Finding hybrid incompatibilities using genome sequences from hybrid populations

Alexandre Blanckaert *1 and Bret A. Payseur 1

1 Laboratory of Genetics, University of Wisconsin-Madison, WI 53706 Madison, United States

* Corresponding author: Alexandre Blanckaert, blanckaert.a@gmail.com
Abstract

Natural hybrid zones offer a powerful framework for understanding the genetic basis of speciation in progress because ongoing hybridization continually creates unfavorable gene combinations. Evidence indicates that postzygotic reproductive isolation is often caused by epistatic interactions between mutations in different genes that evolved independently of one another (hybrid incompatibilities). We examined the potential to detect epistatic selection against incompatibilities from genome sequence data using the site frequency spectrum (SFS) of polymorphisms by conducting individual-based simulations in SLiM. We found that the genome-wide SFS in hybrid populations assumes a diagnostic shape, with the continual input of fixed differences between source populations via migration inducing a mass at intermediate allele frequency. Epistatic selection locally distorts the SFS as non-incompatibility alleles rise in frequency in a manner analogous to a selective sweep. Building on these results, we present a statistical method to identify genomic regions containing incompatibility loci that locates departures in the local SFS compared to the genome-wide SFS. Cross-validation studies demonstrate that our method detects recessive and codominant incompatibilities across a range of scenarios varying in the strength of epistatic selection, migration rate, and hybrid zone age. Our approach takes advantage of whole genome sequence data, does not require knowledge of demographic history, and can be applied to any pair of nascent species that forms a hybrid zone.

Introduction

Under the biological species concept, species formation results from the accumulation of reproductive barriers between lineages (Dobzhansky 1937; Mayr 1942). Reproductive isolation can be mediated by the environment (extrinsic isolation) or by genetic factors (intrinsic isolation). Modes of intrinsic isolation can be broadly categorized as involving processes that precede zygote formation (prezygotic) or involving processes that affect the development of the organism (postzygotic).

A popular explanation for postzygotic isolation is the Bateson-Dobzhansky-Muller (BDM) model (Bateson 1909; Dobzhansky 1936; Muller 1942). The model postulates that reproductive isolation arises when geographically separate populations fix incompatible mutations in genes that normally interact. Hybrids carrying incompatible combinations (referred to as Dobzhansky-Muller incompatibilities or hybrid incompatibilities) suffer reduced viability, reduced fertility, or both. Theoretical examination of the BDM model has produced important insights into speciation, including the extent to which incompatibilities inhibit gene flow between populations (Bank et al. 2012; Blanckaert and Hermisson 2018; Blanckaert et al. 2020) and the rate at which incompatibilities accumulate over time (Orr 1995; Turelli and Orr 2000; Orr and Turelli 2001).

The genes that form hybrid incompatibilities can provide glimpses into the genetic mechanisms responsible for postzygotic isolation (Presgraves 2010; Maheshwari and Barbash 2011). An incompatibility between two genes required to repress transcription of transposable elements and satellite DNA (Satyaki et al. 2014), Hmr and Lhr, kills hybrids between Drosophila simulans and D. melanogaster (Brideau et al. 2006). Dysfunctional interactions between Xmrk and (likely) Rab3d cause lethal melanomas in hybrids between Xiphophorus maculatus and X. helleri (Lu et al. 2020). Xmrk also forms a potential incompatibility with a different gene, Cd97, in a cross between two other species of swordtail fish (X. malinche and X. brichmanni).
Two genes, \textit{Sgs1} and \textit{Msh2}, have recently been shown to largely contribute to the genetic barrier between \textit{Saccharomyces cerevisiae} and \textit{S. paradoxus} \cite{Bozdag2021}, by preventing the formation of crossovers between diverged sequences. Finally, \textit{Prmd9}, a gene that regulates recombination hotspots, causes male sterility in hybrid mice \cite{Mihola2009}.

Although mapping and characterization of hybrid incompatibilities in the lab remains the best way to discover mechanisms of postzygotic isolation, the necessary experiments require substantial time. This framework also focuses on single hybrid phenotypes that are visible and relatively simple, whereas nascent species are often separated by a multitude of complex reproductive barriers. In one example, \textit{Martin and Wainwright} \cite{Martin2013} demonstrated the existence of postzygotic isolation in \textit{Cyprinodon} pupfishes and identified phenotypes affected by incompatibilities. However, given the complexity of the phenotypes, as well as the extrinsic nature of the interactions, the underlying genetic interactions were not investigated.

