PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2012) **367**, 409–421
doi:10.1098/rstb.2011.0249

Review

Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice

Michael W. Nachman[1,*] and Bret A. Payseur[2]

[1]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA*
[2]*Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA*

Recently diverged taxa may continue to exchange genes. A number of models of speciation with gene flow propose that the frequency of gene exchange will be lower in genomic regions of low recombination and that these regions will therefore be more differentiated. However, several population-genetic models that focus on selection at linked sites also predict greater differentiation in regions of low recombination simply as a result of faster sorting of ancestral alleles even in the absence of gene flow. Moreover, identifying the actual amount of gene flow from patterns of genetic variation is tricky, because both ancestral polymorphism and migration lead to shared variation between recently diverged taxa. New analytic methods have been developed to help distinguish ancestral polymorphism from migration. Along with a growing number of datasets of multi-locus DNA sequence variation, these methods have spawned a renewed interest in speciation models with gene flow. Here, we review both speciation and population-genetic models that make explicit predictions about how the rate of recombination influences patterns of genetic variation within and between species. We then compare those predictions with empirical data of DNA sequence variation in rabbits and mice. We find strong support for the prediction that genomic regions experiencing low levels of recombination are more differentiated. In most cases, reduced gene flow appears to contribute to the pattern, although disentangling the relative contribution of reduced gene flow and selection at linked sites remains a challenge. We suggest fruitful areas of research that might help distinguish between different models.

Keywords: genetic hitchhiking; background selection; gene flow; *Mus musculus*; *Oryctolgaus cuniculus*

1. INTRODUCTION

The geographical context and genetic details of how new species arise have been major topics of evolutionary research. Based on geographical patterns of phenotypic variation in birds, Mayr [1] argued that geographical isolation is a common first step in the origin of species. Owing largely to the influence of both Mayr [1] and Dobzhansky [2], allopatric models of speciation have dominated ideas about how new species arise for much of the last 70 years. Over the last dozen years, however, there has been a renewed interest in the possibility that speciation may occur in the presence of gene flow.

This renewed interest in 'speciation with gene flow' comes from at least three places. First, recent theoretical models have demonstrated that selection can drive speciation in the face of gene flow [3–5]. Second, there are now quite a few detailed empirical studies where geographical isolation seems an unlikely explanation for observed patterns, including work on true fruit flies [6,7], cichlids [8], palms [9], sticklebacks [10] and many others (reviewed by Bolnick & Fitzpatrick [11]). In these examples, natural or sexual selection is thought to have driven changes that led to the origin of new species in the absence of geographical isolation. Third, new statistical tools enable us to measure gene flow between recently diverging populations that are not at migration–drift equilibrium and to distinguish this gene flow from incomplete lineage sorting as causes of shared variation [12–17]. Some of these methods have been used quite widely. Although a majority of 49 studies that used an isolation-with-migration (IM) model yielded no evidence of gene exchange, many detected significant gene flow [18]. Finding evidence of gene flow does not mean that speciation occurred in sympatry. There are a range of other possibilities, including parapatric models, allopatric models followed by secondary contact and a variety of more complicated scenarios.

Some taxa may have experienced multiple periods of contact and isolation as ranges expanded and contracted. Periods of contact can offer opportunities for gene exchange, provided that reproductive isolation is not complete. Determining the timing of gene flow is a challenging task [19]. Finding evidence of gene

* Author for correspondence (nachman@u.arizona.edu).

One contribution of 13 to a Theme Issue 'Patterns and processes of genomic divergence during speciation'.

flow also does not mean that speciation was necessarily driven by selection. Mutations that arise and fix (either through drift or selection) in allopatric populations can lead to negative epistatic interactions upon secondary contact, as suggested by Bateson [20], Dobzhansky [2] and Muller [21] (BDM incompatibilities). Such BDM incompatibilities are expected to restrict gene flow, but if some hybrids survive and reproduce, then some genomic regions may be permeable to gene flow, even as other regions differentiate [22].

A key challenge now is to understand the conditions under which gene flow may occur as young taxa diverge. Can we make explicit predictions about which genomic regions are likely to be important in the early stages of speciation? The rate of recombination is a central component of models used to interpret variation in levels of differentiation across the genome. Here, we review speciation models and population-genetic models that invoke variation in recombination rate. We outline specific predictions that follow from these models. We then review methods for evaluating these predictions. Finally, we assess these models in the light of empirical data from rabbits and mice.

2. RECOMBINATION RATE VARIATION AND MODELS OF DIFFERENTIATION

(a) *Speciation models with gene flow*

Ten years ago, several authors proposed the idea that chromosomal rearrangements that distinguish recently diverged taxa might facilitate the origin of reproductive isolation in the face of hybridization [23–28]. The rationale is that hybrids formed between taxa that differ by chromosomal rearrangements, such as inversions, will experience suppressed recombination in rearranged regions of the genome. Genes contributing to reproductive isolation may accumulate in such regions and may continue to diverge in the face of hybridization. These models are therefore fundamentally *genic* models of speciation (in contrast to earlier chromosomal models of speciation, which depend on the underdominance of the chromosomal rearrangements themselves; [29]). These models predict that the genomes of recently diverged species will be divided into high-differentiation and low-differentiation regions. For example, in crosses between *Drosophila persimilis* and *Drosophila pseudoobscura*, genes contributing to reproductive isolation map preferentially to the few inverted regions of the genome [24], and these regions also show greater genetic differentiation than collinear regions of the genome [30]. The models of Noor *et al.* [24], Rieseberg [25] and Navarro & Barton [23] differ in detail but share the key feature that chromosomal rearrangements suppress recombination when heterozygous.

Even between taxa with collinear genomes, genes contributing to isolation are expected to accumulate in regions of suppressed recombination [22,26] and consequently, gene flow is expected to be reduced in such regions. It is important to recognize the role of selection in this situation: gene flow is reduced because 'isolation alleles' are selectively removed when introduced into the sister species. Because these models are concerned with the origin of reproductive isolation between incipient species, we refer to them collectively as 'speciation models'.

(b) *Population-genetic models without gene flow*

In contrast to these speciation models, there are a number of models that seek to explain how selection at linked sites can affect levels and patterns of genetic variation within a single population. We refer to these as 'population-genetic models'. Selection in regions with reduced recombination can reduce levels of genetic variation within populations in two primary ways. First, a new beneficial mutation may quickly rise in frequency owing to positive selection, and result in the associated fixation of linked neutral sites (i.e. 'genetic hitchhiking'; [31]). Positive selection may also act on standing variation, recurrent mutations or alleles introduced by migration (i.e. 'soft sweeps'; [32–34]), and in this case, the reduction in levels of variation at linked sites is expected to be weaker. Second, purifying selection against new deleterious mutations may reduce levels of variation at linked sites through a process termed background selection [35]. Many studies have demonstrated that genomic regions with reduced recombination have lower levels of genetic variation [36–38], though distinguishing between the potential causes of these patterns has proved remarkably difficult [39].

Regardless of the cause of reduced variation in regions of low recombination within populations, this pattern will lead to increased estimates of differentiation between populations when statistics such as F_{ST} are used [40,41], even without gene flow. Because F_{ST} is based on a comparison of within-population diversity to between-population diversity, anything that reduces the former will inflate F_{ST} [41]. A negative correlation between F_{ST} and recombination rate has been observed in empirical studies ranging from a handful of loci [42–44] to the whole genome [45].

(c) *Predictions from models*

The earlier-mentioned models can be used to generate explicit predictions about patterns of genetic variation and differentiation in the early stages of divergence. Both the speciation models and the models of selection at linked sites predict that differentiation will be greater in regions of low recombination. However, differentiation can be measured in a variety of ways, and there are subtle differences that may provide a means of distinguishing between alternative models (table 1).

