Biological Journal of the Linnean Society, 2018, **XX**, 1–23. With 4 figures.

What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species

C. RYAN CAMPBELL, J. W. POELSTRA and ANNE D. YODER*

Department of Biology, Duke University, Durham, NC 27708, USA

Received 8 January 2018; revised 24 April 2018; accepted for publication 25 April 2018

As is true of virtually every realm of the biological sciences, our understanding of speciation is increasingly informed by the genomic revolution of the past decade. Investigators can ask detailed questions relating to both the extrinsic (e.g. inter- and intra-population and ecological interactions) and intrinsic (e.g. genome content and architecture) forces that drive speciation. Technologies ranging from restriction-site associated DNA sequencing (RADseq), to whole genome sequencing and assembly, to transcriptomics, to CRISPR are revolutionizing the means by which investigators can both frame and test hypotheses of lineage diversification. Our review aims to examine both extrinsic and intrinsic aspects of speciation. Genome-scale data have already served to fundamentally clarify the role of gene flow during (and after) speciation, although we predict that the differential propensity for speciation among phylogenetic lineages will be one of the most exciting frontiers for future genomic investigation. We propose that a unified theory of speciation will take into account the idiosyncratic features of genomic architecture examined in the light of each organism's biology and ecology drawn from across the full breadth of the Tree of Life.

ADDITIONAL KEYWORDS: barrier loci – coalescence – ecological speciation – genome scans – genomic islands – lineage diversification – reproductive isolation – sympatric speciation.

INTRODUCTION

It is an exciting time to be an empiricist engaged in the genetics and genomics of speciation. Combined with the enduring power of field and laboratory studies, genomic analysis is allowing investigators to rigorously test long-standing questions regarding the sources of and selective pressures underlying reproductive barriers, the genomic architecture associated with speciation, and the roles of ecology, geography and demography in speciation across the Tree of Life. The process of lineage diversification and the mechanisms that promote it have been of fundamental interest from the very outset of the formalized theory of evolution by natural selection (Darwin, 1858, 1859). In Darwin's view, natural selection was the driving force of speciation, intrinsically augmented by ecological conditions. The introduction of Mayr's (1942) Biological Species Concept, however, laid bare the apparent difficulties of establishing reproductive isolation (RI) without prolonged geographical separation. Although these two views of speciation, sympatric versus allopatric, were initially considered to be fundamentally opposed, it is now appreciated that they are actually endpoints on a continuum. Foundational work by Guy Bush and colleagues (Bush, 1994, 1998), together with genetic and genomic approaches, has clarified that gene flow among diverging species is often a facet of speciation. We are now in the position to consider the relative influence of geography, ecology and selection in driving the speciation process. Moreover, genome-scale data have pointed to the role of genomic architecture in predisposing certain lineages towards divergence, and others towards stasis.

Thus, we have reached a point at which forces that are both extrinsic and intrinsic to the organism are equally tractable for investigation. Even so, the frontier is vast and the unknown significantly outweighs the known. The genetic and genomic data that have thus far been generated are phylogenetically restricted, and have a strong bias towards a limited number of model systems which accordingly imposes a biased organismal perspective (for an insightful review, see Scordato *et al.*, 2014). Although constraining at present, this bias should not

*Corresponding author. E-mail: anne.yoder@duke.edu

be surprising given that model organisms tend to be those best characterized genomically, thus conferring benefits to the study of closely related lineages with decreasing benefits as phylogenetic distance increases. As we discuss below, however, taxonomic bias in available genomic resources is rapidly giving way to a broader phylogenetic perspective as more genomes are being sequenced (Fig. 1A) at a higher standard of quality (Fig. 1B) both qualitatively and quantitatively (Fig. 1C) with important features such as detailed annotation (Fig. 1D). Thanks to remarkable advances in sequencing technologies and *de novo* genome assembly (Alkhateeb & Rueda, 2017; Jackman *et al.*, 2017; Kamath *et al.*, 2017; Paten *et al.*, 2017; Vaser *et al.*, 2017; Worley, 2017), it is no longer the case that the availability of a closely related model species and reference genome are essential to the generation of genome-scale data and analysis of non-model organisms (Box 1). Moreover, genome-scale data have pointed to the role of genomic architecture in predisposing certain lineages towards divergence, and others towards stasis.

It is our aim in this review to examine the history, recent developments and future directions of the field now generally referred to as 'speciation genomics'. Given the enormity of the field, it is not our intent (nor a realistic goal) to provide an exhaustive overview of the relevant literature. Rather, our primary goal is to illustrate the many ways that technological advances for characterizing the genome are serving to enhance understanding of the interacting extrinsic and intrinsic forces that drive speciation. Scordato *et al.* (2014) described 'internal interactions', wherein natural and sexual selection jointly influence divergence in sexual traits and preferences, are considerably more common than cases wherein 'external interactions' are driven by ecological context and transmission efficiency of sexual trait signals. Here, we define *extrinsic features* as those wherein the environment (described as any feature external to the individual organism, including conspecifics) impacts the action of the genome during speciation, and *intrinsic features* as those that are specific to an organism's internal features, most notably, the structure of its genome. This makes for

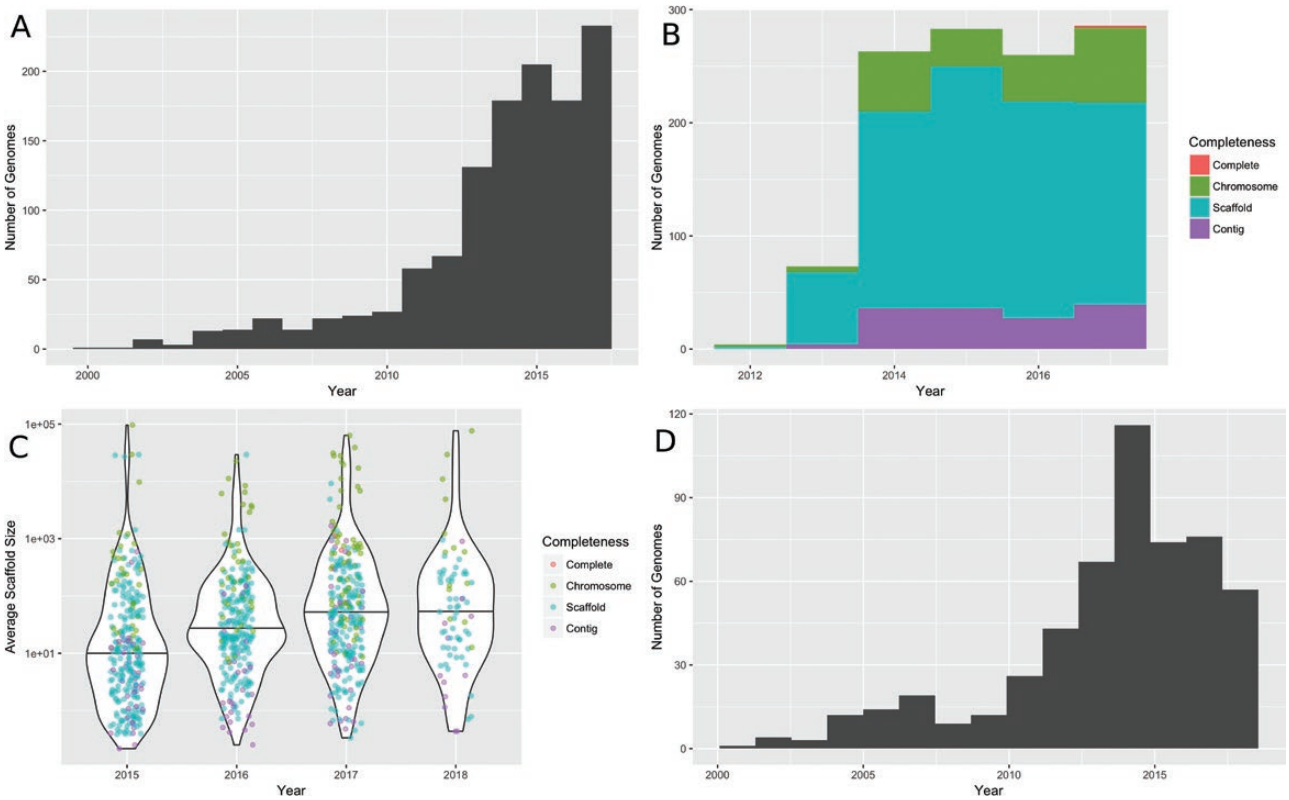

Figure 1. Genome completeness: number and quality of plant and vertebrate genomes uploaded to the National Center for Biotechnology Information (NCBI) over time. (A) Overall number of genomes uploaded per year since 2000. (B) Genomes modified since 2012, displayed by NCBI's assessment of completeness. (C) Violin plots of average scaffold size (genome size/number of scaffolds) by year of genomes modified since 2015; horizontal bar marks the median. (D) Number of genomes that are currently annotated, by original release date.

BOX 1: (TODAY'S) STATE-OF-THE ART GENOMIC APPROACHES

Since high-throughput (also called next-generation or second-generation) sequencing technologies opened up the possibility of genomic characterization any organism in 2005, reference genomes have been assembled for hundreds of non-model organisms (Fig. 1), and with ever-decreasing sequencing costs, whole-genome resequencing projects using population samples have now become commonplace. Besides simply determining the DNA sequence, second-generation sequencing technology has also been widely adopted to identify DNA–protein interactions (Chip-seq) and methylation patterns (BS-seq), and to quantify gene expression (RNA-seq). The last, in particular, is a powerful tool for speciation researchers, because transcriptomic data improve genome annotation (Trapnell *et al.*, 2010; Li *et al.*, 2011) and can be an alternative or complement to genome scans for identifying barrier loci (Wang, Gerstein & Snyder, 2009; Jeukens *et al.*, 2010; Poelstra *et al.*, 2014; Ritchie *et al.*, 2015; Rafati *et al.*, 2018), which may also themselves represent regulatory divergence (Mack & Nachman, 2017).

Short read lengths are a key shortcoming of second-generation sequencing, which has made it difficult to assemble repetitive regions, characterize structural variants and directly observe haplotypes. Third-generation technologies are serving to overcome this limitation by directly sequencing long reads [5–15 kbp both for Pacific Biosciences Single-molecule Real Time (SMRT) sequencing and Oxford Nanopore Technologies Nanopore Sequencing], or by using novel mapping technologies such as the optical mapping system of BioNano Genomics (Servin *et al.*, 2013), the Hi-C approach by Dovetail Genomics (Lieberman-Aiden *et al.*, 2009), and the linked read approach by 10X Genomics (Greer *et al.*, 2017; Yeo *et al.*, 2018). For an overview of third-generation approaches (see Lee *et al.*, 2016) and for sequencing assembly advances, see (Phillippy, 2017).

Access to larger and better characterized regions of the genome and segregating variation therein will be beneficial for all genomics projects, yet the relevance of the advances that third-generation technologies offer to speciation genomics is still to be fully realized. With longer contigs and high-quality assemblies, genomic subtleties with potentially profound impacts on speciation are likely to be revealed. Structural variation such as duplications and inversions

BOX 1: *Continued*

may disproportionately affect speciation, whereas haplotypic information will aid in the inference of gene flow and selection to reconstruct speciation histories and identify barrier loci. In addition, improved assembly of highly repetitive, heterochromatic regions such as centromeres (e.g. Ichikawa *et al.*, 2017; Larsen *et al.*, 2017) may be important because a significant number of hitherto identified hybrid incompatibility genes encode proteins that interact with heterochromatin (Ting *et al.*, 1998; Brideau *et al.*, 2006; Bayes & Malik, 2009; Thomae *et al.*, 2013), probably due to the high concentration of selfish elements there (Castillo & Barbash, 2017). Repeats themselves have also been identified as the focal incompatibility locus (Ferree & Barbash, 2009). Furthermore, centromeres have also been linked to speciation outside of the context of postzygotic incompatibilities, due to their tendency to have particularly low recombination rates (Stump *et al.*, 2005; Carneiro *et al.*, 2009; Noor & Bennett, 2009).

a convenient, if not entirely inclusive, framework for examining speciation via genomic and genetic analysis. By focusing on the interaction of individuals within populations, and the impacts of environment on these interactions, we can explicitly examine the extrinsic 'demography' of speciation, whereas by focusing on features of genome structure and content, we can examine the internal 'architecture' of speciation.

