Interactions Between Natural Selection and Recombination Shape the Genomic Landscape of Introgression

Maud Duranton* and John E. Pool

Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA

***Corresponding author:** E-mail: durantonmaud@gmail.com.

Associate editor: Kelley Harris

Abstract

Hybridization between lineages that have not reached complete reproductive isolation appears more and more like a common phenomenon. Indeed, speciation genomic studies have now extensively shown that many species' genomes have hybrid ancestry. However, genomic patterns of introgression are often heterogeneous across the genome. In many organisms, a positive correlation between introgression levels and recombination rate has been observed. It is usually explained by the purging of deleterious introgressed material due to incompatibilities. However, the opposite relationship was observed in a North American population of *Drosophila melanogaster* with admixed European and African ancestry. In order to explore how directional and epistatic selection can impact the relationship between introgression and recombination, we performed forward simulations of whole *D. melanogaster* genomes reflecting the North American population's history. Our results revealed that the simplest models of positive selection often yield negative correlations between introgression and recombination such as the one observed in *D. melanogaster*. We also confirmed that incompatibilities tend to produce positive introgression–recombination correlations. And yet, we identify parameter space under each model where the predicted correlation is reversed. These findings deepen our understanding of the evolutionary forces that may shape patterns of ancestry across genomes, and they strengthen the foundation for future studies aimed at estimating genome-wide parameters of selection in admixed populations.

Key words: **forward simulations, introgression, recombination, pairwise incompatibilities, directional selection.**

Introduction

Speciation has long been seen as a gradual process during which two populations evolve barriers to gene flow until reproductive isolation is complete (Orr 1995; Coyne and Orr 2004). Now, many speciation genomic studies have shown that it is more of a dynamic process that can involve more than two populations (Novikova et al. 2016; Van Belleghem et al. 2021) and where hybridization is a common phenomenon (Payseur and Rieseberg 2016) that might also sometimes favor speciation (Jacobsen and Omland 2011; Schumer et al. 2015; Schumer, Rosenthal, et al. 2018; Schumer, Xu, et al. 2018; Blanckaert and Bank 2018). There is indeed a grey zone of speciation where diverging lineages can still exchange genes (Roux et al. 2016). Therefore, many species genomes are in fact a mosaic made of different ancestry. Introgression, that is the exchange of genetic material between diverging populations through hybridization, can have different effects on the receiving genome. In most cases, if introgression has a fitness consequence at all, it is probably deleterious either because introgressed DNA fragments can carry deleterious variants (especially if the donor population is small and thus has a high genetic load) (Bierne et al. 2002) or because they carry locally adaptive alleles not fit for the receiving population's environment (Arnegard et al. 2014). The presence of Bateson–Dobzhansky–Muller incompatibilities (BDMIs) (Bateson 1909; Dobzhansky 1982; Muller 1942), in which an allele from one population at one locus has poor fitness when combined with an allele from a second population at a second locus, can also make introgression deleterious. Indeed, BDMIs are thought to be a major mechanism that contributes to the buildup of reproductive isolation and are expected to naturally arise when populations are diverging without gene flow (Masly and Presgraves 2007; Presgraves 2010). Nonetheless, incoming gene flow can also be beneficial through adaptive introgression. This process allows linked combinations of beneficial alleles that have already been tested by natural selection to enter a new genetic background (Fisher 1937; Racimo et al. 2017; Martin and Jiggins 2017). There are now several examples of species that have acquired adaptive phenotypes through introgression such as mimicry in *Heliconius* butterflies (The Heliconius Genome Consortium 2012), seasonal camouflage in the snowshoe hares (Jones et al. 2018), and altitude adaptation (Huerta-Sánchez et al. 2014) and malaria resistance in humans.

Article

A consequence of selection that acts on introgressed alleles is that patterns of introgression are often heterogeneous along the genome (Martin and Jiggins 2017). Understanding how hybrid genomes and patterns of introgression are modulated by natural selection is a major goal in evolutionary biology. To do so, it is necessary to take into account variations of the recombination rate along the genome, as recombination can modulate the strength of selection through linkage disequilibrium (Nachman and Payseur 2012; Harris and Nielsen 2016; Juric et al. 2016). In the absence of selection, no consistent relationship between recombination rate and introgression level is expected. In contrast, for many species, a positive correlation has been observed between introgression and recombination, such as in human with Neanderthal ancestry (Sankararaman et al. 2016; Juric et al. 2016; Schumer, Xu, et al. 2018), in hybrid populations of swordtail fishes (Schumer, Xu, et al. 2018), within populations of Heliconius butterfly (Martin et al. 2019) and between subspecies of house mice (Janoušek et al. 2015). These positive correlations are predicted by theory when introgression is deleterious. In highly recombining regions, neutral introgressed alleles can more easily recombine away from deleterious ones and persist in the genome (Barton and Bengtsson 1986). Whereas in low recombination regions where linkage disequilibrium is higher, deleterious variants when removed bring with them more neutral ones (Charlesworth et al. 1993), which leads to greater reductions in introgression levels.

A contrasting relationship between recombination rate and introgression level was observed for the *Drosophila* Genetic Reference Panel (DGRP). Like most North American populations of *Drosophila melanogaster*, this North Carolina population resulted from an admixture of populations likely originating from Europe and Africa (David and Capy 1988; Caracristi and Schlötterer 2003; Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). This admixture was estimated to have occurred about 1,600 generations ago, based on the lengths of ancestry tracts (Pool 2015; Corbett-Detig and Nielsen 2017). Within the DGRP sample of 205 sequenced inbred strain genomes, a strong negative correlation has been observed between the minor African ancestry and recombination (Pool 2015; Corbett-Detig and Nielsen 2017), in contrast to the predictions of both neutral and BDMI models. This pattern serves as a motivating observation for the present study.

To our knowledge, the only published model predicting a negative correlation between introgression and recombination rate is based on simulations in which diverging populations experience substantial genetic load and each one may fix a distinct set of deleterious variants during an isolation of 2*N* generations (Kim et al. 2018). In that study, negative correlations were observed in models involving recessive deleterious variants (generating heterosis in admixed populations) and/or introgression flowing from a larger population into a smaller one. However, it is not clear that such a model should be responsible for the pattern observed in the DGRP population, in which admixture occurred between large source populations that only diverged on the order of 0.01*N* generations ago (Sprengelmeyer et al. 2020).

Here, we hypothesize that positive selection in admixed populations can also generate negative correlations between recombination rate and introgression level, even if both populations contribute the same number of favored variants. The key reason is that favored variants from the minor population will traverse a greater frequency change, and thus have a greater impact on ancestry levels (compared with major population favored variants), and this imbalance will be greater in low recombination regions where random neutral sites are more likely to be linked to selected sites. As a conceptual example, suppose minor and major populations initially admix in 10% versus 90% proportions. And simplistically, assume that a random neutral site has a 50% chance of being linked to a selected variant if it is in a low recombination region, but only a 10% chance of linkage to a selected variant if it is in a high recombination region. If selection is fully successful in favoring beneficial variants, then the expected minor population ancestry proportion in low recombination regions is: (50% of sites not linked to selection × 10% minor population ancestry) + (25% of sites linked to favored major population variant × 0% minor population ancestry) + (25% of sites linked to favored minor population variant × 100% minor population ancestry) = 30%. Whereas the expected minor population ancestry for high recombination regions is: (90% of sites not linked to selection × 10% minor population ancestry) + (5% of sites linked to favored major population variant × 0% minor population ancestry) + (5% of sites linked to favored minor population variant × 100% minor population ancestry) = 14%. This predicted excess of minor population ancestry in low recombination regions due to positive selection is qualitatively similar to the genetic load model (Kim et al. 2018) but in contrast to expectations from incompatibility models (Martin and Jiggins 2017).

In this study, we use simulation to identify the relationships between recombination and introgression generated by positive selection in an admixed population. We also perform similar simulation analyses using models of pairwise incompatibilities, in part to test whether any specific BDMI scenarios depart from the general expectation of a positive recombination–introgression correlation. We simulated whole genomes of *D. melanogaster* within two source populations that admix in different proportions to generate a third one that evolves on its own for 1,600 generations. After this time, we analyzed patterns of introgression along the genome and the relationship between the minor population ancestry and recombination rate. For the model of directional selection, we varied the number of favored variants present in the population, the fraction of variants present in each population, the average selection coefficient of the variants, and the admixture proportion. For the model with pairwise incompatibilities, we varied the average fitness impact of the incompatibilities, the dominance of the variants, and the admixture proportions. We show that models with only positive

variants often yield negative correlations between minor ancestry and recombination (particularly in cases where the two source populations contribute equal numbers of favored variants) but that depending on the parameters, positive correlations can also be generated. Furthermore, we also show that spatial correlation of population ancestry level along the genome may help distinguish between models of fewer strong versus many weak favored variants. We also confirmed that incompatibilities tend to produce positive introgression–recombination correlations, but we discovered that this correlation can also be reversed in scenarios where one population carries dominant incompatibility partner alleles and the other carries recessive partner alleles. These results will guide the interpretation of evolutionary genomic data from admixing/introgressing taxa, while also improving prospects for the development of novel inference methods to estimate evolutionary parameters from such cases.

