GENOMICS OF HYBRIDIZATION

A genomic perspective on hybridization and speciation

BRET A. PAYSEUR* and LOREN H. RIESEBERG†‡

*Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI 53706, USA, †Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, ‡Department of Biology, Indiana University, Bloomington, IN 47405, USA

Abstract

Hybridization among diverging lineages is common in nature. Genomic data provide a special opportunity to characterize the history of hybridization and the genetic basis of speciation. We review existing methods and empirical studies to identify recent advances in the genomics of hybridization, as well as issues that need to be addressed. Notable progress has been made in the development of methods for detecting hybridization and inferring individual ancestries. However, few approaches reconstruct the magnitude and timing of gene flow, estimate the fitness of hybrids or incorporate knowledge of recombination rate. Empirical studies indicate that the genomic consequences of hybridization are complex, including a highly heterogeneous landscape of differentiation. Inferred characteristics of hybridization differ substantially among species groups. Loci showing unusual patterns – which may contribute to reproductive barriers – are usually scattered throughout the genome, with potential enrichment in sex chromosomes and regions of reduced recombination. We caution against the growing trend of interpreting genomic variation in summary statistics across genomes as evidence of differential gene flow. We argue that converting genomic patterns into useful inferences about hybridization will ultimately require models and methods that directly incorporate key ingredients of speciation, including the dynamic nature of gene flow, selection acting in hybrid populations and recombination rate variation.

Keywords: evolutionary genomics, gene flow, hybridization, introgression, reproductive barriers, speciation

Received 12 November 2015; revision received 18 January 2016; accepted 25 January 2016

Introduction

The scope of hybridization plays a special role in understanding how new species are born. The capacity of lineages to hybridize is often used as a basis for species identification (Dobzhansky 1937; Mayr 1963). More generally, the dynamics of gene flow between diverging populations describe key elements of the speciation process.

Hybridization leaves detectable footprints in genomes, raising the prospect that it can be characterized from genomic patterns of variation. In particular, genomic data provide avenues to answer two basic questions about speciation: (i) what is the history of gene flow between nascent species and (ii) what genetic

Correspondence: Bret A. Payseur, Fax: +1 608 262 2976; E-mail: bret.payseur@wisc.edu

barriers maintain their integrity? Although these questions were addressed before the genomic era, the ability to rapidly and affordably survey genomic variation among individuals and populations has ushered in new hope of answering them (Sousa & Hey 2013; Seehausen *et al.* 2014). Because of recombination and independent assortment during meiosis, diversity patterns at the very large number of loci sampled by genomic data sets can be viewed (ideally) as replicated outcomes of hybridization history, enabling increasingly accurate reconstruction of that history. At the same time, inspection of the genomic distribution of variation can pinpoint specific regions with unusual patterns that might contribute to reproductive isolation or adaptive introgression.

Genomic studies of hybridization generally follow a recipe with two steps. First, genomewide patterns of variation are catalogued. If each of the sampled

populations is suspected or determined to belong to one diverging lineage (*e.g.* species) or another, differentiation between these populations is quantified. If the sampled populations likely reflect recent hybridization (*e.g.* in hybrid zones), other measures of admixture – such as geographic clines in allele frequency across transects – might be used. The second step is to compare observed genomic distributions to those expected under one or more models to draw inferences about hybridization and speciation.

Genomewide portraits of differentiation and admixture are shaped by a suite of evolutionary processes. Parameters of particular interest from the speciation perspective include the form, frequency, magnitude and timing of hybridization, as well as the fitness of hybrids. But several other factors complicate interpretations. Natural selection acting before or after hybridization affects differentiation at targeted mutations and at nearby neutral variants (Wright 1969; Charlesworth *et al.* 1997; Nordborg 1997), thereby inflating heterogeneity in differentiation among loci, even in the absence of gene flow (Noor & Bennett 2009; Nachman & Payseur 2012; Cruickshank & Hahn 2014; Burri *et al.* 2015). Recombination rate – a parameter that varies both across the genome and among species (Smukowski & Noor 2011) – determines the genomic scale over which patterns of differentiation and admixture vary. Shared variation among populations may reflect unsorted ancestral polymorphism (incomplete lineage sorting) rather than hybridization (Clark 1997; Wakeley & Hey 1997).

In this paper, we synthesize emerging trends in the genomics of hybridization between diverging lineages. After reviewing analytical methods that can be used to characterize hybridization, we summarize findings from empirical studies conducted on the genomic scale. Based on this survey, we prioritize future research to maximize insights into the process of speciation.

Methods for characterizing hybridization in the context of speciation

Translating genomic data sets into useful inferences about hybridization and speciation is a formidable task. Recognizing the increasing practicality of collecting high-quality genomic data, we focus here on analytical prospects and challenges. We refer the reader to recent reviews for guidance on important issues related to sequencing, genotyping and variant calling (Nielsen *et al.* 2011; Altmann *et al.* 2012; Arnold *et al.* 2013; Buerkle & Gompert 2013; Mardis 2013; Narum *et al.* 2013; Schlötterer *et al.* 2014; Olson *et al.* 2015).

We surveyed the literature for analytical methods that have either been applied in genomic studies of hybridization between diverging lineages or have the potential to be applied in such studies. Both methods developed specifically to understand speciation and methods developed for other purposes were included.

A wide range of methods is available (Table 1). The collection of approaches samples major descriptors of genetic variation: levels of diversity and divergence, site frequency spectra, haplotype variation and phylogenetic relationships. Methods vary in their characterization of hybridization. Defining gene flow as the movement of alleles between populations, different methods provide access to contrasting levels of gene flow. For example, coalescent approaches usually assume low rates, whereas analyses of clines consider rates high enough to interfere with selection. Some strategies aim to simply detect gene flow, whereas others strive to reconstruct its magnitude and/or timing. Depending on the approach, estimated rates of gene flow are genomewide averages or locus-specific values. Hybridization is modelled as an instantaneous event or as a continuous phenomenon. For several parametric approaches, there exists a trade-off in computational cost between increasing the number of individuals and increasing the number of loci sampled. For example, methods that fit the isolation with migration model using Markov chain Monte Carlo handle many individuals but are slow with a large number of loci (Hey & Nielsen 2004, 2007; Hey 2010; Sethuraman & Hey 2015), whereas analytical likelihood methods applied to this model can use entire genomes but are restricted to a few individuals (Lohse *et al.* 2011, 2015, 2016). Principal components analysis (PCA) (a nonparametric approach) can be efficiently applied to whole-genome data sets from large numbers of individuals (Patterson *et al.* 2006; Price *et al.* 2006), but it provides no direct insights into the causes of resulting patterns.

Methods differ in the frequency of usage by researchers studying the genomics of speciation (Table 2). The most popular strategy interprets heterogeneity in summary statistics of between-population differentiation – usually F_{ST} – as evidence of differential gene flow across the genome. Another commonly used class of approaches searches for signs of gene flow in genomic patterns of shared-derived variants between populations or species. In genomic studies of hybridization, the 'ABBA-BABA test' (Green *et al.* 2010; Durand *et al.* 2011) is a widespread application of such phylogenetic thinking. Interest in this test and its derivatives (Green *et al.* 2010; Durand *et al.* 2011; Patterson *et al.* 2012; Martin *et al.* 2015b; Pease & Hahn 2015), referred to as 'D-statistics', was stimulated by its original application (Green *et al.* 2010), which revealed that ancestors of modern humans hybridized with Neanderthals. Several approaches commonly used to infer demographic