The present deluge of data is progressively placing within reach new, testable hypotheses and improving our understanding of the underlying speciation process. It is the intersection of theory, empiricism and technology that promises to yield remarkable insights into the most fundamental of all evolutionary processes: the genetic and genomic underpinnings of the diversification of organismal linages through time and space.

THE DEMOGRAPHY OF SPECIATION

Speciation and gene flow

Over the past two decades we have moved from a largely geographical allopatric view of speciation to a far more nuanced and complex understanding that harkens back to Darwin and draws meaningfully from the theories of both natural and sexual selection. Whereas allopatric speciation requires only geographical isolation plus time to produce species-level lineage divergence, sympatric speciation is thought

to mostly rely on ecologically mediated natural selection and prezygotic isolation 'acting differently in different places' (Turelli, Barton & Coyne, 2001: 332). Specifically a large shift has occurred in the appreciation of the occurrence and role of gene flow during as well as after speciation. First, it has become clear that gene flow can be overcome, even in the initial stages of divergence, in particular when ecologically based divergent selection is strong. As stated by Nosil (2008): '[s]peciation with gene flow could be common'. Second, it has become clear that introgressive hybridization between substantially diverged populations is commonplace (Sankararaman *et al.*, 2014; Coyner, Murphy & Matocq, 2015; Morii *et al.*, 2015; Árnason *et al.*, 2018; Schumer *et al.*, 2018), and that despite cases of lineage merging or speciation reversal (Campagna *et al.*, 2014; Kearns *et al.*, 2018), such hybridization events often contribute to adaptation (Pardo-Diaz *et al.*, 2012; Racimo *et al.*, 2015; Richards & Martin, 2017) and can contribute to the formation of new lineages (Seehausen, 2004; Abbott *et al.*, 2013; Lamichhaney *et al.*, 2017).

With the advantage of genomic characterization via high-throughput sequencing combined with recent developments in statistical methods, investigators can now, for nearly any species of interest, estimate parameters to describe the demographic aspects of speciation history with unprecedented resolution (Ellegren *et al.*, 2012; Ellegren, 2014; Fan & Meyer, 2014; Gaither *et al.*, 2015; Malinsky *et al.*, 2015; Gante *et al.*, 2016; Kang *et al.*, 2016; Schmitz *et al.*, 2016; Toews *et al.*, 2016; Berner & Roesti, 2017). These aspects are collectively often referred to as the Isolation-with-Migration (IM) model and provide critical information on the rates, direction and possibly timing of gene flow, divergence times (which should be co-estimated with gene flow) and population size trajectories. While challenges remain, as we discuss below, these parameters can together be used to answer key questions pertaining to speciation events, such as: did populations diverge in isolation, in the face of continuous gene flow, or has secondary contact and gene flow been recent? Was speciation associated with a severe population bottleneck (e.g. peripatric speciation)? How rapidly did observed phenotypic divergence or genetic incompatibilities evolve? The answers to these questions are fundamental to understanding the complex interplay among ecology, geography and natural selection in driving lineage diversification.

Characterizing the demography of speciation

A major advance in population and species-level inferences has been facilitated by the use of single nucleotide polymorphisms (SNPs) drawn from across the genome. Reduced representation libraries such as RADseq can be produced at low cost, do not require a reference genome and are often sufficient when questions are focused on estimating genome-averaged historical demography. In fact, because RADseq can generate sequencing data for tens of individuals at a fraction of the cost a single whole genome, the reduced constraint on the number of individuals that can be sequenced is highly beneficial for certain approaches, such as those that rely on the site frequency spectrum. For other applications that require large numbers of individuals, pooled sequencing (e.g. Pool-seq; Schlötterer *et al.*, 2014) and low-coverage sequencing (Nielsen *et al.*, 2011) also offer low-cost whole-genome perspectives, despite sacrificing individual-level genotypes (for an overview see Fuentes-Pardo & Ruzzante, 2017).

Although the advantages of anonymous genome-wide SNPs are many, the need for assembled and annotated genome-scale data persists when the questions being asked require either functional or structural information. In turn, this necessitates increasing levels of theoretical sophistication. Put succinctly by Sousa & Hey (2013: 404), as we accumulate more and more comparative genomic data 'we find our best models and tools for explaining patterns of variation were designed for a simpler time and smaller data sets'. Tremendous progress has nevertheless been made in the development of approaches to estimate demography and gene flow. Three areas in particular are worth pointing out:

- First, several methods can now be used to estimate parameters of IM models from the site frequency spectrum (SFS), that is, the distribution of allele frequencies aggregated across the available sequencing data specific to populations (Gutenkunst *et al.*, 2009; Excoffier *et al.*, 2013; Lohse *et al.*, 2016; Kern & Hey, 2017). These methods are fast and can be easily applied to any genomic dataset. Yet, SFS approaches work best for relatively large sample sizes (i.e. many individuals) due to the highly condensed summarization of the data, and thus concerns exist that they may not always be able to distinguish between competing models (Terhorst & Song, 2015; Lapierre, Lambert & Achaz, 2017).
- Second, coalescent theory, first proposed by Kingman (1982), is now one of the most widely used population genetic models and forms the backbone for many current demographic inference methods. One of the central realizations that has come from coalescent theory is that a species tree will typically contain a distribution of varying gene genealogies. Given that the addition of further unlinked loci never fails to add information with respect to the underlying population history, this has further clarified the applications of coalescent theory for

interpreting the demographic signals contained within large-scale genomic data. Even though coalescent approaches are still unable to use all the genealogical information contained within a genome given the difficulties of fully incorporating recombination, an approximation of the coalescent with recombination, the Sequentially Markov Coalescent (Wiuf & Hein, 1999; McVean & Cardin, 2005), has formed the basis for recent methods that estimate populations size changes through time from high-coverage whole-genome sequences (Li & Durbin, 2009; Schiffels & Durbin, 2014). Moreover, this approach has been used to estimate the age of genomic admixture blocks (Rasmussen *et al.*, 2014), therefore adding tremendous sophistication and precision to temporal estimates of lineage diversification.

- Third, a class of formal tests for admixture that was first developed to test for Neanderthal ancestry in modern humans (Patterson *et al.*, 2012) can be easily computed from any genome sequencing data. This test has been widely adopted and extended, and provides a simple and standardized way to test for admixture between sets of two (f_2 statistics), three (f_3), four (f_4, f_d, D) and five (f_4-ratio, D_{FOIL}) populations or species (Patterson *et al.*, 2012; Martin, Davey & Jiggins, 2015; Pease & Hahn, 2015).

Together, these methods, as applied to high-throughput sequencing data sets, have provided abundant evidence for the pervasiveness of gene flow during all stages of speciation, while also elucidating the demographic speciation histories for numerous study systems (Pinho & Hey, 2010; Feder, Egan & Nosil, 2012; Sousa & Hey, 2013). Despite this rapid progress towards explicitly integrating demographic parameters into our understanding of speciation dynamics, major challenges remain. For example, we do not generally have the resolution to create a hypothesis-free description of the rate, direction and magnitude of gene flow through time, instead often relying on summary measures or being forced to choose from a limited set of hypothesized models. One of the great remaining challenges is to distinguish ancestral population structure from ongoing gene flow. This requires that we disentangle gene flow and divergence time for very recently diverged populations, and also, that we must co-estimate demography and selection. Fundamentally, given the high variance inherent in the coalescence process, even making optimal use of the information contained in high-quality genome assemblies may not provide the resolution necessary to satisfactorily address all the challenges described above. The development of ever-more sophisticated models of the coalescent process thus remains one of the main challenges in the field of speciation genomics.

THE ROLE OF INTRINSIC GENOMIC FEATURES IN SPECIATION

Genomic architecture and speciation predisposition

While extrinsic factors such as available ecological opportunity and within- and between-population dynamics probably explain much variation in diversification rates, intrinsic factors such as underlying features of lineage-specific genome structure, require exploration if we are to understand the phylogenetic propensities for rapid speciation. That is, how often do certain features of genomic 'architecture', such as a genome ploidy, rates and patterns of recombination, and inversion frequency facilitate speciation above and beyond the extrinsic organismal effects of divergence?

The large and well-established effects of sex chromosomes in systems with a heterogametic and homogametic sex are one testament to the role that genome architecture can have on speciation. More generally, because speciation often relies on gene–gene interactions, such as very explicitly in Dobzhansky–Muller (D-M) genetic incompatibilities (Box 3), mechanisms of epistasis and rates of recombination probably impact the probability of speciation. A case for lineage-specific genomic features to promote speciation has for instance been made for ray-finned fishes (e.g. Taylor *et al.*, 2003; Venkatesh, 2003; Rennison, Owens & Taylor, 2012; Cortesi *et al.*, 2015). Volff (2005) proposed the importance of an early stage of whole genome tetraploidization and subsequent rediploidization, along with the 'amazing diversity' of sex determination systems and plasticity of sex chromosomes. After sequencing several lineages of ray-finned fishes, Brawand *et al.* (2014) found evidence of accelerated evolution among regulatory regions, microRNAs (miRNA) and transposable element (TE) insertions. Even so, the exact *mechanistic* impacts of these genomic features on speciation are currently unknown. In finding a path forward, we can look to these processes in model organisms, such as work in yeast linking chromosomal architecture and species formation (Leducq *et al.*, 2016), and next examine whether similar mechanisms may be at play in various rapidly radiating lineages. In this light, genomic characteristics such as an excess of gene duplications, rapid mutation rates, novel miRNAs, high numbers of TEs, and genome-wide diversifying selection on coding and regulatory elements have each been proposed to play differential roles in setting the genomic stage for rapid evolutionary transitions (Brawand *et al.*, 2014).

Genome and gene duplications

As a specific example, duplications at the genome (Fig. 2A) and gene (Fig. 2B) level have frequently been

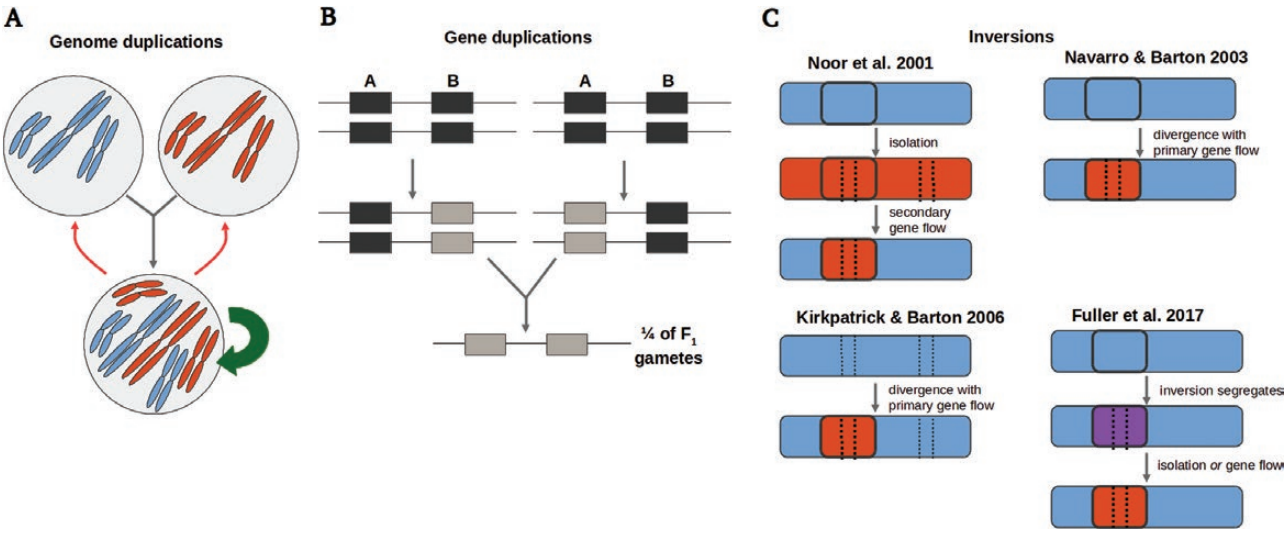

Figure 2. Structural genomic impacts on speciation. (A) Genome duplications: shown is the process of allopolyplodization, where hybridization between two species with 2*N* chromosomes (blue versus red chromosomes) produces individuals with 4*N* chromosomes that are not interfertile with either parent species. (B) Gene duplications: when an ancestral pair of duplicated genes (copies marked A and B) is differentially resolved in two isolated populations, a different copy retains functionality (dark boxes: functional copies, light boxes: non-functional copies) in each population. Upon hybridization between the two lineages, one-quarter of the F1 gametes and one-sixteenth of the F2 zygotes will not carry any functional copy. (C) Inversions can contribute to speciation in several ways due to local reduction of recombination. Inversions are depicted as boxes with thick dark lines within the larger boxes, which represent a stretch of a chromosome. Pairs of dotted lines are interacting genes: co-adapted gene complexes or genetic incompatibilities. Blue represents low genetic divergence between populations, orange represents high divergence and purple (as in the model of Fuller *et al.* 2017) represents high divergence segregating within populations. In contrast to the other models, higher divergence within inverted regions compared to collinear regions is not a consequence of (differential) gene flow.

implicated as both drivers and maintainers of speciation (e.g. Lynch, Force & Travis, 2000; Otto & Whitton, 2000; Taylor, Van de Peer & Meyer, 2001; Soltis, Soltis & Tate, 2004; Roth *et al.*, 2007; Evans, 2008). A clear case of clade-specific genomic architecture associated with speciation can be seen in the differential rates of polyploid speciation in plants versus animals. While this is rare in animals, the most recent study on this subject estimated that as many as 15% of speciation events in angiosperms (and double as many in ferns) are accompanied by ploidy increase (Wood *et al.*, 2009). This difference in propensity for polyploidization (and associated speciation events) does not appear to be due to differences in the initial polyploidization step, and instead is more likely to be related to limitations to regain a balanced genome (Wertheim, Beukeboom & Zande, 2013). For instance, difficulties may arise in many animals after polyploidization due to the nature of their genetic sex determination systems as well as the disruption of dosage compensation for differentiated sex chromosomes (Orr, 1990; Otto & Whitton, 2000; Wertheim *et al.*, 2013).