The key prediction from all three speciation models is that gene flow will be lower in genomic regions experiencing less recombination. Thus, fewer shared polymorphisms are expected in such regions, and levels of differentiation should be greater. Figure 1*a* shows gene genealogies in regions of low and high recombination, in the absence of genetic hitchhiking or background selection. Gene flow is assumed to occur in regions of high recombination but not in regions of low recombination because of linkage to genes involved in isolation. Under these conditions, nucleotide diversity (π) is expected to equal $4N_e\mu$ at equilibrium in the low-recombination region, while π will be larger in the high-recombination region. If the lineages have been separated for a short time, then ancestral polymorphisms may still be segregating (not shown) and thus populations may not be at

Table 1. Alternative predictions from models of speciation with gene flow and population-genetic models of selection at linked sites for patterns of variation within and between species in genomic regions of low- and high-recombination rates.

measure of genetic variation[a]	predictions of speciation models		predictions of models of selection at linked sites	
	low recomb	high recomb	low recomb	high recomb
π	lesser	greater	lesser	greater
D_a	greater	lesser	greater	lesser
F_{ST}	greater	lesser	greater	lesser
D_{xy}	greater	lesser	same or less	same or more

[a]Nucleotide diversity, π, is a measure of genetic variation within populations and provides an estimate of the neutral mutation parameter, $4N_e\mu$, where N_e is the effective population size and μ is the neutral mutation rate [46]. D_a, F_{ST}, and D_{xy} are different measures of differentiation between populations [46].

equilibrium. Nonetheless, gene flow is still expected to increase levels of nucleotide diversity. Without gene flow, net nucleotide divergence, D_a [46], provides an estimate of $2\mu t$, where μ is the neutral mutation rate and t is the time of separation of the lineages. With gene flow, D_a provides an underestimate of this quantity. Similarly, D_{xy} [46]—the average number of pairwise differences between alleles sampled from the two populations—provides an estimate of $2\mu t + 4N_e\mu$, where N_e is the ancestral population size, in cases without gene flow, but provides an underestimate of this quantity in cases with gene flow.

A key prediction from models of selection at linked sites is that positive or negative selection will influence patterns of genetic variation at linked sites more in regions of low recombination than in regions of high recombination. Figure 1*b* shows gene genealogies in regions of low and high recombination, in the absence of gene flow but in the presence of selection at linked sites due to genetic hitchhiking or background selection. In low-recombination regions, selection at linked sites is expected to reduce levels of nucleotide variation, whereas in regions of high recombination, such effects will be negligible. Most theoretical treatments of the effects of selection at linked sites on neutral differentiation have focused on F_{ST}. Using simulations, Charlesworth *et al.* [47] showed that background selection increases F_{ST} between a pair of populations connected by a low degree of gene flow. The effect is primarily caused by the reduction in diversity within populations. Hu & He [48] used two- and three-locus models to demonstrate that the elevation in F_{ST} generated by background selection also applies with an arbitrary number of populations (in the island model) and is closely tied to levels of linkage disequilibrium (LD) between neutral and deleterious alleles. Slatkin & Wiehe [49] showed that genetic hitchhiking causes transient increases in F_{ST} by distorting neutral allele frequencies. It should be noted that another contributing factor to the positive correlation between within-population diversity and recombination rate—the association of recombination with mutation [50]—predicts no correlation between F_{ST} and recombination rate.

The effects of selection at linked sites on other measures of differentiation have received less attention. As a consequence of reduced nucleotide diversity, D_a will be increased in regions of low recombination relative to regions of high recombination. Importantly, however, D_{xy} is expected to behave differently under models of selection at linked sites compared with models of speciation (table 1 and figure 1) [51]. Gene flow will reduce D_{xy} in regions of high recombination (figure 1*a*). Selection at linked sites will generally not affect D_{xy}, unless selection is also operating in the ancestral population, in which case the time to coalescence for alleles from the two daughter populations will be shorter in regions of low recombination, thus reducing D_{xy} in such regions. It remains to be seen whether measures of differentiation developed to reduce dependence on within-population diversity [52,53] show different signals under speciation with gene flow and selection at linked sites.

Table 2 provides additional tests that may help us to distinguish speciation models from models of selection at linked sites as explanations for increased differentiation in regions of low recombination. Several newer methods for detecting gene flow, described below, may be useful for identifying the expected increase in introgression in regions of high recombination under speciation models. In addition, some tests for selection may help identify the expected signatures of selection in low-recombination regions. It is important to bear in mind, however, that some traditional tests of selection, which can be quite powerful in other situations, may be uninformative in this context. For example, a lower ratio of polymorphism to divergence is expected in low-recombination regions under models of selection at linked sites as well as under models of reduced gene flow. Thus, rejection of the null model using the Hudson–Kreitman–Aguade test [54] does not help us to distinguish between these competing sets of models.

3. METHODS FOR MEASURING INTROGRESSION

Assuming that migration follows an island model and that migration–drift equilibrium has been reached, the number of migrants per generation (Nm) can be related directly to levels of differentiation from the expectation $F_{ST} = 1/(4N\text{m} + 1)$. However, recently diverged lineages are generally not expected to be at equilibrium [55]. Under these conditions, connecting

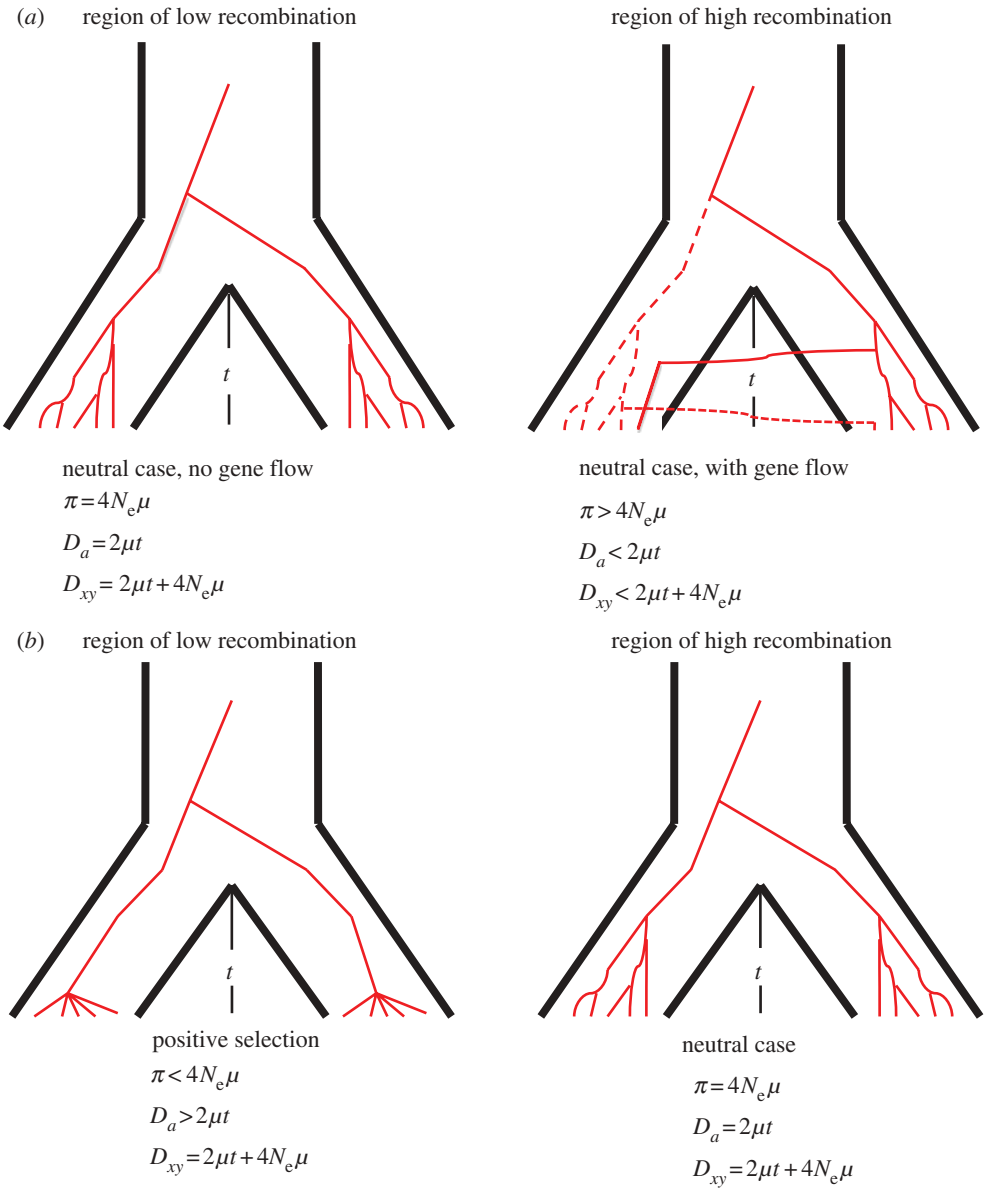

Figure 1. Comparison of expected patterns of variation in regions of low recombination and regions of high recombination under (*a*) models of speciation with gene flow and (*b*) under population-genetic models of selection at linked sites without gene flow. See also legend to table 1.

