936[2)]
1	Evidence wood	148	136	111	153	109	115	140	146
2	**Evidence wood**	**148**	**136**	**111**	**153**	**109**	**113**	**140**	**146**
3	Evidence wood	148	135	111	155	109	113	142	148
1	**Reference wood**	**148**	**136**	**111**	**153**	**109**	**113**	**140**	**146**
2	Reference wood	147	135	110	155	109	116	142	146

[a]Sampled material comprises
Evidence wood—pieces of wood seized onsite by the Police or the Forest Guard Service
Reference wood—a stump from the forest
Identical profiles are highlighted in bold
[1)]Nuclear microsatellite markers
[2)]Chloroplast microsatellite markers

Table 19.2 Genetic profiles of evidence and reference wood samples of Norway spruce with 9 nuclear[1)] microsatellite loci and 1 mitochondrial[2)] DNA marker

No.	Material[a]	EATC1 B02[1)]		EATC1 E3[1)]		EATC2 G05[1)]		SpaGG3[1)]		SpaGC1[1)]		EATC2 B02[1)]		EATC1 G2[1)]		SpaGC2[1)]		SpaG2[1)]		*nad*1[2)]
1	**Evidence wood**	**195**	**195**	**132**	**135**	**222**	**228**	**126**	**130**	**94**	**98**	**194**	**197**	**210**	**213**	**83**	**87**	**90**	**92**	**c**
1	**Reference wood**	**195**	**195**	**132**	**135**	**222**	**228**	**126**	**130**	**94**	**98**	**194**	**197**	**210**	**213**	**83**	**87**	**90**	**92**	**c**
2	Reference wood	193	201	129	135	240	246	126	130	90	96	182	182	207	213	105	111	92	92	a
3	Reference wood	213	217	123	123	210	213	146	148	102	104	191	194	180	189	131	133	100	104	a

[a]Sampled material comprises
Evidence wood—pieces of wood seized onsite by the Police or the Forest Guard Service
Reference wood—a stump from the forest
Identical profiles are highlighted in bold
[1)]Nuclear microsatellite markers
[2)]Chloroplast microsatellite markers

In 2014, **four Norway spruce wood samples** (Table 19.2) from the Forest District in N. were analyzed with the help of nine microsatellite loci, i.e., EATC1B02, EATC1E3, EATC2G05, SpAGG3, SpAGC1, EATC2B02, EATC1G2, SpAGC2, and SpAG2 (Pfeiffer et al. 1997) according to the LMB procedures. The mitochondrial *nad1* gene (Sperisen et al. 2001) variation was examined in DNA 1000 chip electrophoresis in Bioanalyser® (USA).

Genetic profiles of **four silver fir wood samples** (Table 19.3) seized in 2012 in the S. Forest District were established on the basis of four nuclear microsatellite loci SF1, SFb4, SF333, and SF239 (Cremer et al. 2006), as well as two chloroplast microsatellite loci Pt30204 and Pt71936 (Vendramin and Ziegenhagen 1997) with Phire® Plant Direct PCR Kit (Thermo Scientific, USA; ABO Ltd., Poland) according to manufacturer's instructions.

In 2013, the Phire® Plant Direct PCR Kit was also used to analyze the genetic profiles of **six European larch wood samples** (Table 19.4) seized in the Forest District of K., taking into account four nuclear microsatellite loci bcLK263, bcLK211, bcLK225, and bcLK228 according to the description of Isoda and Watanabe (2006). For larch samples, additional taxonomic identification was performed, thanks to chloroplast and mitochondrial markers described by Acheré et al. (2004).

For the investigated wood material, the polymerase chain reactions (PCRs) were conducted in Veriti 96 Thermal Cycler (Life Technologies™, USA), and the quality of DNA prior amplification was checked with NanoDrop® ND-1000 spectrophotometer (Wilmington, USA). The PCR products were analyzed with the 3500 Genetic Analyzer (Life Technologies™, USA) using the 3500 Data Collection Software and GeneMapper® version 5 (Life Technologies™, USA). Mean *PIC* (polymorphism information content) values were established for each set of markers in MolKin software version 2.0 (Gutiérez et al. 2005).

19.6 Identity of Compared Wood Samples

In order to avoid the accidental identity of two compared wood samples with the material of reference in the stand, the probability of identity P_{ID} is calculated according to Hedrick (2000):

$$P_{G_{hj}|k_j} = \prod_{i \in \Omega_h}^{\text{samples in } \Omega_h} P_{g_{ij}|k_j}.$$

where $P_{g_{ij}|k_j}$ signifies frequencies in ij alleles in k loci.

The P_{ID} estimator illustrates the probability that two individuals drawn at random from a population will have the same genotype at multiple loci and is generally used to assess the statistical confidence of the marker system for individual

Table 19.3 Genetic profiles of evidence and reference wood samples of European silver fir with 4 nuclear[1)] and 2 chloroplast[2)] microsatellite DNA markers

No.	Material[a]	SF1[1)]		SFb4[1)]		SF333[1)]		SF239[1)]		Pt30204[2)]	Pt71936[2)]
1	Evidence wood	214	218	144	172	166	168	123	123	142	149
2	Evidence wood	218	220	146	174	166	166	108	112	142	149
3	**Evidence wood**	**222**	**224**	**148**	**174**	**164**	**168**	**108**	**110**	**142**	**149**
1	**Reference wood**	**222**	**224**	**148**	**174**	**164**	**168**	**108**	**110**	**142**	**149**

[a]Sampled material comprises
Evidence wood—pieces of wood seized onsite by the Police or the Forest Guard Service
Reference wood—a stump from the forest
Identical profiles are highlighted in bold
[1)]Nuclear microsatellite markers
[2)]Chloroplast microsatellite markers

Table 19.4 Genetic profiles of evidence and reference wood samples of European larch 4 nuclear[1)] microsatellite DNA markers

No.	Material[a]	bcLK228[1)]		bcLK225[1)]		bcLK211[1)]		bcLK263[1)]	
1	Evidence wood	179	195	160	184	207	209	184	240
2	**Evidence wood**	**179**	**199**	**160**	**184**	**211**	**213**	**184**	**240**
3	Evidence wood	179	195	160	184	207	209	184	240
1	Reference wood	199	199	160	184	209	209	184	240
2	Reference wood	177	199	158	158	207	209	182	182
3	**Reference wood**	**179**	**199**	**160**	**184**	**211**	**213**	**184**	**240**

[a]Sampled material comprises
Evidence wood—pieces of wood seized onsite by the Police or the Forest Guard Service
Reference wood—a stump from the forest
Identical profiles are highlighted in bold
[1)]Nuclear microsatellite markers
[2)]Chloroplast microsatellite markers

identification. In practice, the $P_{ID} = 0.001$ for a set of DNA markers means 0.1 % of probability for the identical genotype of tree existing in the forest. Therefore, the genetic profiles of wood from material of evidence and material of reference are identical with 99.9 %.