At a finer scale, gene duplications may also promote speciation via the resolution of duplicates. The 'differential resolution' of gene copies after a gene duplication event (i.e. a different gene copy degenerates in each of two diverging populations) may represent a powerful, general mechanism underlying hybrid dysfunction (Lynch *et al.*, 2000). Indeed, this process has been demonstrated (Bikard *et al.*, 2009; Mizuta, Harushima & Kurata, 2010) to cause hybrid incompatibilities (Fig. 2B). More generally, rapid evolution of gene duplicates, for instance due to a reduction of purifying selection for one of the gene copies, may render them probable candidates for barrier loci. For example, it was recently discovered that a duplication in a crucial photosynthesis gene is at the root of hybrid lethality between two sympatric species of *Mimulus* (Brandvain & Matute, 2018; Zuellig & Sweigart, 2018). These findings utilized gene mapping and gene expression experiments which required the genomic characterization of both species. In practical terms, this work offers a prime example of the premise that sequencing the genome of a given species, or that of a phylogenetically related lineage (Gnerre *et al.*, 2011), is fundamental to understanding the biological mechanisms that underlie the genomic loci that differentiate species, which in turn may provide mechanistic insight into the speciation process as a whole.

Several studies have pointed to the role of copy number variation (CNV) in the process of speciation. An interrogation of the pig (*Sus scrofa*) genome, along with several related species, showed that CNVs are evolving faster than SNPs and often contain olfactory receptor (OR) genes which could be vital to mate recognition (Paudel *et al.*, 2015). In a similar example, genes that determine butterfly chemosenses (ionotropic receptors, IRs) were identified as divergent between species pairs (van Schooten *et al.*, 2016). The effects of CNVs on speciation can also be indirect. Recently a study of several genera found that CNVs on sex chromosomes were responsible for rapid changes to the sex ratio (O'Neill & O'Neill, 2018). These changes happen very quickly, making them responsible for the development of hybrid incompatibilities if two populations are in allopatry, ultimately leading to speciation. All of this work relies on the accurate assembly of genomes and high-quality sequencing to measure genome-scale changes between a small sample of individuals, work that has not been possible, or more crucially scalable, until very recently.

We are, however, still far from being able to explain differences in polyploidization propensities in any detail (Soltis *et al.*, 2010). In particular, with short-read next-generation sequencing technologies, it has been difficult to accurately assemble duplicated regions of the genome (Ellegren, 2014). These technical barriers are steadily falling away, however, as sequencing technologies continue to become more efficient, accurate and affordable. By improving genome characterization, improved technology has conferred new power to investigate gene duplications as indicators of speciation, even in species for which assembled genomes are as yet unavailable. Along with the increasing number of organisms with sequenced whole genomes, the recent improvement in long-read sequencing technologies and single cell sequencing is allowing for the identification of duplicated genes in non-model genomes (Larsen, Heilman & Yoder, 2014). Longer reads, such as those produced by single-molecule sequencing, are able to differentiate gene copies by their surrounding genomic sequence, and consequently, make genome characterization for even difficult regions of the genome possible without a reference sequence (Jiao & Schneeberger, 2017). Furthermore, the recent adoption of techniques such as optical mapping, Hi-C and linked reads (see Box 1) now makes it possible to accurately assemble repetitive regions across hundreds of kilobases. These technical advances are thus rapidly shifting the field of speciation genomics towards greater methodological and theoretical sophistication.

Chromosomal inversions

The potential of structural genomic features to promote speciation (Noor *et al.*, 2001a) was perhaps first appreciated with the observation that chromosomal inversions may play a special role in facilitating hybrid sterility, and thus incipient RI (Fig. 2C). Chromosomal inversions can promote RI by two fundamental means. The first is structural: heterozygous inversions may (partially) prevent proper chromosome pairing during meiosis, such that hybrids between populations fixed for alternative orientations suffer from reduced fertility (White, 1978). This hypothesis was tested in *Drosophila* as early as 1933 in Dobzhansky's (1933) classic work, but runs into difficulties explaining why such an inversion would spread in the first place (reviewed by Hoffmann & Rieseberg, 2008). The second means by which inversions may promote speciation is by suppressing recombination. In the face of gene flow, this prevents the uncoupling of allelic combinations present in the inversion, and these combinations may include, for instance, co-adapted gene complexes, male trait and female preference combination, and D-M incompatibilities. Given that recombination is the central challenge for modellng speciation-with-gene-flow (Felsenstein, 1981), extreme recombination suppression such as in inversions may be expected to facilitate this potentially common mode of speciation.

Noor *et al.* (2001a) demonstrated that inversions create linkage groups among genes that cause sterility among a pair of *Drosophila* species (*D. persimilis* and *D. pseudoobscura*), prompting the development of a model wherein hybrid incompatibility genes (Noor *et al.*, 2001b; Rieseberg, 2001) accumulate indiscriminately throughout the genome during allopatric divergence, but are retained only in inversions when gene flow is resumed during secondary contact (Fig. 2C). Two related models posit that inversions may also promote speciation with primary gene flow, by allowing adaptations and incompatibilities to disproportionately build up in inverted regions (Navarro & Barton, 2003), or by allowing inversions with co-adapted loci to spread (Kirkpatrick & Barton, 2006; Charlesworth & Barton, 2018). The preceding models invoke gene flow to explain the widely observed pattern of substantially higher divergence within as opposed to outside of inversions. It was recently shown, however, that fixed inversions between the same pair of *Drosophila* species as in Noor's landmark studies, all segregated long before speciation, indicating that ancestrally segregating inversions may be prone to accumulating incompatibilities regardless of the presence of gene flow (Fuller *et al.*, 2017). This is reminiscent of two recent studies on Atlantic cod (*Gadus morhua*), where an ancient inversion is associated with parallel divergence in migratory phenotypes on both sides of

the Atlantic Ocean (Kirubakaran *et al.*, 2016; Sinclair-Waters *et al.*, 2018).

In the European corn borer moth, *Ostrinia nubilalis*, a chromosomal inversion on the Z-chromosome is associated with the accumulation of adaptive alleles and genetic differentiation across nearly 20% of the length of the chromosome (Wadsworth, Li & Dopman, 2015; Yasukochi *et al.*, 2016). The authors posit that in lepidopterans, chromosomal divergence may involve two phases: first, a transient origin through local adaptation, and second, a stable persistence through differential introgression, and a similar scenario may well play out in other groups as well (Conflitti *et al.*, 2015). In addition to paracentric inversions, high rates of other chromosomal rearrangements such as pericentric inversions, reciprocal translocations, fusions and polyploidization appear to be evolving at high rates in several groups of 'notorious speciators' such as *Mimulus* (Fishman *et al.*, 2013), fish (Cioffi *et al.*, 2015) and butterflies (Sichova *et al.*, 2015; Arias, Van Belleghem & McMillan, 2016). Remarkably, in a recent comparative study, the number of fixed inversions between closely related species of songbirds was most strongly predicted by whether or not the species overlap in their geographical range (Hooper & Price, 2017).

Although it is clear that inversions can promote speciation through recombination suppression, this still leaves open whether their contributions differ qualitatively from strong selection or non-inversion-related variation in recombination rates across the genome. In a simulation study, Feder & Nosil (2009) found that strong selection acting on these genes was just as effective in driving divergence as were the large differences between inverted and co-linear regions of diverging genomes. Further simulations refined this result by showing that the effects of inversions were most pronounced when fixed in populations prior to secondary contact, with subsequent RI maintained by adaptive change involving many genes with small fitness effects (Feder, Nosil & Flaxman, 2014). Charlesworth & Barton (2018) recently showed that, as may be intuitively expected, the propensity of an inversion to promote speciation depends strongly on the magnitude of the reduction in recombination rate, which may be small when co-adapted loci are already tightly linked (see also Ortiz-Barrientos & James, 2017). It should also be noted, however, that very few genetic elements have been identified within inversions that contribute to RI. Thus, although spontaneous inversions remain one of the most compelling genomic features associated with rapid speciation, the precise mechanisms by which this is accomplished remain elusive.

GENOMIC DIFFERENTIATION AND BARRIER LOCI

Genomic differentiation islands

For decades, the prevailing view was that RI developed as a byproduct of independent evolution through the progressive substitution of incompatible alleles in geographically isolated populations leading to speciation via postzygotic genetic incompatibilities (i.e. D-M incompatibilities; Via, 2001). Conversely, under a model of speciation with gene flow, lineages are expected to show 'profound genetic similarity' (Via, 2001: 381) differing only at a few loci, presumably those conferring RI. As Nosil, Harmon & Seehausen (2009: 145) aptly state 'although selection often initiates the process of speciation, it often fails to complete it'. Wu, (2001: 887) was among the first to state that speciation reflects 'a process of emerging genealogical distinctness, rather than a discontinuity affecting all genes simultaneously'. Under this view of 'genic speciation' (Wu, 2001; Wu & Ting, 2004), the process is driven by selection on specific regions of the genome, and RI is frequently incomplete until long after categorical speciation (Gourbière & Mallet, 2010).

The genic view of speciation has gained momentum with the model of 'genomic islands of speciation' (Feder *et al.*, 2012; Malinsky *et al.*, 2015). Originally formulated in an empirical setting (Turner, Hahn & Nuzhdin, 2005), the concept of genomic islands has been more broadly conceived as a case wherein certain regions of the genome (typically, loci under strong selection) will show patterns of divergent evolution even in the face of considerable gene flow. Moreover, surrounding areas of the genome, even if evolving neutrally, can show similar patterns of population divergence (as measured by F_{ST}) via the process of divergence hitchhiking (DH; Via, 2012). Theoretically, speciation can thus proceed from a stage wherein genomic islands are small and dispersed throughout the genome, to a later stage wherein genome-wide divergence will occur and the genomic islands are erased (Feder & Nosil, 2010). Under this model, early-stage population divergence is predicted to be characterized by highly heterogeneous genomic divergence, with barrier loci residing in highly differentiated regions of the genome. For the empiricist, this is an attractive model: these predictions provide ideal circumstances for the identification of barrier loci using genome scans. Furthermore, due to increased frequencies of hemiplasy, a genic speciation process has important consequences for the interpretation of phylogenetic and comparative analyses (Box 2).