The examination of patterns of gene flow in natural hybrid populations is a compelling alternative to genetic dissection of reproductive barrier phenotypes in the lab. This framework features the advantages of focusing on fitness effects of hybridization in the wild, applying to many nascent species pairs that cannot be crossed in captivity, avoiding assumptions about the phenotypic basis of isolation, and considering hybrid individuals after many generations of admixture. In addition, genes identified using natural hybrid populations are likely involved in speciation, whereas incompatible genes identified between species that no longer naturally hybridize might have diverged after speciation was complete.

Several statistical methods have been developed with the goal of characterizing natural selection against hybrids (reviewed by \textit{Payseur} \cite{Payseur2010}, \textit{Payseur and Rieseberg} \cite{Payseur2016}). A strategy based on geographic clines \cite{Porter1997} builds on theory that equates changes in allele frequency across transects of a hybrid zone to a balance between selection and migration \cite{Barton1985, Szymura1986}. A framework based on genomic clines searches for loci that show collections of genotypes that depart from genome-wide hybrid indices \cite{Rieseberg1999, Gompert2009, Gompert2011}. Building on empirical studies that identified unlinked pairs of loci showing strong associations between conspecific alleles in collections of hybrids \cite{Payseur2005, Teeter2008, Schumer2014, Pool2015}, another approach looks for loci displaying stronger admixture disequilibrium than the remainder of the genome \cite{Schumer2016}. Collectively, these analytical strategies for identifying loci that experience selection against hybrids face challenges. Although methods that focus on geographic clines or genomic clines have found loci that likely reduce fitness in hybrids, these methods were not designed for application to whole genome sequences. Despite the intuitive appeal of using admixture disequilibrium to detect epistatic selection, simulations suggest this approach is unreliable, even when population structure is accounted for \cite{Schumer2016}. As genomic data from hybrid populations continue to accumulate, new methods that locate hybrid incompatibilities using genome sequences would accelerate progress toward deciphering the genetics of speciation.

A potentially general signature of selection against incompatibilities comes from consideration of the effects of epistatic selection in hybrid populations. When a heterospecific combination of alleles at two loci decreases hybrid fitness, selection reduces the number of individuals with this genotype. As a result, the population frequencies of the compatible alleles increase. Neutral mutations linked to each compatible allele should also spread through the population in a
manner partly analogous to a selective sweep (Maynard Smith and Haigh, 1974). Therefore, we might expect the site frequency spectrum of polymorphisms in genomic regions containing hybrid incompatibilities to depart from the genome-wide site frequency spectrum (Braverman et al. 1995; Simonsen et al. 1995).

In this paper, we present a method to locate hybrid incompatibilities using the site frequency spectrum. Following the logic of successful approaches developed to detect selective sweeps in non-hybrid populations ( Nielsen, 2005; Williamson et al., 2007), we deploy the full site frequency spectrum (rather than summarizing it) and we use the genomic spectrum as a null model (bypassing the need to specify a demographic model). Simulations of hybrid populations demonstrate that our method finds incompatibilities across a range of scenarios involving epistatic selection, migration, and genetic drift.

New approaches

We developed a new method to detect epistatic selection in natural hybrid populations by exploiting information contained in the site frequency spectrum (SFS). We computed the local unfolded SFS along the genome over regions of 500kb and calculated the probability of observing the local SFS given the global one using a multinomial distribution. Through bootstrapping, we established the distribution of windows of the chromosome that are found within the 1% lower tail of the probability of observing the local SFS given the global SFS. A region was defined as an outlier if it fulfills the two following conditions: the window itself appears often enough in the 1% tail of the distribution and this window and its neighbors appear on average often enough in the 1% tail of the distribution.

Methods

Simulations

We modeled a hybrid zone consisting of one hybrid population and two parental populations, using SLiM v3.4 (Haller and Messer 2019). Initially, the hybrid population was absent. The two parental populations, each of size $N_p = 5000$, began with the same genome except for a fixed difference at each of two loci (derived alleles $A$ and $B$), and diverged in allopatry for $n = 50000$ generations (burn-in phase). New mutations arose at rate $\mu = 10^{-9}$ and were assumed to be neutral. After this initial phase, the hybrid population, of size $N = 10000$, was formed in a single generation as a 50:50 combination of the two parental populations (without depleting them).