Table 1 Genomic methods for detecting and characterizing hybridization

Method	Characterization of hybridization						Focal pattern of variation	References
	Rate of gene flow	Timing of gene flow	Variable gene flow across genome	Variable gene flow across time	Individual ancestry proportions (genomewide)	Individual, locus-specific ancestries		
Geographic clines	Yes	No	Yes	No	No	No	Geographic gradient of allele frequencies across populations (hybrid zone)	Barton & Hewitt (1985); Szymura & Barton (1986); Porter et al. (1997)
Genomic clines	No	No	Yes	No	No	No	Individual genotypes (hybrid zone)	Szymura & Barton (1986); Rieseberg et al. (1999); Gompert & Buerkle (2009, 2011); Fitzpatrick (2013)
Structure/Structurama/Frappe/Admixture/FastStructure	No	No	No	No	Yes	No	Individual genotypes	Pritchard et al. (2000); Falush et al. (2003); Huelsenbeck & Andolfatto (2007); Hubisz et al. (2009); Raj et al. (2014); Tang et al. (2005); Alexander et al. (2009)
Principal components analysis (PCA)	No	No	No	No	No	No	Individual genotypes	Patterson et al. (2006); Price et al. (2006)
HapMix/Recombination via ancestry switch probabilities (RASPberry)	No	No	Yes	No	Yes	Yes	Individual haplotypes	Price et al. (2009); Wegmann et al. (2011)
Ancestry tract lengths/shared haplotype lengths	Yes	Yes	No	Yes	No	No	Individual haplotypes	Pool & Nielsen (2009); Patterson et al. (2012); Gravel (2012); Harris & Nielsen (2013)
Markov Chain Monte Carlo – fit to isolation with migration model	Yes	No	No	No	No	No	Numbers of unique, shared and divergent polymorphisms	Hey & Nielsen (2004); Becquet & Przeworski (2007); Hey & Nielsen (2007); Wang & Hey (2010); Sousa et al. (2013); Sethuraman & Hey (2015)
Coalescent hidden Markov model (CoalHMM) – fit to isolation with migration model	Yes	Yes	No	No	No	No	Individual haplotypes	Mailund et al. (2012)
Blockwise likelihood – fit to isolation with migration model	Yes	Yes	No	No	No	No	Numbers of unique, shared and divergent polymorphisms	Lohse et al. (2011); Lohse & Frantz (2014); Lohse et al. (2016)
Diffusion approximations for demographic inference (dadi) – fit to isolation with migration model	Yes	No	No	No	No	No	Joint site frequency spectrum	Gutenkunst et al. (2009)

Table 1 *Continued*

Method	Characterization of hybridization						Focal pattern of variation	References
	Rate of gene flow	Timing of gene flow	Variable gene flow across genome	Variable gene flow across time	Individual ancestry proportions (genomewide)	Individual, locus-specific ancestries		
Sequentially Markov conditional sampling distribution – fit to isolation with migration model	Yes	Yes	No	Yes	No	No	Individual haplotypes	Steinrücken et al. (2016)
fineSTRUCTURE	No	No	No	No	Yes	No	Individual haplotypes	Lawson et al. (2012)
TreeMix	No	No	No	No	No	No	Joint site frequency spectrum	Pickrell & Pritchard (2012)
Divergence time heterogeneity	No	No	Yes	No	No	No	Numbers of unique, shared and divergent polymorphisms	Yang (2010); Garrigan et al. (2012)
Approximate Bayesian computation	Yes	Yes	Yes	Yes	No	No	Array of summary statistics	Beaumont et al. (2002); Robinson et al. (2014)
Genomic outliers (summary statistics, e.g. F_{ST})	No	No	Yes	No	No	No	Summaries of population differentiation	Beaumont & Balding (2004)
ABBA-BABA/D-statistics	No	No	No	No	No	No	Pattern of shared-derived changes	Green et al. (2010) Durand et al. (2011) Patterson et al. (2012) Martin et al. (2015b) Pease & Hahn (2015)
Phylogenetic discordance	No	No	No	No	No	No	Pattern of shared-derived changes	Ané et al. (2007) Meng & Kubatko (2009)
Phylogenetic networks	No	No	No	No	No	No	Pattern of shared-derived changes	Yu et al. (2012) Yu et al. (2013) Yu et al. (2014a) Liu et al. (2014a)

Table 2 Inferences about hybridization and speciation from genomic data

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
Butterflies	81 individuals across *Heliconius melpomene*, *H. timareta*, *H. elevatus*	Whole-genome sequencing RAD sequencing	Higher gene flow (2–5% of genome) between sympatric than between allopatric *H. melpomene* vs. *H. timareta*; elevated gene flow in B/D colour region	D-statistics	No	Exchange of protective colour pattern genes between species; *H. elevatus* formed by hybridization	Heliconius Genome Consortium (2012)
Butterflies	31 individuals (total) from 5 *H. melpomene* populations, 1 *Heliconius cydno* population, 1 *H. timareta* population	Whole-genome sequencing	Pervasive phylogenetic discordance, structured by geography; fraction of genome introgressed higher in older comparisons; reduced differentiation between sympatric pairs; higher ratio of sympatric/allopatric F_{ST} on Z chromosome	Phylogenetic discordance D-statistics F_{ST} outliers	No	Large-scale gene flow between *H. melpomene* and *H. cydno/timareta* lineage throughout much of their divergence history	Martin *et al.* (2013)
Butterflies	51 individuals (total) from *H. melpomene*, *H. cydno*, *H. timareta*, *Heliconius heurippa*	RAD sequencing	Admixture up to 13% in some individuals; F_{ST} highly correlated across loci among sympatric pairs but not parapatric pairs; some differentiated regions associated with geography rather than taxon	Structure F_{ST} outliers	No	Gene flow between sympatric *H. melpomene* and *H. cydno/timareta*; possible hybrid origin for *H. heurippa*	Nadeau *et al.* (2013)
Butterflies	116 *Lycaides idas*, 76 *Lycaides melissa*, 186 putative hybrids	RAD sequencing	Among hybrids, outlier F_{ST} loci enriched for *L. idas* ancestry; differentiation correlated with genomic cline parameters	F_{ST} outliers Genomic clines	No	Adaptive divergence and genetic drift in allopatry contribute to reproductive isolation; complex genetic basis to reproductive isolation	Gompert *et al.* (2012)
Chipmunks	4 *Tamias ruficaudus* 4 *Tamias amoenus*	Exome sequencing	Gene genealogies show species monophyly	Phylogenetic discordance D-statistics	No	Little to no role for hybridization between these	Good *et al.* (2015)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/ introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
						sympatric species, despite earlier introgression of mtDNA	
Cichlids	10 individuals from each of *Pundamilia nyererei, P. pundamilia, P. sp.* 'pink and fin', *Mbipia mbipi, M. Lutea*	RAD sequencing	When SNPs are partitioned by levels of differentiation, species relationships change (e.g. high-differentiation SNPs group by colour)	F_{ST} outliers Structure	No	At least two intergeneric hybridization events; hybridization generates new species and contributes to adaptive radiation; selection plays an important role in speciation	Keller *et al.* (2013)
Cichlids	24 species in four sympatric crater lake radiations and 29 outgroups	RAD sequencing	Weak support for monophyly of sympatric radiations; species within a crater vary in shared ancestry with outgroups	Structure PCA Treemix	No	Complex colonization histories within crater lakes; gene flow with neighbouring rivers after initial colonization of lakes; questions status as compelling model for sympatric speciation	Martin *et al.* (2015a)
Cut-throat trout	94 individuals from 5 populations of *Onchorhynchus clarkia lewisi*, westslope cut-throat trout	RAD sequencing	1–10% population admixture; several superinvasive alleles (at much higher frequencies than genomic background)	Frequency of introgressed alleles at diagnostic loci	No	Candidate invasiveness genes	Hohenlohe *et al.* 2013;
Drosophila	3 *Drosophila pseudoobscura* (1 each for 3 subspecies), 1 *Drosophila persimilis*, 1 *D. miranda*	Partial genome sequencing	Elevated sequence divergence near inversions; diversity correlated with divergence correlated between *D. pseudoobscura* and *D. persimilis* but not with divergence between *D. pseudoobscura* and *D. miranda*	D_{XY}	No	Inversions crucial for persistence of species in the face of recent gene flow; secondary contact between *D. persimilis* and *D. pseuduoobscura* homogenized collinear regions; evidence for	Kulathinal *et al.* (2009)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
						unidirectional gene flow less than 200 kya, but not on the X chromosome	
Drosophila	1 *Drosophila simulans*, 1 *D. mauritiana*, 1 *D. sechellia*	Whole-genome sequencing	Evidence of gene flow between *D. simulans* and island endemics at up to 4.6% (autosomes) and 2.2% (X chromosome) of genome	Divergence time heterogeneity	No	'Islands of introgression, in genomes otherwise consistent with isolation'; introgression reduced on the X chromosome	Garrigan *et al.* (2012)
Drosophila	4 *Drosophila mojavensis* (2 from each of 2 populations); 1 *D. arizonae*	Whole-genome sequencing	Overall divergence similar in sympatric and allopatric comparisons, with elevated divergence in both in inverted regions	Blockwise likelihood	Yes	Postdivergence gene flow, ceasing at 270 kya; same timing in collinear and inverted regions, but reduced gene flow in inversions; inversions originated close to time of initial split	Lohse *et al.* (2015)
Flycatchers	10 *Ficedula albicollis* (collared), 10 *F. hypoleuca* (pied)	Whole-genome sequencing	50 clusters of elevated F_{ST} (overrepresented at chromosome ends); reduced nucleotide diversity (within species), and elevated LD – but reduced Dxy; higher overall divergence on Z	F_{ST} outliers	No	Species divergence driven by directional selection at many loci	Ellegren *et al.* (2012)
Flycatchers	10 *Ficedula albicollis* (collared), 10 *F. hypoleuca* (pied)	Whole-genome sequencing	Substantial fraction of shared polymorphisms (22%) with fewer fixed differences (3%)	Approximate Bayesian Computation	Yes	Recent origin of flycatcher species, suggesting rapid evolution of postzygotic isolation; allopatric speciation followed by secondary, recent gene flow; 0.16–0.36	Nadachowska-Brzyska *et al.* (2013)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/ introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
Flycatchers	79 *F. albiocollis*, 79 *F. hypoleuca*, 20 *F. speculigera*, 20 *F. semitorquata*	Whole-genome sequencing	High heterogeneity in differentiation across the genome; concordant genomic regions of differentiation across species pairs	F_{ST} outliers Dxy outliers D-statistics	No	Linked selection within species rather than variation in gene flow among species explains islands of differentiation; migrants per generation *F. hypoleuca* into *F. albiocollis*, no gene flow in the other direction	Burri *et al.* (2015)
Flycatchers	2 *Zimmerius viridiflavus*, 3 *Z. chrysops* 5 individuals from mosaic population 1 *Z. acer* 1 *Z. gracilipes*	RAD sequencing	1.1% of the alleles in the mosaic population represent introgression from *Z. chrysops*; introgressed regions enriched in functions involving cell projection and plasma membranes	D-statistics	No	Introgression may account for variation in plumage coloration	Rheindt *et al.* (2014)
Frogs	45 *Craugastor augusti* 1 *C. tarahumaraensis* 1 *C. uno*	RAD sequencing	High lineage differentiation between evolutionarily significant units – numerous private alleles and high F_{ST}	D-statistics	No	Two episodes of introgression occurred in the same direction between lineages but at different times; higher variation in introgressed lineages than in independently evolving lineages	Streicher *et al.* (2014)
Hominins	*Homo neanderthalensis*, 5 *H. sapiens*	Whole-genome sequencing	East Asians share more SNPs than sub-Saharan Africans with *H. neanderthalensis*	D-statistics	No	Gene flow from *H. neanderthalensis* into non-African *H. sapiens* occurred before split of Eurasian groups	Green *et al.* (2010)
Hominins				Haplotypes	Yes		