In line with the genomic islands model, highly heterogeneous patterns of divergence across the genome have indeed been a ubiquitous feature

BOX 2: 'MESSY SPECIATION' AND GENEALOGICAL VARIATION ACROSS THE GENOME

Genomic approaches have clarified many phylogenetic relationships that were previously unclear or controversial, and have provided an enormous increase in resolution and precision to delimit species and population structure within species. This is because the limited information present in single gene fragments and the high variance of the coalescence process necessitates a multitude of independent loci ('gene trees') to accurately infer the underlying genealogy ('species tree'). However, the high variance and stochasticity of the coalescence process can create extensive genealogical variation across the genome, such that genealogies underlying traits of interest are not always likely to follow the inferred species tree – a phenomenon originally called hemiplasy (Avise, Robinson & Kubatko, 2008). Hemiplasy is more likely under precisely some of the patterns in speciation that genomic approaches have helped to uncover, and will thus be necessary to take into account especially in analyses of trait evolution (Hahn & Nakhleh, 2016; Wu *et al.*, 2017). More generally, relying on a single bifurcating tree for species delimitation, phylogenetic hypotheses and subsequent comparative analyses may not always be appropriate given the often multifaceted nature of speciation and the complexities of gene-tree/species-tree reconstruction.

These 'messy' aspects of speciation include, first, gene flow during and after speciation, which produces additional variance in genealogies, and is now believed to be a widespread phenomenon. Second, multiple speciation events that occur in rapid succession, or even simultaneously (Bolnick, 2006; Kautt, Machado-Schiaffino & Meyer, 2016), such as in rapid radiations, result in high proportions of incomplete lineage sorting, the second source of genealogical discordance. Third, in a number of intriguing instances of incipient speciation, strong discordance was found between overall genomic ancestry clines and clines for phenotypes that are thought to represent major isolating barriers (Poelstra *et al.*, 2014; Vijay *et al.*, 2016; Harris *et al.*, 2017; Semenov *et al.*, 2017). In such cases, counter-intuitive patterns that may be uncovered by genomic approaches can certainly complicate attempts at species delimitation, and again, have demonstrated the inherent complexity of speciation.

of what has been identified as the differentiation landscape. However, it has also become clear that the interpretation of these landscapes is highly complicated (Noor & Bennett, 2009; Nachman & Payseur, 2012; Cruickshank & Hahn, 2014; Wolf & Ellegren, 2017). When comparing measures of relative (F_{ST}) and absolute divergence (d_{XY}), Cruickshank & Hahn (2014) found little evidence that 'islands of divergence' are actually produced by lack of introgression via gene flow. Rather, they conclude that differentiation islands represent genomic areas of reduced diversity, which are produced by the effects of linked selection regardless of gene flow. Specifically, variation across the genome in recombination rates and the density of functional elements interacts with selection to produce variation in genetic diversity, and regions of low diversity will automatically show higher levels of relative divergence when populations are isolated (Burri, 2017b). Moreover, given that most fitness effects of new mutations are negative, the effects of background (i.e. negative) selection on patterns of diversity across the genome is expected to be substantial, perhaps larger than those of positive selection (Stephan, 2010), further reducing the likelihood that such regions are commonly important for speciation. Finally, differences in effective population size, sex-linked regions of the genome and the interaction between the two (Belleghem *et al.*, 2018) underline the need for further development of methods that co-estimate selection and all aspects of historical demography.

The emerging consensus appears to be that 'islands of differentiation' are more commonly caused by processes unrelated to speciation, and thus do not by themselves provide evidence for a genic process of speciation. A further problem with the genomic-islands-of-speciation metaphor, and the underlying model, is that in many cases such islands need not form at all during speciation, For instance, when speciation proceeds without flow or is underpinned by polygenic adaptation (Feder & Nosil, 2010; Feder *et al.*, 2012), this model is not particularly relevant. Nevertheless, in systems where speciation may genuinely be characterized as genic, that is with high levels of gene flow and a limited number of barrier loci, such islands may be likely to contain barrier loci. As a recent example, in a comparison of more than 100 populations from 11 species of stick insects (genus *Timema*), investigators identified a strong correlation between genomic islands and localized differentiation of loci underlying colour differences under ecological selection (Riesch *et al.*, 2017). Furthermore, regions with low recombination rates may not only be likely to generate spurious signals of differentiation, but

BOX 3: GENETIC INCOMPATIBILITIES AND RULES OF SPECIATION

Genetic incompatibilities reduce or nullify hybrid fertility and viability. Given that they tend to be slow to develop, and during hybridization events act *after* other barriers to gene flow, they may in many taxa not be as important as prezygotic barriers. Even so, they are the only barriers held to be irreversible. The D-M model posits that interactions between two or more loci that each diverged between two populations are responsible for genetic incompatibilities, which circumvents the need to invoke negative effects of these allelic changes when they occurred *within* each population. 'Haldane's Rule' and the 'Large X-effect' have been described as 'the most consistent empirical patterns in speciation genetics' (Demuth, 2014; see also Delph & Demuth, 2016; Irwin, 2018) and are both related to D-M incompatibilities. They also both involve sex chromosomes (e.g. Johnson & Lachance, 2012; Irwin, 2018), underlining the role of chromosome-level peculiarities in speciation, which has long been recognized (Coluzzi *et al.*, 1977).

In the earliest descriptions of hybrid sterility, Haldane (1922) observed that it was typical for the heterogametic sex to be the one to manifest hybrid sterility as observed in mammalian males (XY) and avian females (ZW). In a related phenomenon, it has been observed that X-chromosome genes in *Drosophila* and most other animals cause infertility in hybrid males at a far greater rate than autosomal genes (Presgraves, 2010), with 60% of X-chromosome genes causing infertility in hybrid males versus the 18% for all the non-sex chromosomes (Masly & Presgraves, 2007). Not surprisingly, sex chromosomes have been implicated for harbouring an excess of genes with sex-biased expression and thus predisposing features for facilitating speciation (Yoshida *et al.*, 2014). This theory has recently been investigated, and supported, to an extremely granular level in mice (Larson *et al.*, 2016). The view across the eukaryotic tree of life suggests that speciation rates are lower in lineages without differentiated sex chromosomes (Phillips & Edmands, 2012), presumably correlated with the lower levels of postzygotic isolation in organisms without sex chromosomes, even when levels of overall genetic divergence are similar (Lima, 2014).

as discussed previously in the context of inversions, regionally low recombination rates may also oppose introgression and promote speciation (Carneiro, Ferrand & Nachman, 2009; Nachman & Payseur, 2012; Schumer *et al.*, 2014; Janoušek *et al.*, 2015; Berner *et al.*, 2017; Ortiz-Barrientos & James, 2017; Samuk *et al.*, 2017).

Noting that current theory is presently dominated by a limited number of model species, perhaps biased by a 'adaptationist perspective', Wolf & Ellegren (2017: 97) call for a cautious approach for interpreting genomic islands as signals of divergent selection. In their comparison of 67 published empirical studies, these authors found that general conclusions are necessarily hampered by a number of confounding factors, including (but not limited to) differential genome quality, differing life history strategies amongst the lineages examined, as well as differing methodologies such as the chosen genome-scan window size and the methods for identifying outliers. This comparison across a wide phylogenetic range therefore suggests that identifying the genomic causes and consequences of divergent genomic islands will require a more fine-scaled approach. We therefore echo Ellegren's (2014) prediction that as the field of speciation genomics continues to develop, enhanced genome characterization will provide a richer understanding of the interaction of genotype and phenotype as targets for divergent selection.

IDENTIFYING BARRIER LOCI

Among the motivations for identifying 'genomic islands' is to associate patterns of divergence with functional genomic mechanisms that may potentially be driving divergence. One such functional class has been described as 'barrier loci'. Conceptualized as any locus that contributes to RI and meets the criteria of pre-speciation divergence and measurable effect size (Nosil & Schluter, 2011), the hunt for 'barrier loci' has been active (Ravinet *et al.*, 2017). The term 'barrier' makes clear that the locus in question, although contributing to the process, may not by itself be sufficient for irreversible lineage divergence, and 'locus' implies that the genetic element in question does not need to be a gene. By determining the specific identity of barrier loci, we can hope that this will move us closer to answering a range of long-standing questions, such as what are the number and effect sizes of barrier loci, what types of genomic regions are involved, what types of mutations are required, and under which evolutionary forces have they evolved (Nosil & Schluter, 2011)? However, arguing that RI is an effect rather than a cause of speciation, some have suggested that instead of primarily focusing on RI and the genes contributing to it, more attention should be given to the causes and consequences of diverging phenotypes, i.e. 'speciation phenotypes' (Shaw & Mullen, 2011). In the context of speciation genomics, both approaches nevertheless come down to establishing links between genotypes

and phenotypes, although detecting loci that underlie RI may be more straightforward in natural populations and using top-down approaches such as genome scans (see Fig. 3 for an overview of approaches that can be used across different types of systems).

Until recently, the identification of barrier loci was predominantly focused on genomic regions contributing to postzygotic isolation. In striking contrast to the current enthusiasm for the role of adaptation and ecology in speciation, most loci that have so far been linked to such hybrid incompatibilities appear to have evolved in response to internal genetic conflicts as well as neutral mutational mechanisms and recombination hotspots, of which PRDM9, discussed below, is a probable example (Maheshwari & Barbash, 2011; Presgraves, 2010). Nevertheless, prezygotic isolation may be more likely than postzygotic isolation to be a consequence of ecological and sexual sources of selection, and prezygotic barrier loci are now also beginning to be identified (Ding *et al.*, 2016). For non-model organisms, especially those that cannot be crossed in laboratory settings, genome scans are the most widely applicable and currently most commonly used method to identify candidate barrier loci (Ravinet *et al.*, 2017). Genome scans examine genetic variation across the genome to find regions with unusual patterns such as strongly elevated genetic differentiation between populations (Lewontin & Krakauer, 1973; Beaumont & Balding, 2004). Ever-decreasing sequencing costs mean that population-level whole-genome resequencing projects are feasible for many non-model organisms, and we stress that given the difficulties described below, this should often be the approach of choice. As discussed in the previous section, differentiation landscapes are commonly highly heterogeneous, with many regions of high differentiation, although most of these do not appear to be directly relevant to speciation. The difficulty in separating highly differentiated genomic regions that harbour barrier loci from those that do not is illustrated by the striking overlap in expected patterns at the genomic level: both are likely to disproportionately represent regions with low recombination, a high density of functional elements and signatures of selection.

Several approaches may help to identify the processes underlying the formation of a given genomic region that stands out for its high levels of differentiation. First, Cruickshank & Hahn (2014) suggested using absolute (i.e. d_{xy}) rather than relative (e.g. F_{ST}) measures of divergence, and this has been widely adopted. Nevertheless, d_{xy} has very little power for recently diverged lineages (Burri, 2017a), may be masked by linked selection (Burri, 2017a), and may also be susceptible to demographic changes (Belleghem *et al.*, 2018). Second, comparative approaches that examine differentiation landscapes across several populations or species pairs, including pairs that are known not to exchange genes, enable

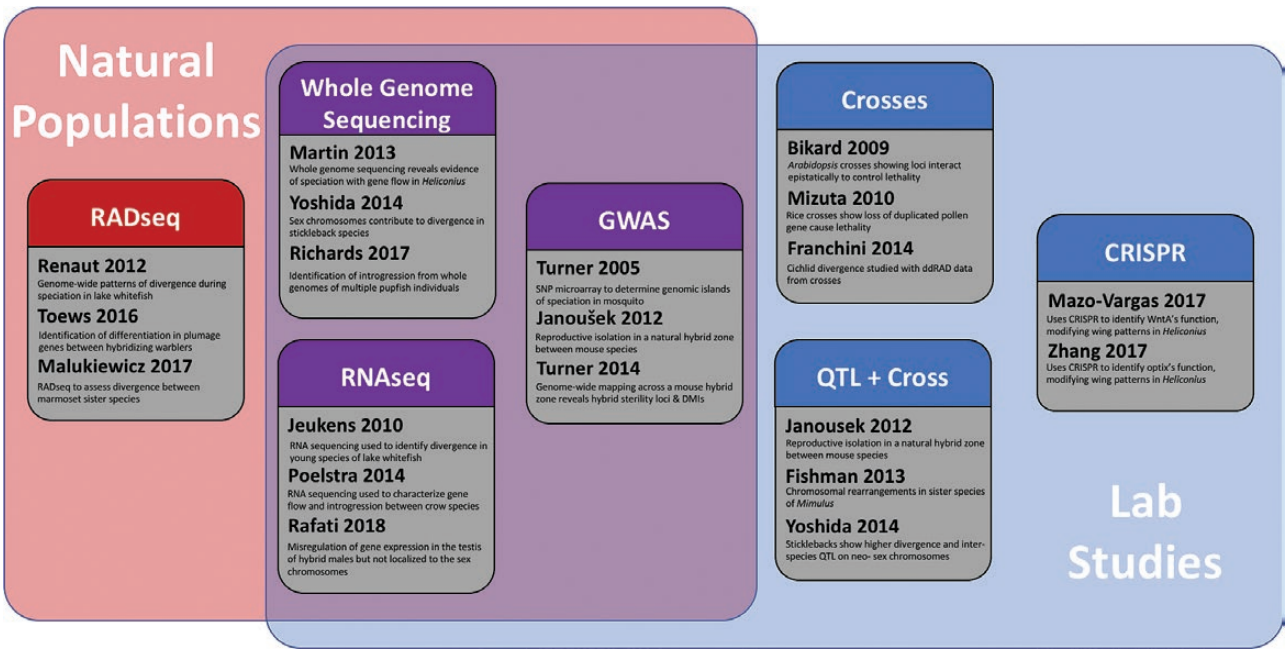

Figure 3. Methods of studying genomics of speciation: common methods of studying the genomics of speciation, with example publications. Each method is assigned to the category of study system in which it is most applicable. The methods are arranged, from left to right, in increasing order of cost and sophistication. The red methods are suitable for natural populations, the blue methods are suitable for laboratory studies and the purple methods are useful in both systems.