The hybrid population received migrants at rate $m = 0.005$ from each parental population (for a total of $2m = 0.01$). There was no migration from the hybrid population to the parental populations. The hybrid population (a sink) formed a barrier to gene flow between the parental populations (sources), which could reflect restrictions on migration imposed by the environment or reduced fitness in hybrids (Barton and Bengtsson 1986).

Individuals were diploid with genomes composed of one chromosome of length $L = 500Mb$. Recombination occurred during meiosis at a uniform rate of $r = 5 \times 10^{-9}$ per base pair. Loci $A$ and $B$ were situated at positions 175Mb and 325Mb, respectively, and their derived alleles
interacted epistatically to reduce fitness by $\epsilon = -0.1$. To model incompatibility associated with local adaptation, we also considered fitness disadvantages $s_a$ and $s_b$ for the two ancestral alleles (default $s_a = s_b = 0$). Fitness effects were multiplicative, with the fitness of an individual given by:

$$w = (1 + s_a)^2 - X_A (1 + s_b)^2 - X_B (1 + h_{X_A X_B} \epsilon)^{X_A X_B}$$

(1)

In this equation, $X_A$ and $X_B$ indicate the number of derived alleles (0, 1, or 2) at locus $A$ and $B$, respectively. Here $h_{X_A X_B}$ represents the dominance coefficient for the epistatic interaction. The recessive and codominant cases only differ for the double heterozygote individual $AaBb$: $h_1 = 0$ if the incompatibility is recessive and $h_1 = 1$ if the incompatibility is codominant. For all other cases ($k \in 0, 2, 4$), $h_k = 1$ for both recessive and codominant cases. Fitness was equal to the probability that an individual survived to adulthood. All surviving individuals contributed equally to the next generation.

We sampled 50 individuals without replacement from the hybrid population at generations 100, 1 000, 5 000 and 10 000 (with the counter beginning at 0 when the hybrid population was formed). Sampled individuals were not removed from the population.

We used two approaches to make the simulations computationally feasible. We scaled up recombination rates and mutation rates by a factor of 10 to represent a larger chromosome ($L_r$ and $L_\mu$ were kept constant). In addition, we used a single burn-in in the parental populations for all simulations that assumed the same genomic architecture. The range of parameter values used for simulations is provided in Table 1; values given in the paragraphs above were treated as defaults. Our source code is available on zenodo (doi:10.5281/zenodo.4614847).

Site frequency spectrum and outlier detection

To detect incompatibilities, we computed the unfolded site frequency spectrum (SFS) in non-overlapping 500kb windows (for a total of 1000 windows in the genome) for samples of 50 diploid individuals. We also tabulated the genome-wide SFS. Single nucleotide polymorphisms (SNPs) with more than two alleles were ignored. For each window, we calculated the probability of observing the local SFS using a multinomial distribution with parameters estimated from the genome-wide SFS. This probability was estimated $n_b = 1000$ times by bootstrap resampling of the 100 sampled sequences. For each bootstrap replicate, we recorded the 1st percentile of the probability distribution of the local SFS given the genome-wide SFS. Then, we constructed the distribution across bootstrap replicates of the position of the 1st percentile along the chromosome. We counted the number of times each window appeared in the 1st percentile of the distribution ($\kappa$). If $\kappa$ was larger than a threshold value ($\text{thr}_1$, $\kappa > \text{thr}_1$), the window was identified as a potential outlier. To further reduce false positives, we required that windows adjacent to a potential outlier also appeared in the 1st percentile, reasoning that the distortion generated by selection was likely to extend beyond 500kb. Therefore, for each candidate outlier, we calculated $\kappa_d$, the average number of times a window in the $i - d, i + d$ region was found in the 1st percentile of the distribution. A candidate outlier was classified as a true outlier if $\kappa_d$ was larger than a second threshold ($\text{thr}_2$, $\kappa_d > \text{thr}_2$). Given that the initial filter was done on each window independently ($\kappa > \text{thr}_1$), we considered as false positive any outlier that did not include incompatibility loci $A$ or $B$ in the $i - d, i + d$ region of detection. Therefore, there
were $4 \times d + 2$ windows that were considered true positives and $1000 - 4 \times d - 2$ considered false positives. Due to dependence on $d$ of the numbers of true positives and false positives, larger values of $d$ will perform better. Due to the bootstrap step, there were at most $10000/\text{thr}_1$ possible candidate outliers (ranging from 25 if $\text{thr}_1 = 400$ to 10 if $\text{thr}_1 = 1000$). Therefore, we used this value ($10000/\text{thr}_1$) in Bonferroni corrections for multiple testing.