Results

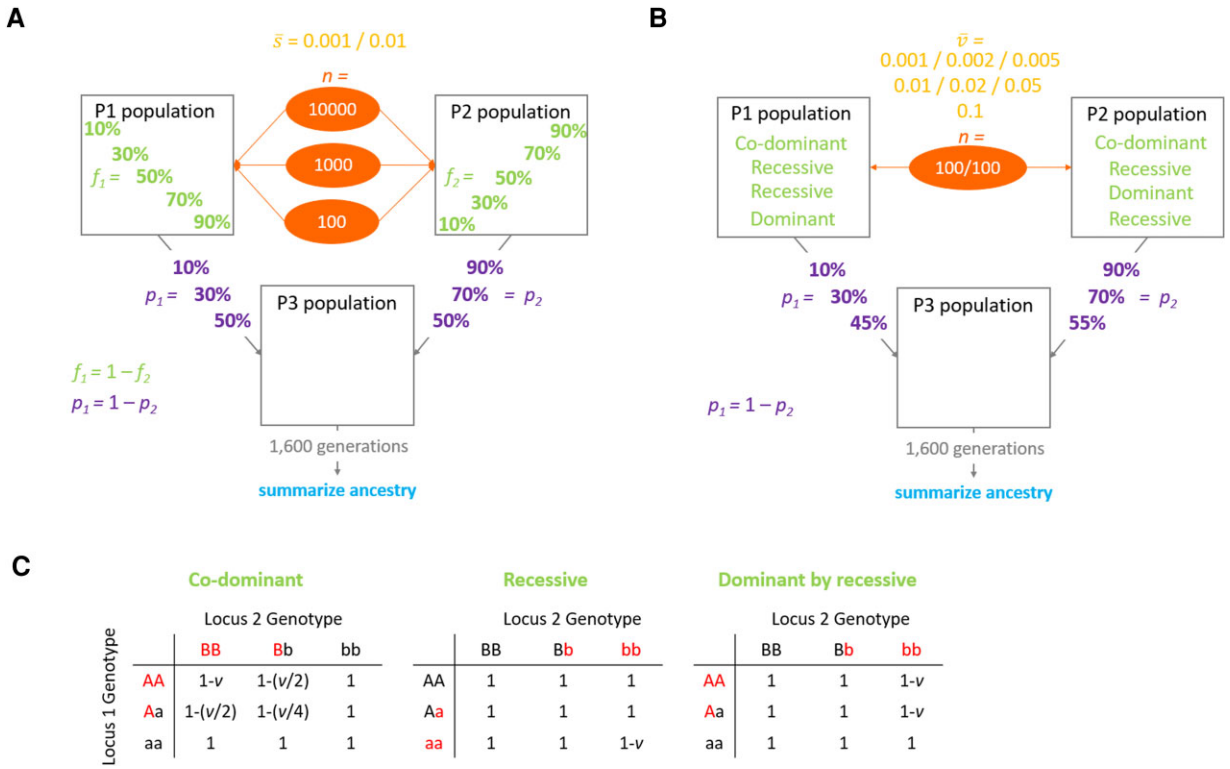
Directional Selection: General Patterns

Rather than assuming introgression to be uniformly beneficial or deleterious, we simulated a potentially more realistic model in which two source populations (P1 and P2) each contain a certain fraction of the total number of positively selected variants that will be favored in an admixed population (P3). Selected variants are assumed to act additively and to be fixed differences between source populations. Although we describe this as a model of positive selection, it should also be relevant to scenarios where each source population has accumulated deleterious variants at different loci (in which case the other population's allele is favored). Although our motivating case was from *Drosophila*, test simulations confirmed that we could scale down the population size to 10,000 with no meaningful influence on average genomic ancestry landscapes (supplementary fig. S1, Supplementary Material online). Sex ratio was also found not to have any notable impact on these results, and was otherwise fixed at 50% females/males.

We allowed P1 and P2 to admix in different proportions (p, with $p_1 = 1 - p_2$) to form P3. We also varied the average selection coefficient of the beneficial variants ($\bar{s}$), along with the total number of favored variants (n) and the fraction of favored variants present in each population (f, with $f_1 = 1 - f_2$) (figs. 1A and 2A). In scenarios with equal numbers of favored variants (i.e., same fitness, $f_1 = 50\%$) and unequal admixture proportions from each source population ($p_1 = 10\%$, 30%, 70%, or 90%), we found that negative correlations between recombination rate and minor population ancestry were consistently produced (fig. 2*B*, *C*, *F*, and *G* yellow; and likewise in supplementary fig. S2 and table S1, Supplementary Material online). Other than cases with the most pervasive selection ($\bar{s} = 0.01$ and $n =$ 10,000), all models with equal beneficial proportions and unequal admixture produced strongly negative Pearson's correlation coefficients between −0.475 and −0.814 for the autosomes, and generally similar results for the X chromosome (supplementary figs. S3 and S4, Supplementary Material online). These results are in line with the predictions described in Introduction—the rise of minor population alleles has a relatively greater impact on ancestry compared with favored major population alleles, and this effect is greatest in regions of low recombination.

In contrast, scenarios in which the admixing source populations contributed unequal numbers of beneficial variants produced more diverse correlations between recombination and minor population ancestry, with Pearson's r ranging from −0.786 to 0.705 for the autosomes and similar results for the X chromosome (supplementary figs. S3 and S4, Supplementary Material online). Positive correlations were observed in some scenarios if the ancestry that was minor at the end of the simulations was predominantly disfavored over time, as proposed by Martin and Jiggins (2017). This pattern was often observed if an initially major (or co-equally contributing) source population had a higher fitness (i.e., contributed most of the beneficial variants; fig. 2*F*–*H* light and dark blue; as in supplementary figs. S2*F*–*H*, S3, and S4, Supplementary Material online when $f_1 = 10\%$ or $f_1 =$ 30%), or if the minor population contributed most favored variants but the selection was pervasive enough to change it into the major ancestry by the time of sampling (supplementary fig. S2*F*–*H*, Supplementary Material online light and dark orange and supplementary fig. S3*C*, Supplementary Material online when $f_1 = 90\%$ or $f_1 =$ 70%). This switch in major/minor ancestries between the beginning and the end of the simulation was more likely to happen if selection was strong, and in regions of low recombination. Due to the autosomes' lower recombination rate than the X chromosome in the *D. melanogaster* genome, ancestry reversals happened relatively more frequently on the autosomes (see supplementary table S1 and figs. S3 and S4, Supplementary Material online when $p_1 = 90\%$, 70% and $f_1 = 10\%$, 30%). These results are all consistent with the general expectation that positively selected variants will bring more linked neutral variation with them in regions of low recombination (Maynard Smith and Haigh 1974). In addition, models of asymmetric fitness between admixing source populations may lead to synergistic frequency increases between beneficial variants from the more fit population, particularly in regions of lower recombination. In light of both of these factors, the more fit population's ancestry should increase more rapidly in low recombination regions, potentially generating a positive correlation between the minor ancestry proportion and recombination.

Nonetheless, we found some particular cases where even if the initially major population had a higher number of beneficial variants, there was a negative correlation between minor ancestry proportion and recombination, especially in cases with fewer but stronger favored variants. This is true for three models on the X and the autosomes:



C

Co-dominant

Locus 1 Genotype	Locus 2 Genotype: BB	Bb	bb
AA	1-*v*	1-(*v*/2)	1
Aa	1-(*v*/2)	1-(*v*/4)	1
aa	1	1	1

Recessive

Locus 1 Genotype	Locus 2 Genotype: BB	Bb	bb
AA	1	1	1
Aa	1	1	1
aa	1	1	1-*v*

Dominant by recessive

Locus 1 Genotype	Locus 2 Genotype: BB	Bb	bb
AA	1	1	1-*v*
Aa	1	1	1-*v*
aa	1	1	1

Fig. 1. Schematic representation of the models used in the simulations. Two source populations P1 and P2 are mixed in different proportions (p, admixture proportion) to form a third population P3 that evolves on its own. After 1,600 generations population P3 ancestry is analyzed to determine which one of P1 or P2 is the minority ancestry at the end of the simulations. (*A*) Model of directional selection where a certain number of positively selected variants (n, number of variants) is distributed between P1 and P2 populations in different proportions (f, fraction of beneficial variants). Variants' selection coefficients are drawn from an exponential distribution with a predetermined mean value ($\bar{s}$). (*B*) Model of pairwise incompatibility where the P1 population contains 100 fixed alleles each involved in a pairwise negative interaction with one of 100 alleles fixed within the P2 population. Each negative interaction reduces the fitness of the individual who carries it by a certain value (v) drawn from an exponential distribution with a predetermined mean value ($\bar{v}$). Three different models of dominance were simulated: co-dominant incompatibilities, recessive incompatibilities, and dominant-by-recessive incompatibilities (where either P1 or P2 contributes exclusively dominant incompatibility partner variants, and the other contributes only recessive variants). (*C*) Fitness effects of all potential two-locus genotypes are given for each incompatibility model investigated. Here, the variants that may experience an incompatibility are shown in red, while the universally compatible variants are in black.

when $n = 100$, $\bar{s} = 0.01$, and (A) $p_1 = 10\%$ and $f_1 = 10\%$, (B) $p_1 = 10\%$ and $f_1 = 30\%$, and (C) $p_1 = 30\%$ and $f_1 = 30\%$, see supplementary figs. 3*B* and 4*B*, Supplementary Material online) and for two other models only on the X chromosome (when $n = 1{,}000$, $\bar{s} = 0.01$, and [A] $p_1 = 10\%$ and $f_1 = 30\%$ *and* [B] $p_1 = 30\%$ *and* $f_1 = 30\%$, see supplementary fig. S3*D*, Supplementary Material online). These results may reflect the competition between processes favoring positive correlations between recombination and minor population ancestry (previous paragraph) and the unequal magnitudes of beneficial frequency changes that favor negative correlations (Introduction). To explore more deeply this particular scenario and see how the number of beneficial variants influences the correlation between minor ancestry and recombination, we simulated a wider range of variation for the number of beneficial variants for $p_1 = 10\%$, $f_1 = 30\%$, and $\bar{s} = 0.01$ (supplementary fig. S5, Supplementary Material online). We can see that in this model, the number of variants does not influence the sign of the correlation as it is always negative, but only influences its strength (supplementary fig. S5*B*, Supplementary Material online). However, when the number of beneficial variants became very high ($n > 1{,}000$), the major ancestry goes to fixation, and therefore, there is no correlation between minor ancestry and recombination (supplementary fig. 5*B*, Supplementary Material online). In this model, selection tends to overall favor the major ancestry that has a higher fitness, which on its own should generate a positive correlation. However, selection is also favoring some minor ancestry alleles, and neutral alleles associated with them are going to have a more important frequency change in low compared with high recombination regions, generating a negative correlation between minor ancestry and recombination. Finally, the small difference in correlation observed between the X chromosome and the autosome can be explained by the higher rate of recombination of the X chromosome.