In all coniferous samples studied for forensic purposes, the P_{ID} estimator ranged from 0.0001 (in Scots pine wood) to 0.011 (in European silver fir wood), proving the very low probability of identical genotypes among 100 randomly examined trees in the stand. These data strongly supported the decision taken by several District Courts in Poland, as far as the identification of wood samples was proved with the high probability (approximately 98.89–99.99 %).

19.7 DNA Profiles as a Strong Proof in the Court Cases

The essence of comparative studies based on DNA analysis is to provide a set of genetic markers with high discrimination power between individuals. Current SSR-based multiplex kit taking into account biallelic single nucleotide polymorphism (SNPs) ensures almost 100 % of likelihood for the match between two human profiles with an error ratio of 1 to billion (Goodwin et al. 2011). Human population is sufficient to analyze c.a. 9 STR loci with the probability close to $1{:}10^6$. Due to the high polymorphism level of the SSR loci in woody species, the minimum of 4 microsatellite DNA loci is sufficient to exclude or confirm the identity of investigated coniferous material at the level of 99.99 %.

Based on 6 nuclear and 8 chloroplast DNA markers of **Scots pine wood** from the G. Forest District, the genetic identity of sample n° 2 (material of evidence) and wood sample n° 1 (material of reference) was concluded (Table 19.1). The final decision sent to the court was supported by the genetic identity between those samples proved with a high probability of 99.99 % ($P_{ID} = 0.0001$).

A wood comparison study performed for **Norway spruce wood** from the N. Forest District revealed the genetic identity between evidence material n° 1 and reference material n° 1 (Table 19.2) with 99.98 % of probability (P_{ID} = 0.002).

The DNA profile comparison between **European silver fir wood** samples from the S. Forest District proved the genetic identity of evidence between the material n° 3 and the reference material n° 1 (Table 19.3) with 98.90 % of probability (P_{ID} = 0.011).

Based on 4 nuclear DNA markers of **European larch wood** from the K. Forest District, the genetic identity between the sample n° 2 (material of evidence) and the wood sample n° 3 (material of reference) was determined (Table 19.4), with the high probability of 98.89 % (P_{ID} = 0.0102).

All selected nuclear DNA markers were characterized by the high level of polymorphism content, with the mean *PIC* = 85.6 % for microsatellite nuclear DNA loci and *PIC* = 95.4 % for cytoplasmic DNA (*PIC* = 40.0 % for mitochondrial and *PIC* = 65.7 % for chloroplast DNA loci).

19.8 Selection Pressure and Environment-Driven Forces

The genomes of many species undergo long evolutionary processes resulting in Hardy–Weinberg equilibrium (HWE) observed in numerous forest tree populations. That means the genotype frequencies can be predicted from the frequencies of the SSR alleles. The effects of genetic drift (more pronounced in small, isolated populations) and self-pollination phenomenon are less frequent in wind-pollinated pine, spruce, fir, and larch stands. Moreover, the DNA loci are less prone to selection pressure and environment-driven forces.

Coniferous forests cover mostly areas in the Northern Hemisphere where they have a huge environmental and economic value, especially in wood industry. However, despite the high importance of this group of woody plants, the knowledge of their genomes is still limited. The genomes are not well characterized, although there are some reports pointing that gene families in conifers are much bigger than those in their angiosperm equivalents, which additionally contain a number of pseudogenes and are rich in high repetitive elements (Ahuja and Neadle 2005; Buschiazzo et al. 2012).

19.9 Genome Complexity and SSR Markers

Due to the big size of the conifer genome (ca. 20–30 Gb), its molecular analysis is a major challenge. However, the genome size is not the only obstacle in the molecular analysis, but also the large effective population size, the high heterozygosity, and the low substitution rate connected to their long life span. One of the hypotheses of an evolution of conifer genome suggests that the basal number of 12 chromosomes

($2n = 24$) slowly expanded according to the activity of transposable elements (LTR—long terminal repeats), which are shared in many conifer genomes. This expansion probably started early and in opposite to angiosperms, and the LTR copies remained in the genomes of conifers, because of less effective mechanism of removal of transposable elements in conifer than in other organisms (Nystedt et al. 2013). Moreover, huge chromosomes can be formed as a result of chromosome replication with genes separated with long fragments of transposable elements, pseudogenes, and highly polymorphic noncoding regions with low recombination frequencies. Conservative structure of the genome, marginal DNA rearrangements, and lack of the whole-genome duplication can explain a high degree of conservation and low phenotypic variation. However, the high degree of the adaptability of different conifers in varied ecosystems may be closely related to the high genetic variability of their huge genome sizes.

Nowadays, the SSR markers are widely used for the analysis of the population diversity, gene flow, parentship, construction of linkage maps, etc. Therefore, the knowledge of their importance, the evolution, and the rate of mutation is very valuable in many fields of research. Thanks to the DNA profiles established on the basis of minimum 4 microsatellite nuclear DNA loci, and at least one cytoplasmic (mitochondrial or chloroplast) DNA marker, the determination of the DNA profiles provided fast and reliable comparison between material of evidence (wood, needles) and material of reference (tree stumps) in the forest. Confirmed by the present and previous data obtained during the illegal logging investigation in Poland, the low probability of occurrence (P_{ID} value) of the identical genotype among randomly chosen trees in a forest stand proved the appropriateness of all nuclear and cytoplasmic DNA markers designed in the Forest Research Institute (FRI) for Scots pine, Norway spruce, silver fir, and European larch wood material investigation (Nowakowska 2011). Similar markers assigned for European larch identification may be successfully used for Japanese larch (*L. kaempferi* Sorg.) wood genetic identification (unpublished data). Still, the most crucial in the forest practice against illegal logging remains the proper manner of wood collection, avoiding material deterioration or contamination. Detailed instructions for seizing wood samples for DNA analyses were presented in a video training film entitled "DNA analysis of wood in combating timber illegal trade" (www.ibles.pl).