**Evaluation of method performance**

For each simulated scenario, we initially calculated the power of the method, the false positive rate and the proportion of false positives among the outliers for combinations of $\{\text{thr}_1, \text{thr}_2, d\}$ using all available simulations, with $\text{thr}_1$ any value in $\{400, 450, \ldots, 1000\}$, $\text{thr}_2$ any value in $\{0, 20, \ldots, 300\}$ and $5 \leq d \leq 19$. We determined which $\{\text{thr}_1, \text{thr}_2, d\}$ provided the highest power for each scenario. As mentioned in the previous section, $d$ determines the number of true positive regions in the genome, and therefore larger $d$ will always have higher power. To compensate for this effect, we also determined the best $\{\text{thr}_1, \text{thr}_2, d\}$, using two additional optimizing metrics, $\frac{\text{power}}{(2d+1)}$ and $\frac{\text{power}}{(2d+1)^2}$ and introducing a penalty for larger values of $d$. If the optimizing metrics were identical, we used resolution (lower is better), then proportion of false positives in outliers (smaller is better), and finally $\text{thr}_2$ (larger is better) to determine the best combination of $\{\text{thr}_1, \text{thr}_2, d\}$.

Based on these results, we examined further those scenarios for which our method showed at least 50% power, a false positive rate below $\frac{5 \times \text{thr}_1}{10000}$, and a proportion of false positives among outliers below 5%. For these scenarios, we preserved only combinations of $\{\text{thr}_1, \text{thr}_2, d\}$ that passed the criteria described above (a false positive rate below $\frac{5 \times \text{thr}_1}{10000}$) and a proportion of false positives among outliers below 5% and determined the best combination by averaging the optimizing metric over the selected scenarios.

To further evaluate the method, we performed cross-validation. Using a randomly chosen subset of the simulations (25%) from the restricted group described above, we determined which combination $\{\text{thr}_1, \text{thr}_2, d\}$ provided the best results. We then measured performance on the remaining dataset using the values of $\text{thr}_1$, $\text{thr}_2$ and $d$ estimated in the previous step, providing cross-validation of our approach. We repeated this step 100 times with power and $\frac{\text{power}}{(2d+1)}$ as optimizing metrics. For each optimizing metric, the mode of the distribution of best $\{\text{thr}_1, \text{thr}_2, d\}$ was considered to be the best value to detect incompatibilities. In addition to choosing the best combination $\{\text{thr}_1, \text{thr}_2, d\}$ for each iteration, we also considered keeping any combination of $\{\text{thr}_1, \text{thr}_2, d\}$ that performed almost as well (with the optimizing metric being within 5% of the best one). This approach allowed us to capture the possible existence of a plateau in the $\{\text{thr}_1, \text{thr}_2, d\}$ space, therefore obtaining a more robust estimate of an effective $\{\text{thr}_1, \text{thr}_2, d\}$ combination, at the slight cost of power.

**Results**

**Characterization of the site frequency spectrum in a hybrid population**

Populations in a hybrid zone display a SFS that departs from what is expected in classical Wright–Fisher populations (where the relative proportion of SNPs with a frequency $f$ is proportional to $\frac{1}{f}$ (Wright 1938), as illustrated in Figure 1A. Indeed, when initially formed, the hybrid
Table 1: Scenarios and parameter values examined