Negative correlations between recombination and minor population ancestry were also observed if the initially minor population contributed more beneficial alleles but did not rise to become the major ancestry at the time of sampling (see fig. 2*F*–*H* light and dark orange and supplementary figs. 3*A* and 4*A*, Supplementary Material

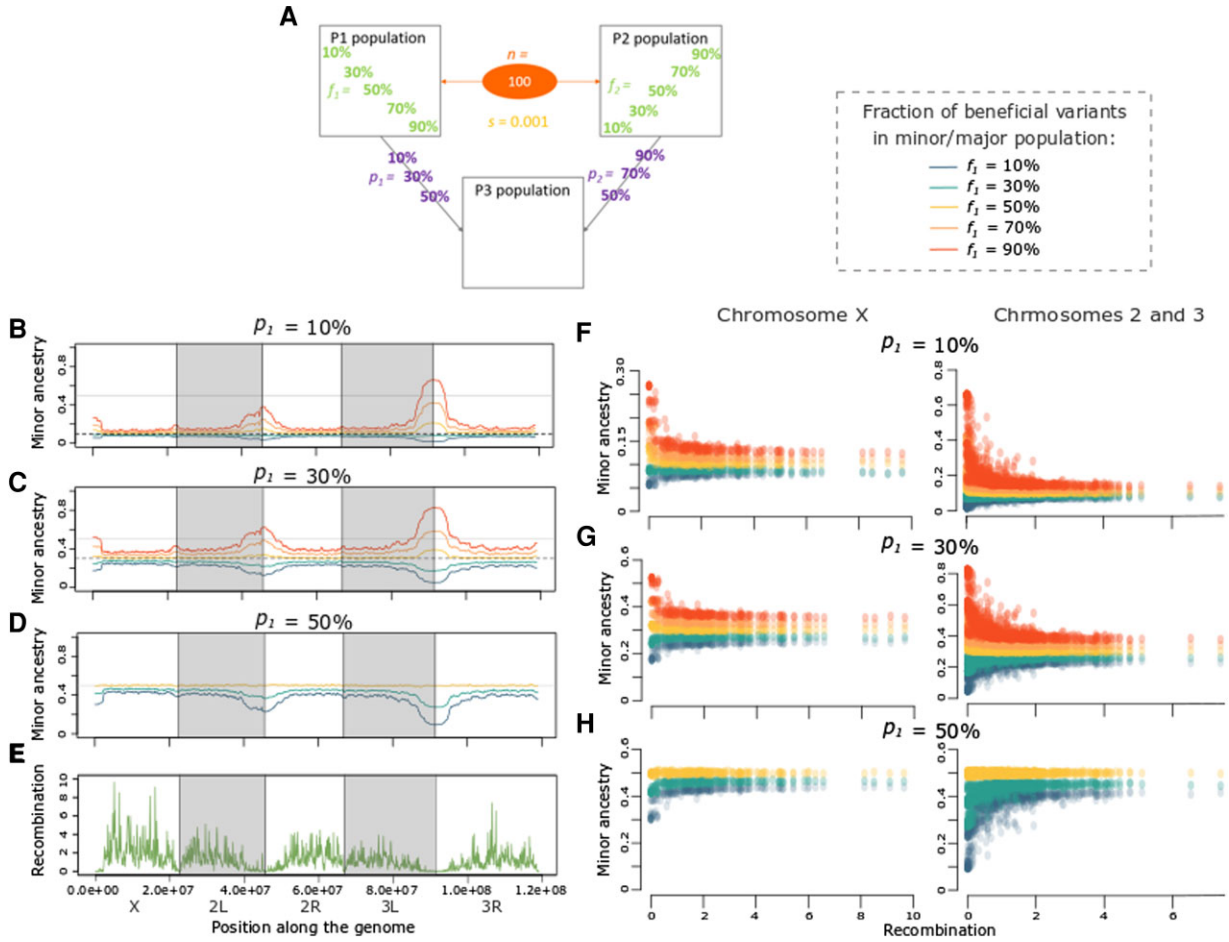


FIG. 2. Directional selection can produce negative or positive correlations between recombination rate and minor population ancestry. (*A*) Depiction of the specific parameter space of directional selection presented in the following panels, in terms of admixture proportions (p) and beneficial variant fractions (f) for cases with 100 selected variants (n) with average $\bar{s} = 0.001$. The left graphs depict the average minor ancestry proportion (across 1,000 replicates) along the genome when p_1 is: (*B*) 10%, (*C*) 30%, and (*D*) 50%. (*E*) Recombination rate (sex-averaged crossover frequency in cM/Mb) along the genome of *Drosophila melanogaster*. The right panels depict the relationship between minor ancestry proportion and recombination for the X chromosome and the autosomes when p_1 is: (*F*) 10%, (*G*) 30%, and (*H*) 50%. Colors in *B*–*D* and *F*–*H* represent the fraction of beneficial variants in P1 (f_1), as indicated on the right of *A*. All of the correlations shown in *F*–*H* are significant ($P < 0.05$) except for panel *H*'s X-linked $f_1 = f_2 = 50\%$ case (Supplementary figs. S3 and S4, Supplementary Material online).

online when $p_1 = 10\%$, 30% and $f_1 = 90\%$, 70%). Here, both differences in population fitness and the unequal magnitudes of beneficial frequency changes lead the minor population ancestry to increase, particularly in regions of low recombination, thus generating negative correlations. In summary, the correlations between recombination rate and minor population ancestry generated by positive selection in admixing populations are complex. Negative correlations are observed when major and minor source populations contribute equal numbers of beneficial variants, and also when the minor population contributes more beneficial variants but remains the minor ancestry. In contrast, positive correlations are usually generated if a major (or equally admixing) population contributes more beneficial variants, or if a favored minor population becomes the major population.

Number Versus Strength of Selected Sites

The above results suggest that the correlation between recombination rate and minor population ancestry holds information about the nature of selection in admixed populations. Such information may be useful in extending the evolutionary parameter inferences that are possible from population genomic data. However, some parameters may have superficially similar influences on population ancestry in high versus low recombination regions. Two evolutionarily important but potentially confounded parameters are the number versus strength of selected sites. Our simulations offer the possibility to compare models with the same genome-wide average level of selection ($n \times \bar{s}$) with different number of positively selected variants (fig. 3 and supplementary table S2, Supplementary Material online). Overall, the sign and strength of the correlation between minor ancestry and recombination are comparable between cases with the same average selection ($n\bar{s} = 1$, supplementary fig. S3*B* and *C*, Supplementary Material online) except for the particular scenario where the minor population has a low fraction of beneficial variants that we explored more deeply above. Therefore, the value of the correlation alone does not seem