Table 2. Tests that may help distinguish between speciation and population-genetic models as explanations for greater differentiation in genomic regions with low recombination.

test	expected patterns under speciation models with gene flow	expected patterns under models of selection at linked sites
non-equilibrium estimates of gene flow	higher gene flow in regions of high recombination than in regions of low recombination	gene flow independent of recombination rate
haplotype-based estimates of gene flow	higher gene flow in regions of high recombination than in regions of low recombination	gene flow independent of recombination rate
tests of selection based on distribution of allele frequencies	deviations from a neutral frequency distribution associated with alleles introduced by migration from sister lineage and thus greater in regions of high recombination than in regions of low recombination	greater skew towards an excess of rare alleles in regions of low recombination than in regions of high recombination; applies if genetic hitchhiking events are sufficiently common and recent or if background selection occurs in small populations
tests of selection based on ratio of polymorphism to divergence	lower ratio of polymorphism/divergence in low-recombination regions than in high-recombination regions	lower ratio of polymorphism/divergence in low-recombination regions than in high-recombination regions

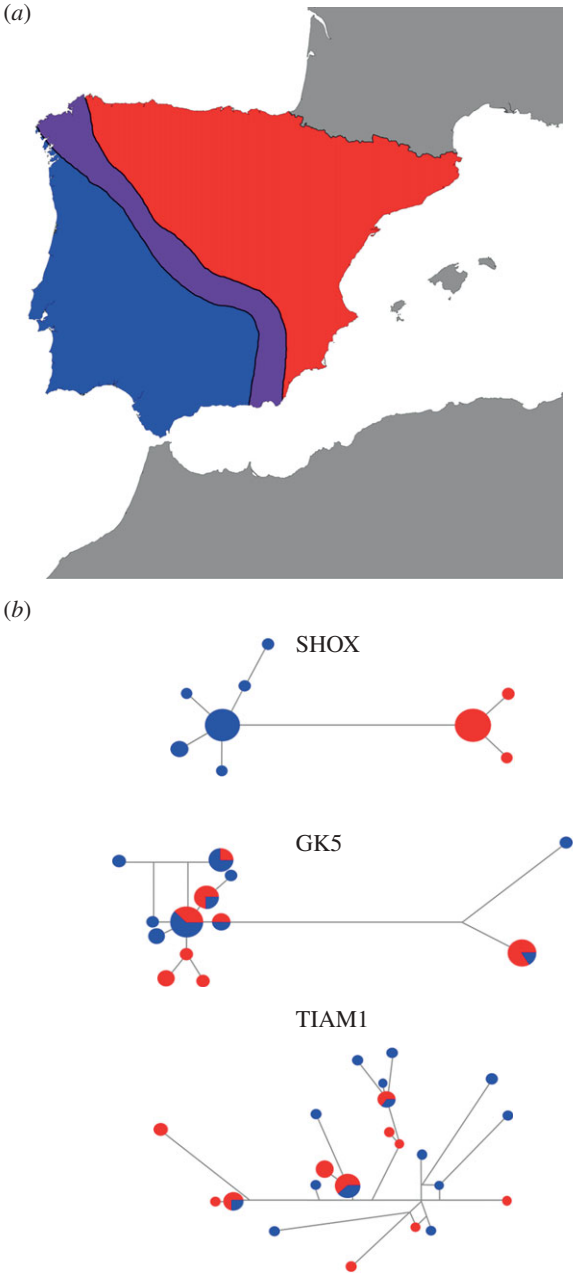

Figure 2. (*a*) Geographical distribution of rabbit subspecies in the Iberian peninsula: *Oryctolagus cuniculus algirus* (blue) and *Oryctolagus cuniculus cuniculus* (red). The approximate location of the hybrid zone is shown in purple. (*b*) Examples of gene genealogies in rabbits for one X-linked locus (SHOX, $F_{ST} = 0.907$) and two autosomal loci (GK5, $F_{ST} = 0.120$ and TIAM1, $F_{ST} = 0.008$). Colours correspond to subspecies of origin.

patterns of differentiation to levels of gene flow is difficult since shared polymorphism may arise as a consequence of migration or from unsorted polymorphisms that remain from the ancestral population. Several solutions to this problem have emerged in the last decade.

The most widely used set of methods were developed by Hey, Nielsen and Wakeley, and are known as IM models [12–14]. The original model of Nielsen & Wakeley [12] includes a single ancestral population that splits into two descendent populations which exchange genes. The model includes six parameters: three for population size (the ancestral population and each contemporary population), divergence time and migration rates into each population. Parameters are estimated jointly in the model using a Markov chain Monte Carlo (MCMC) simulation that incorporates stochastic variation in gene genealogies. The method provides estimates of parameter values as well as tests of alternative models in a likelihood framework. Subsequent extensions to this basic model now allow the inclusion of multiple loci, relax the assumption of constant population size and permit the examination of more than two populations [13,14,56,57].

Another method for estimating gene flow in a non-equilibrium context uses the joint allele-frequency spectrum for two or more populations. For example, Gutenkunst *et al.* [16] used forward simulations with a diffusion approximation to derive demographic parameters for four human populations in a likelihood framework.

A third approach to detecting gene flow under non-equilibrium conditions makes use of specific patterns in haplotype structure or LD. Machado *et al.* [58] pointed out that LD should be higher between linked single nucleotide polymorphisms (SNPs) introduced by recent gene flow, since they will be younger in the population, on average, than SNPs that remain from the ancestral population. In this test, the levels of LD for shared SNPs are compared with the levels of LD for SNPs that are exclusive to each population. More recently, Pool & Nielsen [59] developed a simulation method based on the erosion of migrant tract lengths. This method allows migration parameters to be estimated in a likelihood framework.

Another approach, approximate Bayesian computation (ABC), combines Bayesian inference with coalescent simulations in which the data are represented by a set of summary statistics [60–62] (reviewed by Beaumont [63]). Because it uses summaries of the data, ABC is computationally fast and can handle large datasets and complex demographic scenarios. ABC methods are highly flexible, requiring only the ability to simulate datasets under models of interest. Posterior distributions of parameters, such as rates of gene flow, can be reconstructed using a variety of techniques, including rejection sampling [61], MCMC without likelihoods [64] or local-linear regression of simulated parameter values on simulated summary statistics [62].

Finally, analyses of clinal patterns in hybrid zones provide information about levels of introgression for different loci [65]. In general, for tension zones, cline width is proportional to dispersal rate and inversely proportional to the strength of selection acting on a locus. Thus, loci on which selection is strong will introgress little and have narrow clines, whereas loci on which selection is weak will introgress more and have wide clines. In recently formed hybrid zones, the signature of ongoing gene flow may dominate that of selection at linked sites, making hybrid zones good places to examine the role of recombination in models of speciation with gene flow. Patterns observed in hybrid zones may reflect introgression over different time scales than patterns observed in allopatric populations.

4. EMPIRICAL PATTERNS IN RABBITS AND MICE

The predictions outlined above can be tested in two groups of mammals with recently separated subspecies that hybridize in nature, rabbits and mice.

(a) *Rabbits*

The European rabbit (*Oryctolagus cuniculus*) consists of two major subspecies: *Oryctolagus c. cuniculus*—which is distributed in the northeastern portion of the Iberian peninsula and southern France—and *Oryctolagus c. algirus*—which is distributed throughout the southwestern part of the Iberian peninsula (figure 2*a*). These two lineages are believed to have diverged in allopatry approximately two million years ago [66]. Throughout the Pleistocene, they likely underwent periods of isolation and secondary contact as climatic changes allowed for range contraction and expansion. Today, their ranges meet in a zone that runs diagonally across the Iberian peninsula from the northwest to the southeast. Phenotypic differences between the subspecies are slight [67].

Genetic variation within and between the two subspecies has been studied using mtDNA [66,68], serological typing [69], allozymes [70], microsatellites [71] and DNA sequences of autosomal and X-linked loci [72–74].