19.10 Perspectives of DNA Profiling Methods to Investigate Illegal Logging in Forests

The investigation focusing on DNA markers of wood resulted in designing the efficient method to be used for forensic purposes. The presented methodology helps to identify any single tree species forming major forest stands in Poland, i.e., Scots pine (*P. sylvestris*), Norway spruce (*P. abies*), European silver fir (*A. alba*), and European larch (*L. decidua*), with the high probability c.a. of 99.99 %. The

comparison based on detailed DNA patterns can be used for diagnostics of individual wood samples, taking into account the random probability of identical trees existing in the stand.

Forensic wood material (shafts, logs, sawmill assortment, and stumps) has been proved to be appropriate for the DNA-based identification studies. In order to prohibit the placing on the EU market illegally harvested timber and products derived from such timber, the genetic tool based on DNA (SSR and cytoplasmic markers) helps to identify and match the timber logs in question with the stumps found in the forest with very high probability, practically close to 100 %. It is also possible to match logs with needles (leaves), branches, or roots left on the spot of illegal harvesting. Thanks to this tool the EU traders who place their timber products on the EU market for the first time can check if wood was legally harvested.

The methods based on DNA profiling applied to investigate illegal logging in European forests may contribute to the international actions like Forest Law Enforcement, Governance and Trade (FLEGT) focusing on wood trade between overseas countries and Europe, as well as promoting wood from certified and sustainable managed forests. It is also consistent with the assumptions of the European Parliament Directive on Timber Regulation (EUTR), which came into effect in 2013 to stop the circulation of illegally logged wood in the European Union.

In future, developed genetic techniques can help to identify the wood imported from the disqualified forest regions in the world (e.g., from the forest not meeting the sustainable forest management). Also, they may help to facilitate the traceability of timber products by economic operators in this part of the supply chain referred to as traders in the EU Timber Regulation.

Acknowledgments The research concerning the forensic genetics cited in this chapter was funded by the General Directorate of State Forests (grants BLP-333 and BLP-384). The Polish Forest Service Guard and the Police are acknowledged for wood sample supply. Grateful thanks are expressed to our colleagues from FRI (IBL, Poland), especially Malgorzata Borys, M.Sc., Jolanta Bieniek, and Dr Marcin Klisz for wood processing and the DNA analyses.

References

Acheré V, Faivre Rampant P, Pâques LE, Prat D (2004) Chloroplast and mitochondrial molecular tests identify European x Japanese larch hybrids. Theor Appl Genet 108:1643–1649

Ahuja MR, Neale DB (2005) Evolution of genome size in conifers. Silvae Genet 54:126–137

Asif MJ, Cannon CH (2005) DNA extraction from processed wood: a case study for the identification of an endangered timber species (*Gonystylus bancanus*). Plant Mol Biol Reporter 23:185–192

Buschiazzo E, Ritland CE, Bohlmann J, Ritland K (2012) Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evol Biol 12:8

Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryote genomes. BioEssays 28:1040–1050

Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109:1204–1214

Cremer E, Liepelt S, Sebastiani F, Buonamici A, Michalczyk IM, Ziegenhagen B, Vendramin GG (2006) Identification and characterization of nuclear microsatellite loci in *Abies alba* Mill. Mol Ecol Notes 6:374–376

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. BRL Focus 12:13–15

Dumolin-Lapègue S, Pemonge MH, Gielly L, Taberlet P, Petit RJ (1999) Amplification of oak DNA from ancient and modern wood. Mol Ecol 8:2137–2140

Echt CS, May-Marquardt P (1997) Survey of microsatellite DNA in pine. Genome 40:9–17

Hedrick PW (2000) Genetics of population. Jones and Bartlett Publ Boston

Goodwin W, Linacre A, Hadi S (2011) An introduction to Forensic genetics, 2nd ed. Wille-Blackwell Publ., pp 198

Gutiérez JP, Royo LJ, Álvarez I, Goyache F (2005) MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. J Heredity 96(6):718–721

Isoda K, Watanabe A (2006) Isolation and characterization of microsatellite loci from *Larix kaempferi*. Mol Ecol Notes 6:664–666

Jaramillo-Correa JP, Aguirre-Planter E, Eguiarte LE, Khasa DP, Bousquet J (2013) Evolution of an ancient microsatellite hotspot in the conifer mitochondrial genome and comparison with other plants. J Mol Evol 76:146–157

Karhu A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17:259–265

Marco J, Artajona J, Larrechi MS, Rius FX (1994) Relationship between geographical origin and chemical composition of wood for oak barrels. Am J Enol Viticulture 45:192–200

Nowakowska JA (2011) Application of DNA markers against illegal logging as a new tool for the forest guard service. Folia Forestalia Pol Series A 53(2):142–149

Nowakowska JA, Pasternak T (2014) The wood DNA analysis in criminal proceedings (in Polish). ORWLP Bedoń (Poland) Editor: CILP (Forest state holding), pp 127. ISBN: 978-83-63895-18-1

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-Ch, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson Å, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Gil RG, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterisation of microsatellites in Norway spruce (*Picea abies* K.). Genome 40:411–419

Provan J, Soranzo N, Wilson NJ, McNicol JW, Forrest GI, Cottrell J, Powell W (1998) Gene-pool variation in Caledonian and European Scots pine (*Pinus sylvestris* L.) revealed by chloroplast simple-sequence repeats. Proc R Soc Lond B 265:1697–1705

Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2006) Extraction, amplification and characterization of wood DNA from Dipterocarpaceae. Plant Mol Biol Reporter 24:45–55

Schmidt A, Doudrick RL, Heslop-Harrison JS, Shmidt T (2000) The contribution of short repeats of low sequence complexity to large conifer genomes. Theor Appl Genet 101:7–14

Seton C (1988) Life for sex killer who sent decoy to take genetic test. The Times (London)

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiate*. Genome 37:977–983

Soranzo N, Provan J, Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Mol Ecol 7:1247–1263

Sperisen C, Büchler U, Gugerli F, Mátyás G, Gebourek T, Vendramin GG (2001) Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Mol Ecol 10:257–263

Tibbits JFG, McManus LJ, Spokevicius AV, Bossinger G (2006) A rapid method for tissue collection and high throughput of genomic DNA from mature trees. Plant Mol Biol Reporter 24:1–11

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Mol Ecol 5(4):595–598

Vendramin GG, Ziegenhagen B (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. Genome 40(6):857–864

Yaman B, Akkemik Ü (2009) The use of dendrochronological method in illegal tree cutting in Turkey: a case study. Baltic For 15(1):122–126

Chapter 20
Domestication and Genetics: What a Comparison Between Land and Aquatic Species Can Bring?