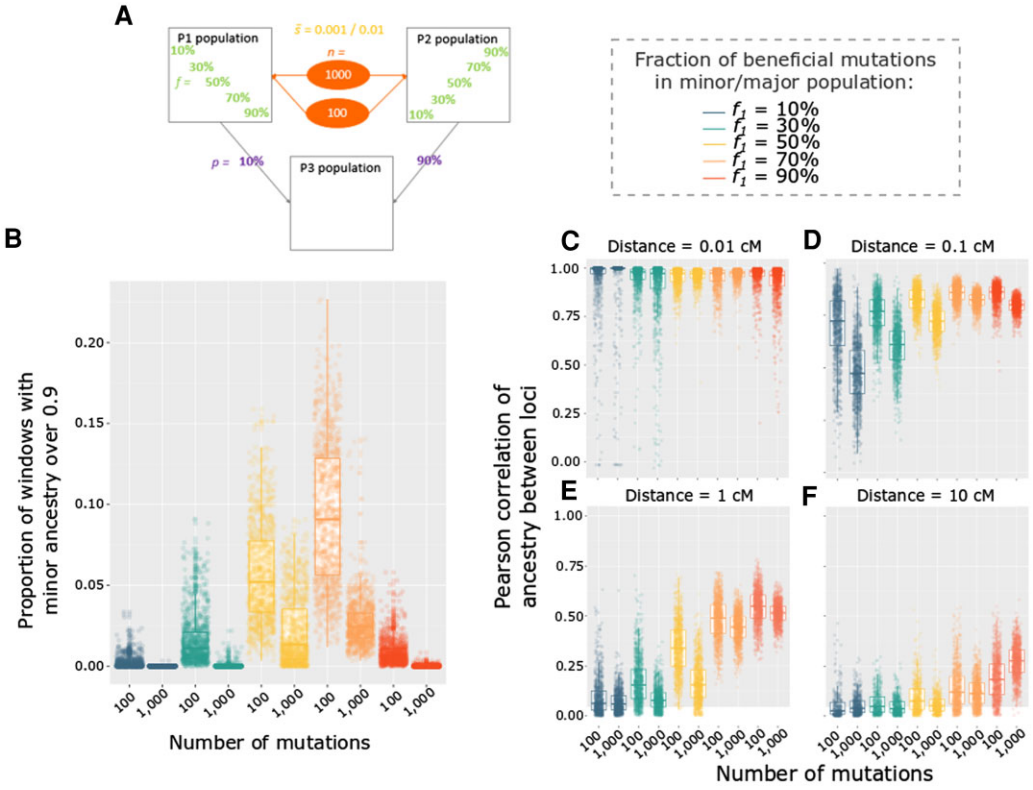


FIG. 3. Differences between models of directional selection with the same level of selection genome-wide but different numbers of selected variants. (*A*) Depiction of the specific parameter space of directional selection presented in the following panels: models with either $n = 100$ variants and an average $\bar{s} = 0.01$, or $n = 1{,}000$ variants and an average $\bar{s} = 0.001$, were examined for different beneficial variant fractions (f_1) for cases in which $p_1 = 10\%$. (*B*) Distribution among replicates of the proportion of genomic windows with minor ancestry above 0.9. Spatial correlations of ancestry along chromosomes, comparing between the same models as described above. Each point represents a replicate and measures of the Pearson correlation involving all 100 kb windows separated by a distance of: (*C*) 0.01 cM, (D) 0.1 cM, (*E*) 1 cM, and (*F*) 10 cM. Colors in *C*–*F* represent the fraction of beneficial variants contributed by P1 population (f_1), as indicated on the right of *A*.

to offer a promising way to distinguish between models with a low or high number of selected sites. We further investigated patterns of P2 ancestry along the genome for two models with an average summed selection coefficient of 1 (either $n = 100$ and $\bar{s} = 0.01$ or $n = 1{,}000$ and $\bar{s} = 0.001$), since models with more pervasive selection (i.e., summed selection coefficient of 10) often resulted in complete fixation of one ancestry (supplementary table S2, Supplementary Material online). We can see that population ancestry patterns tend to be smoother (meaning that there are less extreme values) when there is a smaller number of variants with on average a higher selection coefficient compared with a case with a higher number of variants with a low selection coefficient (fig. 3*B*). When we focus more precisely on the spatial correlation of population ancestry along the genome, we can see that at a short genetic distance, 0.01 cM, there is a very high spatial correlation for both models and we cannot distinguish the two kinds of model (fig. 3*C* and supplementary table S2, Supplementary Material online). In the same way, at a very high genetic distance, 10 cM, spatial correlations are very low and there are almost no differences between the models (fig. 3*F* and supplementary table S2, Supplementary Material online). However, in between these two distances, there is a clear tendency for spatial correlations to be higher for models with a lower compared with higher number of positively selected variants (fig. 3*D* and *E* and supplementary table S2, Supplementary Material online). It also seems that for both kind of models, correlations tend to be stronger as the proportion of P2 is higher. Therefore, the genetic distance at which it is easier to distinguish between few versus many variants seems to depend on the admixture proportion, meaning that future inferential studies may wish to consider spatial correlations of population ancestry along chromsosomes at multiple scales of genetic distance.

Number of Generations Since Admixture

To reflect the time since admixture of North American *D. melanogaster* populations, we initially focused on introgression patterns after 1,600 generations of admixture. In order to determine how the correlation between minor ancestry and recombination is influenced by the time since admixture, we also looked at shorter and longer times. Focusing on a particular model with $n = 100$ and $p_1 = 10\%$, we varied both the selection strength ($\bar{s} = 0.01$ or $\bar{s} = 0.001$) and proportion ($f_1 = 10\%$, 30%, 50%, 70%, and 90%) of favored variants present in each population

(fig. 4A and supplementary fig. 6A, Supplementary Material online).

Results showed that in most cases, the correlation tends to be stably either positive or negative through time (fig. 4B and C and supplementary fig. S6B and C, Supplementary Material online), reflecting the balance of factors described above. However, for some cases, the correlation can reverse over time, in two different ways. First, the correlation can start negative and become positive if there is a major ancestry switch over time (fig. 4B orange dots). In this particular case, the minor population has the larger proportion of beneficial variants and increases more rapidly in low recombination regions, creating a negative correlation. After ~400 generations, the minor ancestry becomes the major one and thus, the correlation is reversed. We note that this scenario did not produce an ancestry shift and correlation reversal for the X chromosome due to higher recombination rates (supplementary fig. S6B, Supplementary Material online). Second, the correlation can start positive and become negative through time, when the minor population has a small proportion of beneficial variants (fig. 4 and supplementary fig. S6, Supplementary Material online). Here, selection favoring the major ancestry will initially be more prominent, pushing its greater number of beneficial variants from their initial 90% frequency to fixation, with greater linkage effects in low recombination regions creating a positive correlation between minor ancestry and recombination. After the beneficial variants from the major population have largely been fixed, selection will continue to favor beneficial variants from the minor population as they rise from 10% frequency toward fixation. As before, linkage effects of this frequency shift will be greater in low recombination regions, but because the minor ancestry is now being favored, it will generate a negative correlation. Unlike the opposite correlation shift, this positive to negative transition was observed in at least one case for both selection coefficients and for both the autosomes and the X chromosome (fig. 4 and supplementary fig. 6, Supplementary Material online). With stronger selection ($\bar{s} = 0.01$), correlation shifts were associated with a smaller minority of favored variants from the minor population (10%, vs. 30% with $\bar{s} = 0.001$). As recombination is higher along the X chromosome and thus selection is more efficient, the switch from a positive to a negative correlation happened sooner.

Pairwise Incompatibilities

In our second model, both source populations contain 100 fixed variants that are each involved in pairwise negative epistatic interactions with one of the 100 variants fixed in the other population (fig. 1B and C). As for the directional selection model, we confirmed that population size could be scaled down to 10,000 and that skewed sex ratios had little effect on ancestry patterns (supplementary fig. 7, Supplementary Material online). We simulated three different dominance cases: co-dominant, recessive, and dominant by recessive (in which one population contributes only dominant partner variants and the other recessive; fig. 1C). We then varied the average fitness impact of the incompatibility (where $\bar{v}$ is the mean of exponentially distributed values). Finally, as for the directional selection model, we varied the admixture proportion of the two populations (p) (fig. 4A).

As expected, we showed that in most cases, there is a positive correlation between the minor ancestry proportion and recombination (fig. 5, supplementary table S3 and figs. S8 and S9, Supplementary Material online). Indeed, neutral introgressed variants are more likely to recombine away from incompatible loci in highly recombining regions and therefore to persist there. For example, recessive incompatibilities produced fairly strong positive correlations with autosomal Pearson's r values between 0.493 and 0.619 (supplementary fig. 8D, Supplementary Material online). However, for one model in particular—when the major population only carries recessive variants and the minor one only dominant variants—the relationship between the minor ancestry proportion and recombination is sometimes negative (fig. 4E and J, supplementary table S3 and figs. 8C and 9C, Supplementary Material online). For example, on the autosomes, when the dominant population starts at 30% and the average fitness effect is between 0.001 and 0.01, there is a negative relationship, with Pearson r between −0.509 and −0.578 (fig. 4E and J; supplementary fig. S8C, Supplementary Material online). When the fitness effect is stronger, the minor population at the end of the simulation is no longer the dominant population but the recessive one, and in that case, the relationship is positive (supplementary fig. 8C, Supplementary Material online). For this same 30% dominant model on the X chromosome, there is also a band of incompatibility fitness effects that produce negative correlations (Pearson r between 0.393 and 0.650), but it is shifted toward stronger incompatibilities (supplementary fig. 9C, Supplementary Material online). For the X only, we find that there is also a range of parameter space generating negative correlations when the initial admixture proportion of the dominant population is 45% (supplementary fig. 9C, Supplementary Material online), whereas for the autosomes, a major ancestry switch prevents a negative correlation (supplementary fig. S8C, Supplementary Material online), underscoring the complementary signals of autosomal and X-linked data.