Patterns of genetic differentiation vary widely among loci (figure 2*b*). Deep haplotype divergence—with two major clades—is observed at many loci. For example, the mitochondrial cytochrome b gene contains two major lineages (one in each subspecies) with a sequence divergence of 11.9 per cent [66]. Similar patterns of deep divergence between subspecies are seen for the Y chromosome [75], most X-linked loci and many autosomal loci [72,74]. Despite this strong differentiation, F_{ST} varies from nearly 0 to 1 when all surveyed loci are considered. In general, high F_{ST} values are seen at loci that show two divergent haplogroups corresponding to the two subspecies (e.g. SHOX, figure 2*b*). Loci with low F_{ST}, however, are of two sorts. Some loci also have two divergent haplogroups, but both haplogroups are seen in both subspecies (e.g. GK5, figure 2*b*), a pattern consistent with a period of isolation followed by gene flow [74]. Other loci with low F_{ST} do not have two divergent haplotypes (e.g. TIAM1, figure 2*b*), a pattern that could reflect greater mixing and recombination after secondary contact, unsorted ancestral polymorphism or some combination of both.

Several studies have attempted to disentangle the relative contributions of unsorted ancestral variation from gene flow as explanations for shared polymorphisms between the rabbit subspecies. The rate of lineage sorting for neutral genes is a function of the effective population size [76]. Assuming a sex ratio of 1, mitochondrial and Y-linked genes experience an N_e that is one-fourth as large as N_e for autosomal genes, whereas X-linked loci experience an N_e that is three-fourths that of autosomal loci. As a result, we might expect patterns of differentiation to be the highest for mtDNA and Y-linked genes, intermediate for X-linked genes and lowest for autosomal loci. In fact, this is what is observed: F_{ST} for the Y is 0.93 [77], F_{ST} for the mitochondrial *Cytb* gene is 0.83 [66], average F_{ST} for 27 X-linked loci is 0.45 [74] and average F_{ST} for 17 autosomal loci is 0.15 [74]. At face value, this suggests that much of the variation among loci in patterns of differentiation might be accounted for simply by differences in rates of lineage sorting. But the situation is more complicated. Rates of recombination also vary among the Y, mtDNA, X and autosomes. The mitochondrial genome and most of the Y chromosome are effectively free of recombination, and the X chromosome experiences less recombination than the autosomes (since it does not recombine in males outside of the pseudo-autosomal region). Thus, if regions of lower recombination have lower rates of gene flow, similar patterns might arise even without differences in rates of lineage sorting. A good test of this idea would come from comparisons among loci on the same chromosome that experience different rates of recombination. Unfortunately, the rabbit genetic map is insufficiently detailed to provide accurate estimates of the local recombination rate for different genomic regions.

Without knowledge of recombination rates, loci near centromeres of metacentric chromosomes can be compared with loci in the middle of chromosome arms. In many organisms, recombination is suppressed near the centromeres of metacentric chromosomes [78]. Two studies have taken this approach in rabbits. In the first, two loci near the centromere were compared with two loci far from the centromere on the X chromosome [72]. The centromeric loci had an average F_{ST} of 0.75, whereas the telomeric loci had an average F_{ST} of 0.02. In the second study, five autosomal loci near centromeres were compared with five autosomal loci near telomeres [73]. All telomeric loci showed low levels of differentiation (mean $F_{ST} = 0.06$), whereas three centromeric loci showed high differentiation (mean $F_{ST} = 0.47$ for these three loci) and two showed little differentiation (mean $F_{ST} = 0.06$ for these two loci). Importantly, levels of LD were significantly higher within subspecies for loci near centromeres than for loci near telomeres, suggesting that these two types of loci experience low- and high-recombination rates, respectively.

Is the pattern of higher F_{ST} at centromeric loci driven by reduced gene flow or by selection at linked sites? Figure 3*a* compares F_{ST} and the relative node depth (RND) for these loci. RND is D_{xy} divided by D_{xy} to an outgroup, in this case *Lepus granatensis*; this ratio corrects for differences in mutation rate among loci [79]. Both F_{ST} and RND are higher for the genes near centromeres, consistent with greater gene flow at telomeric loci than at centromeric loci, as shown in figure 1*a*. In addition, using likelihood ratio tests, IM models incorporating migration [56] are a significantly better fit to the data than models without migration [74]. Moreover, estimates of gene flow at individual loci using IM are higher for loci near telomeres than for loci near centromeres [73]. These observations suggest that reduced gene flow contributes to the higher differentiation at loci with lower recombination rates.

Does selection at linked sites also contribute to this pattern? Most loci harbour an excess of rare alleles [73], with negative values of Tajima's D [80] and Fu & Li's D [81]. The fact that this pattern is seen at nearly all loci is more consistent with a demographic explanation (such as a population expansion) than with a selective

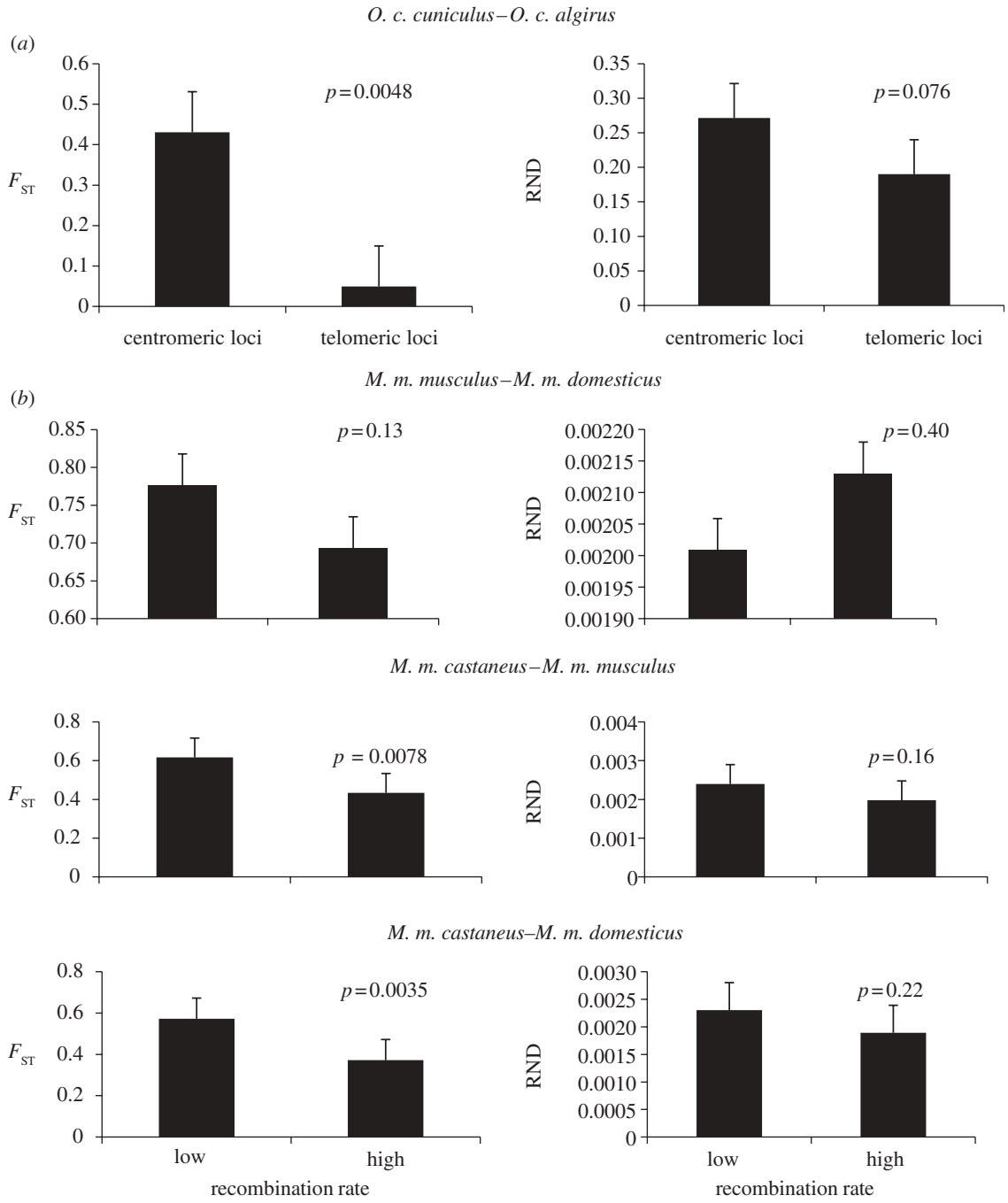

Figure 3. (*a*) Comparison of patterns of differentiation between centromeric loci (presumed to experience low recombination) and telomeric loci (presumed to experience high recombination) between subspecies of rabbits. (*b*) Comparison of patterns of differentiation between low-recombination loci and high-recombination loci between subspecies of house mice.

explanation. Most of these negative values are not significant, although a few are significantly lower than expected under a neutral, equilibrium model [73]. Importantly, however, the average value of Tajima's D or of Fu and Li's D is not lower at centromeric loci than at telomeric loci [73], as might be expected under a simple hitchhiking model. It is important to realize that these results provide only a weak test of selection, and do not rule out a role for more complicated forms of positive selection or a role for background selection.