Fabrice Teletchea

Abstract Domestication, which is by definition a long and endless process, was one of the most significant cultural and evolutionary transitions of human history. In land, the domestication of animals started about 12,000 years ago and resulted in an apparent dichotomy between domesticated and wild animals. Nevertheless, new findings suggest that long-term gene flow between wild and captive land animal populations was much more common than previously assumed challenging assumptions about genetic bottlenecks during domestication, expectations about monophyletic origins, and interpretations of multiple independent domestication events. Besides, it raises new questions regarding ways in which behavioral and phenotypic domestication traits were maintained, and just what a domestic population was. In contrast to land animals, the onset of the domestication of aquatic species is a recent phenomenon, which started in the 1980s for most species. Hence, today there are still lots of exchanges between wild and captive individuals, and thus, captive fish have only slightly changed from their wild congeners. To better describe the diverse strategies for fish production, a new classification was recently developed comprising five levels of domestication with 1 being the least domesticated to 5 being the most domesticated. The recent domestication of fish species, and the diversity of domestication levels, provides a unique opportunity to better understand how genetic variability evolves during the early phases of fish domestication that could also be useful to discuss both the domestication history of land animals and concepts, such as domestication itself and the differences between wild and domesticated animals.

F. Teletchea (✉)
Research Unit Animal and Functionalities of Animal Products (URAFPA), University of Lorraine—INRA, 2 Avenue de la Forêt de Haye, BP 172, 54505 Vandoeuvre-lès-Nancy, France
e-mail: fabrice.teletchea@univ-lorraine.fr

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0_20

20.1 Introduction

Domestication is traditionally defined as "the process by which a population of animals becomes adapted to man and to the captive environment by some combinations of genetic changes occurring over generations and environmentally induced developmental events occurring during each generation" (Price 1984, 1999). Such a definition implies that domestication is a long and endless process during which captive individuals will become more adapted to humans and captive conditions and consequently progressively modified from their wild counterparts. Thus, domestication should not be conflated with taming because the former leads to permanent genetic modifications of a bred lineage, while the latter is only conditioned behavioral modification of an individual (Driscoll et al. 2009). Nevertheless, even though domestication is probably studied for centuries, the word is still confusing and poorly defined, chiefly because of the inherent difficulty in assigning state terms to a process involving long-term and continuous change (Denis 2004; Dobney and Larson 2006; Driscoll et al. 2009; Vigne 2011; Larson and Burger 2013; Zeder 2015).

During domestication, five main genetic processes are involved in the evolution of captive animals (Mignon-Grasteau et al. 2005; Zeder 2012). These include two uncontrolled processes that are due to the limited size of the population (known as inbreeding) and the random changes in gene frequencies (genetic drift). Then, there are two partially controlled processes resulting from the selection imposed on captive populations that cannot be ascribed to active selection (natural selection in captivity) and the relaxation of natural selection in captivity that can be expected to accompany the transition from wild to captive populations. At last, the fifth genetic process is controlled, known as active selection, because changes are directional (e.g., increase of milk production) (Price 1999; Mignon-Grasteau et al. 2005).

During the course of domestication, captive animals will become a domestic species. Yet, there is still no consensus today on what a domestic animal species is (Dobney and Larson 2006; Zeder 2006; Driscoll et al. 2009; Zeder 2015). Most often, such a species is defined as a group of animals bred in captivity and modified from their wild ancestors to become more useful to humans (Diamond 2002). In addition, other authors considered that a domestic animal must display morphological and physiological variations never observed in the wild and also that some individuals at least would not survive in the wild (Balon 2004). An obvious example of such a domestic species is the dog *Canis lupus familiaris*, which exhibits variation of both phenotypic and behavioral traits that exceed those observed within the whole family and even order (Trut et al. 2009; Galibert et al. 2011). Nevertheless, Dobney and Larson (2006) recently highlighted that the terms 'wild' and 'domestic' are not complementaries such as true/false, but rather they represent the extremes of a process. In others words, no threshold clearly separates wild from captive/domestic animals (Zeder 2006). This chiefly explains why it is so difficult to make a list of all domestic animal species (Denis 2004), particularly for fish (Bilio 2008; Klinger et al. 2013; Teletchea and Fontaine 2014). Interestingly

enough, captive/domestic animals can go back to the wild and reproduce (known as feral animals), which illustrates that for an animal to be domesticated is neither a permanent nor a "non-return" state (Vigne 2011; Lorenzen et al. 2012; Zeder 2012).

The main goals of the present article are to briefly describe and compare the domestication of land and aquatic species in order to better understand how genetic variability evolves during the early phases of domestication and then to discuss the concept of domestication itself and the differences between wild and domesticated animals.

20.2 Domestication of Mammals

20.2.1 Brief History of Domestication on Land

Domestication on land, which probably started around 12,000 years ago, was one of the most significant cultural and evolutionary transitions of human history (Diamond 2002; Mannion 1999; Driscoll et al. 2009; Vigne 2011; Zeder 2015). Domestication was indeed a core component of a major change in the way of life of an increasing number of human societies throughout the world, in a process called Neolithisation (Driscoll et al. 2009; Vigne 2011; Larson and Fuller 2014). This process constitutes also a fundamental change in the evolution of the biosphere, mainly due to the development of agriculture, which is now responsible for the transformation of approximately 40 % of the earth's surface (Diamond 2002; Vigne 2011; Teletchea and Fontaine 2014).

During domestication, the very few species that were eventually kept were profoundly modified by humans, including brain size, fat content, body size and proportions, floppy ears, altered coat color, decreased flight responses, reduced fear of humans, lower motivation for foraging, increased sociality, earlier reproduction, and modification of endocrine and metabolic systems (Price 1999; Mignon-Grasteau et al. 2005; Dobney and Larson 2006; Taberlet et al. 2011; Vigne 2011; Zeder 2012; Larson et al. 2014; Larson and Fuller 2014; Marshall et al. 2014). Among the most modified animals used for human consumption, there are the five major mammal species (cattle, pig, sheep, goat, and horse), which represent nearly 94 % of mammalian livestock today (Diamond 2002). These five major species all possess hundreds of well-defined breeds used for a variety of purposes with differing levels of performance (e.g., dairy cattle versus meat cattle breeds) (Mignon-Grasteau et al. 2005; Groeneveld et al. 2010). For instance, sheep *Ovis aries*, of which *O. musimon*/*O. orientalis*, *O. ammon*, and *O. vignei* are probably the wild ancestors, display the highest number of breeds, comprised between 850 and 1409 (Taberlet et al. 2011; Teletchea and Fontaine 2014). This results that nowadays there is an apparent dichotomy between the very few domesticated (produced in farms) and wild (from hunting) animals used for human consumption; for some species, the wild ancestors is even extinct or nearly extinct (Vigne 2011).