To try to understand why there is a negative relationship in the above parameter space, we used two-locus analytical models to predict the frequency change of two autosomal loci involved in an incompatibility with one allele being dominant and the other one recessive. We fixed the fitness reduction due to the incompatibility to 0.01 whether the dominant allele is homozygous or heterozygous, and took into account the probability that either one or both partners may be on the X chromosome. We did these calculations for different proportions of admixture and determined after 1,600 generation which ancestry

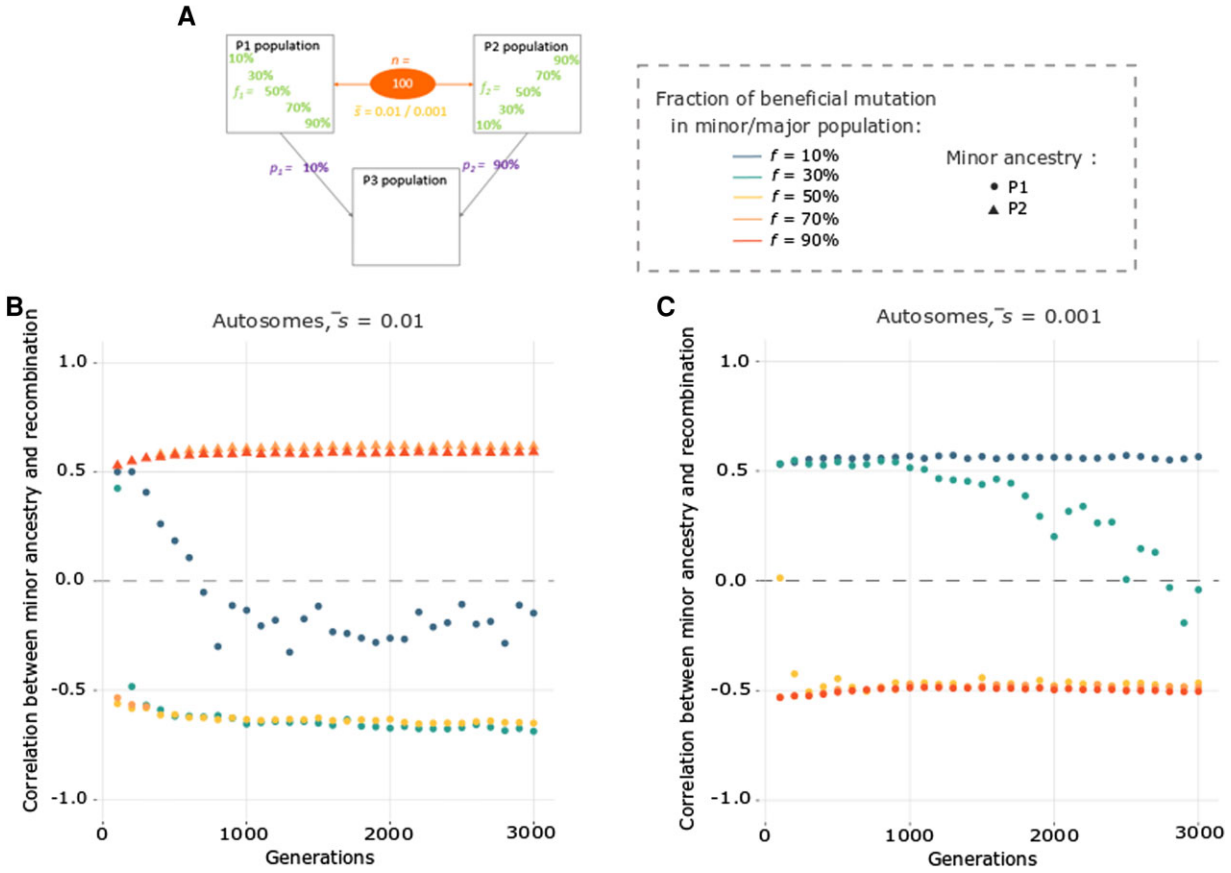


FIG. 4. Correlation between minor ancestry and recombination is usually stable through time. (*A*) Depiction of the specific parameter space of directional selection for which we looked at the correlation between minor ancestry and recombination after different admixture times (every 100 generations from 100 to 3,000) in terms of admixture proportions ($p_1 = 10\%$), beneficial variant fractions (f_1), number of selected variants ($n = 100$), and average selection coefficient ($\bar{s} = 0.01$ or $\bar{s} = 0.001$). (*B* and *C*) Correlation between minor ancestry and recombination for the autosomes, depending on the time since admixture when (*B*) $\bar{s} = 0.01$ and (*C*) $\bar{s} = 0.001$. Colors of the points represent the fraction of beneficial variants contributed by each source population and their shape denotes the minority ancestry at the end of the simulations, as indicated on the right of *A*. Corresponding X chromosome results are given in supplementary figure S6, Supplementary Material online.

(the dominant or recessive one) had increased during this time (fig. 6). Indeed, similar to the model of directional selection, there was a positive correlation between minor ancestry and recombination when the minor ancestry decreased over time and a negative one when it increased (unless the major ancestry switched). We confirmed that (with a fitness reduction of 0.01) there is indeed a particular parameter space where the recessive population starts between 60% and 70% and is still the major population at the end of the simulations even if its ancestry tends to decrease. This means that during the 1,600 generations, the dominant ancestry tends to increase but not enough to become the major population, which explains why there is a negative relationship between the minor ancestry proportion and recombination. If the recessive ancestry starts in a lower proportion (below 60%) then it is decreasing over time and remains the minor population explaining why there is a positive correlation between recombination and introgression (fig. 6). If the recessive ancestry starts in a higher proportion (over 70%) then it is increasing over time, meaning that the minor ancestry proportion is decreasing and thus generates a positive correlation between the minor ancestry proportion and recombination (fig. 6).

These results can be explained by the interaction between the probability for an allele to be in a deleterious genotype and the absolute frequency shift that it undergoes. For example, in a model with the recessive ancestry starting at 10%, the dominant allele is going to be in a deleterious genotype 1% of the time, whereas the recessive one is 10% of the time. Purifying selection is thus stronger on the recessive allele, and its (minority) ancestry will decrease more strongly in low recombination regions, leading to a positive correlation between introgression and recombination (fig. 6). However, when the recessive population starts at 70%, the recessive allele is deleterious 36% [$0.7 \times (1–0.7^2)$] of the time and the dominant one is 49% (0.7^2), and yet overall, the (minority) dominant ancestry has increased, leading to a negative correlation. Here, we suggest that the same two competing factors are at work as discussed for the directional selection model, where in this case, the minority (dominant) ancestry

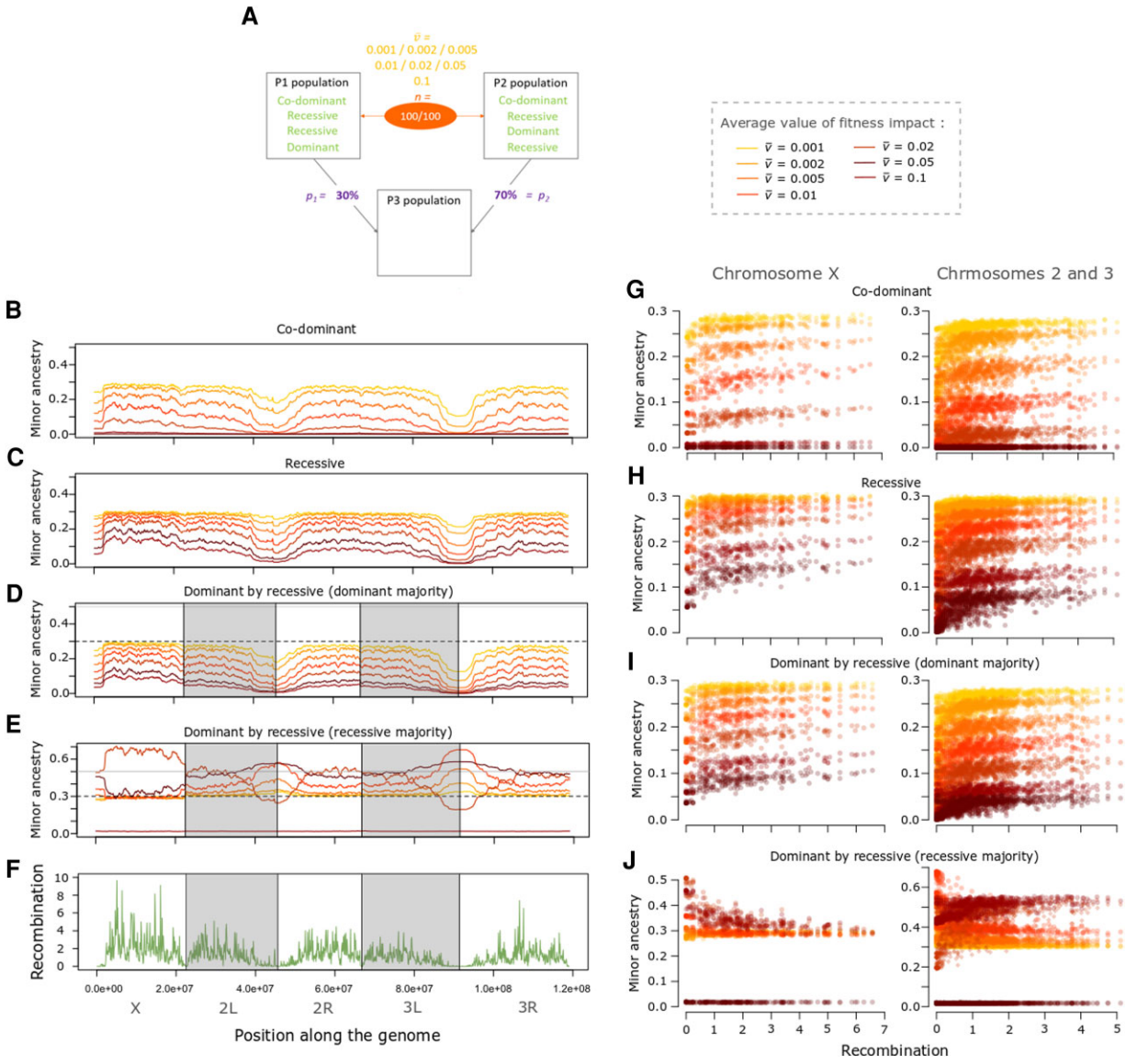


Fig. 5. Incompatibilities usually generate positive correlations between recombination rate and minor population ancestry. (*A*) Depiction of the specific parameter space of the incompatibility presented in the following panels, for an admixture model in which $p_1 = 30\%$, and there are 100 pairs of incompatible variants (n). Results are shown for several mean fitness effects ($\bar{v}$) and dominance scenarios. The left graphs depict the average minor ancestry proportion (across 1,000 replicates) along the genome when incompatibility fitness effects are: (*B*) co-dominant, (*C*) recessive, (*D*) dominant in the major population and recessive in the minor, or (*E*) recessive in the minor population and dominant in the major. (*F*) Recombination rate (sex-averaged crossover frequency in cM/Mb) along the genome of *D. melanogaster*. The right panels depict the relationship between minor ancestry proportion and recombination for the X chromosome and the autosomes when variants are: (*G*) co-dominant, (*H*) recessive, (*I*) dominant in the major population and recessive in the minor, or (*J*) recessive in the major population and dominant in the minor. Colors in *B*–*D* and *F*–*H* represent the mean fitness effect of incompatibilities, as indicated on the right of *A*.