| Parameters       | Name codominant case     | Name recessive case      |
|------------------|--------------------------|--------------------------|
| $\epsilon = -0.1$, $s = 0$, $m = 0.005$ | Default                  | Default rec.             |
| $\epsilon = -0.02$, $s = 0$, $m = 0.005$ | Low ep.                  | Low ep. rec.             |
| $\epsilon = -0.5$, $s = 0$, $m = 0.005$ | High ep.                 | High ep. rec.            |
| $\epsilon = -0.1$, $s = 0$, $m = 0.0005$ | Low mig.                 | Low mig. rec.            |
| $\epsilon = -0.5$, $s = 0$, $m = 0.05$ | High mig.                | High mig. rec.           |
| $\epsilon = 0$, $s = 0$, $m = 0.005$ | SL-sel.                  | NA                       |
| $\epsilon = 0$, $s = -0.02$, $m = 0.005$ | High SL-sel.             | NA                       |
| $\epsilon = -0.1$, $s = -0.02$, $m = 0.05$ | SL-sel. and ep.          | SL sel. and ep. rec.     |

population possesses a unique SFS (Fig. 1B), characterized by a large proportion of SNPs at frequency 0.5; this proportion depends on the level of divergence between the two parental populations. These SNPs correspond to fixed mutations, private to each parental population (since the hybrid population is formed by a symmetric contribution of the two parental populations). The constant influx of fixed differences between parental populations into the hybrid zone via migration helps to maintain the proportion of SNPs with frequency 0.5, whereas drift spreads the distribution around 0.5 (Figure 1D). In the absence of selection, the SFS of a (sink) hybrid population can be decomposed into three components: one part resulting from migration-drift balance and the initial composition of the hybrid population, and two parts inherited from the parental populations via migration (here, these two parts are indistinguishable due to the symmetric contribution of the parental populations to the hybrid population). In the absence of migration, only the first component remains: the SFS of the initial hybrid population slowly converges towards the SFS of a Wright-Fisher population under drift. After 1 000 generations, the isolated hybrid population still displays an excess of alleles at intermediate frequency (Fig. 1C), though this effect is far more diffuse than in the presence of migration and almost vanishes after 10 000 generations (Fig. S1). In addition, it is worth noting that while the SFS in the parental populations does not change, the part of the SFS in the hybrid population generated by migration (equivalent to Fig. 1B) changes as the parental populations continue to diverge and accumulate fixed differences.

Epistatic selection against hybrid genotypes shifts the SFS in the vicinity of incompatibility loci (Fig. 2A for the default scenario; see Fig. S2-S17 for other scenarios). The distortion is spatially broad, extending approximately 10-20 Mb around incompatibility loci (Fig. 2A). With epistatic selection against incompatible alleles A and B, genetic backgrounds containing the ancestral a allele or the ancestral b allele have a marginal fitness advantage. As a result, derived alleles at neutral SNPs linked to a or b increase in frequency away from 0.5, creating a detectable, local signature in the SFS. A similar distortion of the SFS is generated by single-locus selection but this distortion is stronger (assuming $s = \epsilon$) because single-locus selection acts independently of the genetic background (Fig. S3) and therefore is always present.
Figure 1: Neutral site frequency spectrum (SFS) in a hybrid zone. The neutral expectation of an isolated population is given in red and proportional to $1/f$ (Wright, 1938). A/ SFS for the parental population at the time of formation of the hybrid population. The relative proportion of fixed mutations ($\frac{\# \text{fixed}}{\# \text{polymorphic}}$) has been added to the right. B/ Predicted SFS for the hybrid population at generation 0. C/ SFS for an isolated hybrid population at generation 1 000. D/ Site frequency spectrum for a sink hybrid population at generation 1 000 ($m = 0.005$). SFS for both the isolated hybrid populations and the sink hybrid population at different time points (100, 5 000 and 10 000) can be found in Figure S1.

**Finding genomic regions with distortions in the site frequency spectrum generated by epistatic selection**

To quantify the distortion of the SFS near incompatibility loci, we used the multinomial distribution to compute the probability of observing the local SFS (over a 500kb window) given the global SFS calculated for the entire chromosome. Due to the resulting high dimensionality, the probability of observing a particular SFS was always extremely low (Fig. 2B). For reference, the probability of observing exactly once each number between 1 and 100 when drawing 100 numbers at random (all having the same probability of being chosen) is $9.3310^{-43}$. Nevertheless, the probability of observing a SFS near an incompatibility locus is many orders of magnitude lower than the probability of observing one that is similar to the global SFS (Fig. 2B). Most of the windows with the lowest probabilities of observing the local SFS are found near incompatibility loci, $A$ and $B$. However, due to chance, some regions of the chromosome may have a local SFS that differs from the rest of the chromosome (for example, on Fig. 2B to the left of locus $A$, around position 125Mb).