Classically, studies on the domestication process, following Darwin's seminal work, were heavily skewed toward the central roles of human intentionality, directed or controlled breeding of individuals, and genetic isolation of captive herds from wild relatives (Marshall et al. 2014). Inversely, the early steps of domestication were rarely addressed and notably the possible links between the first captive generations of animals and their wild counterparts. However, new ethnoarcheological, archaeological, and genetic findings focusing on the onset of domestication suggest that long-term gene flow between wild and captive land animal populations was much more common than previously assumed since Darwin (Marshall et al. 2014; Larson et al. 2014). Such findings challenge assumptions about genetic bottlenecks during domestication, expectations about monophyletic origins, and interpretations of multiple domestication events (Marshall et al. 2014). Besides, it raises new questions regarding ways in which behavioral and phenotypic domestication traits were maintained, and just what a domestic population was (Marshall et al. 2014). These very recent developments are further described below based on the domestication of cattle *Bos taurus/B. indicus*; for other species such as donkey *Equus asinus*, horse *Equus caballus*, or pig *Sus scrofa*, see among others Groeneveld et al. (2010) and Marshall et al. (2014).

20.2.2 Domestication of Cattle

Cattle are one of the five major mammal species farmed throughout the world. Today, nearly 1.4 billion individuals belonging to more than 800 breeds constitute the global livestock, which are major source of milk, meat, hides, and draft powder (Felius et al. 2011; Teletchea and Fontaine 2014). The wild ancestor of all domesticated cattle is a group of races of the aurochs *Bos primigenius* (Woodford 2000; Taberlet et al. 2011). The aurochs was once throughout Europe and has a range that extended through North Africa and the middle-east to southeast Asia and China (Woodford 2000; Taberlet et al. 2011). It lived in the European forests until its extinction in 1627 in a Polish park and thus could also have been crossed with some domestic cattle during several millennia (Woodford 2000: Groeneveld et al. 2010; Taberlet et al. 2011; Schibler et al. 2015). The presence of two mitochondrial DNA haplogroups was generally interpreted as an indication of two main independent domestication events (probably through a population bottleneck) that dated back ca. 8–11,000 years ago, the one in the Fertile Crescent leading to taurine cattle (*Bos taurus*) and the other in the Indian subcontinent leading to zebu (*Bos indicus*) (Groeneveld et al. 2010; Taberlet et al. 2011; Marshall et al. 2014; Schibler et al. 2015). Even though zebu differs from taurine cattle by the presence of a prominent hump, these two species fully interbreed and there were numerous cases of taurine–zebu admixture notably over Europe, southwest Asia and Africa (Groeneveld et al. 2010; Felius et al. 2011; Taberlet et al. 2011). However, Larson and Burger (2013) recently suggest that only taurine cattle were domesticated in the Neolithic core zone of Anatolia and the Near East, while zebu may not have resulted from

independent domestication, but instead from the introgression of wild zebu populations into taurine cattle that were transported eastward.

During thousands of years, cattle slowly became adapted to local environments and fulfilled the needs of farmers (Taberlet et al. 2011). During this period, gene flow among different phenotypes was possible, leading to relatively high effective population sizes and preventing genetic drift at the regional scale (Taberlet et al. 2011). Besides, it is also likely that other species were crossed with cattle in some parts of the world, including yak (*Bos grunniens*) in Nepal or banteng (*Bos javanicus*) in southeast Asia and Indonesia, which also contribute to maintain or increase genetic variability (Groeneveld et al. 2010). However, this situation changed dramatically about two hundred years ago with the emergence of breed concept (Taberlet et al. 2011). Since that time, stronger selection pressures have been applied to local populations, isolating breeds from each other, which could have resulted in genetic drift and inbreeding and arguably in a decrease of fitness (Groeneveld et al. 2010; Taberlet et al. 2011; Wiener and Wilkinson 2011). Nevertheless, gene flow between neighboring regions did not completely stop as deliberate upgrading was realized in order to improve production characteristics by using bulls of other populations from the same or a different country; such practices are well documented in Europe (Felius et al. 2011). More recently, the number of males involved in reproduction schemes has drastically decreased with the development of artificial insemination, leading to another strong reduction of their effective population size and genetic drift and loss of alleles (De Roos et al. 2008; Taberlet et al. 2011; Wiener and Wilkinson 2011). For instance, the global Holstein cattle have an effective population size of about 50 (Taberlet et al. 2011). This strong decrease of the effective population size might explain the strong reduction in fertility as well as the genetic diseases observed in this breed (Taberlet et al. 2011). One extreme case of low genetic variability is a feral British breed, Chillingham cattle, for which 24 out of 25 microsatellite loci were found homozygous (Wiener and Wilkinson 2011). Inversely, numerous cattle breeds still have substantial nucleotide diversity, indicating a large ancestral effective population size (Wiener and Wilkinson 2011). In the past decades, a few of the most productive breeds were imported throughout the world at the expense of local, seemingly less productive populations (Felius et al. 2011). Consequently, within a few decades, we might lose most of the highly valuable farm animal genetic resources that humanity has gradually selected over the past millennia (Felius et al. 2011; Taberlet et al. 2011).

20.2.3 Conclusions

Summing up, the domestication of land animal species is much more complex than previously thought and probably includes several millennia gene flow within and between wild and domestic populations, and only recently, over the past centuries, intensive breeding practices, particularly in the last 50 years, which lead to modern

breeds (Larson et al. 2014; Marshall et al. 2014). This chiefly explains why it is so difficult to decipher the earliest phases of domestication based on the genetic analysis of modern livestock (Larson and Burger 2013; Larson et al. 2014; Marshall et al. 2014). Moreover, because captive/domestic populations were not as closed (or separated from their wild counterparts or other species) as usually assumed and intentional breeding of females was largely absent during the early stages of domestication, this raises many, still unanswered, questions regarding ways in which behavioral and phenotypic domestication traits were maintained in the long term and just a domestic population was (Larson et al. 2014; Marshall et al. 2014). The history of fish domestication might contribute to provide new answers to these outstanding questions.