Finally, the rabbit subspecies do form a hybrid zone in the central portion of the Iberian peninsula. Analyses of clinal patterns of variation for loci near telomeres and for loci near centromeres could provide additional insight into the role of gene flow in preventing differentiation in some genomic regions. Speciation models predict that centromeric loci should have steeper clines than telomeric loci.

(b) *House mice*

House mice consist of three subspecies: *Mus musculus musculus*, *Mus musculus domesticus* and *Mus musculus castaneus* (also referred to as species by some authors). The subspecies diverged from one another within a short period of time within the last 500 000 years [82]. Their phylogenetic relationships are difficult to resolve, although available evidence suggests that *M. m. musculus* and *M. m. castaneus* are sister subspecies [83]. *Mus musculus domesticus* is found in the Mediterranean region and Western Europe, *M. m. musculus* occurs in

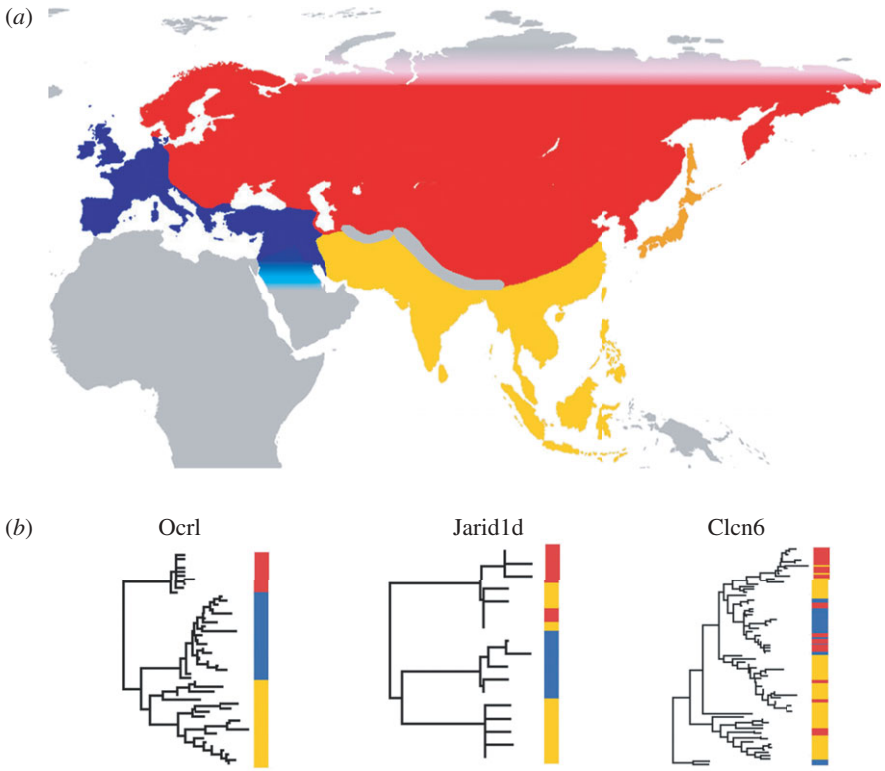

Figure 4. (*a*) Geographical distribution of house mouse subspecies: *Mus musculus domesticus* (blue), *Mus musculus musculus* (red) and *Mus musculus castaneus* (yellow). *Mus musculus molossinus* (orange) is found in Japan and is believed to derive from hybridization between *M. m. musculus* and *M. m. castaneus*. Ranges reflect inferred distributions before expansions associated with humans during the last few hundred years. (*b*) Examples of gene genealogies in house mice for one X-linked locus (Ocrl, $F_{ST} = 0.867$), one Y-linked locus (Jarid1d, $F_{ST} = 0.907$) and one autosomal locus (Clcn6, $F_{ST} = 0.434$).

Eastern Europe and Northern Asia, and *M. m. castaneus* is found in southeast Asia (figure 4). A well-defined hybrid zone between *M. m. domesticus* and *M. m. musculus* occurs in central Europe, running from Denmark to the Black Sea. *Mus musculus molossinus* is found in Japan, and seems to be derived from hybridization between *M. m. musculus* and *M. m. castaneus*. House mice are commensal with humans [84]. In historical times, house mice primarily from Western Europe have been spread in association with humans throughout much of the world, including the Americas, Africa, Australia and many oceanic islands [85].

The genetic basis of speciation in house mice has been studied extensively both in the laboratory and in natural populations. Most of this work has focused on reproductive isolation between *M. m. domesticus* and *M. m. musculus*, though some studies have included comparisons with *M. m. castaneus* and *M. m. molossinus*. In the laboratory, crosses between *M. m. musculus* and *M. m. domesticus* (or inbred strains such as C57BL/6J, which are largely of *M. m. domesticus* origin) have revealed that hybrid males suffer from sterility or reduced fertility, whereas hybrid females are usually fully fertile or show only slight reductions in fertility [86–90]. The only known hybrid sterility gene in vertebrates was recently identified in mice as *Prdm9* [91]. Interestingly, this gene also underlies variation in fine-scale recombination rate in both mice and humans [92–94].

Genetic differentiation in wild mice has been studied both in hybrid and in allopatric populations. The hybrid zone between *M. m. musculus* and *M. m. domesticus* is a result of secondary contact following the spread of mice into Europe within the last 3000 years [95]. Most loci in this hybrid zone display concordant clines, although inter-locus variation in cline width and midpoint is observed [96–100]. Notably, the X chromosome and the Y chromosome show steeper clines than the autosomes in some transects, suggesting a role for these chromosomes in reproductive isolation [101,102]. At some autosomal loci, alleles from one subspecies are observed well into the range of the sister species, suggesting that gene flow between subspecies occurs across the hybrid zone [96,99].

Several studies have analysed multiple autosomal or X-linked loci in allopatric populations [82,103–107]. As in rabbits, gene genealogies vary widely among loci in mice (figure 4*b*). At some loci, such as *Ocrl* on the X chromosome, the three subspecies form three fully sorted groups of haplotypes. Other loci, such as *Jarid1d* on the Y chromosome, have three divergent lineages, but some haplotypes from one subspecies cluster with a lineage that is otherwise restricted to a different subspecies (figure 4*b*). This pattern is similar to the pattern seen at loci such as *Gk5* in rabbits (figure 2*b*), and is probably best explained by gene flow. Finally, at some loci the three subspecies are intermingled on the gene genealogy (e.g. *Clcn6*, figure 4*b*). As in rabbits, F_{ST} varies from nearly 0 to 1 among loci.

Several observations suggest that shared polymorphisms result at least partly from gene flow [59,82]. In comparisons between *M. m. musculus* and *M m. domesticus*, the X chromosome has fewer shared polymorphisms and higher F_{ST} than the

autosomes [82,105,106]. Geraldes *et al.* [82] analysed sequence data from eight loci and found that likelihood ratio tests in an IM framework rejected a strict allopatric model of speciation in each of the three comparisons between pairs of subspecies. Moreover, estimates of gene flow in these models were higher for autosomal loci compared with X-linked loci. Lower gene flow on the X is also supported by hybrid zone studies, where loci on the X chromosome have narrower cline widths, on average, than do autosomal loci [96,101,108]. This general agreement between large-scale patterns of differentiation on the X versus autosomes in allopatric and hybrid populations suggests that reduced gene flow contributes to higher differentiation on the X chromosome. Importantly, this agrees well with numerous laboratory crosses that implicate much of the X chromosome in hybrid male sterility [87,88,109,110]. Thus, genes underlying reproductive isolation are associated with genomic regions showing greater levels of differentiation, as predicted by speciation models.