increases because the greater magnitude of its frequency change outweighs its lesser mean fitness. In this case, the greater frequency reduction experienced by recessive incompatible alleles in this scenario may outweigh their lesser fitness consequences on average. A further consideration is that both incompatible variants decrease in frequency, selection against the minority dominant variants will weaken more quickly, as many fewer homozygous recessive genotypes will be created at the partner locus. For other initial frequencies, there is also the possibility to have a switch of major/minor ancestry between the beginning and the end of the simulations. When the recessive population starts at 55%, the recessive allele is deleterious 38% of the time and the dominant one 30%. Here, the dominant ancestry increased enough to become the major ancestry, and there is a positive correlation between introgression and recombination (fig. 6).

Discussion

In this paper, we aimed at understanding how patterns of introgression along the genome are influenced by forms of positive and negative selection. We simulated two populations of *D. melanogaster* that are mixed in different

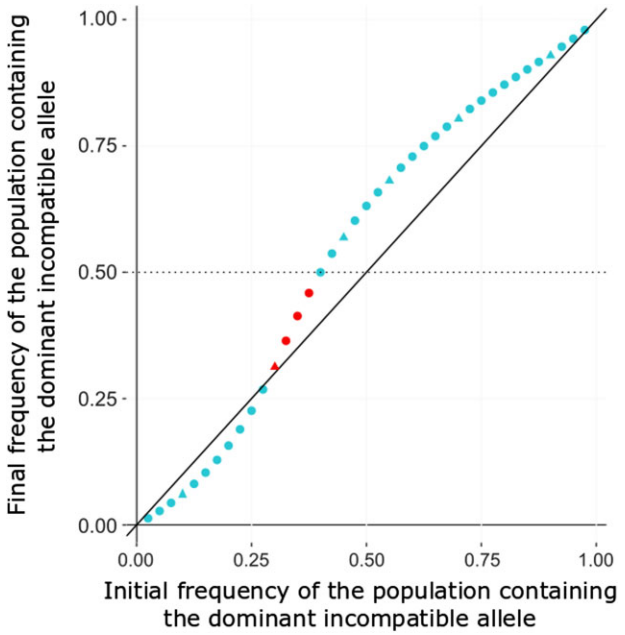


Fig. 6. Theoretical expectation of the mean two-locus ancestry proportion for the population contributing dominant incompatible variants under a dominant-by-recessive model. In this plot, the *x*-axis represents the initial admixture proportion for the population contributing the dominant incompatibility partner allele, and the *y*-axis is its predicted ancestry proportion after 1,600 generations of admixture (averaged across both incompatibility loci). Hence, the final frequency depicted depends on both selection against the dominant incompatibility allele and selection against the other population's recessive allele at the incompatibility partner locus. The black line of equation $x = y$ indicates whether the dominant-contributing population's ancestry proportion has increased or decreased over time. Results were calculated using a simple two-locus fitness model in which incompatible two-locus genotypes led to a relative fitness reduction of 0.01. Either or both partner loci could fall on the X chromosome with probability 20%, and were otherwise autosomal. No linkage between partner loci was modeled. Triangles represent frequencies that were included in our simulation study and red points indicate scenarios in which the predicted correlation between minor ancestry and recombination is negative.

proportion to form a third one. The source populations contained fixed variants that were either unconditionally positively selected or else involved in pairwise negative epistatic interaction with the variants fixed in the other source population. We showed that positive selection can sometimes generate negative correlations between the minor ancestry and recombination and that incompatibilities generally result in positive correlations, but that for either model, the correlation can be reversed in some scenarios.

Directional Selection

While non-neutral introgression is sometimes modeled as being universally beneficial or deleterious, our directional selection simulations allowed for potentially competing selection on unconditionally favored variants from both source populations. This model may be applicable to cases of adaptive differentiation between source populations, and potentially also cases of accumulated deleterious variants between source populations (although we do restrict our attention to additive selection). Under this model, we showed that when major and minor source populations contribute equal numbers of beneficial variants to an admixed population, negative correlations between recombination rate and minor ancestry proportion are generated (supplementary figs. S3 and S4, Supplementary Material online). We attribute this observation primarily to the greater ancestry frequency change entailed by fixing a minor population allele compared with a major population allele, and the greater impact of this imbalance on neutral variation in regions of low recombination.

In contrast, when there is a fitness difference between the two source populations (meaning here a difference in the fraction of variants present in the population f), we usually recover the expected relationship between minor ancestry and recombination (supplementary figs. S3 and S4, Supplementary Material online). The population with the higher fitness is more often under positive selection, and so its ancestry increases in frequency over time (particularly in low recombination regions), whether it started as the major or minor contributor. Indeed, increasing the number of beneficial variants or the average value of the selection coefficient mainly increases the probability to have a switch between the minor ancestry at the beginning and the end of the simulations. Hence, either positive or negative correlations can be generated in cases of asymmetric fitness. Interestingly, we demonstrate that these results are not usually influenced by the time since admixture, as the correlation tends to be stable through time, except for particular cases where there is an ancestry switch or under particular parameters when selection tends to favor the major ancestry first and then the minor (fig. 4 and supplementary fig. S6, Supplementary Material online).

These results are comparable to what was observed in the presence of additive deleterious variants (Kim et al. 2018). When the receiving population has a lower fitness than the donor, introgressed alleles can be favored in low recombination regions (Kim et al. 2018), whereas this phenomenon is reversed if introgression is from a less fit into a more fit population. Thus, positive and negative correlations between minor ancestry and recombination can be generated by both the presence of deleterious and beneficial variants, depending on the relative fitness of the admixing population.

However, not all results can be explained by fitness differences between the two admixing populations. For example, when deleterious variants are recessive, an increase of introgression can be observed in low recombination regions when populations have the same fitness and even if the donor population has a lower fitness than the receiving one (Kim et al. 2018). This is explained by the fact that introgressed haplotypes generate associative overdominance by preventing the expression of recessive deleterious variants (Ohta and Kimura 1970; Kim et al. 2018).

As introgressed blocks are longer in low recombination regions, negative recombination–introgression correlations are produced. In contrast to the deleterious model, we find that a positive selection model can generate negative recombination–introgression correlations without fitness asymmetry or recessiveness. Although we have focused on additively beneficial variants, future studies may wish to examine the interplay between dominance and positive selection in admixing populations as well.

Pairwise Incompatibilities

In our model with source populations each containing 100 fixed variants involved in incompatibilities, we showed that when all variants are co-dominant or recessive, we always recover the expected correlation between minor ancestry and recombination, that is a positive one (supplementary figs. S8*A*, *D* and 9*A*, *D*, Supplementary Material online). Indeed, introgressed neutral variants can more easily recombine from those involved in incompatibilities and thus segregate in highly recombining regions. We also tested another model of dominance, where one population only contains dominant variants and the other only recessive variants. This particular scenario might happen in nature if one population went through a particular event of local adaptation and assuming that adaptation tended to favor either dominant or recessive variants. Indeed, alleles of local adaptation may sometimes be involved in genetic incompatibilities, especially if the two populations have been evolving in different environments. We find that in the particular parameter space in which the population containing the recessive variants starts as the major contributor, the correlation between minor ancestry and recombination can actually be negative (fig. 3*E*, *J* and supplementary figs. S8 and S9, Supplementary Material online).

To complement our simulation results, we also used mathematical predictions to estimate the expected frequency of each allele involved in the two-locus incompatibility for each of the 1,600 generations of our simulated admixture case to understand how this negative correlation is generated. We found that when the recessive population starts between 0.6 and 0.7, a negative correlation is indeed expected (fig. 6). Indeed, in this particular model of dominance, there is an interaction between the probability of an allele to be in a deleterious genotype (which determines the strength of purifying selection on it) and the frequency of this allele. And yet, even in some cases in which a (dominant) minority allele is deleterious on average (which would normally entail a positive correlation), we find a negative correlation instead. Here, we suggest that just with the directional selection model, the difference in the magnitude of frequency changes encountered by majority versus minority alleles can sometimes outweigh the mean fitness of each ancestry and reverse the typically expected correlation. In this case, the majority population's incompatible variants have farther to fall. And as these recessive majority alleles decrease, selection against the minority population's dominant incompatibility will be created. This correlation reversal parameter space is probably too particular to be widespread in nature. Nonetheless, the variability and magnitude and even direction of recombination–introgression correlations we observed under incompatibility models further underscores the complex task of making inferences about selection models in admixed populations.

What About the Particular Case of *D. melanogaster*?