20.3 Domestication of Fish

20.3.1 Brief History of Domestication of Fish

In contrast to land animals, the onset of domestication of aquatic species is a recent phenomenon; even though some evidence of fish farming dated back ca. 3500 years ago (Liao and Huang 2000; Balon 2004). Most farming of fish species indeed started in the last century (Teletchea and Fontaine 2014). Consequently, for numerous species, there are still today lots of exchanges between wild and captive individuals, and thus, captive fish have only slightly changed from their wild congeners (Gjedrem et al. 2012; Lind et al. 2012; Teletchea and Fontaine 2014). This also implies that the apparent dichotomy between domesticated and wild land animal species is not reliable for aquatic species. In other words, there is a gradual transition from wild to farmed aquatic animals, sometimes making it difficult to determine when capture (fisheries) ends and aquaculture begins (Bilio 2008; Klinger et al. 2013; Teletchea and Fontaine 2014).

In order to better describe the various fish production strategies, Teletchea and Fontaine (2014) proposed a new classification based on the level of human control over the life cycle of species in captivity and independence from wild inputs. This classification comprises five levels of domestication with 1 being the least domesticated to 5 being the most domesticated (Table 20.1). Among the 250 species recorded in the Food and Agriculture Organization (FAO) database in 2009, 70 % belong to the first three levels of domestication, and for 114 species (levels 1 and 2), major bottlenecks preclude closing their life cycle in captivity and thus represent initial experiments in farming with no foreseeable lasting results (Teletchea and Fontaine 2014). In contrast, only a few species, or more accurately populations, can be considered truly domesticated, similar to cattle or sheep. Such results contradict those of Duarte et al. (2007) who had concluded that the farmed fish species listed in the FAO database were all domesticated. Yet, the trend documented by Duarte et al. (2007) simply reflects the growth of aquaculture globally,

Table 20.1 Domestication levels of fish species used in aquaculture (modified from Teletchea and Fontaine 2014)

Level	Description	# species
5	Selective breeding program is used focusing on specific goals	30
4	Entire life cycle closed in captivity without wild inputs	45
3	Entire life cycle closed in captivity with wild inputs	61
2	Part of the life cycle closed in captivity: several bottlenecks	75
1	First trials of acclimatization to the culture environment	39
0	Capture fisheries	Thousands

Note that reaching a particular level does not necessarily imply that the entire production is based on this level

or what Zeder (2015) called resource management, but not the history of domestication that could be written, perhaps, after another century for each domesticated species (Hedgecock 2012).

Compared to land animals, the literature on fish domestication is relatively scant (e.g., Ruzzante 1994; Balon 2004; Huntingford 2004; Teletchea and Fontaine 2014); more than 800 papers dealing with land domestication were published in 2013 only (Zeder 2015). The number of studies focusing on the evolution of genetic variability during domestication in fish species is even more limited (Vandeputte et al. 2009a, b; Gjedrem et al. 2012; Vandeputte 2012; Lind et al. 2012). Overall, there might be less than 100 studies. For instance, Porta et al. (2007) assessed the genetic structure of four representative broodstocks from southern Spain of Senegalese sole *Solea senegalensis* (level 3 in Teletchea and Fontaine's classification) by means of eight microsatellites loci. They found a drastic reduction of the genetic variability on only one generation in three of the four analyzed stocks due to the incorporation of family-related G1 (first-generation progeny) individuals in their breeder groups, without taking into account for their selection any genetic criteria. Similar losses of genetic variability have also been found in farmed Atlantic halibut *Hippoglossus hippoglossus* (level 4) and turbot *Psetta maxima* (level 5) when comparing the F1 generation to their founder wild populations (Damancher and Garcia-Vasquez 2011). The case of European sea bass *Dicentrarchus labrax* is further described below because it is arguably the fish species for which most of the domestication history and genetic variability are both relatively well known (Vandeputte 2012; Hillen et al. 2014).

20.3.2 Domestication of European Sea Bass

The European sea bass is now second only to sea bream (*Sparus aurata*) in terms of global aquaculture production (with more than 153,000 tonnes in 2012) in the Mediterranean Sea (Hillen et al. 2014). The wild sea bass is found in coastal waters of the northeastern Atlantic Ocean and the Baltic Sea and throughout the

Mediterranean and Black Sea (Hillen et al. 2014). Three main distinct subpopulations have been determined genetically: northeastern Atlantic, eastern and western Mediterranean (Vandeputte 2012; Hillen et al. 2014).

Even though extensive polyculture of this species (based on the farming of wild caught juveniles) has a long history in coastal lagoons around the Mediterranean Sea, the domestication of this species started only during the late 1960s in France and Italy (Chatain and Chavanne 2009; Hillen et al. 2014). Intensive aquaculture started rapidly when the entire life cycle was closed in captivity (level 3 in Teletchea and Fontaine's classification). Further improvements in the control of the life cycle (controlled spawning, larval rearing) lead to the takeoff of the production (Chatain and Chavanne 2009). Nevertheless, most farms still rely today on wild broodstock for reproduction or, to lesser extent, from first-generation (F1) individuals and rarely from selected F2 or F3 fish (Chatain and Chavanne 2009; Vandeputte 2012; Novel et al. 2013; Hillen et al. 2014). Noteworthy, it has recently been demonstrated that some domesticated or selected (for growth for instance) populations of sea bass already display different behavior (swimming activity) and responses to acute stress (Millot et al. 2011; Benhaïm et al. 2013), yet they display similar growth (Vandeputte et al. 2009a, b). Besides, broodstock can be sourced locally or imported from distant population or hatcheries (Hillen et al. 2014). At last, sea bass can escape from sea cages and return to the wild and could interbreed with wild sea bass or compete for food or suitable habitat (Hillen et al. 2014). This example illustrates how complicated is the beginning of domestication of a fish species, with probably different independent domestication events, mixing of wild and captive individuals, and several exchanges between different regions. Such practices would evidently complicate the tasks of genetics that would like to understand the domestication history of this species based on the genetic data of domestic sea bass sampled, for instance, in 2100. However, for other species, such as the Atlantic salmon *Salmo salar*, it is more likely that most of current production is based only on wild salmon domesticated in Norway in the mid-1970s (Gjedrem 2010). For this species, the domestication history would probably be much straightforward to reconstruct based on samples analyzed at the end of this century.