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
	Homo neanderthalensis, 1000 *H. sapiens*	Whole-genome sequencing	Rate of exponential decay of linkage disequilibrium among SNPs in *H. sapiens* genomic regions introgressed from *H. neanderthalensis* fits a model of recent gene flow			Last gene flow from *H. neanderthalensis* into *H. sapiens* occurred 37–86 kya	Sankararaman *et al.* (2012)
Hominins	*Homo neanderthalensis*, 'Denisovans', 42 *H. sapiens*	Whole-genome sequencing	Higher levels of *H. neanderthalensis* ancestry in East Asians than in Europeans	D-statistics	No	At least two separate episodes of hybridization between *H. neanderthalensis* and *H. sapiens*, including one after split between East Asians and Europeans	Wall *et al.* (2013)
Hominins	*Homo neanderthalensis*, 'Denisovans', *H. sapiens*	Whole-genome sequencing	1.5–2.1% of non-African *H. sapiens* genomes derived from *H. neanderthalensis*	D-statistics	No	3–5 cases of interbreeding among extinct hominin populations, including an unknown species; complex history of hybridization leaving signatures in small fractions of the genome	Prüfer *et al.* (2014)
Hominins	*Homo neanderthalensis*, 1004 *H. sapiens*	Whole-genome sequencing	*H. sapiens* genomic regions of reduced *H. neanderthalensis* ancestry include X chromosome and genic regions	Conditional random field	No	*H. neanderthalensis* X chromosome contributed to hybrid sterility with *H. sapiens*	Sankararaman *et al.* (2014)
Hominins	*Homo neanderthalensis*, 3 *H. sapiens*	Whole-genome sequencing	3.4–7.9% *H. neanderthalensis* admixture in Eurasia	Blockwise likelihood	Yes	*H. neanderthalensis* – *H. sapiens* admixture higher than other studies	Lohse & Frantz (2014)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/ introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
Horses	6 individuals from each of *Equus quagga quagga, E. quagga boehmi, E. greyvi E. zebra hartmannae, E. kiang, E. hemionus onager, E. africanus somaliensis, E. asinus asinus, E. caballus*	Whole-genome sequencing	Four episodes of gene flow detected: one during the early divergence of the *Equus* in North America, and three between contemporary *Equus* lineages in the Old World; D-statistics associated with chromosomal changes	D-statistics CoalHMM	No	Gene flow occurred despite extensive variation in chromosome number, although rearrangements inhibited gene flow; speciation in the face of gene flow	Jónsson *et al.* (2014)
House mice	1301 mice from two transects of hybrid zone between *Mus musculus domesticus* and *M. m. musculus*	SNP array	Asymmetric patterns of linkage disequilibrium across transects	Geographic clines	No	Movement of hybrid zone over time	Wang *et al.* (2011)
House mice	679 mice from two transects of hybrid zone between *Mus musculus domesticus* and *M. m. musculus*	SNP array	Signatures of epistatic selection at many loci; disproportionate number of selected loci on the X chromosome; some selected loci overlap in two transects, although most do not	Genomic clines	No	Central and distal portions of X chromosome contribute to reproductive isolation; complex genetic basis for reproductive isolation, arising in allopatry	Janoušek *et al.* (2012)
House mice	Mice from three transects of hybrid zone between *Mus musculus domesticus* and *M. m. musculus*	SNP array	Introgression positively correlated with recombination rate and negatively correlated with differentiation	Genomic clines	No	Some reproductive barriers shared across hybrid zone; genome properties (including recombination	Janoušek *et al.* (2015)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
House mice	8 wild-derived inbred lines of *Mus musculus castaneus*, 7 *M. m. domesticus*, 8 *M. m. musculus*	Transcriptome sequencing	X more highly differentiated than autosomes; higher differentiation in low-recombination regions in comparison between *M. m. castaneus* and *M. m. musculus*; some high-differentiation regions overlap with previously identified loci involved in hybrid male sterility	F_{ST} outliers; Dxy outliers; Numbers of unique, shared, and divergent polymorphisms	No	rate) influence speciation; Differences in demographic history among subspecies may account for differences in divergence patterns among subspecies pairs; X chromosome contributes disproportionately to reproductive isolation	Phifer-Rixey *et al.* (2014)
House mice	2 *Mus spretus* 20 *M. musculus domesticus*	SNP array	Phylogenetic signals of introgression from *M. spretus* on a few chromosomes of each *M. m. domesticus* individual	Phylogenetic networks	Yes	At least three hybridization events, including one 50 years ago connected to warfarin resistance	Liu *et al.* (2015)
Lousewart	1 *Pedicularis przewalskii*, 1 *P. cyathophylla*, 1 *P. superba*, 2 *P. cyathophylloides*, 2 *P. thamnophila*, 5 *P. rex*	RAD sequencing	8.7–27% of *P. thamnophila* genome estimated to be derived from *P. rex*	D-statistics	No	Introgression among nearly all taxa in the *P. rex – thamnophila* clade	Eaton & Ree (2013)
Manakins	48 *Manacus candei*, 52 *M. vitellinus*, 104 putative hybrids	Genotype by sequencing	Gene flow highly variable across genome; rapid decay of genomic autocorrelations; F_{ST} correlated with gene flow estimated in hybrids, but genomic	F_{ST} outliers PCA Genomic clines	No	Barriers to gene flow caused by many alleles with small effects; some evidence that selection drives reproductive isolation, but isolation	Parchman *et al.* (2013)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
Mosquitoes	*Anopheles gambiae, A. arabiensis, A. merus*	RAD sequencing	Extensive shared polymorphism; higher divergence in X chromosome inversions; outliers not strongly colocalized	Structure PCA Dxy outliers F_{ST} outliers	No	X chromosome inversions involved in reproductive isolation is not simply a consequence of adaptive change	O'Loughlin *et al.* (2014)
Mosquitoes	*Anopheles coluzzii, A. gambiae, A. arabiensis, A. quadriannulatus, A. melas, A. merus*	Whole-genome sequencing	Strong phylogenetic discordance between X chromosome and autosomes	Phylogenetic concordance	No	Phylogenetic relationships on the X chromosome represent the species history, revealing pervasive gene flow on the autosomes	Fontaine *et al.* (2014)
Oaks	4 *Quercus virginiana*, 4 *Q. geminate*, 4 *Q. minima*, 3 *Q. sagraeana*, 5 *Q. oleoides*, 3 *Q. brandegei*, 4 *Q. fusiformis*, 7 outgroup samples	RAD sequencing	Substantial heterogeneity in the presence of admixture between species; six of seven species show evidence of admixture with one or more congeners; significant admixture limited to samples in close proximity	Structure Treemix D-statistics dadi	Yes	Oak species boundaries and ranges have been evolutionarily stable	Eaton *et al.* (2015)
Pigs	*S. scrofa, S. verrucosus, S. cebifrons*	Whole-genome sequencing	23% of genome admixed	Blockwise likelihood	Yes	Asymmetric hybridization occurring long before arrival of humans	Frantz *et al.* (2014)
Rabbits	6 *Oryctolagus cuniculus cuniculus*, 6 *Oryctolagus cuniculus algirus*, 1 *Lepus timidus*	Targeted capture of intron sequence; transcriptome sequencing	Substantial fraction of shared polymorphisms (31.2%), with fewer fixed differences (0.3%); higher differentiation on the X chromosome and near centromeres; islands of differentiation highly variable in size, but	F_{ST} outliers	No	Many fixed differences probably result from positive selection; complex genetic basis to reproductive isolation	Carneiro *et al.* (2014)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
Poplars	7 genotypes each from *Populus alba, P. tremula*	RAD sequencing	Divergence estimates strongly autocorrelated along chromosomes; many autocorrelations involve low divergence blocks, including incipient sex chromosome; usually contain small numbers of genes	F_{ST} outliers; Genomic autocorrelation	No	*Populus* species have porous genomes and exhibit high levels of introgression across incipient sex chromosome	Stölting *et al.* (2013)
Poplars	498 *Populus trichocarpa*, 10 *P. balsamifera*	SNP array	Admixture limited to drainages where the species' ranges overlap	Admixture; PCA	No	Hybridization contributes to clinal variation in allele frequencies observed in *P. trichocarpa*.	Geraldes *et al.* (2014)
Sunflowers	40 *H. annuus*, 25 *H. petiolaris*, 28 *H. argophyllus*, 14 *H. debilis*	Transcriptome sequencing	Higher differentiation and more fixed differences in nonhybridizing species pairs than in hybridizing species pairs, despite closer phylogenetic relationship of the former; islands of genetic differentiation small and found in areas of low recombination	F_{ST} outliers; Genomic autocorrelation	No	Interspecific gene flow has less impact than local genome architecture on the clustering of divergent loci; complex genetic basis to reproductive isolation	Renaut *et al.* (2013)
Sunflowers	40 *H. annuus*, 28 *H. argophyllus*	Transcriptome sequencing	Elevated F_{ST} and D within rearrangements and adjacent regions (within 5 cM of chromosomal break points)	F_{ST} outliers	No	Chromosomal translocations and inversions reduce interspecific gene flow following secondary contact	Barb *et al.* (2014)
Sunflowers	100 genotypes representing 20 subpopulations of *Helianthus petiolaris* (strong intrinsic and	RAD sequencing	Split between dune and nondune ecotypes estimated between 440–10 274 ya; Ongoing asymmetric introgression; $N_e m$	F_{ST} outliers; Structure; MCMC fit to isolation with migration	Yes	Possible genomic signature of reinforcement: two F_{ST} outlier regions show greater divergence in	Andrew *et al.* (2013); Andrew & Rieseberg (2013)