Qualitatively, we now know of three types of selection in admixed populations that are capable of generating a negative correlation observed between the minor ancestry proportion and recombination rate such as that observed in a North American DGRP population of *D. melanogaster* (Pool 2015). As previously demonstrated by Kim et al. (2018), an accumulation of deleterious variation between populations can lead to this outcome, especially if this variation is recessive and/or introgression is from a larger population into a smaller one. Importantly, the simulations of Kim et al. involved relatively smaller populations ($N_e = 10{,}000$) that had been separated for $2N_e$ generations before introgression occurred. Whereas in the case of our fly population, not only have effective population sizes been generally much larger (aside from a moderate out-of-Africa bottleneck), which should selection against deleterious variation, but also the divergence between the African and European source population of the DGRP only occurred within the past $0.01N_e$ generations (Sprengelmeyer et al. 2020). Hence, it is not clear that a sufficient quantity of deleterious variation should be expected to have fixed between these fly populations on such a time scale, in order for this model to generate a strong negative correlation.

Our study has identified two other ways in selection might generate a negative correlation. Regarding selection against deleterious incompatibilities, we identified a particular scenario involving consistent differences in the dominance of incompatibility alleles between source populations. Incompatibilities may well exist between the African and European source populations of the admixed DGRP population. European and African populations of *D. melanogaster* show signs of partial reproductive isolation (Wu et al. 1995; Yukilevich and True 2008; Lachance and True 2010; Kao et al. 2015), and a surprising abundance of "ancestry disequilibrium" between unlinked loci in the DGRP population (Pool 2015) may point to the presence of epistatic selection of this nature. The dominance asymmetry required for this model to generate a negative correlation is conceivable—for example, if incompatibilities were primarily due to adaptation in the European source population's lineage, and this adaptation mainly involved recessive variants. However, it is not clear that incompatibilities should primarily be of such a type.

Finally, we show that across a somewhat broader parameter space, models of positive selection can generate negative correlations between recombination rate and

minor population ancestry. In the context of the DGRP population, where it appears that admixture occurred between a majority European source population and a minority African source population, a negative correlation would probably be expected under positive selection under the simplest scenario in which these source populations contributed equal numbers of beneficial variants. The same prediction would also hold if a greater share of beneficial variants came from the African source population. Hence, the positive selection hypothesis may be the strongest candidate to explain the strong negative correlation observed in the DGRP fly population (where African ancestry ranges from 30% in low recombination regions to 10% in high recombination regions). However, we must emphasize that our study is not designed to identify a specific causative model and that these different mechanisms are not mutually exclusive and they may all play a role in the observed relationship between recombination and ancestry.

Conclusion

We have shown that positive selection in admixed populations can generate negative correlations between recombination and minor population ancestry, but that this correlation can be reversed if positive selection generally favors major population alleles or leads to a change in the major ancestry. These results are somewhat analogous to the findings of Kim et al. (2018) for models involving deleterious variation, which was found to produce mainly negative correlations but with some exceptions. However as outlined above, the mechanisms underlying the predictions of these two directional selection models are somewhat different. We also confirm that positive correlations between recombination and introgression are generally expected from pairwise incompatibility models, but we find that here too, there are exceptions (specifically involving asymmetric dominance between source populations). Our results build on previous research to establish a more complete picture of the effects of selection on recombination–ancestry relationships (fig. 7).

Given that any of the three models discussed above can generate either positive or negative correlations between recombination and ancestry, one obvious conclusion is that the direction of such a correlation is not sufficient to infer a causative selection model. In some cases, aspects of species biology and population history may inform the plausibility of different models (e.g., degree of reproductive isolation, effective population size, population genetic time scale of isolation), as discussed above in the case of *D. melanogaster*. Otherwise, we suggest that future inferential studies will need to consider not only quantitative characteristics of this relationship, but also complementary aspects of genomic variation in admixed populations. Our analysis of spatial correlations in ancestry proportion (in the context of separating the number and strength of selected sites) provides just one example of such an approach. Our results also imply that joint consideration of autosomal and X-linked correlations may improve model inference. Further work is needed to identify additional features of ancestry variation in admixed populations that help to differentiate the above models. Nevertheless, the above findings bring new perspectives on how positive and negative selection interacts with recombination to shape hybrid genomes and suggest promising directions for estimating the parameters involved.

Materials and Methods

Basic Structure of the Model

To study the relationship between introgression and recombination, we simulated admixture between two populations using forward simulations in SLiM 3.5 (Haller and Messer 2019), assuming different selective effects. All files and scripts needed to run the simulations can be found at https://github.com/Mduranton/simulations_introgression_recombination. We wanted our model to reflect admixture between the European and African source populations of *D. melanogaster* in North America, as represented by genomic data from the *Drosophila* Genetic Reference Population (Mackay et al. 2012). Therefore, we modeled the structure of *D. melanogaster* genome (119 Mb in total), focusing on the two major autosomes (2 and 3) and the sex chromosome (X), using the recombination map estimated by Comeron et al. (2012) with recombination only occurring within females. In the first generation before admixture, two populations with a sex ratio of 0.5 are created. Then depending on the model, a certain number of fixed variants are randomly positioned along the genome of each population. As the X chromosome represents ~20% of the total genome size, on average 20% of variants occur on the X chromosome and the rest randomly fall along chromosomes 2 and 3. Since all variants are fixed, the size of the two source populations does not matter; thus, to improve computational efficiency, we choose to only start the simulations with 10 individuals in each population. In the second generation, a third population of 10,000 individuals with a sex ratio of 0.5 is created from a certain fraction of population P1 and P2, depending on the model simulated. For computational feasibility, we choose to use only 10,000 individuals, but we also simulated different population sizes to confirm that this parameter has no noticeable influence on the relationship between introgression and recombination (see supplementary figs. S1*A*, *B*, *D*, *F* and S7*A*, *B*, *D*, *F*, Supplementary Material online). In the same way, as the actual sex ratio of *D. melanogaster* is not clearly identified, we simulated different sex ratios to confirm that sex ratio does not strongly influence the relationship between recombination and introgression (see supplementary figs. S1*A*, *C*, *E*, *F* and S7*A*, *C*, *E*, *F*, Supplementary Material online). On the third generation, population P1 and P2 are removed. Finally, population P3 evolves on its own during 1,600 generations, which is approximately the time since admixture started between the European and African population of *D. melanogaster* in North America (Pool 2015;

Model:	General Correlation:	Exceptions:	Sources:
Incompatibilities	Positive	Major/minor source populations contribute recessive/dominant alleles	This study; Martin & Jiggins, 2017
Positive Selection	Negative	Major source population favored; or minor population becomes majority ancestry	This study
Genetic Load	Negative	Load is additive and introgression is from smaller into larger population (as with introgression from Neanderthals into modern humans)	Kim *et al.*, 2018 Harris & Nielsen, 2016 Juric *et al.*, 2016

Fig. 7. General relationships between minor population ancestry and recombination rate depending on the selection model, and known exceptions.

Corbett-Detig and Nielsen 2017). As we are only interested here in how the fixed variants present in the two first populations impact the relationship between introgression and recombination, the mutation rate was set to zero for all our simulations. We simulated 1,000 whole-genome replicates for each set of parameters.

In order to evaluate the level of introgression along the genome, we tracked ancestry along the genome using a feature of SLiM, tree sequence files, that record genealogies all along the simulations (Haller et al. 2019). Tree sequence files were then analyzed using a custom python script and tools from pyslim v0.600 (Kelleher et al. 2018) and msprime v0.7.4 (Kelleher et al. 2016) that allow to trace the ancestry of different portions of the genome back to population P1 or P2 for each individual. In order to have the same effective population size along the whole genome, we only analyzed ancestry within females (the homogametic sex in *D. melanogaster*). Ancestry estimates were then averaged across individuals and replicates. We focused on the relationship between recombination and the minor ancestry proportion, meaning the minority ancestry at the end of the simulations, to reflect what can be observed in natural populations where the initial admixture proportions are not known.

Directional Selection

We created a model with only positively selected variants, with four different parameters (fig. 1A). The mean of the exponential distribution of selection coefficient, $\bar{s}$, took two different values: 0.001 or 0.01. The number of variants present in the model, n, was either 100, 1,000, or 10,000. In one model, we specifically used a wider range of variants number to see how this could influence the correlation between minor ancestry and recombination; here, we also simulated $n = 10$, 20, 50, 200, 500, 2,000, and 5,000. The proportion of variants fixed in population P1, f_1, was, respectively, either 10%, 30%, 50%, 70%, or 90%. Lastly, p_1, the proportion of population P1 contributing to the creation of population P3, was, respectively, either 10%, 30%, or 50%. When variants were (randomly) positioned along the genome, they were assigned a selection coefficient sampled from the exponential distribution previously defined. Depending on the parameters selected, the population contributing the most to population P3 could either have a higher or lower number of positive variants than the population that contributed the least. The combination of all the different parameters represents 78 different scenarios that were all replicated 1,000 times. In order to see how the correlation between minor ancestry and recombination evolves through time, we chose specific models for which to examine admixture patterns after different times since admixture (i.e., every 100 generations from 100 to 3,000 generations since admixture).