20.3.3 Conclusions

The recent domestication of fish species and the diversity of domestication levels provide a unique opportunity to better understand how genetic variability evolves during the early phases of domestication. Such information could be useful to test alternative hypotheses such as "multiple independent domestication events versus a single domestication event followed by several introgressions" scenarios (Larson and Burger 2013; Larson et al. 2014; Marshall et al. 2014). In the first scenario, the process of domestication is common and emerges easily. In the second, domestication is rare, but gene flow between wild and domestic populations is common. Determining which of these two is more likely has significant ramifications for our

understanding of the frequency and nature of the process itself (Larson and Fuller 2014). Besides, it could allow testing experimentally how a captive population can change and adapt to humans and captive conditions even though there are still some exchanges with wild congeners. At last, it might be interesting to apply the concept of domestication level, which focused only on the original process that led from wild animals to an early domestic population (Larson and Burger 2013) developed for fish species (Table 20.1) to land animals in order to better describe the difference between farmed or captive land animals (including those living in zoos, such as rhinoceroses or elephants; see Pelletier et al. 2009) and also within species over time (e.g., how the domestication level has evolved during time for cattle).

20.4 General Conclusions

Even though same words and concepts are used when describing the domestication of land and aquatic species, they do present strong differences (Table 20.2).

First, the pathways in which wild animals became captive, then domestic, differ. In land, three main pathways have been described: (i) a commensal pathway, in which the animal first moves into an anthropogenic habitat and later develops a two-way partnership with humans (e.g., dog), (ii) a prey pathway, initiated by a human interest in enhancing the yield or predictability of a resource provided by target species (e.g., sheep), and (iii) a directed pathway, in which humans deliberately set out to domesticate a species (e.g., horse) (Zeder 2012, 2015). Species following the first two pathways tend to possess more traits that make them suitable candidates for domestication. Species on directed pathways, in contrast, likely possess barriers to domestication that require more knowledge on the part of humans to overcome (Zeder 2012, 2015). For aquatic species, this is only the third pathway, directed, that is used and may partly explain why the history of aquaculture shows that farming of numerous fish species corresponds to only one or a few years of trials before being abandoned (Jobling 2010; Teletchea and Fontaine 2014).

Table 20.2 Summary of the domestication history of mammal and fish species (modified from Teletchea 2012; [a]based on Zeder 2012, 2015)

	Mammal species	Fish species
1. Pathways to domestication[a]	Commensal pathway	Directed pathway
	Prey pathway	
	Directed pathway	
2. Beginning of domestication	ca. 10,000 years	ca. 50 years
3. Domestication levels	Most at level 5	All levels
4. Number of farmed species	Few	Numerous

See text for details

Second, the duration of the domestication between land and aquatic species diverges strongly (Table 20.2). This implies that mammal species probably remain at the level 3 for several millennia before reaching the levels 4 and 5 in the past centuries (Tables 20.1 and 20.2). Such a long time has allowed crossing captive populations with their wild counterparts and sometimes with other species, some never domesticated (Larson et al. 2014), which perhaps provide large genetic variability on which the numerous breeds observed today could be developed. Perhaps, if captive populations were closed very early in the domestication history of land species, then the possibility to develop breeds would have been smaller? Inversely, for numerous fish species, selective breeding (level 5) starts as soon as the life cycle is closed (levels 3 and 4), mainly because the theory of breeding and the gains it can generate is now well known (Vandeputte 2012). Therefore, in only few years or decades, it is often possible to domesticate and select specific characters (e.g., growth, flesh quality) for fish species, yet this may also lead in the future to some of the problems or diseases that are observed in certain breeds of mammal species due to too strong bottlenecks? Therefore, these domesticated fish species might develop diseases or malformations in the coming decades linked to genetic drift or inbreeding.

Third, all domestic mammal species have reached the highest level of domestication (level 5), with the development of numerous breeds (Table 20.2). Inversely, the intraspecific variability of fish species has rarely been exploited because their life cycle has only recently been closed in captivity. Therefore, it is possible that, at least for some species, numerous well-defined breeds with differing levels of performance could be developed in the next century.

Fourth, the number of farmed species on land is very small compared to those farmed in water. This is mainly due to the fact that for land, we observed a result, while for fish it is the beginning of a process. Thus, this might imply that in the past millennia, other land species might have been tried but were not eventually kept? This could also mean that for fish, perhaps much less fish species will be farmed and domesticated in the coming century.

In conclusion, because there is no scientific reason to consider the domestication of land and aquatic animals differently (Balon 2004; Teletchea and Fontaine 2014), fruitful exchanges between scholars working on the domestication of land and aquatic animals could bring new insights to both the concept of domestication itself, i.e., the initial process of domestication of a discrete population in time and space (Larson and Fuller 2014), and the differences between wild and domesticated animals. On a more applied view, this could also help to better domesticate both mammal and fish species in the future because as defined in the introduction domestication is a long and endless process, thus farmed animals are still evolving today, particularly in response to changes in technology and husbandry practices, which themselves are evolving and constantly improving.