In contrast to rabbits, local recombination rates can be estimated in mice by comparing dense genetic maps to the reference genome sequence. Recombination rates vary across the genome [111,112], with rates in some regions differing among divergent strains [113]. Is differentiation between subspecies correlated with local recombination rate along the autosomes?

Takahashi *et al.* [104] analysed sequence data from 19 autosomal loci sampled in two inbred strains each of *M. m. musculus*, *M. m. domesticus*, *M. m. castaneus* and *M. m. molossinus*. Despite the small sample size ($n = 2$ for each subspecies), they found a significant negative correlation between F_{ST} and recombination rate, with recombination rate explaining about 40 per cent of the variation in levels of differentiation. This pattern was attributed to genetic hitchhiking or background selection, but the potential contribution of differences in levels of gene flow was not considered. F_{ST} in this analysis was based on the average among all four taxa, and patterns of differentiation between pairs of subspecies were not reported. Harr [105] analysed patterns of differentiation between inbred strains of *M. m. musculus* and *M. m. domesticus* for 10 000 SNPs that were ascertained in other samples and found no association between recombination rate and F_{ST}. Similarly, Teeter *et al.* [99] found no correlation between recombination rate and cline width for 39 markers in the *M. m. musculus*–*M. m. domesticus* hybrid zone.

Geraldes *et al.* [107] analysed sequence data from 27 autosomal loci in population samples of each of the three subspecies. In each of the three pairwise comparisons between subspecies, F_{ST} was higher for genes in regions of low recombination than for genes in regions of high recombination (figure 3*b*). In comparisons between *M. m. castaneus* and *M. m. musculus* or *M. m. domesticus*, RND was higher for the low-recombination genes compared with the high-recombination genes (figure 3*b*), consistent with greater gene flow at the high-recombination genes (figure 1*a*) and similar to the patterns seen in rabbits. In the comparison between *M. m. musculus* and *M. m. domesticus*, RND was lower for the low-recombination genes compared with the high-recombination genes, although differences in F_{ST} were not significant in this comparison. IM models incorporating migration are a significantly better fit to the data than models without migration in all three pairwise comparisons [82], and IM estimates of gene flow at individual loci are higher for high-recombination loci than for low-recombination loci [107]. These results show that *M. m castaneus* is more differentiated from *M. m. musculus* and from *M. m. domesticus* in regions of low recombination than in regions of high recombination and that differences in levels of gene flow account for some of this variation, in agreement with predictions from speciation models.

Does selection also contribute to greater differentiation in regions of low recombination? For each of the three subspecies, Tajima's D and Fu and Li's D are both lower on average for the low-recombination loci than for the high-recombination loci, although these differences are slight and are not significant [107]. Baines & Harr [103] compared patterns of DNA sequence variation on the X chromosome and on the autosomes in ancestral and derived populations of *M. m. musculus* and of *M. m. domesticus*, and found some evidence for positive selection on the X chromosome of derived populations. No comparable study has been conducted on ancestral and derived populations comparing regions of low and high recombination. Studies of selection in high- and low-recombination regions would be useful, especially in comparisons between *M. m. castaneus* and *M. m. musculus* or between *M. m. castaneus* and *M. m. domesticus*, where variation in F_{ST} is associated with variation in recombination rate.

5. FUTURE PROSPECTS

Multiple models predict that genomic regions with little recombination will exhibit relatively high levels of differentiation in recently separated taxa and this prediction is supported by data from rabbits and mice. The ability to distinguish between speciation with gene flow and selection at linked sites as explanations for this pattern would be improved most rapidly by investing in two areas.

First, we need to identify those aspects of genetic variation that best separate the two processes. Speciation with gene flow and selection at linked sites are evolutionarily distinct phenomena, suggesting that careful comparisons will reveal the discordant signatures they leave in patterns of diversity. Analysis of simulated populations in which one or both processes are operating should lead to diagnostic combinations of existing summary statistics (e.g. measures of LD and the frequency spectrum), and could suggest new summary statistics. In addition to identifying more informative measures of variation, simulations should show whether the two models predict different genomic scales of differentiation. For example, we might expect recent gene flow to produce larger regions of high differentiation than selection at linked sites. Furthermore, simulations should help disentangle the relative contributions of recombination rate, selection and gene flow to genomic patterns of differentiation [114]. Simulations will also motivate inferences that are model-based, rather than relying on analytical theory with restrictive assumptions. Informative combinations

of summary statistics could be used in an ABC framework to directly compare the support for speciation with gene flow and selection at linked sites in a dataset.

A second promising approach to distinguishing these models is to integrate studies of population differentiation with genetic studies of reproductive isolation phenotypes [24,115]. Speciation with gene flow specifically predicts that genomic regions with reduced recombination rates will contain genes responsible for reproductive isolation, whereas models of selection at linked sites make no such prediction. Comparing the genomic positioning of loci responsible for reproductive isolation phenotypes with the genomic patterning of differentiation would offer the added benefit of gauging the contributions of the phenotypes being studied to gene flow in nature.

We thank Pennie Liebig for help with the preparation of the manuscript. We thank M. Carneiro and A. Geraldes for comments on the manuscript. This work was supported by NSF and NIH grants to M.W.N. and by NSF and NIH grants to B.A.P.

REFERENCES

1 Mayr, E. 1942 *Systematics and the origin of species*. New York, NY: Columbia University Press.
2 Dobzhansky, T. 1937 *Genetics and the origin of species*. New York, NY: Columbia University Press.
3 Dieckmann, U. & Doebeli, M. 1999 On the origin of species by sympatric speciation. *Nature* **400**, 354–357. (doi:10.1038/22521)
4 Kondrashov, A. S. & Kondrashov, F. A. 1999 Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351–354. (doi:10.1038/22514)
5 Doebeli, M. & Dieckmann, U. 2003 Speciation along environmental gradients. *Nature* **421**, 259–264. (doi:10.1038/nature01274)
6 Feder, J. L., Chilcote, C. A. & Bush, G. L. 1988 Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**, 61–64. (doi:10.1038/336061a0)
7 McPheron, B. A., Smith, D. C. & Berlocher, S. H. 1988 Genetic differences between host races of *Rhagoletis pomonella*. *Nature* **336**, 64–66. (doi:10.1038/336064a0)
8 Schliewen, U. K., Tautz, D. & Paabo, S. 1994 Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–632. (doi:10.1038/368629a0)
9 Savolainen, V. *et al.* 2006 Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210–213. (doi:10.1038/nature04566)
10 Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. 2000 Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308. (doi:10.1126/science.287.5451.306)
11 Bolnick, D. I. & Fitzpatrick, B. M. 2007 Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* **38**, 459–487. (doi:10.1146/annurev.ecolsys.38.091206.095804)
12 Nielsen, R. & Wakeley, J. 2001 Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**, 885–896.
13 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. (doi:10.1534/genetics.103.024182)
14 Hey, J. 2005 On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol.* **3**, e193. (doi:10.1371/journal.pbio.0030193)
15 Becquet, C. & Przeworski, M. 2007 A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* **17**, 1505–1519. (doi:10.1101/gr.6409707)
16 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695. (doi:10.1371/journal.pgen.1000695)
17 Lopes, J. S., Balding, D. J. & Beaumont, M. A. 2009 PopABC: a program to infer historical demographic parameters. *Bioinformatics* **25**, 2747–2749. (doi:10.1093/bioinformatics/btp487)
18 Pinho, C. & Hey, J. 2010 Divergence with gene flow: models and data. *Annu. Rev. Ecol. Evol. Syst.* **41**, 215–230. (doi:10.1146/annurev-ecolsys-102209-144644)
19 Strasburg, J. L. & Rieseberg, L. H. 2011 Interpreting the estimated timing of migration events between hybridizing species. *Mol. Ecol.* **20**, 2353–2366. (doi:10.1111/j.1365-294X.2011.05048.x)
20 Bateson, W. 1909 Heredity and variation in modern lights. In *Darwin and modern science* (ed. A. C. Seward), pp. 85–101. Cambridge, UK: Cambridge University Press.
21 Muller, H. J. 1940 Bearing of the *Drosophila* work on systematics. In *The new systematics* (ed. J. S. Huxley), pp. 185–268. Oxford, UK: Clarendon.
22 Feder, J. L., Gejji, R., Yeaman, S. & Nosil, P. 2012 Establishment of new mutations under divergence and genome hitchhiking. *Phil. Trans. R. Soc. B* **367**, 461–474. (doi:10.1098/rstb.2011.0256)
23 Navarro, A. & Barton, N. H. 2003 Chromosomal speciation and molecular divergence: accelerated evolution in rearranged chromosomes. *Science* **300**, 321–324. (doi:10.1126/science.1080600)
24 Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. 2001 Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12 084–12 088. (doi:10.1073/pnas.221274498)
25 Rieseberg, L. H. 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358. (doi:10.1016/S0169-5347(01)02187-5)
26 Butlin, R. K. 2005 Recombination and speciation. *Mol. Ecol.* **14**, 2621–2635. (doi:10.1111/j.1365-294X.2005.02617.x)
27 Faria, R. & Navarro, A. 2010 Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669. (doi:10.1016/j.tree.2010.07.008)
28 Guerrero, R. F., Rousset, F. & Kirkpatrick, M. 2012 Coalescent patterns for chromosomal inversions in divergent populations. *Phil. Trans. R. Soc. B* **367**, 430–438. (doi:10.1098/rstb.2011.0246)
29 White, M. J. D. 1968 Models of speciation. *Science* **159**, 1065–1070. (doi:10.1126/science.159.3819.1065)
30 McGaugh, S. E. & Noor, M. A. F. 2012 Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Phil. Trans. R. Soc. B* **367**, 422–429. (doi:10.1098/rstb.2011.0250)
31 Maynard Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a fabourable gene. *Genet. Res.* **23**, 23–35. (doi:10.1017/S0016672300014634)
32 Hermisson, J. & Pennings, P. S. 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352. (doi:10.1534/genetics.104.036947)
33 Przeworski, M., Coop, G. & Wall, J. D. 2005 The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323. (doi:10.1554/05-273.1)