Table 2 *Continued*

Species group	Sampling	Genomic data	Genomic patterns of differentiation/ introgression	Analytical method for detecting hybridization	Estimate rate or timing of gene flow?	Inferences about hybridization and speciation from authors	References
	extrinsic isolation between dune and nondune ecotypes); 20 individuals from other species or subspecies		(Nondune to Dune) = 2.06; $N_e m$ (Dune to Nondune) = 5.53.			comparisons between adjacent dune edge and nondune populations than between more distant core dune and nondune populations	
Swordtail fish	60 *Xiphophorus birchmanni*, 60 *X. malinche*; 313 hybrids from two hybrid zones	Genotype by sequencing	Many unlinked pairs of loci show linkage disequilibrium in hybrid populations	Linkage disequilibrium among unlinked loci	Yes	A large number of hybrid incompatibilities contribute to reproductive isolation	Schumer *et al.* (2014b)
Warblers (ring species)	95 individuals from 22 populations (total) from *Phylloscopus trochiloides viridanus*, *P. t. nitidus*, *P. t. ludlowi*, *P. t. trochiloides*, *P. t. obscuratus*, *P. t. plumbeitarsus*	Genotype by sequencing	Two northern forms are distinct genetic clusters connected by ring of genetically intergraded forms	PCA Structure	No	Allopatric divergence played a major role in the formation of ring species; recent hybridization at each area of secondary contact	Alcaide *et al.* (2014)

history within species (especially humans) – including diffusion-based analysis of the site frequency spectrum (Gutenkunst *et al.* 2009) and a host of methods based on individual haplotypes (Pool & Nielsen 2009; Gravel 2012; Mailund *et al.* 2012; Harris & Nielsen 2013) – have seen limited application in the genomic characterization of gene flow between diverging lineages.

Overall, available methods provide useful tools for reconstructing hybridization. The diversity of focal genomic patterns (Table 1) suggests that collectively, methods can access gene flow occurring over a range of timescales. For example, phylogenetic patterns among species can reveal older hybridization events, whereas the distribution of shared haplotype lengths is shaped by recent gene flow. Although no methods incorporate all aspects of genomic variation, many approaches adopt powerful likelihood or Bayesian frameworks that avoid drastic summaries of the data. For example, several strategies are available that analyse the complete (unsummarized), site frequency spectrum across multiple populations.

The ancestries of individuals provide special insights into hybridization history. Importantly, diverse strategies now exist for probabilistically assigning individual ancestry, assuming that appropriate reference populations can be surveyed (Gompert & Buerkle 2013). Since the Bayesian clustering approach Structure revolutionized the inference of ancestry proportions (Pritchard *et al.* 2000), methods have been developed that use the same likelihood (genotype data conditional on ancestry) but different algorithms to achieve substantial increases in computational speed (Tang *et al.* 2005; Alexander *et al.* 2009; Raj *et al.* 2014). With other methods, including HapMix (Price *et al.* 2009) and RASPberry (Wegmann *et al.* 2011), the *locus-specific* histories of individuals can be reconstructed, enabling changes in ancestry to be detected over short chromosomal distances.