the identification of unique differentiation islands in the focal pair (Renaut *et al.*, 2013; Roesti *et al.*, 2015; Vijay *et al.*, 2016; Samuk *et al.*, 2017), which are less likely to simply be the consequence of local genomic features (Burri, 2017b). It should be noted that this approach assumes that landscapes of these genomic features, such as recombination rate, are the same across all population pairs. Third, the genomic features that may shape the differentiation landscape can also be characterized separately. This is increasingly possible in natural populations due to improved genome annotations and improved estimation of local recombination rates using information on linkage disequilibrium (Smukowski & Noor, 2011), the latter owing to techniques such as single-molecule long-read sequencing and linked-read sequencing, which enable better detection of structural variants and also retain haplotype information. Fourth, detailed examination of global and genomically localized gene flow, as well as signatures of selection, may also allow for a better understanding of a given differentiation island.

Despite the promise from more widespread adoption of these approaches, a decade or so of widespread genome scans have made it clear that using such scans to identify barrier loci is in many systems very challenging and may in others not even be feasible (Buerkle, 2017; Jiggins & Martin, 2017; Lindtke & Yeaman, 2017). Before embarking on a genome scan approach for identifying barrier loci, it is thus helpful to consider whether a focal system lends itself to this approach. Systems with relatively low overall divergence, with much ongoing gene flow and within which barrier loci are expected to be few and of large effect are generally most conducive to the identification of candidate barrier loci (Jones *et al.*, 2012; Poelstra *et al.*, 2014; Malinsky *et al.*, 2015; Belleghem *et al.*, 2018). However, if barrier loci are likely to be detected through genome scans only in systems with specific biological features, this may mean that these barrier loci are not a representative subset of all barrier loci. It should also be noted that genome scans in the context of speciation research may be worth pursuing even when there is little scope for direct and precise identification of barrier loci. For instance, such studies provide insight into genome structure and its relation to patterns of differentiation, allow the quantification of gene flow both at the global and the local genomic level, and provide insight into the general architecture (rather than the specific identity) of barriers to gene flow (Jiggins & Martin, 2017).

Hybrid zones – and admixed populations more generally – provide an opportunity to use an alternative set of methods for identifying candidate barrier loci (Gompert, Mandeville & Buerkle, 2017). First, if candidate barrier phenotypes are known, and these segregate within the admixed population, genotype-to-phenotype links can be assessed by genome-wide association methods such as Bayesian Variable Selection Regression (BVSR; Guan & Stephens, 2011; Gompert *et al.*, 2013), genome-wide efficient mixed-model association (GEMMA; Zhou & Stephens, 2012; Turner & Harr, 2014; Delmore *et al.*, 2016) and GenABEL (Aulchenko *et al.*, 2007; Nadeau *et al.*, 2014). Second, loci exhibiting unusually steep clines can be detected by exploiting spatial clines in hybrid zones (Barton & Gale, 1993; Payseur, 2010; Trier *et al.*, 2014; Rafati *et al.*, 2018), and similarly, yet without relying on spatial patterns, genomic clines across many loci in admixed individuals (Lexer *et al.*, 2007; Gompert & Buerkle, 2009). Finally, when local genomic ancestry can be inferred among admixed individuals, the length of continuous ancestry tracts may offer clues to barrier loci (Sedghifar, Brandvain & Ralph, 2016), and ancestry disequilibrium between locus pairs can be used to test for two-locus genetic incompatibilities specifically (Schumer *et al.*, 2014; Schumer & Brandvain, 2016).

If candidate barrier loci are identified, functional approaches are often necessary to validate their effects. Although these approaches will for the foreseeable future remain limited to organisms that can be kept in laboratory settings, manipulated, and in most cases, bred (such as *Drosophila*, Cooper & Phadnis, 2016), the recent breakthrough of CRISPR holds great promise for testing candidate genes much more effectively and in a wider variety of species than other transgenic approaches (Bono, Olesnicky & Matzkin, 2015). The overriding practical requirement for the application of CRISPR for genome editing is that CRISPR/Cas9 elements can be delivered to early-stage embryos, thus yielding the potential for the two repair pathways that are triggered by the double-stranded breaks induced by CRISPR/Cas9 to be exploited for multiple applications. For example, the non-homologous endjoining (NHEJ) repair pathway can induce large deletions that create knockouts, which has, for example, been used in a series of studies that have identified some of the genes underlying several butterfly wing colour pattern traits (Zhang & Reed, 2016; Mazo-Vargas *et al.*, 2017; Zhang *et al.*, 2017a; Zhang, Mazo-Vargas & Reed, 2017b). CRISPR/Cas9-mediated NHEJ can also create other structural variants such as duplications and inversions, which may be particularly useful for testing their potential role in speciation (Bono *et al.*, 2015; Kraft *et al.*, 2015). Furthermore, homology-directed repair can be used to introduce precise genetic modifications. For instance, Ding *et al.* (2016) used CRISPR/Cas9-mediated homology-directed repair both to fine-map and to create mutations within the locus responsible for a courtship song difference between two species of *Drosophila*. This locus has been identified as the insertion of a retro-element in an

intron of *slo*, an ion channel gene, illustrating both the potential for small mutations to have large effects as well as the unpredictable relationship between gene function and impacts on speciation. To summarize, advances in CRISPR techniques are developing at astonishing rates, including methods to directly convert single bases without requiring the formation of double-stranded breaks, using a catalytically impaired CRISPR/Cas9 mutant base editing (Nishida *et al.*, 2016; Gaudelli *et al.*, 2017). Thus, we are at the early stages of a technological revolution with unforeseeable impacts on the field of speciation genomics.

PRDM9: A CRUCIAL BARRIER LOCUS?

Perhaps the most intriguing candidate barrier locus to have emerged from a suite of recombination hotspot modifiers (Johnson, 2010) is PDRM9 (Oliver *et al.*, 2009; Brand & Presgraves, 2016). Known to be strongly associated with recombination hotspots in placental mammals, PRDM9 is a rapidly evolving zinc finger protein with sequence-specific DNA binding and histone methyltransferase activity. As such, it 'neatly wrap[s] genetic, epigenetic, and trans-acting factors known to influence recombination into one intriguing package' (Sandovici & Sapienza, 2010: 1). Although as yet only characterized at the population level in a few natural populations, mice (Kono *et al.*, 2014) and humans (1000 Genomes Project Consortium, 2010), PRDM9 polymorphism is hypothesized to play two fundamental roles in the genome: to yield a diverse spectrum of recombination hotspots and to cause male hybrid sterility. In this regard, it is noteworthy for mediating both recombination rate and hybrid sterility, which in turn raises the tantalizing possibility that these activities are potentially causative to RI and speciation (Kono *et al.*, 2014; Payseur, 2016). Moreover, it appears to be something of a smoking gun connecting rates of recombination directly to rapid rates of sequence evolution associated with strong positive selection (Oliver *et al.*, 2009; Sandovici & Sapienza, 2010; Ponting, 2011; Axelsson *et al.*, 2012; Groeneveld *et al.*, 2012; Gravogl, Schwarz & Tiemann-Boege, 2014; Kono *et al.*, 2014; Schwartz *et al.*, 2014; Padhi *et al.*, 2017).

As support for this hypothesis, in vertebrate groups such as birds that lack PRDM9, interspecific hybridization appears to be more feasible across larger evolutionary distances (Singhal *et al.*, 2015) than in mammals. Oliver *et al.* (2009) found that concerted evolution and positive selection have united to drive rapid evolution of the gene in rodents, producing high levels of sequence variation across 13 rodent genomes. These authors also found that PRDM9 plays a measurable role in determining male sterility both within and among species as divergent as rodents and primates. Broad phylogenetic surveys of PRDM9 suggest that it may be the most rapidly evolving gene in human and other animals (Ponting, 2011), and in a survey of 64 individuals across 18 species of primate, 68 unique alleles were identified (Schwartz *et al.*, 2014). Of particular interest to human evolutionary biology, alignments of Neanderthal and Denisovan genomes reveal that PRDM9 sequences in these extinct species are closely related to present-day alleles in modern humans that are both rare and specific to African populations (Schwartz *et al.*, 2014). However, by far the most intensive and compelling work on PRDM9's role in hybrid sterility has been conducted in mouse (Mihola *et al.*, 2009; Kono *et al.*, 2014; Davies *et al.*, 2016; Smagulova *et al.*, 2016; Zelazowski & Cole, 2016), with increasing evidence that recombination rate and hybrid sterility are linked and phylogenetically widespread, thus pointing to a more general connection to speciation (Payseur, 2016).

Successful attempts have been made to advance PRDM9 research in non-model organisms. The recombination patterns of organisms lacking PRDM9 (e.g. dogs and bees) have been mapped, and in cattle some preliminary attempts have been made to establish PRDM9 as active in holding up species boundaries (Lou *et al.*, 2014). Tarsiers, the most diverged lineage within the primate clade, show high allelic diversity of PRDM9 that is highly congruent with phylogeography, thus suggesting an important role in speciation within the genus *Tarsius*, and by inference, haplorrhine primates (Heerschop *et al.*, 2016). Similarly, the remarkable variation in the zinc finger domain of PRDM9 in goats and sheep, wherein numerous amino acid sites are apparently under strong positive selection, has also been interpreted as evidence of the gene's intriguing role in speciation (Padhi *et al.*, 2017).