To distinguish SFS signatures due to epistatic selection from those due to chance, we ran a bootstrap analysis to identify outliers. We designated as outliers those regions with probabilities
Figure 2: A/ Local site frequency spectrum calculated over regions of 500kb. The X axis corresponds to the position along the chromosome, the Y axis shows SNP allele frequency. The density is indicated by color, with yellow denoting a lack of SNPs with the corresponding frequency and black denoting an abundance of SNPs with that frequency. B/ Probability of observing the local site frequency spectrum conditional on the global site frequency spectrum. Results are shown for simulations with $s = 0, \epsilon = -0.1, m = 0.005, gen. = 1,000$. 

that consistently fell within the lowest 1% of the distribution of probability of observing the local SFS based on the global SFS along the chromosome. Figure 3 illustrates the outcome of the bootstrap analysis for the case of epistatic selection (with $\epsilon = -0.1$). Most of the regions that
Figure 3: Outcome of bootstrap analysis and definition of thresholds. A/ Number of times each window was in the lower 1% tail for the probability of the local SFS being observed given the global SFS in the analysis with 1 000 bootstrap replicates. The red dotted line indicates the first threshold, \( \text{thr}_1 \). In this example, there are 5 windows that satisfy this first criterion. B/ Zoom-in on the two outlier windows detected close to locus A. The blue dashed line indicates the second threshold \( \text{thr}_2 \) and the black line indicates the average number of times a window appears in the lower 1% tail. Both the blue line and the black line are drawn only over the windows of interest, i.e., including \( d = 5 \) windows to the left and 5 windows to the right of the focal window. C/ and D/ Zoom-in on the two outlier windows detected close to locus B. The bootstrap analysis depicted in this figure used the same set of data presented in Figure 2.

are detected as outliers more than 50% of the time map near the incompatibility loci (windows 350 and 650, Fig. 3A). The effects of epistatic selection extend beyond windows containing the incompatibility loci. The bootstrap analysis effectively removes some windows with extreme probabilities while retaining windows adjacent to incompatibility loci. Figure 2 displays an example in which a local SFS in a neutral region around 125Mb has a very low probability of being observed. After bootstrap analysis, the signal partially vanishes: first this region appears in the 1% tail about 40% of the time; second, the signal is isolated.

The probability of observing a given dataset under a multinomial distribution depends on the number of observations as well as how they are partitioned. Just as the binomial probability of observing exactly \( n \) “heads” with \( 2n \) coin flips decreases as \( n \) increases, the multinomial probability of observing a particular SFS decreases as the number of SNPs grows. Although we recover such a relationship between the number of SNPs and the (log) probability of the local SFS, windows with the lowest probabilities are not those with the most SNPs, whether they evolve under neutrality (Fig. S19A), single-locus selection (Fig. S19B) or epistatic selection...
Method performance

For each scenario, we calculated the power, false positive rate and proportion of false positives in outliers for our method, considering diverse combinations of \( \{d, \text{thr}_1, \text{thr}_2\} \). Table 2 shows the combination of \( \{d, \text{thr}_1, \text{thr}_2\} \) that generates a false positive rate below \( \frac{5 \times \text{thr}_1}{10000} \% \) (Bonferroni correction), a proportion of false positive in outliers below 5%, and the highest power. As mentioned in the Methods, since power increases with the size of the region considered as “true”, \( 2 \times d + 1 \), and therefore with \( d \), we also considered two additional optimization metrics:

1. Power corrected by the number of windows around the true location considered as “true” positive \( \left( \frac{\text{power}}{2 \times d + 1} \right) \)
2. The square of this value \( \left( \frac{\text{power}}{(2 \times d + 1)^2} \right) \).

When we ignore the effect of \( d \) on power, we obtain relatively high power across multiple scenarios, finding the incompatibility loci more than 90% of the time, for both recessive and codominant incompatibilities. The incompatibility loci are found in regions ranging in size from 15.5Mb (under weak epistatic selection with codominance) to 19.5Mb (low migration). Reducing the size of the region is possible, at the cost of power. By penalizing for the size of the window of detection, we can narrow the position of the incompatibility loci to less than 10Mb for multiple scenarios. It can be further narrowed but at a large cost of power. For example, in the default scenario, the window can be reduced by approximately two-thirds to 6.5Mb with an associated decrease by half in power from 0.958 to 0.486). Therefore, using \( \frac{\text{power}}{2 \times d + 1} \) as the optimization metric offers a good compromise between power and resolution.