Computational and statistical advances have significantly expanded the hybridization scenarios that may be considered. With approximate Bayesian computation, genomic data can be fit to any model that can be rapidly simulated, allowing (in principle) arbitrarily complex histories of gene flow to be compared. Existing methods also enable testing of specific hypotheses about reproductive barriers. For example, the genomic clines framework (Gompert & Buerkle 2009, 2011) statistically evaluates evidence for selection against gene flow at specific genomic locations through comparison to the genomewide hybrid index.

Despite their potential, available methods for characterizing hybridization suffer from several challenges when applied to genomic studies of speciation. First, options for measuring gene flow often exclude scenarios of interest. Among the subset of methods that estimate gene flow, most assume it happens continuously at an invariant rate. Few approaches reliably reconstruct the timing of hybridization. For example, popular methods based on the isolation with migration framework may produce wide confidence intervals on the timing of gene flow and falsely infer the presence of gene flow under some conditions (Sousa *et al.* 2011; Strasburg & Rieseberg 2011; Cruickshank & Hahn 2014; Hey *et al.* 2015). These restrictions do not match the reality of the speciation process, in which the opportunity for hybridization varies over time due to accumulating reproductive isolation, fluctuating geographic range sizes and other factors. Indeed, temporal changes in gene flow are key ingredients for distinguishing major models of speciation. For example, sympatric and parapatric speciation features gene flow during the earliest stages of divergence, whereas speciation by reinforcement is thought to be triggered by secondary contact and gene flow after a period of divergence in allopatry.

A second challenge with existing methods is the limited treatment of natural selection. Approaches that estimate the rate of gene flow usually assume that population history and neutral mutation account for genomic patterns. Therefore, selection has the potential to bias geneflow estimates. This problem could be severe for lineages with a history of reproductive isolation, especially if many loci contribute to isolation. A more fundamental methodological issue in the context of speciation is the assumption that one rate of gene flow characterizes the entire genome (Sousa *et al.* 2013). Because selection acting on hybrids targets specific loci, rates of introgression could vary significantly across the genome (Barton & Hewitt 1985; Harrison 1990; Payseur 2010); indeed, documenting this variation is a major goal of genomic studies of speciation. Alternatively, genome scan approaches search for locus-specific distortions in summary statistics, but usually leave unestimated rates of gene flow and associated selection coefficients. These shortcomings are not surprising as most methods were designed to analyse population structure within species rather than gene flow among diverging lineages. Although strategies specifically developed to analyse the genomic consequences of mixing in hybrid zones (*e.g.* geographic and genomic clines) enable the detection of selection and differential introgression in hybrids, they downplay the effects of demographic history and do not estimate locus-specific rates of gene flow.

The rate of meiotic recombination is another key parameter to consider in genomic studies of hybridization (Sousa & Hey 2013). Differential gene flow across a chromosome requires recombination. Recombination rate determines the chromosomal scale over which

selection targeting reproductive isolation mutations reduces gene flow at neighbouring loci (Bazykin 1969; Barton 1979; Bengtsson 1985; Barton & Bengtsson 1986; Gavrilets 1997), the focal signature of genomic scan methods. Furthermore, several models of speciation propose that loci involved in reproductive isolation will preferentially accumulate in rearranged (Noor *et al.* 2001; Rieseberg 2001; Navarro & Barton 2003) or collinear (Butlin 2005) chromosomal regions with little recombination, allowing divergence to continue in the face of hybridization. Unfortunately, genomic methods for studying speciation ignore the reality that recombination rates vary across chromosomes on broad and fine scales (McVean *et al.* 2004; Kong *et al.* 2010; Comeron *et al.* 2012; Liu *et al.* 2014b). In addition, some methods assume that recombination is absent over short distances (*i.e.* within loci). As hybridization studies transition to using whole-genome sequences, another practical challenge emerges. How should the genome be partitioned into loci before geneflow analyses are conducted? For example, the typical genomic scan strategy of comparing windows of the same physical size (*e.g.* 10 kb) implicitly assumes that the rate of recombination is invariant. Most methods do not yet account for the expectation that switches in ancestry associated with hybridization will occur over finer scales in genomic regions with higher recombination rates.

Several approaches assume free recombination among polymorphisms, even when they are closely linked. A more general challenge arises from the treatment of unlinked loci as independent, a feature of all available methods. Loci on different chromosomes can behave nonindependently in scenarios of interest. For example, recent hybridization events involving only a few individuals can generate introgression at many loci (Goodman *et al.* 1999).

The role of geography in speciation has long been and continues to be a source of controversy (Mayr 1963; Coyne & Orr 2004; Seehausen *et al.* 2014). Knowledge of the geographic locations of past hybridization events is vital for understanding the connection between gene flow and speciation. With few exceptions (Yang *et al.* 2014), genomic methods for analysing hybridization do not consider geographic information nor do they infer the geographic context of gene flow.

Outcomes of hybridization may vary depending on whether the hybridizing populations diverged in the presence of gene flow or came into contact secondarily after a period of allopatry. But reliable methods for distinguishing between these scenarios are not readily available. Theoretical expectations for patterns of genome divergence following secondary contact or divergence with gene flow appear to be similar under equilibrium conditions (*e.g.* Barton 1979; Barton &

Hewitt 1985; Charlesworth *et al.* 1997; Feder & Nosil 2010). However, given the extent of temporal variation in geographic ranges due to major climate oscillations, nonequilibrium conditions are likely to be common and could impact inferences about hybridization, as well as its consequences. For example, hybrid zones in the Arctic and some temperate regions are thought to have formed repeatedly following interglacial range expansions of hybridizing lineages, in some instances leading to introgression or hybrid speciation (Hewitt 2011).

In summary, researchers may choose from a variety of analytical approaches to draw inferences about hybridization from genomic data. Nevertheless, existing strategies are missing some components that are important in the context of speciation, suggesting that investment in further method development is warranted.

Empirical studies of hybridization and speciation on the genomic scale

To identify emerging empirical patterns in the genomics of hybridization and speciation, we surveyed the literature. We focused on studies that: (i) articulated a specific interest in characterizing hybridization between diverging lineages (rather than simply scanning genomes for evidence of natural selection); (ii) measured variation across the genome at thousands to millions of loci; and (iii) examined species, subspecies or populations with independent evidence for reproductive isolation (rather than focusing on ecotypic differentiation). Both studies that sampled geographically separate populations and those that examined currently hybridizing populations were considered.

Inspection of the studies (listed in Table 2) reveals several themes. Genomic data confirm what is now conventional wisdom: hybridization between diverging lineages is common (Arnold 1997; Rieseberg 1997; Mallet 2005). Many instances of hybridization were previously unreported and unanticipated (*e.g.* genetic exchange between humans and Neanderthals), an indication of the potentially high power of genomic data to detect gene flow, even when it involves extinct populations. The history of hybridization among diverging lineages is sometimes complex, including multiple periods of gene flow as well as asymmetries in its direction. In some cases, genomic data raise the possibility that hybridization facilitated speciation (*e.g.* butterflies, cichlids). Nevertheless, the inference that most of the hybridizing lineages in Table 2 – including a ring species (Alcaide *et al.* 2014) – evolved in isolation for some period of time suggests that geographic separation contributes to reproductive isolation in many organismal lineages (Mayr 1963; Coyne & Orr 2004).