34 Pennings, P. S. & Hermisson, J. 2006 Soft sweeps II: molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23**, 1076–1084. (doi:10.1093/molbev/msj117)
35 Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.
36 Begun, D. J. & Aquadro, C. F. 1992 Levels of naturally-occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520. (doi:10.1038/356519a0)
37 Nachman, M. W., Bauer, V. L., Crowell, S. L. & Aquadro, C. F. 1998 DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**, 1133–1141.
38 Cutter, A. D. & Payseur, B. A. 2003 Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol. Biol. Evol.* **20**, 665–673. (doi:10.1093/molbev/msg072)
39 Hernandez, R. D., Kelley, J. L., Elyashiv, E., Melton, S. C., Auton, A., McVean, G., Sella, G. & Przeworski, M. 2011 Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924. (doi:10.1126/science.1198878)
40 Nei, M. 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. USA* **70**, 3321–3323. (doi:10.1073/pnas.70.12.3321)
41 Charlesworth, B. 1998 Sex chromosomes: evolving dosage compensation. *Curr. Biol.* **8**, R931–R933. (doi:10.1016/S0960-9822(98)00013-X)
42 Begun, D. J. & Aquadro, C. F. 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550. (doi:10.1038/365548a0)
43 Stephan, W., Xing, L., Kirby, D. A. & Braverman, J. M. 1998 A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl Acad. Sci. USA* **95**, 5649–5654. (doi:10.1073/pnas.95.10.5649)
44 Baines, J. F., Das, A., Mousset, S. & Stephan, W. 2004 The role of natural selection in genetic differentiation of worldwide populations of *Drosophila ananassae*. *Genetics* **168**, 1987–1998. (doi:10.1534/genetics.104.027482)
45 Keinan, A. & Reich, D. 2010 Human population differentiation is strongly correlated with local recombination rate. *PLoS Genet.* **6**, e1000886. (doi:10.1371/journal.pgen.1000886)
46 Nei, M. 1987 *Molecular evolutionary genetics*. New York, NY: Columbia University.
47 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. *Genet. Res.* **70**, 155–174. (doi:10.1017/S0016672397002954)
48 Hu, X. S. & He, F. L. 2005 Background selection and population differentiation. *J. Theor. Biol.* **235**, 207–219. (doi:10.1016/j.jtbi.2005.01.004)
49 Slatkin, M. & Wiehe, T. 1998 Genetic hitch-hiking in a subdivided population. *Genet. Res.* **71**, 155–160. (doi:10.1017/S001667239800319X)
50 Hellmann, I., Ebersberger, I., Ptak, S. E., Paabo, S. & Przeworski, M. 2003 A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**, 1527–1535. (doi:10.1086/375657)
51 Noor, M. A. F. & Bennett, S. M. 2009 Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**, 439–444. (doi:10.1038/hdy.2009.151)
52 Hedrick, P. W. 2005 A standardized genetic differentiation measure. *Evolution* **59**, 1633–1638.
53 Jost, L. 2008 G(ST) and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026. (doi:10.1111/j.1365-294X.2008.03887.x)
54 Hudson, R. R., Kreitman, M. & Aguade, M. 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.
55 Whitlock, M. C. & McCauley, D. E. 1999 Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity* **82**, 117–125. (doi:10.1038/sj.hdy.6884960)
56 Hey, J. & Nielsen, R. 2007 Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl Acad. Sci. USA* **104**, 2785–2790. (doi:10.1073/pnas.0611164104)
57 Hey, J. 2010 Isolation with migration models for more than two populations. *Mol. Biol. Evol.* **27**, 905–920. (doi:10.1093/molbev/msp296)
58 Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19**, 472–488.
59 Pool, J. E. & Nielsen, R. 2009 Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181**, 711–719. (doi:10.1534/genetics.108.098095)
60 Tavare, S., Balding, D. J., Griffiths, R. C. & Donnelly, P. 1997 Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505–518.
61 Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. 1999 Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798.
62 Beaumont, M. A., Zhang, W. Y. & Balding, D. J. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035.
63 Beaumont, M. A. 2010 Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* **41**, 379–406. (doi:10.1146/annurev-ecolsys-102209-144621)
64 Marjoram, P., Molitor, J., Plagnol, V. & Tavare, S. 2003 Markov chain Monte Carlo without likelihoods. *Proc. Natl Acad. Sci. USA* **100**, 15 324–15 328. (doi:10.1073/pnas.0306899100)
65 Barton, N. H. & Hewitt, G. M. 1985 Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**, 113–148. (doi:10.1146/annurev.es.16.110185.000553)
66 Branco, M., Ferrand, N. & Monnerot, M. 2000 Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity* **85**, 307–317. (doi:10.1046/j.1365-2540.2000.00756.x)
67 Sharples, C. M., Fa, J. E. & Bell, D. J. 1996 Geographical variation in size in the European rabbit *Oryctolagus cuniculus* (Lagomorpha: Leporidae) in western Europe and North Africa. *Zool. J. Linn. Soc.* **117**, 141–158. (doi:10.1111/j.1096-3642.1996.tb02153.x)
68 Biju-Duval, C., Ennafaa, H., Dennebouy, N., Monnerot, M., Mignotte, F., Soriguer, R., El Gaied, A., El Hili, A. & Mounolou, J. C. 1991 Mitchondrial DNA evolution in Lagomorphs: origin of systematic heteroplasmy, organization of diversity in European rabbits. *J. Mol. Evol.* **33**, 92–102. (doi:10.1007/BF02100200)
69 van der Loo, W., Ferrand, N. & Soriguer, R. C. 1991 Estimation of gene diversity at the *b* locus of the constant region of the immunoglobulin light chain in natural populations of European rabbit (*Oryctolagus cuniculus*) in Portugal, Andalusia and on the Azorean Islands. *Genetics* **127**, 789–799.
70 Ferrand, N. & Branco, M. 2007 The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns

of population differentiation and geographic expansion inferred from protein polymorphism. In *Phylogeography of southern European refugia* (eds S. Weiss & N. Ferrand), pp. 207–235. Amsterdam, The Netherlands: Springer.
71 Queney, G., Ferrand, N., Weiss, S., Mougel, F. & Monnerot, M. 2001 Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). *Mol. Biol. Evol.* **18**, 2169–2178.
72 Geraldes, A., Ferrand, N. & Nachman, M. W. 2006 Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **173**, 919–933. (doi:10.1534/genetics.105.054106)
73 Carneiro, M., Ferrand, N. & Nachman, M. W. 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. (doi:10.1534/genetics.108.096826)
74 Carneiro, M., Blanco-Aguiar, J. A., Villafuerte, R., Ferrand, N. & Nachman, M. W. 2010 Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution* **64**, 3443–3460. (doi:10.1111/j.1558-5646.2010.01092.x)
75 Geraldes, A., Rogel-Gaillard, C. & Ferrand, N. 2005 High levels of nucleotide diversity in the European rabbit (*Orydolagus cuniculus*) SRY gene. *Anim. Genet.* **36**, 349–351. (doi:10.1111/j.1365-2052.2005.01300.x)
76 Tajima, F. 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.
77 Geraldes, A., Carneiro, M., Delibes-Mateos, M., Villafuerte, R., Nachman, M. W. & Ferrand, N. 2008 Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Mol. Ecol.* **17**, 4489–4499. (doi:10.1111/j.1365-294X.2008.03943.x)
78 Kong, A. *et al.* 2002 A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247.
79 Feder, J. L., Xie, X. F., Rull, J., Velez, S., Forbes, A., Leung, B., Dambroski, H., Filchak, K. E. & Aluja, M. 2005 Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *Proc. Natl Acad. Sci. USA* **102**, 6573–6580. (doi:10.1073/pnas.0502099102)
80 Tajima, F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
81 Fu, Y. X. & Li, W. H. 1993 Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.
82 Geraldes, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H. T., Bulatova, N., Ziv, Y. & Nachman, M. W. 2008 Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol.* **17**, 5349–5363. (doi:10.1111/j.1365-294X.2008.04005.x)
83 White, M. A., Ane, C., Dewey, C. N., Larget, B. R. & Payseur, B. A. 2009 Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet.* **5**, e1000729. (doi:10.1371/journal.pgen.1000729)
84 Sage, R. D. 1981 Wild mice. In *The mouse in biomedical research* (eds H. L. Foster, J. D. Small & J. G. Fox), pp. 40–90. New York, NY: Academic Press.
85 Boursot, P., Auffray, J. C., Brittondavidian, J. & Bonhomme, F. 1993 The evolution of house mice. *Annu. Rev. Ecol. Syst.* **24**, 119–152. (doi:10.1146/annurev.es.24.110193.001003)
86 Forejt, J. 1996 Hybrid sterility in the mouse. *Trends Genet.* **12**, 412–417. (doi:10.1016/0168-9525(96) 10040-8)
87 Storchova, R., Gregorova, S., Buckiova, D., Kyselova, V., Divina, P. & Forejt, J. 2004 Genetic analysis of X-linked hybrid sterility in the house mouse. *Mamm. Genome* **15**, 515–524. (doi:10.1007/s00335-004-2386-0)
88 Oka, A., Mita, A., Sakurai-Yamatani, N., Yamamoto, H., Takagi, N., Takano-Shimizu, T., Toshimori, K., Moriwaki, K. & Shiroishi, T. 2004 Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. *Genetics* **166**, 913–924. (doi:10.1534/genetics.166.2.913)
89 Britton-Davidian, J., Fel-Clair, F., Lopez, J., Alibert, P. & Boursot, P. 2005 Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biol. J. Linn. Soc.* **84**, 379–393. (doi:10.1111/j.1095-8312.2005.00441.x)
90 Good, J. M., Handel, M. A. & Nachman, M. W. 2008 Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* **62**, 50–65.
91 Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. 2009 A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375. (doi:10.1126/science.1163601)
92 Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. & de Massy, B. 2010 PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840. (doi:10.1126/science.1183439)
93 Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G. & Donnelly, P. 2010 Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**, 876–879. (doi:10.1126/science.1182363)
94 Parvanov, E. D., Petkov, P. M. & Paigen, K. 2010 PRDM9 controls activation of mammalian recombination hotspots. *Science* **327**, 385–835. (doi:10.1126/science.1181495)
95 Cucchi, T., Vigne, J. D. & Auffray, J. C. 2005 First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biol. J. Linn. Soc.* **84**, 429–445. (doi:10.1111/j.1095-8312.2005.00445.x)
96 Macholan, M., Munclinger, P., Sugerkova, M., Dufkova, P., Bimova, B., Bozikova, E., Zima, J. & Pialek, J. 2007 Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution* **61**, 746–771. (doi:10.1111/j.1558-5646.2007.00065.x)
97 Hunt, W. G. & Selander, R. K. 1973 Biochemical genetics of hybridization in European house mice. *Heredity* **31**, 11–33. (doi:10.1038/hdy.1973.56)
98 Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C., Baird, S. J. E., Bonhomme, F., Dod, B. & Boursot, P. 2005 Inferences of selection and migration in the Danish house mouse hybrid zone. *Biol. J. Linn. Soc.* **84**, 593–616. (doi:10.1111/j.1095-8312.2005.00457.x)
99 Teeter, K. C. *et al.* 2008 Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* **18**, 67–76. (doi:10.1101/gr.6757907)
100 Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W. & Tucker, P. K. 2010 The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* **64**, 472–485. (doi:10.1111/j.1558-5646.2009.00846.x)
101 Tucker, P. K., Sage, R. D., Warner, J., Wilson, A. C. & Eicher, E. M. 1992 Abrupt cline for sex chromosomes in a hybrid zone between 2 species of mice. *Evolution* **46**, 1146–1163. (doi:10.2307/2409762)
102 Vanlerberghe, F., Dod, B., Boursot, P., Bellis, M. & Bonhomme, F. 1986 Absence of Y chromosome

introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. *Genet. Res.* **48**, 191–197. (doi:10.1017/S0016672300025003)
103 Baines, J. F. & Harr, B. 2007 Reduced X-linked diversity in derived populations of house mice. *Genetics* **175**, 1911–1921. (doi:10.1534/genetics.106.069419)
104 Takahashi, A., Liu, Y. H. & Saitou, N. 2004 Genetic variation versus recombination rate in a structured population of mice. *Mol. Biol. Evol.* **21**, 404–409. (doi:10.1093/molbev/msh030)
105 Harr, B. 2006 Genomic islands of differentiation between house mouse subspecies. *Genome Res.* **16**, 730–737. (doi:10.1101/gr.5045006)
106 Salcedo, T., Geraldes, A. & Nachman, M. W. 2007 Nucleotide variation in wild and inbred mice. *Genetics* **177**, 2277–2291. (doi:10.1534/genetics.107.079988)
107 Geraldes, A., Basset, P., Smith, K. & Nachman, M. W. 2011 Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol. Ecol.* **20**, 4722–4736. (doi:10.1111/j.1365-294X.2011.05285.x)
108 Dod, B., Jermiin, L. S., Boursot, P., Chapman, V. H., Nielsen, J. T. & Bonhomme, F. 1993 Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. *J. Evol. Biol.* **6**, 529–546. (doi:10.1046/j.1420-9101.1993.6040529.x)
109 Good, J. M., Dean, M. D. & Nachman, M. W. 2008 A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* **179**, 2213–2228. (doi:10.1534/genetics.107.085340)
110 Good, J. M., Giger, T., Dean, M. D. & Nachman, M. W. 2010 Widespread over-expression of the X chromosome in sterile F1 hybrid mice. *PLoS Genet.* **6**, e1001148. (doi:10.1371/journal.pgen.1001148)
111 Shifman, S., Bell, J. T., Copley, R. R., Taylor, M. S., Williams, R. W., Mott, R. & Flint, J. 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol.* **4**, e395. (doi:10.1371/journal.pbio.0040395)
112 Cox, A. *et al.* 2009 A new standard genetic map for the laboratory mouse. *Genetics* **182**, 1335–1344. (doi:10.1534/genetics.109.105486)
113 Dumont, B. L., White, M. A., Steffy, B., Wiltshire, T. & Payseur, B. A. 2011 Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Res.* **21**, 114–125. (doi:10.1101/gr.111252.110)
114 Nosil, P. & Feder, J. L. 2012 Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B* **367**, 332–342. (doi:10.1098/rstb.2011.0263)
115 Machado, C. A. & Hey, J. 2003 The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond. B* **270**, 1193–1202. (doi:10.1098/rspb.2003.2333)