References

Balon EK (2004) About the oldest domesticates among fishes. J Fish Biol 65:1–27
Benhaïm D, Bégout M-L, Lucas G et al (2013) First insight into exploration and cognition in wild caught and domesticated sea bass (*Dicentrarchus labrax*) in a Maze. Plos One 8:e65872
Bilio M (2008) Plenary lecture: the future of capture and culture fisheries. In: Arai K (ed) Proceedings of the 7th international symposium, keeping and creating diversity in fish production—innovative marine life science for three Es, Edibles Environment and Education, in 21st century, Hokkaido, pp 5–24, 17–19 Nov 2008
Chatain B, Chavanne H (2009) La génétique du bar (*Dicentrarchus labrax* L.). Cah Agric 18: 249–255
Damancher D, Garcia-Vasquez E (2011) Genetic population structure in flatfishes and potential impact of aquaculture and stock enhancement on wild populations in Europe. Rev Fish Biol Fisher 21:441–462
Denis B (2004) La domestication: un concept devenu pluriel. INRA Prod Anim 17:161–166
De Roos APW, Hayes BJ, Spelman RJ et al (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus Cattle. Genetics 179:1503–1512
Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700–707
Dobney K, Larson G (2006) Genetics and animal domestication: new windows on an elusive process. J Zool 269:261–271
Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. P Natl Acad Sci USA 106:9971–9978
Duarte CM, Marbá N, Holmer M (2007) Rapid domestication of marine species. Nature 316: 382–383
Felius M, Koolmees PA, Theunissen B, European Cattle Genetic Diversity Consortium, Lenstra JA (2011) On the breeds of cattle—historic and current classifications. Diversity 3:660–692
Galibert F, Quignon P, Hitte C et al (2011) Toward understanding dog evolutionary and domestication history. CR Biol 334:190–196
Gjedrem T (2010) The first family-based breeding program in aquaculture. Rev Aquacult 2:2–15
Gjedrem T, Robinson N, Rye M (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture 350–353:117–129
Groeneveld LF, Lenstra JA, Eding H, the GLOBALDIV Consortium et al (2010) Genetic diversity in farm animals—a review. Anim Genet 41:6–31
Hedgecock D (2012) Aquaculture, the next wave of domestication. In: Gepts P, Famula TR, Bettinger RL, Brush SB, Damania AB, McGuire PE, Qualset CO (eds) Biodiversity in agriculture, domestication, evolution, and sustainability. Cambridge University Press, Cambridge, pp 538–548
Hillen J, Coscia I, Volckaert F (2014) Aquatrace species leaflet. European sea bass (*Dicentrarchus labrax*). Available at https://aquatrace.eu/documents/80305/c0a6cb45-4d4e-448b-ab6a-ea9e85836a94. Accessed 3 Mar 2015
Huntingford FA (2004) Implications of domestication and rearing conditions for the behavior of cultivated fishes. J Fish Biol 65:122–142
Jobling M (2010) Chapter 4. Farmed fish and their characteristics. In: Le François N, Jobling M, Carter C, Blier P (eds) Finfish aquaculture diversification. CABI, Oxfordshire, pp 88–99
Klinger DH, Turnipseed M, Anderson JL et al (2013) Moving beyond the fished and farmed dichotomy. Mar Policy 38:369–374
Larson G, Burger J (2013) A population genetics view of animal domestication. Trends in Genet 29:197–205
Larson G, Fuller DQ (2014) The evolution of animal domestication. Annu Rev Ecol Syst 66: 115–136

Larson G, Piperno DR, Allaby RG et al (2014) Current perspectives and the future of domestication studies. P Natl Acad Sci USA 111:6139–6146
Lind CE, Ponzoni RW, Nguyen NH et al (2012) Selective breeding in fish and conservation of genetic resources for aquaculture. Reprod Domest Anim 47:255–263
Lorenzen K, Beveridge MCM, Mangel M (2012) Cultured fish: integrative biology and management of domestication and interactions with wild fish. Biol Rev 87:639–660
Liao IC, Huang YS (2000) Methodological approach used for the domestication of potential candidates for aquaculture. Cahiers Options Méditerranéennes 47:97–107
Mannion AM (1999) Domestication and the origins of agriculture: an appraisal. Prog Phys Geog 23:37–56
Marshall FB, Dobney K, Denham T et al (2014) Evaluating the roles of directed breeding and gene flow in animal domestication. P Natl Acad Sci USA 111:6153–6158
Mignon-Grasteau S, Boissy A, Bouix J et al (2005) Genetics of adaptation and domestication in livestock. Livest Sci 93:3–14
Millot S, Péan S, Chatain B et al (2011) Self-feeding behavior changes induced by a first and a second generation of domestication or selection for growth in the European sea bass, *Dicentrarchus labrax*. Aquat Living Resour 24:53–61
Novel P, Porta J, Fernández J et al (2013) Critical points for the maintenance of genetic variability over a production cycle in the European sea bass *Dicentrarchus labrax*. Aquaculture 416–417:8–14
Pelletier F, Réale D, Watters J et al (2009) Value of captive populations for quantitative genetics research. Trends Ecol Evol 24:263–270
Porta J, Maria Porta J, Cañavate P et al (2007) Substantial loss of genetic variation in a single generation of Senegalese sole (*Solea senegalensis*) culture: implications in the domestication process. J Fish Biol 71:223–234
Price EO (1984) Behavioral aspects of animal domestication. Q Rev Biol 59:1–32
Price EO (1999) Behavioral development in animals undergoing domestication. Appl Anim Behav Sci 65:245–271
Ruzzante DE (1994) Domestication effects on aggressive and schooling behavior in fish. Aquaculture 120:1–24
Schibler J, Elsner J, Schlumbaum A (2015) Incorporation of aurochs into a cattle herd in Neolithic Europe: single event or breeding? Sci Rep UK 4:5798. doi:10.1038/srep05798
Taberlet P, Coissac E, Pansu J et al (2011) Conservation genetics of cattle, sheep and goats. CR Biol 334:247–254
Teletchea F (2012) Qu'est-ce qu'un poisson domestique? Implications pour le développement futur de l'aquaculture. Ethnozootechnie 90:7–12
Teletchea F, Fontaine P (2014) Levels of domestication in fish: implications for the sustainable future of aquaculture. Fish Fish 15:181–195
Trut L, Oskina I, Kharlamova A (2009) Animal evolution during domestication: the domesticated fox as model. BioEssays 31:349–360
Vandeputte M (2012) Genetic variation of growth and sex ratio in the European sea bass (*Dicentrarchus labrax* L.) as revealed by molecular pedigrees. Anim Genet. AgroParisTech. English. <NNT : 2012AGPT0057>. <pastel-00957623>
Vandeputte M, Baroillier J-F, Haffray P et al (2009a) Amélioration génétique des poissons: quelles réalisations et quels défis pour demain ? Cah Agric 18:262–269
Vandeputte M, Dupont-Nivet M, Haffray P et al (2009b) Response to domestication and selection for growth in the European sea bass (*Dicentrachus labrax*) in separate and mixed tanks. Aquaculture 286:20–27
Vigne J-D (2011) The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. CR Biol 334:171–181
Wiener P, Wilkison S (2011) Deciphering the genetic basis of animal domestication. P Roy Soc B-Biol Sci 278:3161–3170

Woodford MH (2000) World watch list for domestic animal diversity, 3rd edns, Part 3. Wild relatives of domestic livestock and some suggestions for new domesticants, pp 647–716. FAO, Rome

Zeder MA (2006) Central questions in the domestication of plants and animals. Evol Anthropol 15:105–117

Zeder MA (2012) Pathways to animal domestication. In: Gepts P, Famula TR, Bettinger RL, Brush SB, Damania AB, McGuire PE, Qualset CO (eds) Biodiversity in agriculture, domestication, evolution, and sustainability. Cambridge University Press, Cambridge, pp 227–259

Zeder MA (2015) Core questions in domestication research. P Natl Acad Sci USA. doi:10.1073/pnas.1501711112

Index

P. Pontarotti (ed.), *Evolutionary Biology: Biodiversification from Genotype to Phenotype*, DOI 10.1007/978-3-319-19932-0