A consistent observation across surveyed taxa is rampant heterogeneity in patterns of differentiation and admixture within the genome, sometimes over short chromosomal distances. As discussed above, this patchwork reflects a combination of evolutionary factors, including incomplete lineage sorting, effects of selection before and after hybridization, and recombination rate variation, as well as gene flow and selection targeting hybrids. That heterogeneity occurs over a fine genomic scale suggests that some aspects of hybridization can only be captured by examining the entire genome. In those studies that sample hybrid zones, this heterogeneity raises the prospect that boundaries between nascent species are semipermeable, with a subset of loci maintaining reproductive barriers (Harrison 1986, 1990; Wu 2001). In those studies that sample geographically separate populations, this conclusion is more difficult to reach: the null model of a low rate of gene flow that is constant across the genome could still produce strong heterogeneity if only a few regions introgress. From a phylogenetic perspective, the genomes of recently diverged species are complex mosaics of alternating histories, highlighting the difficulty of reconstructing species relationships, even with genomic data. Indeed, extensive hybridization challenges basic assumptions of phylogenetic methods, suggesting that some species histories are better represented as phylogenetic networks (Yu *et al.* 2014).

Genomic regions that display high differentiation (in allopatric populations) or narrow clines (in hybrid zones) could contain genetic changes that confer reproductive isolation. Under this assumption, the studies in Table 2 collectively raise a few genomic themes that could characterize reproductive barriers between species. First, the number of loci that appear to maintain differentiation in the face of gene flow is usually high, suggesting a complex genetic basis to reproductive isolation. An alternative explanation for this pattern is that many loci are falsely labelled as 'genomic outliers' because significance thresholds are based on null models that ignore or mischaracterize demographic history.

A second observation is that although outlying loci are scattered throughout the genome, there is evidence for a bias towards regions of low recombination. In several species, this pattern takes the form of higher differentiation near centromeres, which are known or suspected to recombine less. In others, the relationship between outlier location and recombination rate is more obvious. Cline width and recombination rate are positively correlated across the genome in a hybrid zone between two subspecies of house mice (Janoušek *et al.* 2015). Selection to remove long genomic blocks from one species in hybrids was postulated to explain a negative correlation between absolute divergence and recombination rate in monkey flowers (Brandvain *et al.* 2014). Rearranged chromosomal regions exhibit higher differentiation than collinear regions in several species pairs. Perhaps these results indicate a role for suppressed recombination in the evolution of reproductive isolation. Alternatively, they could simply reflect increased power to detect isolation loci using linked markers in low-recombination regions. In some cases, the pattern appears to be caused by selection at linked sites within lineages rather than barriers to gene flow (Cruickshank & Hahn 2014; Burri *et al.* 2015). Regardless, these results reveal the importance of considering variation in local recombination rate when interpreting genomic patterns of differentiation and admixture.

A third trend among surveyed species is the overrepresentation of highly differentiated loci on the sex chromosomes relative to the autosomes, seen in both X-Y and Z-W systems. This pattern agrees with results from controlled crosses in the laboratory, which often associate isolation phenotypes – especially hybrid sterility and hybrid inviability – with the sex chromosomes (Coyne & Orr 2004). Whether this genomic tendency indicates that the sex chromosomes harbour a higher density of loci involved in reproductive isolation (Masly & Presgraves 2007), loci with stronger phenotypic effects on isolation, or both, remains to be seen. Interpretations of this pattern must also recognize the complications inherent in setting neutral expectations for the sex chromosomes, which can experience different rates of genetic drift and migration than the autosomes.

The collection of empirical studies (Table 2) reveals considerable interest in using genomic data to understand the role of natural selection in speciation. The degree to which selection can increase divergence in the face of gene flow remains controversial and difficult to determine (Kirkpatrick & Barton 2006; Via 2012; Andrew & Rieseberg 2013; Flaxman *et al.* 2013, 2014; Yeaman 2013; Cruickshank & Hahn 2014; Feder *et al.* 2014; Seehausen *et al.* 2014). In contrast, the idea that reproductive isolation is a by-product of local adaptation in allopatry is widely accepted, but genetic tests of this hypothesis remain rare. Those genomic studies that jointly consider currently allopatric populations *and* currently hybridizing populations provide a special opportunity to evaluate this influential prediction. If the genetic changes that block gene flow are initially driven to fixation by positive selection, genomic regions with restricted introgression in hybrid zones could display higher differentiation between nascent species. Genomic cline width and differentiation are negatively correlated in lycaenid butterflies and in manakins, suggesting that adaptive evolution contributes to reproductive isolation (Gompert *et al.* 2012; Parchman *et al.* 2013). Nevertheless, the weakness of these correlations leaves open the

possibility that genetic drift plays an important role in the establishment of reproductive barriers.

Despite the trends described above, the genomic landscape of divergence appears to differ substantially among species pairs. Estimated hybridization rates vary widely. Whereas some species show evidence for extensive gene flow with a few genomic regions maintaining divergence (*e.g.* rabbits, mosquitoes), others show widespread differentiation with limited pockets of introgression (*e.g. Drosophila simulans* vs. island endemics). The genomic scale of heterogeneity also varies. Although these comparisons are complicated by contrasting study designs, some observed disparities probably reflect biological differences among species pairs in the timing and mode of speciation.

This empirical survey also reveals challenges that face genomic studies of hybridization. From the perspective of speciation, a primary motivation for analysing genomic data is its potential for characterizing the magnitude and timing of gene flow among diverging lineages. Only a handful of studies (Table 2) reported estimates of these important quantities. The most widely used measure of locus-specific differentiation was F_{ST}, a summary statistic that compares variation between populations to variation within populations (Wright 1951). Although F_{ST} enables comparisons of loci that differ in within-population diversity, it suffers from interpretive challenges when used to measure gene flow across the genome. Reduced gene flow, which maintains between-population divergence, is confounded with selection at linked sites, which decreases within-population diversity (Charlesworth 1998; Noor & Bennett 2009; Nachman & Payseur 2012; Cruickshank & Hahn 2014). Furthermore, most studies identified unusually differentiated loci as those with values that simply fell in the tail of the genomic distribution, a strategy complicated by the fact that completely neutral genomic distributions also have tails. Other studies used neutral simulations to construct distributions of summary statistics expected in the absence of selection. This approach ignores the likely possibility that selection against gene flow (and other forms of selection) occurs at many genomic locations in hybrids.

Furthermore, Table 2 reveals biases in sampling strategy. The majority of studies examined populations that are currently allopatric, with no *a priori* evidence of hybridization. Compared to hybrid zones, allopatric populations are usually easier to find and investigators may be more likely to sample them for purposes other than studying gene flow. Nevertheless, connecting genomic patterns with hybridization is much simpler in populations that are currently hybridizing in visible ways (*e.g.* hybrid zones). Other biases are taxonomic. Animal species – especially vertebrates – were the most popular subjects. It is too early to identify differences between plants and animals from genomic studies of hybridization. But the presence of heteromorphic sex chromosomes seems to alter the genetic architecture of reproductive barriers and to increase the speed with which they accumulate, and most plants lack heteromorphic sex chromosomes (Rieseberg 2001; Coyne & Orr 2004; Levin 2012; Phillips & Edmands 2012; Lima 2014). Likewise, it has long been known that the frequency of hybridization tends to be higher in species with external fertilization, such as plants and fishes, than in taxa with internal fertilization (Stebbins 1959). These and other related observations should motivate genomic studies of hybridization in organismal groups that differ in the genetics of sex determination and/or other key features of reproductive biology.

Recommendations and future directions

Our overview of methods and empirical studies suggests fruitful directions for research on the genomics of hybridization and speciation. First, model-based estimation of the magnitude and timing of hybridization should be prioritized. Reconstructing the history of gene flow between diverging lineages is a necessary step towards understanding the causes and consequences of speciation, and it may provide key insights. For example, the alternative scenarios of speciation with gene flow and allopatric speciation followed by hybridization during secondary contact could be distinguished by quantifying the dynamics of gene flow. Accurate estimates of the magnitude and timing of hybridization are also needed to interpret the results from scans that seek to identify genomic regions with restricted gene flow, which are growing in popularity. Because the expected genomic distribution of differentiation depends on the level and timing of hybridization, the determination that certain loci experienced reduced gene flow ultimately requires a demographic model including these parameters. At present, some authors seem to assume that evidence of hybridization elsewhere in the genome is sufficient to demonstrate that a differentiated genomic region is resistant to gene flow. But a wide spectrum of differentiation among loci is expected, even in the absence of reproductive barriers.