Pairwise Incompatibilities

In this model, populations contained fixed, otherwise neutral variants that are involved in negative epistatic interactions and therefore deleterious when present with the interacting allele (fig. 1B). Incompatibilities were always generated between one allele fixed within population P1 and an allele fixed in population P2, in order to model BDMIs (Dobzhansky 1982; Muller 1942; Coyne and Orr 2004). For example, suppose population P1 is fixed for ancestral allele "A" and derived allele "b" at the first and second locus, respectively, whereas population P2 is fixed for derived allele "a" and ancestral allele "B." Alleles "a" and "b" that have never been in the same genotype can prove to be deleterious when interacting. In our model, each population had 100 variants of this kind, generating 100 pairwise incompatibilities. As variants were randomly positioned along the genome, incompatible loci were not necessarily present on the same chromosome and incompatibilities could also be generated between an autosome and the X chromosome. There were three different parameters in our model. The average of the exponential distribution of the fitness reduction generated by the pairwise

incompatibility, $\bar{v}$, took seven different values: 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, or 0.1. The proportion of population P1 and P2 contributing to the creation of population P3, p, was either 10–90%, 30–70%, or 45–55%. In this model, population P1 was always the population that initially contributed the least to P3 population. Finally, d represents the dominance, as we simulated three different types of incompatibility: models where all incompatible alleles are co-dominant, or all recessive, or where one population contained only dominant alleles and the other only recessive ones (fig. 1C). For the last model, the population containing recessive alleles could either be the major or minor contributor to the admixed population. For each pairwise incompatibility, a value of fitness reduction (v) was sampled from the exponential distribution previously defined. The fitness of each individual was then calculated depending on its genotype. For the co-dominant model, individuals that were homozygous at both incompatible loci (aabb) had a fitness of $1 - v$, individuals that were homozygous at one locus and heterozygous at the other (Aabb or aaBb) had a fitness of $1 - \frac{1}{2}v$, and individuals that were heterozygous at both loci (AaBb) had a fitness of $1 - \frac{1}{4}v$. For the recessive model, only individuals that were homozygous at both loci (AAbb) had a fitness reduction equal to v. Finally, for the dominant by recessive model, individuals that were homozygous with the incompatible allele at the recessive locus have a fitness reduction equal to v if they presented one or two incompatible alleles at the dominant partner locus (AAbb or Aabb if A and b are the dominant and recessive incompatible variants, respectively). The combination of all the different parameters represents 84 different scenarios that were all replicated 1,000 times.

We also generated mathematical predictions of the evolution of alleles' frequency under the dominant-by-recessive model. To do so, we calculated the expected frequency of two unlinked variants, one dominant and one recessive, after 1,600 generations of admixture considering different initial frequencies and a fitness reduction of 0.01 for incompatible two-locus genotypes. Mirroring our simulations, we assume there is 20% chance that each variant is located on the X chromosome, and the sex ratio is 0.5. We then estimated the frequency shift for each of four possibilities (both partners autosomal, recessive autosomal by dominant X-linked, dominant autosomal by recessive X-linked, and both X-linked) and then estimated an average final allele frequency weighted by the probability of each scenario.

Analyses of Ancestry

In order to determine if the correlation between minor ancestry proportion and recombination is positive or negative, we measured the correlation coefficient between recombination rate and introgression for each scenario using the minor ancestry proportion average across the 1,000 replicates. The minor ancestry was measured as a proportion between 0 and 1. Recombination rate values averaged between the sexes, estimated in cM/Mb for 100 kb windows, were obtained from Comeron et al. (2012). We used the Pearson product–moment correlation and its 95% confidence interval. In addition, knowing that the correlation is not linear, we measured for each replicate, the ratio between the average ancestry of the 50 lowest recombination windows over the average ancestry of the 50 highest recombination windows. We then looked at the distribution of this ratio to obtain its median and 95% confidence interval.

To distinguish between models with the same selection coefficient on average ($\bar{s} \times n$) but with a different number of variants, we compared for different models the level of genomic spatial correlation of ancestry at the end of the simulations (i.e., the correlation of individual ancestry between two windows a given genetic distance apart). We chose to measure this correlation for different genetic distances (0.01, 0.1, 1, and 10 cM) to see if there is an optimal distance to distinguish between a model with few variants and a high selection coefficient versus a model with more mutation and a low selection coefficient. To do so, we used a custom script that scans chromosome by chromosome and identifies for each 100 kb window a partner window that is at the closest cM distance to the one we defined. We allowed for a difference from the desired distance not greater than half of the target distance. Once all the pairs of windows were identified, we measured for each of a model's replicates the correlation between all the pairs using a Pearson correlation, which gave us a distribution of correlation coefficients for each scenario.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank the University of Wisconsin-Madison Center for High Throughput Computing (CHTC) for computational resources and assistance. This research was funded by US NIH grant R35 GM136306.

Data Availability

SLiM code used to generate the simulations for this paper are avilable in a GitHub repository (https://github.com/Mduranton/simulations_introgression_recombination).

References

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

Arnegard ME, McGee MD, Matthews B, Marchinko KB, Conte GL, Kabir S, Bedford N, Bergek S, Chan YF, Jones FC, *et al.* 2014 Genetics of ecological divergence during speciation. *Nature* **511**:307–311.

Barton N, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity* **57**:357–376.
Bateson W. 1909. Heredity and variation in modern lights. In: Seward AC, editor. *Darwin and modern science*. Cambridge University Press. p. 85–101.
Bergland AO, Tobler R, González J, Schmidt P, Petrov D. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol Ecol*. **25**:1157–1174.
Bierne N, Lenormand T, Bonhomme F, David P. 2002. Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow? *Genet Res*. **80**:197–204.
Blanckaert A, Bank C. 2018. In search of the Goldilocks zone for hybrid speciation. *PLoS Genet*. **14**:e1007613.
Caracristi G, Schlötterer C. 2003. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol* **20**:792–799.
Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**:1289–1303.
Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genet*. **8**: e1002905.
Corbett-Detig R, Nielsen R. 2017. A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLOS Genet*. **13**:e1006529.
Coyne JA, Orr AH. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.
David JR, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet*. **4**:106–111.
Dobzhansky T. 1982. *Genetics and the origin of speciesn. Columbia University Press*.
Duchen P, Živković D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals African and European Admixture in the North American *Drosophila melanogaster* population. *Genetics* **193**:291–301.
Fisher RA. 1937. The wave of advance of advantageous genes. *Ann Eugenics*. **7**:355–369.
Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. 2019. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Mol Ecol Resour*. **19**:552–566.
Haller BC, Messer PW. 2019. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol Biol Evol*. **36**:632–637.
Harris K, Nielsen R. 2016. The genetic cost of Neanderthal introgression. *Genetics* **203**:881–891.
Huerta-Sánchez E, Jin X, Asan Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, *et al.* 2014 Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**: 194–197.
Jacobsen F, Omland KE. 2011. Increasing evidence of the role of gene flow in animal evolution: hybrid speciation in the yellow-rumped warbler complex. *Mol Ecol*. **20**:2236–2239.
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. *Mol Biol Evol*. **32**: 1208–1220.
Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. 2018 Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**:1355–1358.
Juric I, Aeschbacher S, Coop G. 2016. The strength of selection against neanderthal introgression. *PLoS Genet*. **12**: e1006340.
Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the southeastern United States and Caribbean Islands. *Mol Ecol*. **24**: 1499–1509.
Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol*. **12**:e1004842.
Kelleher J, Thornton KR, Ashander J, Ralph PL. 2018. Efficient pedigree recording for fast population genetics simulation. *PLoS Comput Biol*. **14**:e1006581.
Kim BY, Huber CD, Lohmueller KE. 2018. Deleterious variation shapes the genomic landscape of introgression. *PLoS Genet*. **14**: e1007741.
Lachance J, True JR. 2010. X-autosome incompatibilities in *Drosophila melanogaster*: tests of Haldane's rule and geographic patterns within species. *Evolution* **64**:3035–3046.
Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM. *et al*., 2012. The Drosophila melanogaster genetic reference panel. *Nature* **482**: 173–178.
Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol*. **17**:e2006288.
Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Curr Opin Genet Dev*. **47**:69–74.
Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biol*. **5**: e243.
Maynard Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23**:23–35.
Muller H. 1942. Isolating mechanisms, evolution, and temperature. *Biol Symp*. **6**:71–125.
Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc B* **367**:409–421.
Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM, *et al.* 2016 Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat Genet*. **48**:1077–1082.
Ohta T, Kimura M. 1970. Development of associative overdominance through linkage disequilibrium in finite populations. *Genet Res*. **16**:165–177.
Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**:1805–1813.
Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol*. **25**:2337–2360.
Pool JE. 2015. The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol Biol Evol*. **32**:3236–3251.
Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nat Rev Genet*. **11**:175180.
Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol*. **34**:296–317.
Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N, Moritz C. 2016 Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLOS Biol*. **14**:e2000234.
Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr Biol*. **26**:1241–1247.
Schumer M, Cui R, Rosenthal GG, Andolfatto P. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLOS Genet*. **11**:e1005041.
Schumer M, Rosenthal GG, Andolfatto P. 2018. What do we mean when we talk about hybrid speciation? *Heredity* **120**:379.
Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, *et al.* 2018 Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**:656–660.

Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE. *et al*., 2020. Recurrent collection of Drosophila melanogaster from wild African environments and genomic insights into species history. *Mol Biol Evol.* **37**:627–638.

Van Belleghem SM, Cole JM, Montejo-Kovacevich G, Bacquet CN, McMillan WO, Papa R, Counterman BA. 2021 Selection and isolation define a heterogeneous divergence landscape between hybridizing *Heliconius* butterflies. *Evolution* **75**:2251–2268.

Wu CI, Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML. 1995 Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. *Proc Natl Acad Sci U S A.* **92**: 2519–2523.

Yukilevich R, True JR. 2008. African morphology, behavior and phermones underlie incipient sexual isolation between us and Caribbean *Drosophila melanogaster*. *Evolution* **62**: 2807–2828.