As new understanding of the molecular evolutionary dynamics of PRDM9, and recombination hotspots in general, emerges, an ever more nuanced view of its role in speciation is developing. Although it has long been known that the locations of recombination hotspots are highly mobile and are rarely conserved even between closely related species (e.g. Ptak *et al.*, 2005; Winckler *et al.*, 2005), the impact on hybrid sterility is only now coming to light. For example, Davies *et al.* (2016) found that hybrid sterility between two mouse lineages could be instantaneously reversed by 'humanizing' the PRDM9 allele. When the PRDM9 array was genetically engineered in one lineage to represent the human sequence, the genomic position of recombination hotspots was accordingly rearranged, and surprisingly, yielded fertile male hybrids. Thus, one of the key findings of this study is that although PRDM9 shows a direct involvement in hybrid infertility, the effects are likely to be evolutionarily transient. In other words, increased divergence of PRDM9 is likely to mean a decreased role in the maintenance of species boundaries, thereby

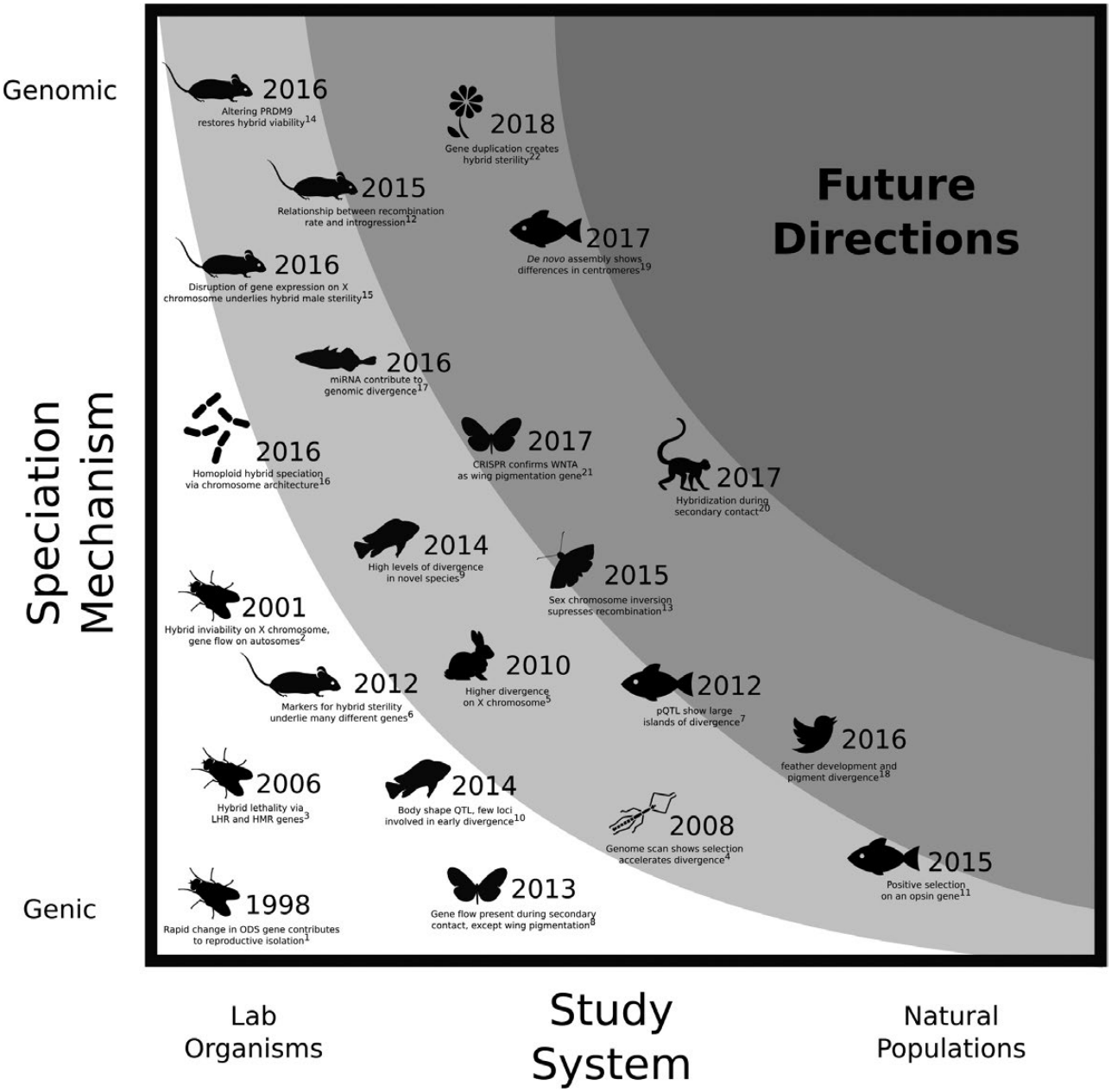

Figure 4. Speciation study spectrum: graphical representation of where recent published studies of speciation fall along two continua: 1 – whether the focal specie was studied as a lab organism or a natural population (bottom axis of diagram) and 2 – the complexity that underlies the nature of the genetic mechanism responsible for reproductive isolation, ranging from genic to genomic (left axis of diagram). Studies are represented by an icon depicting species, year of publication, and a short summary with superscript notation of the full citation. The references (in chronological order) are as follows: 1, Ting *et al.* (1998); 2, Noor *et al.* (2001a); 3, Brideau *et al.* (2006); 4, Nosil; Egan & Funk (2008); 5, Carneiro *et al.* (2010); 6, Janoušek *et al.* (2012); 7, Renaut *et al.* (2012); 8, Martin *et al.* (2013); 9, Fan *et al.* (2014); 10, Franchini *et al.* (2014); 11, Gaither *et al.* (2015); 12, Janoušek *et al.* (2015); 13, Wadsworth *et al.* (2015); 14, Davies *et al.* (2016); 15, Larson *et al.* (2016); 16, Leducq *et al.* (2016); 17, Rastorguev *et al.* (2016); 18, Toews *et al.* (2016); 19, Ichikawa *et al.* (2017); 20, Malukiewicz *et al.* (2017); 21, Mazo-Vargas *et al.* (2017); 22, Zuellig & Sweigart (2018).

suggesting that there may be a phylogenetic distance 'sweet spot' wherein PRDM9 can strongly impact propensity for speciation, but with diminishing impact as phylogenetic distance increases. It will be fascinating to explore this phenomenon in an array of non-model species across a greater phylogenetic breadth.

LOOKING TO THE FUTURE: IS THERE HOPE FOR A UNIFIED THEORY OF SPECIATION?

We have long known that organisms are hierarchically distributed across the tree of life, existing in 'bins' that biologists attempt to define as species. These bins have boundaries of varying completeness and clarity, made porous by hybridization and introgression. Asking how these biological 'edges' are formed, and how they are maintained, are among the most basic questions in evolutionary biology. The relationship between genomic differentiation and lineage diversification is profoundly complex, and can range from circumstances wherein speciation is virtually instantaneous owing to possibly random genomic events such as chromosomal inversions, to scenarios of rapid speciation *in situ* owing to strong environmental selection, to speciation on evolutionary timescales wherein differentiation and RI slowly build in geographical isolation. Thus, it is not surprising that few, if any, rules have been identified to formalize the role of the genome in speciation.

Intrinsic genomic features such as inversions, gene duplications, recombination patterns and higher order architecture have all been implicated in speciation. In many cases, these discoveries occurred initially in lab-based model organisms with well-characterized genomes and tractable life histories. Identification of the genetics underlying RI has been most feasible for postzygotic incompatibilities between pairs of genes in long-studied species such as *Drosophila*. We are now reaching a point, however, wherein the field is rapidly expanding outward and is discovering the more complex genomic underpinnings of speciation in a wider array of species (Fig. 4). As genomic resources have spread to evolutionarily proximate species, mechanisms of speciation are being described in non-model species and natural populations. Accordingly, our view of speciation has become richer and more complex. The interplay among underlying features of the genome, patterns and processes of speciation, and the ecological surroundings of species will continue to emerge as knowledge of non-model genomics increases, and the field will push ever further toward insights into natural, non-model populations with complex speciation stories – the 'unexplored corner' suggested in Figure 4.

A unified field of speciation genomics will thus require a multi-pronged approach to speciation dynamics that takes into account intrinsic features of genomic architecture examined in the light of each organism's extrinsic biology and ecology. There is an essential place for targeted studies that illuminate the role of specific genes or structural variants, while many valuable insights can also be gained through genome scan comparisons, although caution must be applied. Consequently, the continued development of theory and competing models will always be relevant in order to make sense of what will be an ever-increasing torrent of empirical data. By examining the role of the genome in contrasting models of speciation, we will attain powerful insight into the differential effects of historical constraint in the face of ecological opportunity. It is the interplay between these two forces that has and continues to produce species diversity across the Tree of Life.

ACKNOWLEDGEMENTS

We thank John Allen for the invitation to submit this review, and are especially grateful for his patience and understanding as we made our way through the exponentially expanding literature on this subject. Five anonymous reviewers provided constructive input which substantially improved the manuscript. The work was supported by Duke University start-up funds and NSF DEB1354610 to ADY and a Katherine Stern Dissertation fellowship to CRC. ADY also gratefully acknowledges support from the John Simon Guggenheim Memorial Foundation and the Alexander von Humboldt Foundation.

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JB, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26:** 229–246.

Alkhateeb A, Rueda L. 2017. Zseq: an approach for preprocessing next-generation sequencing data. *Journal of Computational Biology* **24:** 746–755.

Arias CF, Van Belleghem S, McMillan WO. 2016. Genomics at the evolving species boundary. *Current Opinion in Insect Science* **13:** 7–15.

Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A. 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science Advances* **4:** eaap9873.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23:** 1294–1296.

Avise JC, Robinson TJ, Kubatko L. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Systematic Biology* **57:** 503–507.

Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindblad-Toh K; LUPA Consortium. 2012. Death of

PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Research* **22:** 51–63.

Barton NH, Gale KS. 1993. Genetic analysis of hybrid zones. In: Harrison RG, ed. *Hybrid zones and the evolutionary process.* Oxford: Oxford University Press. 13–45.

Bayes JJ, Malik HS. 2009. Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science* **326:** 1538–1541.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13:** 969–980.

Belleghem SMV, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA, Jiggins Chris D, Martin SH. 2018. Patterns of Z chromosome divergence among Heliconius species highlight the importance of historical demography. *Molecular Ecology*. doi:10.1111/mec.14560

Berner D, Ammann M, Spencer E, Rüegg A, Lüscher D, Moser D. 2017. Sexual isolation promotes divergence between parapatric lake and stream stickleback. *Journal of Evolutionary Biology* **30:** 401–411.

Berner D, Roesti M. 2017. Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. *Molecular Ecology* **26:** 6351–6369.

Bikard D, Patel D, Le Metté C, Giorgi V, Camilleri C, Bennett MJ, Loudet O. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323:** 623–626.

Bolnick DI. 2006. Multi-species outcomes in a common model of sympatric speciation. *Journal of Theoretical Biology* **241:** 734–744.

Bono JM, Olesnicky EC, Matzkin LM. 2015. Connecting genotypes, phenotypes and fitness: harnessing the power of CRISPR/Cas9 genome editing. *Molecular Ecology* **24:** 3810–3822.

Brand CL, Presgraves DC. 2016. Evolution: on the origin of symmetry, synapsis, and species. *Current Biology* **26:** R325–R328.

Brandvain Y, Matute DR. 2018. When genes move, genomes collide. *PLoS Genetics* **14:** e1007286.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken B, Alföldi J, Amemiya C, Azzouzi N, Baroiller JF, Barloy-Hubler F, Berlin A, Bloomquist R, Carleton KL, Conte MA, D'Cotta H, Eshel O, Gaffney L, Galibert F, Gante HF, Gnerre S, Greuter L, Guyon R, Haddad NS, Haerty W, Harris RM, Hofmann HA, Hourlier T, Hulata G, Jaffe DB, Lara M, Lee AP, MacCallum I, Mwaiko S, Nikaido M, Nishihara H, Ozouf-Costaz C, Penman DJ, Przybylski D, Rakotomanga M, Renn SCP, Ribeiro FJ, Ron M, Salzburger W, Sanchez-Pulido L, Santos ME, Searle S, Sharpe T, Swofford R, Tan FJ, Williams L, Young S, Yin S, Okada N, Kocher TD, Miska EA, Lander ES, Venkatesh B, Fernald RD, Meyer A, Ponting CP, Streelman JT, Lindblad-Toh K, Seehausen O, Di Palma F. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513:** 375–381.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. *Science* **314:** 1292–1295.

Buerkle CA. 2017. Inconvenient truths in population and speciation genetics point towards a future beyond allele frequencies. *Journal of Evolutionary Biology* **30:** 1498–1500.

Burri R. 2017a. Dissecting differentiation landscapes: a linked selection's perspective. *Journal of Evolutionary Biology* **30:** 1501–1505.

Burri R. 2017b. Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters* **1:** 118–131.

Bush GL. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology & Evolution* **9:** 285–288.

Bush GL. 1998. The conceptual radicalization of an evolutionary biologist. In *Endless forms: species and speciation*. Oxford: Oxford University Press, 425–438.

Campagna L, Kopuchian C, Tubaro PL, Lougheed SC. 2014. Secondary contact followed by gene flow between divergent mitochondrial lineages of a widespread Neotropical songbird (*Zonotrichia capensis*). *Biological Journal of the Linnean Society* **111:** 863–868.

Carneiro M, Blanco-Aguiar JA, Villafuerte R, Ferrand N, Nachman MW. 2010. Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution* **64:** 3443–3460.

Carneiro M, Ferrand N, Nachman MW. 2009. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **181:** 593–606.

Castillo DM, Barbash DA. 2017. Moving speciation genetics forward: modern techniques build on foundational studies in Drosophila. *Genetics* **207:** 825–842.

Charlesworth B, Barton NH. 2018. The spread of an inversion with migration and selection. *Genetics* **208:** 377–382.

Cioffi MD, Bertollo LAC, Villa MA, de Oliveira EA, Tanomtong A, Yano CF, Supiwong W, Chaveerach A. 2015. Genomic organization of repetitive DNA elements and its implications for the chromosomal evolution of channid fishes (Actinopterygii, Perciformes). *PLoS One* **10:** e0130199.