Second, natural selection should be directly incorporated into methods for characterizing hybridization between diverging lineages. Even partial reproductive isolation can generate strong selection against hybrids that shapes variation throughout the genome (Barton 1983; Barton & Bengtsson 1986). Accounting for the effects of selection is therefore a necessary step towards accurately reconstructing the history of hybridization. Furthermore, model-based characterization of selection

could provide insights into genetic mechanisms of speciation. By estimating the contributions of individual loci to hybrid fitness, the genetic architecture of reproductive barriers in nature, as well as the identity of genomic regions and candidate genes that underlie adaptive introgression or heterosis, could be revealed. It may be possible to compare the fit of observed genomic patterns to models featuring different architectures. For example, selection against heterozygotes at single loci is predicted to impede gene flow at linked variants more strongly than epistatic selection against two-locus, hetero-specific genotypes ('hybrid incompatibilities') (Gavrilets 2004; Gompert & Buerkle 2011; Lindtke & Buerkle 2015). The genomic clines approach (Gompert & Buerkle 2009, 2011) provides a good foundation for incorporating selection into genomic studies of hybridization in the context of hybrid zones. For currently allopatric populations that hybridized in the past, model-based assignment of loci to one of a few classes with different rates of gene flow (due to selection) (Sousa *et al.* 2013) may offer a promising start towards the construction of more realistic methods.

Building geographic information into genomic methods for analysing hybridization is another worthwhile goal. This strategy would recognize that the prior probabilities associated with past hybridization are functions of current geographic ranges. It would also enable tests of speciation models that depend on geography. For example, reinforcement is expected to generate 'inverse clines' in which greater divergence for the trait/loci under reinforcing selection is expected between populations at the centre than at the ends of the cline (Caisse & Antonovics 1978; Andrew & Rieseberg 2013). Finally, incorporating geographic information could help identify the selective forces responsible for initial divergence and the formation of reproductive isolation, an important goal (Shaw & Mullen 2011). In this regard, clinal models of adaptation with gene flow (Endler 1977; Novembre & Di Rienzo 2009) could provide useful guidelines.

Even without these methodological advances, authors of empirical studies should be explicit about the models they are considering and the quantitative patterns they expect to find. In general, it is important to bear in mind that hybrid genomes are outcomes of a complex mixture of processes. As a result, intuition about expected patterns will sometimes mislead. Studies that embrace the complications inherent in the speciation process – rather than automatically attributing observed genomic patterns to differential gene flow generated by hybridization – should be prioritized.

Our survey suggests other guidelines for genomic studies of hybridization and speciation. Reproductive isolation should be directly measured whenever possible. Genomic comparisons alone are unlikely to identify genes responsible for reproductive barriers. Understanding how genetic variants restrict gene flow will ultimately require characterizing their functional effects on specific isolation phenotypes. More broadly, the interpretation of genomic patterns will be enriched by comparison to knowledge of reproductive barriers. For example, the over-representation of sex chromosome loci in genetic dissections of isolation traits provides a potential explanation for higher differentiation of sex-linked loci in genomic studies of hybridization. Likewise, a recent critique of the evidence for homoploid hybrid speciation showed that while admixture was well-documented in most putative hybrid species, only a handful of studies connected hybridization to reproductive barrier formation – a key criterion for hybrid speciation (Schumer *et al.* 2014a).

Genomic studies of hybridization should stimulate the development of new organismal models for speciation. Most information about mechanisms of speciation still comes from a small subset of species, but Table 2 suggests that the dynamics and outcomes of hybridization could differ significantly across taxa. Examining those systems featuring multiple species pairs that collectively sample a range of divergence times will enable the effects of hybridization to be evaluated at different stages of speciation. Furthermore, species that currently hybridize offer the most direct insights into the genomic consequences of hybridization, indicating that these species should be prioritized in speciation studies.

Finally, the limits to inferring hybridization from genomic data should be examined and embraced. The space of potential hybridization scenarios is expansive and which subset of historical information is recorded in genomes is highly stochastic. Determining what we can realistically hope to learn about speciation from genomic data – rather than treating it as a panacea – is an important next step.

Acknowledgements

References

Alcaide M, Scordato ESC, Price TD, Irwin DE (2014) Genomic divergence in a ring species complex. *Nature*, **511**, 83–85.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Altmann A, Weber P, Bader D *et al.* (2012) A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Human Genetics*, **131**, 1541–1554.

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482.

Andrew RL, Kane NC, Baute GJ, Grassa CJ, Rieseberg LH (2013) Recent nonhybrid origin of sunflower ecotypes in a novel habitat. *Molecular Ecology*, **22**, 799–813.

Ane C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution*, **24**, 412–426.

Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, Oxford.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RAD-seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Barb JG, Bowers JE, Renaut S *et al.* (2014) Chromosomal evolution and patterns of introgression in Helianthus. *Genetics*, **197**, 969–979.

Barton NH (1979) Gene flow past a cline. *Heredity*, **43**, 333–339.

Barton NH (1983) Multi-locus clines. *Evolution*, **37**, 454–471.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity*, **57**, 357–376.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Bazykin AD (1969) Hypothetical mechanism of speciation. *Evolution*, **23**, 685–687.

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

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, **17**, 1505–1519.

Bengtsson BO (1985) The flow of genes through a genetic barrier. In: *Evolution: Essays in Honor of John Maynard Smith* (eds Greenwood PJ, Slatkin M), pp. 31–42. Cambridge University Press, Cambridge.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression between Mimulus nasutus and Mimulus guttatus. *PLoS Genetics*, **10**, e1004410.

Buerkle C, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, **22**, 3028–3035.

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

Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.

Caisse M, Antonovics J (1978) Evolution in closely adjacent plant populations. IX. Evolution of reproductive isolation in clinal populations. *Heredity*, **40**, 371–384.

Carneiro M, Albert FW, Afonso S *et al.* (2014) The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genetics*, **10**, e1003519.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155–174.

Clark AG (1997) Neutral behavior of shared polymorphism. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7730–7734.

Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1002905.

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

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.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York, New York.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**, 2587–2601.

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

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

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, Wacholder AC *et al.* (2014) Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *The Journal of Heredity*, **105**(Suppl), 810–820.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Flaxman SM, Feder JL, Nosil P (2013) Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, **67**, 2577–2591.

Flaxman SM, Wacholder AC, Feder JL, Nosil P (2014) Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, **23**, 4074–4088.

Fontaine MC, Pease JB, Steele A *et al.* (2014) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**, 1258524.

Frantz LAF, Madsen O, Megens H-J, Groenen MAM, Lohse K (2014) Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian *Sus* species during the Plio-Pleistocene climatic fluctuations. *Molecular Ecology*, **23**, 5566–5574.

Garrigan D, Kingan SB, Geneva AJ *et al.* (2012) Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Research*, **22**, 1499–1511.

Gavrilets S (1997) Hybrid zones with Dobzhansky-type epistatic selection. *Evolution*, **51**, 1027–1035.

Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, New Jersey.

Geraldes A, Farzaneh N, Grassa CJ *et al.* (2014) Landscape genomics of *Populus trichocarpa*: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, **68**, 3260–3280.

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, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

Gompert Z, Buerkle CA (2013) Analyses of genetic ancestry enable key insights for molecular ecology. *Molecular Ecology*, **22**, 5278–5294.

Gompert Z, Lucas LK, Nice CC *et al.* (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Good JM, Vanderpool D, Keeble S, Bi K (2015) Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, **69**, 1961–1972.

Goodman SJ, Barton NH, Swanson G, Abernathy K, Pemberton JM (1999) Introgression through rare hybridization: a genetic study of a hybrid zone between red and sika deer (genus Cervus) in Argyll, Scotland. *Genetics*, **152**, 355–371.

Gravel S (2012) Population genetics models of local ancestry. *Genetics*, **191**, 607–619.

Green RE, Krause J, Briggs AW *et al.* (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710–722.

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.

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics*, **9**, e1003521.

Harrison RG (1986) Pattern and process in a narrow hybrid zone. *Heredity*, **56**, 337–349.

Harrison RG (1990) Hybrid zones: windows on evolutionary processes. *Oxford Surveys in Evolutionary Biology*, **7**, 69–128.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–638.

Hey J (2010) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution*, **27**, 921–933.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and D-persimilis. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hey J, Chung Y, Sethuraman A (2015) On the occurrence of false positives in tests of migration under an isolation-with-migration model. *Molecular Ecology*, **24**, 5078–5083.

Hohenlohe PA, Day MD, Amish SJ *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. *Genetics*, **175**, 1787–1802.

Janoušek V, Wang L, Luzynski K *et al.* (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.

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.

Jónsson H, Schubert M, Seguin-Orlando A *et al.* (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 18655–18660.

Keller I, Wagner CE, Greuter L *et al.* (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**, 2848–2863.

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

Kong A, Thorleifsson G, Gudbjartsson DF *et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, **467**, 1099–1103.

Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genetics*, **5**, e1000550.

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genetics*, **8**, e1002453.

Levin DA (2012) The long wait for hybrid sterility in flowering plants. *New Phytologist*, **196**, 666–670.

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

Lindtke D, Buerkle CA (2015) The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, **69**, 1987–2004.

Liu KJ, Dai J, Truong K *et al.* (2014a) An HMM-based comparative genomic framework for detecting introgression in eukaryotes (T Lengauer, Ed,). *PLoS Computational Biology*, **10**, e1003649.

Liu EY, Morgan AP, Chesler EJ *et al.* (2014b) High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. *Genetics*, **197**, 91–106.

Liu KJ, Steinberg E, Yozzo A *et al.* (2015) Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 196–201.

Lohse K, Frantz LA (2014) Neandertal admixture in Eurasis confirmed by maximum-likelihood analysis of three genomes. *Genetics*, **196**, 1241–1251.

Lohse K, Harrison RJ, Barton NH (2011) A general method for calculating likelihoods under the coalescent process. *Genetics*, **189**, 977–987.

Lohse K, Clarke M, Ritchie MG, Etges WJ (2015) Genome-wide tests for introgression between cactophilic Drosophila implicate a role of inversions during speciation. *Evolution*, **69**, 1178–1190.

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

Mailund T, Halager AE, Westergaard M *et al.* (2012) A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genetics*, **8**, e1003125.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Mardis ER (2013) Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, **6**, 287–303.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, **23**, 1817–1828.

Martin CH, Cutler JS, Friel JP *et al.* (2015a) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, **69**, 1406–1422.

Martin SH, Davey JW, Jiggins CD (2015b) 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**, e243.

Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge, Massachusetts.

McVean GA, Myers SR, Hunt S *et al.* (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, **304**, 581–584.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology*, **75**, 35–45.

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.

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

Nadeau NJ, Martin SH, Kozak KM *et al.* (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, **22**, 814–826.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, **57**, 447–459.

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

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

Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12084–12088.

Nordborg M (1997) Structured coalescent processes on different time scales. *Genetics*, **146**, 1501–1514.

Novembre J, Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. *Nature Reviews. Genetics*, **10**, 745–755.

O'Loughlin SM, Magesa S, Mbogo C *et al.* (2014) Genomic analyses of three malaria vectors reveals extensive shared polymorphism but contrasting population histories. *Molecular Biology and Evolution*, **31**, 889–902.

Olson ND, Lund SP, Colman RE *et al.* (2015) Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Frontiers in Genetics*, **6**, 235.

Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, **22**, 3304–3317.

Patterson NJ, Price AL, Reich D (2006) Population structure and eigenanalysis. *Plos Genetics*, **2**, e190.

Patterson N, Moorjani P, Luo Y *et al.* (2012) Ancient admixture in human history. *Genetics*, **192**, 1065–1093.

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

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

Phifer-Rixey M, Bomhoff M, Nachman MW (2014) Genome-wide patterns of differentiation among house mouse subspecies. *Genetics*, **198**, 283–297.

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

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967.

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, **181**, 711–719.

Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM (1997) The Pontia daplidice-edusa hybrid zone in northwestern Italy. *Evolution*, **51**, 1561–1573.

Price AL, Patterson NJ, Plenge RM *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

Price AL, Tandon A, Patterson N *et al.* (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genetics*, **5**, e1000519.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prüfer K, Racimo F, Patterson N *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Renaut S, Grassa CJ, Yeaman S et al. (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, **4**, 1827.

Rheindt FE, Fujita MK, Wilton PR, Edwards SV (2014) Introgression and phenotypic assimilation in Zimmerius Flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, **63**, 134–152.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**, 359–389.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, **16**, 351–358.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Robinson JD, Bunnefeld L, Hearn J, Stone GN, Hickerson MJ (2014) ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology*, **23**, 4458–4471.

Sankararaman S, Patterson N, Li H, Pääbo S, Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genetics*, **8**, e1002947.

Sankararaman S, Mallick S, Dannemann M et al. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, **507**, 354–357.

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.

Schumer M, Rosenthal GG, Andolfatto P (2014a) How common is homoploid hybrid speciation? *Evolution*, **68**, 1553–1560.

Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P (2014b) High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife*, **3**, e02535.

Seehausen O, Butlin RK, Keller I et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Sethuraman A, Hey J (2015) IMa2p - parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Molecular Ecology Resources*, **16**, 206–215.

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

Smukowski CS, Noor MAF (2011) Recombination rate variation in closely related species. *Heredity*, **107**, 496–508.

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

Sousa VC, Grelaud A, Hey J (2011) On the nonidentifiability of migration time estimates in isolation with migration models. *Molecular Ecology*, **20**, 3956–3962.

Sousa VC, Carneiro M, Ferrand N, Hey J (2013) Identifying loci under selection against gene flow in isolation-with-migration models. *Genetics*, **194**, 211–233.

Stebbins GL (1959) The role of hybridization in evolution. *Proceedings of the American Philosophical Society*, **103**, 231–251.

Steinrücken M, Kamm JA, Song YS (2016) Inference of complex population histories using whole genome sequences from multiple populations. *Bioarxiv*. doi:10.1101/026591.

Stölting KN, Nipper R, Lindtke D et al. (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, **22**, 842–855.

Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. *Molecular Ecology*, **20**, 2353–2366.

Streicher JW, Devitt TJ, Goldberg CS et al. (2014) Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, **23**, 3273–3291.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution*, **40**, 1141–1159.

Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology*, **28**, 289–301.

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.

Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics*, **145**, 847–855.

Wall JD, Yang MA, Jay F et al. (2013) Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics*, **194**, 199–209.

Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics*, **184**, 363–379.

Wang L, Luzynski K, Pool JE et al. (2011) Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Molecular Ecology*, **20**, 2985–3000.

Wegmann D, Kessner DE, Veeramah KR et al. (2011) Recombination rates in admixed individuals identified by ancestry-based inference. *Nature Genetics*, **43**, 847–853.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright S (1969) *Evolution and the Genetics of Populations, vol. 2, The Theory of Gene Frequencies*. University of Chicago Press, Chicago, Illinois.

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

Yang Z (2010) A likelihood ratio test of speciation with gene flow using genomic sequence data. *Genome Biology and Evolution*, **2**, 200–211.

Yang W-Y, Platt A, Chiang CW-K et al. (2014) Spatial localization of recent ancestors for admixed individuals. *G3*, **4**, 2505–2518.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E1743–E1751.

Yu Y, Degnan JH, Nakhleh L (2012) The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics*, **8**, e1002660.

Yu Y, Barnett RM, Nakhleh L (2013) Parsimonious inference of hybridization in the presence of incomplete lineage sorting. *Systematic Biology*, **62**, 738–751.

Yu Y, Dong J, Liu KJ, Nakhleh L (2014) Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 16448–16453.