Coluzzi M, Sabatini A, Petrarca V, Di Deco MA. 1977. Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nature* **266:** 832–833.

Conflitti IM, Shields GF, Murphy RW, Currie DC. 2015. The speciation continuum: ecological and chromosomal divergence in the *Simulium arcticum* complex (Diptera: Simuliidae). *Biological Journal of the Linnean Society* **115:** 13–27.

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467:** 1061.

Cooper JC, Phadnis N. 2016. A genomic approach to identify hybrid incompatibility genes. *Fly* **10:** 142–148.

Cortesi F, Musilova Z, Stieb SM, Hart NS, Siebeck UE, Malmstrom M, Torresen OK, Jentoft S, Cheney KL, Marshall NJ, Carleton KL, Salzburger W. 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 1493–1498.

Coyner BS, Murphy PJ, Matocq MD. 2015. Hybridization and asymmetric introgression across a narrow zone of contact between *Neotoma fuscipes* and *N. macrotis* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society* **115:** 162–172.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23:** 3133–3157.

Darwin C. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society* **1858:** 45–52.

Darwin C. 1859. *On the origin of species by means of natural selection*. London: John Murray.

Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Diaz R, Preece C, Li R, Bitoun E, Brick K, Green CM, Camerini-Otero RD, Myers SR, Donnelly P. 2016. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* **530:** 171–176.

Delmore KE, Toews DP, Germain RR, Owens GL, Irwin DE. 2016. The genetics of seasonal migration and plumage color. *Current Biology* **26:** 2167–2173.

Delph LF, Demuth JP. 2016. Haldane's rule: genetic bases and their empirical support. *The Journal of Heredity* **107:** 383–391.

Demuth J. 2014. Do sex chromosomes affect speciation rate?. *BioEssays* **36:** 632.

Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536:** 329–332.

Dobzhansky T. 1933. On the sterility of the interracial hybrids in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences of the United States of America* **19:** 397–403.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29:** 51–63.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, Uebbing S, Wolf JB. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491:** 756–760.

Evans BJ. 2008. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Frontiers in Bioscience* **13:** 4687–4706.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9:** e1003905.

Fan S, Meyer A. 2014. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Frontiers in Genetics* **5:** 163.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28:** 342–350.

Feder JL, Nosil P. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63:** 3061–3075.

Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64:** 1729–1747.

Feder JL, Nosil P, Flaxman SM. 2014. Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. *Frontiers in Genetics* **5:** 295.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35:** 124–138.

Ferree PM, Barbash DA. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. *PLoS Biology* **7:** e1000234.

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution* **67:** 2547–2560.

Franchini P, Fruciano C, Spreitzer ML, Jones JC, Elmer KR, Henning F, Meyer A. 2014. Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology* **23:** 1828–1845.

Fuentes-Pardo AP, Ruzzante DE. 2017. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Molecular Ecology* **26:** 5369–5406.

Fuller Z, Leonard C, Young R, Schaeffer S, Phadnis N. 2017. The role of chromosomal inversions in speciation. *bioRxiv*. doi:10.1101/211771

Gaither MR, Bernal MA, Coleman RR, Bowen BW, Jones SA, Simison WB, Rocha LA. 2015. Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology* **24:** 1543–1557.

Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. 2016. Genomics of speciation and introgression in Princess cichlid fishes from Lake Tanganyika. *Molecular Ecology* **25:** 6143–6161.

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551:** 464–471.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 1513–1518.

Gompert Z, Buerkle CA. 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18:** 1207–1224.

Gompert Z, Lucas LK, Nice CC, Buerkle CA. 2013. Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. *Evolution* **67:** 2498–2514.

Gompert Z, Mandeville EG, Buerkle CA. 2017. Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* **48:** 207–229.

Gourbière S, Mallet J. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution* **64:** 1–24.

Gravogl Y, Schwarz T, Tiemann-Boege I. 2014. Characterizing the zinc finger binding interactions of PRDM9 with the DNA of recombination hotspots. *FEBS Journal* **281:** 297.

Greer SU, Nadauld LD, Lau BT, Chen J, Wood-Bouwens C, Ford JM, Kuo CJ, Ji HP. 2017. Linked read sequencing resolves complex genomic rearrangements in gastric cancer metastases. *Genome Medicine* **9:** 57.

Groeneveld LF, Atencia R, Garriga RM, Vigilant L. 2012. High diversity at PRDM9 in chimpanzees and bonobos. *PLoS One* **7:** e39064.

Guan Y, Stephens M. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics* **5:** 1780–1815.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* **5:** e1000695.

Hahn MW, Nakhleh L. 2016. Irrational exuberance for resolved species trees. *Evolution* **70:** 7–17.

Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12:** 101–109.

Harris RB, Alström P, Ödeen A, Leaché AD. 2018. Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (*Motacilla*). *Molecular Phylogenetics and Evolution* **120:** 183–195.

Heerschop S, Zischler H, Merker S, Perwitasari-Farajallah D, Driller C. 2016. The pioneering role of PRDM9 indel mutations in tarsier evolution. *Scientific Reports* **6:** 34618.

Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* **39:** 21–42.

Hooper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology & Evolution* **1:** 1526–1534.

Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, Kumagai M, Inoue Y, Uchida Y, Irie N, Takeda H, Morishita S. 2017. Centromere evolution and CpG methylation during vertebrate speciation. *Nature Communications* **8:** 1833.

Irwin DE. 2018. Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*. doi:10.1111/mec.14537

Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Research* **27:** 768–777.

Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK. 2015. Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution* **32:** 1208–1220.

Janoušek V, Wang L, Luzynski K, Dufková P, Vyskočilová MM, Nachman MW, Munclinger P, Macholán M, Piálek J, Tucker PK. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology* **21:** 3032–3047.

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular Ecology* **19:** 5389–5403.

Jiao WB, Schneeberger K. 2017. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36:** 64–70.

Jiggins CD, Martin SH. 2017. Glittering gold and the quest for Isla de Muerta. *Journal of Evolutionary Biology* **30:** 1509–1511.

Johnson NA. 2010. Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics* **26:** 317–325.

Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences* **1256:** E1–22.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, Zhang H, Pollen AA, Howes T, Amemiya C, Baldwin J, Bloom T, Jaffe DB, Nicol R, Wilkinson J, Lander ES, Di Palma F, Lindblad-Toh K, Kingsley DM; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484:** 55–61.

Kamath GM, Shomorony I, Xia F, Courtade TA, Tse DN. 2017. HINGE: long-read assembly achieves optimal repeat resolution. *Genome Research* **27:** 747–756.

Kang L, Settlage R, McMahon W, Michalak K, Tae H, Garner HR, Stacy EA, Price DK, Michalak P. 2016. Genomic signatures of speciation in sympatric and allopatric Hawaiian picture-winged Drosophila. *Genome Biology and Evolution* **8:** 1482–1488.

Kautt AF, Machado-Schiaffino G, Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of nicaraguan crater lake cichlids. *PLoS Genetics* **12:** e1006157.

Kearns AM, Restani M, Szabo I, Schrøder-Nielsen A, Kim JA, Richardson HM, Marzluff JM, Fleischer RC, Johnsen A, Omland KE. 2018. Genomic evidence of speciation reversal in ravens. *Nature Communications* **9:** 906.

Kern AD, Hey J. 2017. Exact calculation of the joint allele frequency spectrum for isolation with migration models. *Genetics* **207:** 241–253.

Kingman JFC. 1982. The coalescent. *Stochastic Processes and their Applications* **13:** 235–248.

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173:** 419–434.

Kirubakaran TG, Grove H, Kent MP, Sandve SR, Baranski M, Nome T, De Rosa MC, Righino B, Johansen T, Otterå H, Sonesson A, Lien S, Andersen Ø. 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology* **25:** 2130–2143.

Kono H, Tamura M, Osada N, Suzuki H, Abe K, Moriwaki K, Ohta K, Shiroishi T. 2014. Prdm9 polymorphism unveils mouse evolutionary tracks. *DNA Research* **21:** 315–326.

Kraft K, Geuer S, Will Anja J, Chan Wing L, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim Daniel M, Kragesteen Bjørt K, Spielmann M, Mundlos S, Lupiáñez Darío G, Andrey G. 2015. Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Reports* **10:** 833–839.

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2017. Rapid hybrid speciation in Darwin's finches. *Science* eaao4593.

Lapierre M, Lambert A, Achaz G. 2017. Accuracy of demographic inferences from the site frequency spectrum: the case of the Yoruba population. *Genetics* **206:** 439–449.

Larsen PA, Harris RA, Liu Y, Murali SC, Campbell CR, Brown AD, Sullivan BA, Shelton J, Brown SJ, Raveendran M, Dudchenko O, Machol I, Durand NC, Shamim MS, Aiden EL, Muzny DM, Gibbs RA, Yoder AD, Rogers J, Worley KC. 2017. Hybrid de novo genome assembly and centromere characterization of the gray mouse lemur (*Microcebus murinus*). *BMC Biology* **15:** 110.

Larsen PA, Heilman AM, Yoder AD. 2014. The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genomics* **15:** 720.

Larson EL, Keeble S, Vanderpool D, Dean MD, Good JM. 2016. The composite regulatory basis of the large X-effect in mouse speciation. *Molecular Biology and Evolution* **34:** 282–295.

Leducq JB, Nielly-Thibault L, Charron G, Eberlein C, Verta JP, Samani P, Sylvester K, Hittinger CT, Bell G, Landry CR. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1:** 15003.

Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz M. 2016. Third-generation sequencing and the future of genomics. *bioRxiv*: 048603.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74:** 175–195.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF. 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.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J, Zhang J, Jiang A, Li J, Zhou C, Zhang J, Liu Y, Sun X, Zhao H, Niu Z, Lou P, Xian L, Shen X, Liu S, Zhang S, Zhang M, Zhu L, Shuai S, Bai L, Tang G, Liu H, Jiang Y, Mai M, Xiao J, Wang X, Zhou Q, Wang Z, Stothard P, Xue M, Gao X, Luo Z, Gu Y, Zhu H, Hu X, Zhao Y, Plastow GS, Wang J, Jiang Z, Li K, Li N, Li X, Li R. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics* **45:** 1431–1438.

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K. 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics* **12:** 540.

Lieberman-Aiden E, Berkum NLv, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326:** 289–293.

Lima TG. 2014. Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation. *Nature Communications* **5:** 4743.

Lindtke D, Yeaman S. 2017. Identifying the loci of speciation: the challenge beyond genome scans. *Journal of Evolutionary Biology* **30:** 1478–1481.

Lohse K, Chmelik M, Martin SH, Barton NH. 2016. Efficient strategies for calculating blockwise likelihoods under the coalescent. *Genetics* **202:** 775–786.

Lou YN, Liu WJ, Wang CL, Huang L, Jin SY, Lin YQ, Zheng YC. 2014. Histological evaluation and Prdm9 expression level in the testis of sterile male cattle-yaks. *Livestock Science* **160:** 208–213.

Lynch M, Force AG. 2000. The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist* **156:** 590–605.

Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends in Genetics* **33:** 68–80.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annual Review of Genetics* **45:** 331–355.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350:** 1493–1498.

Malukiewicz J, Guschanski K, Grativol AD, Oliveira MA, Ruiz-Miranda CR, Stone AC. 2017. Application of PE-RADSeq to the study of genomic diversity and divergence of two Brazilian marmoset species (*Callithrix jacchus* and *C. penicillata*). *American Journal of Primatology* **79:** 1–12.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23:** 1817–1828.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution* **32:** 244–257.

Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biology* **5:** 1890–1898.

Mayr E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. Cambridge: Harvard University Press.

Mazo-Vargas A, Concha C, Livraghi L, Massardo D, Wallbank RWR, Zhang L, Papador JD, Martinez-Najera D, Jiggins CD, Kronforst MR, Breuker CJ, Reed RD, Patel NH, McMillan WO, Martin A. 2017. Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. *Proceedings of the National Academy of Sciences USA* **114:** 10701–10706.

McVean GA, Cardin NJ. 2005. Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **360:** 1387–1393.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323:** 373–375.

Mizuta Y, Harushima Y, Kurata N. 2010. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proceedings of the National Academy of Sciences USA* **107:** 20417–20422.

Morii Y, Yokoyama J, Kawata M, Davison A, Chiba S. 2015. Evidence of introgressive hybridization between the morphologically divergent land snails *Ainohelix* and *Ezohelix*. *Biological Journal of the Linnean Society* **115:** 77–95.

Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367:** 409–421.

Nadeau NJ, Ruiz M, Salazar P, Counterman B, Medina JA, Ortiz-Zuazaga H, Morrison A, McMillan WO, Jiggins CD, Papa R. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research* **24:** 1316–1333.

Navarro A, Barton NH. 2003. Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. *Science* **300:** 321–324.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews. Genetics* **12:** 443–451.

Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**: pii: aaf8729.

Noor MA, Grams KL, Bertucci LA, Almendarez Y, Reiland J, Smith KR. 2001a. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution* **55:** 512–521.

Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103:** 439.

Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001b. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences USA* **98:** 12084–12088.

Nosil P. 2008. Speciation with gene flow could be common. *Molecular Ecology* **17:** 2103–2106.

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62:** 316–336.

Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* **24:** 145–156.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. *Trends in Ecology & Evolution* **26:** 160–167.

O'Neill MJ, O'Neill RJ. 2018. Sex chromosome repeats tip the balance towards speciation. *Molecular Ecology*. doi:10.1111/mec.14577

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genetics* **5:** e1000753.

Orr HA. 1990. "Why polyploidy is rarer in animals than in plants" revisited. *The American Naturalist* **136:** 759–770.

Ortiz-Barrientos D, James ME. 2017. Evolution of recombination rates and the genomic landscape of speciation. *Journal of Evolutionary Biology* **30:** 1519–1521.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* **34:** 401–437.

Padhi A, Shen BT, Jiang JC, Yang Z, Liu GE, Li M. 2017. Ruminant-specific multiple duplication events of PRDM9 before speciation. *BMC Evolutionary Biology* **17**.

Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO, Jiggins CD. 2012. Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genetics* **8:** e1002752.

Paten B, Novak AM, Eizenga JM, Garrison E. 2017. Genome graphs and the evolution of genome inference. *Genome Research* **27:** 665–676.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* **192:** 1065–1093.

Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Crooijmans RP, Groenen MA. 2015. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* **16:** 330.

Payseur BA. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10:** 806–820.

Payseur BA. 2016. Genetic links between recombination and speciation. *PLoS Genetics* **12:** e1006066.

Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology* **64:** 651–662.

Phillippy AM. 2017. New advances in sequence assembly. *Genome Research* **27:** xi–xiii.

Phillips BC, Edmands S. 2012. Does the speciation clock tick more slowly in the absence of heteromorphic sex chromosomes? *BioEssays* **34:** 166–169.

Pinho C, Hey J. 2010. Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics* **41:** 215–230.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf

JB. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344:** 1410–1414.

Ponting CP. 2011. What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics* **27:** 165–171.

Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nature Reviews. Genetics* **11:** 175–180.

Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans (vol 37, pg 429, 2005). *Nature Genetics* **37:** 445–445.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. *Nature Reviews. Genetics* **16:** 359–371.

Rafati N, Blanco-Aguiar JA, Rubin CJ, Sayyab S, Sabatino SJ, Afonso S, Feng C, Alves PC, Villafuerte R, Ferrand N, Andersson L, Carneiro M. 2018. A genomic map of clinal variation across the European rabbit hybrid zone. *Molecular Ecology* **27:** 1457–1478.

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. 2014. Genome-wide inference of ancestral recombination graphs. *PLoS Genetics* **10:** e1004342.

Rastorguev SM, Nedoluzhko AV, Sharko FS, Boulygina ES, Sokolov AS, Gruzdeva NM, Skryabin KG, Prokhortchouk EB. 2016. Identification of novel microRNA genes in freshwater and marine ecotypes of the threespined stickleback (*Gasterosteus aculeatus*). *Molecular Ecology Resources* **16:** 1491–1498.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30:** 1450–1477.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4:** 1827.

Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, Bernatchez L. 2012. Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367:** 354–363.

Rennison DJ, Owens GL, Taylor JS. 2012. Opsin gene duplication and divergence in ray-finned fish. *Molecular Phylogenetics and Evolution* **62:** 986–1008.

Richards EJ, Martin CH. 2017. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genetics* **13:** e1006919.

Riesch R, Muschick M, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR, de Carvalho CF, Safran RJ, Sandoval CP, Feder JL, Gries R, Crespi BJ, Gries G, Gompert Z, Nosil P. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology and Evolution* **1:** 82.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16:** 351–358.

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews. Genetics* **16:** 85–97.

Roesti M, Kueng B, Moser D, Berner D. 2015. The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications* **6:** ncomms9767.

Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* **308:** 58–73.

Samuk K, Owens GL, Delmore KE, Miller SE, Rennison DJ, Schluter D. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology* **26:** 4378–4390.

Sandovici I, Sapienza C. 2010. PRDM9 sticks its zinc fingers into recombination hotspots and between species. *F1000 Biology Reports* **2**.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507:** 354–357.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46:** 919–925.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Reviews. Genetics* **15:** 749–763.

Schmitz J, Noll A, Raabe CA, Churakov G, Voss R, Kiefmann M, Rozhdestvensky T, Brosius J, Baertsch R, Clawson H, Roos C, Zimin A, Minx P, Montague MJ, Wilson RK, Warren WC. 2016. Genome sequence of the basal haplorrhine primate *Tarsius syrichta* reveals unusual insertions. *Nature Communications* **7:** 12997.

Schumer M, Brandvain Y. 2016. Determining epistatic selection in admixed populations. *Molecular Ecology* **25:** 2577–2591.

Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P. 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife* **3:** doi: 10.7554/eLife.02535.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, Przeworski M. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* doi: 10.1126/science.aar3684.

Schwartz JJ, Roach DJ, Thomas JH, Shendure J. 2014. Primate evolution of the recombination regulator PRDM9. *Nature Communications* **5:** 4370.

Scordato ES, Symes LB, Mendelson TC, Safran RJ. 2014. The role of ecology in speciation by sexual selection: a systematic empirical review. *The Journal of Heredity* **105**(Suppl 1): 782–794.

Sedghifar A, Brandvain Y, Ralph P. 2016. Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology* **25:** 2559–2576.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19:** 198–207.

Semenov GA, Scordato ESC, Khaydarov DR, Smith CCR, Kane NC, Safran RJ. 2017. Effects of assortative mate choice on the genomic and morphological structure of a hybrid zone between two bird subspecies. *Molecular Ecology* **26:** 6430–6444.

Servin B, Zhu B, Sayre B, Bian C, Sweeney D, Zhang G, Tosser-Klopp G, Kijas J, Womack J, Liang J, Chen J, Wang J, Li J, Wang J, Xie M, Xiao N, Cockett N, Zeng P, Moore R, Zhao R, Pan S, Yang S, Zhao S, Faraut T, Wang W, Chen W, Zhang W, Nie W, Wang W, Du X, Liu X, Xu X, Dong Y, Zhang Y, Hou Y, Shen Y, Jiang Y, Li Y. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology* **31:** 135.

Shaw KL, Mullen SP. 2011. Genes versus phenotypes in the study of speciation. *Genetica* **139:** 649–661.

Sichova J, Volenikova A, Dinca V, Nguyen P, Vila R, Sahara K, Marec F. 2015. Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evolutionary Biology* **15:** 89.

Sinclair-Waters M, Bradbury IR, Morris CJ, Lien S, Kent MP, Bentzen P. 2018. Ancient chromosomal rearrangement associated with local adaptation of a post-glacially colonized population of Atlantic Cod in the northwest Atlantic. *Molecular Ecology* **27:** 339–351.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, Griffith SC, McVean G, Przeworski M. 2015. Stable recombination hotspots in birds. *Science* **350:** 928–932.

Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. 2016. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes & Development* **30:** 266–280.

Smukowski CS, Noor MA. 2011. Recombination rate variation in closely related species. *Heredity* **107:** 496–508.

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS. 2010. What we still don't know about polyploidy. *Taxon* **59:** 1387–1403.

Soltis DE, Soltis PS, Tate JA. 2004. Advances in the study of polyploidy since plant speciation. *New Phytologist* **161:** 173–191.

Sousa V, Hey J. 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews. Genetics* **14:** 404–414.

Stephan W. 2010. Genetic hitchhiking versus background selection: the controversy and its implications. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365:** 1245–1253.

Stump AD, Fitzpatrick MC, Lobo NF, Traoré S, Sagnon NF, Costantini C, Collins FH, Besansky NJ. 2005. Centromere-proximal differentiation and speciation in Anopheles gambiae. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 15930–15935.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Research* **13:** 382–390.

Taylor JS, Van de Peer Y, Meyer A. 2001. Genome duplication, divergent resolution and speciation. *Trends in Genetics* **17:** 299–301.

Terhorst J, Song YS. 2015. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 7677–7682.

Thomae AW, Schade GO, Padeken J, Borath M, Vetter I, Kremmer E, Heun P, Imhof A. 2013. A pair of centromeric proteins mediates reproductive isolation in Drosophila species. *Developmental Cell* **27:** 412–424.

Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282:** 1501–1504.

Toews DP, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology* **26:** 2313–2318.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28:** 511–515.

Trier CN, Hermansen JS, Sætre GP, Bailey RI. 2014. Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. *PLoS Genetics* **10:** e1004075.

Turelli M, Barton NH, Coyne JA. 2001. Theory and speciation. *Trends in Ecology & Evolution* **16:** 330–343.

Turner LM, Harr B. 2014. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife* **3: doi:** 10.7554/eLife.02504.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* **3:** 1572–1578.

van Schooten B, Jiggins CD, Briscoe AD, Papa R. 2016. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. *BMC Genomics* **17:** 254.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27:** 737–746.

Venkatesh B. 2003. Evolution and diversity of fish genomes. *Current Opinion in Genetics & Development* **13:** 588–592.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16:** 381–390.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **367:** 451–460.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JB. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications* **7:** 13195.

Volff JN. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* **94:** 280–294.

Wadsworth CB, Li X, Dopman EB. 2015. A recombination suppressor contributes to ecological speciation in *Ostrinia* moths. *Heredity* **114:** 593–600.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics* **10:** 57–63.

Wertheim B, Beukeboom LW, van de Zande L. 2013. Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. *Cytogenetic and Genome Research* **140:** 256–269.

White MJD. 1978. *Modes of speciation*. San Francisco: W. H. Freeman.

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308:** 107–111.

Wiuf C, Hein J. 1999. Recombination as a point process along sequences. *Theoretical Population Biology* **55:** 248–259.

Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews. Genetics* **18:** 87–100.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 13875–13879.

Worley KC. 2017. A golden goat genome. *Nature Genetics* **49:** 485–486.

Wu C-I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* **14:** 851–865.

Wu CI, Ting CT. 2004. Genes and speciation. *Nature Reviews. Genetics* **5:** 114–122.

Wu M, Kostyun JL, Hahn MW, Moyle L. 2017. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *bioRxiv*: 201376.

Yasukochi Y, Ohno M, Shibata F, Jouraku A, Nakano R, Ishikawa Y, Sahara K. 2016. A FISH-based chromosome map for the European corn borer yields insights into ancient chromosomal fusions in the silkworm. *Heredity* **116:** 75–83.

Yeo S, Coombe L, Chu J, Warren RL, Birol I. 2018. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* **34:** 725–731.

Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A, Fujiyama A, Kitano J. 2014. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genetics* **10:** e1004223.

Zelazowski MJ, Cole F. 2016. X marks the spot: PRDM9 rescues hybrid sterility by finding hidden treasure in the genome. *Nature Structural & Molecular Biology* **23:** 267–269.

Zhang L, Martin A, Perry MW, van der Burg KR, Matsuoka Y, Monteiro A, Reed RD. 2017a. Genetic basis of melanin pigmentation in butterfly wings. *Genetics* **205:** 1537–1550.

Zhang L, Mazo-Vargas A, Reed RD. 2017b. Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proceedings of the National Academy of Sciences of the United States of America* **114:** 10707–10712.

Zhang L, Reed RD. 2016. Genome editing in butterflies reveals that *spalt* promotes and Distal-less represses eyespot colour patterns. *Nature Communications* **7:** 11769.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44:** 821–824.

Zuellig MP, Sweigart AL. 2018. Gene duplicates cause hybrid lethality between sympatric species of *Mimulus*. *PLoS Genetics* **14:** e1007130.