When migration is high (10% of the hybrid population is replaced by individuals from the parental sources each generation), we fail to detect the incompatibility loci. In this case, the SFS reflects mainly the genome-wide effects of migration, rather than the balance between migration and selection (Fig. S12). Even when selection against hybrids is extremely strong (F1 hybrid fitness is reduced by 50% in the codominant case, and by 94% for the double homozygote \( AABB \)), incompatibility loci are detected only 40% of the time and with a rather low resolution (in a 16.5Mb region). In that case, due to the presence of perpetual strong migration and selection, both incompatible alleles remain at intermediate frequencies (\( \approx 0.30 \) for all 4 time points). There is not enough time for the parental haplotype to fully break down and therefore the distortion of the SFS extends over a rather large region (more than 50Mb, see Fig. S14), making it a challenge to pinpoint the incompatibilities themselves. When migration is not too strong (1% of the hybrid population is replaced per generation), weakly selected incompatibility loci (\( \epsilon = -0.02 \)) can be potentially detected (\( \approx 43\% \) within a 18.5Mb region or \( \approx 25\% \) within a 7.5Mb).

When there is both single-locus selection against alleles \( a \) and \( b \) and epistatic selection against alleles \( A \) and \( B \), power is reduced by about 40% compared to epistatic selection alone (from 0.958 to 0.548 for the codominant case and from 0.910 to 0.575 for the recessive one, when power is the optimizing metric). This is also true with \( \frac{\text{power}}{2 \times d + 1} \) as the optimization metric (Table 2).

Finally, we emphasize that while we obtain a variety of best combinations \( \{d, \text{thr}_1, \text{thr}_2\} \) for the diverse scenarios, we still find a common combination of \( \{d, \text{thr}_1, \text{thr}_2\} \) where the method performs close to its best across scenarios.
| Optimizing metric: | power | power \(2d+1\) | power \(5xthr_1\) % | power \(2d+1\) |
|-------------------|-------|----------------|-----------------|----------------|
| Scenario          | \(d, thr_1, thr_2\) | \(d, thr_1, thr_2\) | \(d, thr_1, thr_2\) | \(d, thr_1, thr_2\) |
| Default           | 16, 400, 20 0.958 | 9,800, 40 0.763 | 6,950, 80 0.486 |
| Default rec.      | 15, 400, 40 0.910 | 9,700, 80 0.722 | 5,950, 100 0.293 |
| Low epis.         | 18, 400, 60 0.431 | 7,750, 100 0.254 | 5,850, 120 0.16  |
| Low epis. rec.    | 19, 500, 60 0.261 | 7,750, 120 0.153 | 5,800, 160 0.111 |
| High epis.        | 18, 500, 20 0.948 | 9,950, 60 0.611  | 9,950, 60 0.611  |
| High epis. rec.   | 16, 400, 20 0.986 | 9,800, 100 0.770 | 6,950, 80 0.481  |
| Low mig.          | 19, 450, 20 0.915 | 6,950, 120 0.669 | 6,950, 120 0.669 |
| Low mig. rec.     | 18, 650, 40 0.836 | 6,950, 100 0.579 | 6,950, 100 0.579 |
| High mig.         | 16, 400, 100 0.099 | 16, 400, 100 0.099 | 5,800, 160 0.024 |
| High mig. rec.    | 16, 400, 100 0.07  | 5,800, 100 0.03  | 5,800, 100 0.03  |
| High mig. high epis.| 16, 400, 100 0.406 | 16, 400, 100 0.406 | 5,950, 100 0.068 |
| High mig. high epis. rec. | 18, 400, 80 0.353 | 18, 400, 80 0.353 | 5,900, 120 0.055 |
| SL-sel. and epis. | 16, 550, 60 0.548 | 13,750, 160 0.466 | 7,950, 300 0.154 |
| SL-sel. and epis. rec. | 16, 400, 60 0.575 | 9,700, 100 0.405 | 5,850, 200 0.205 |

Table 2: Best combination of \(d, thr_1, thr_2\) and power for the different scenarios. We consider here three different optimization metrics: power (left columns), \(\frac{\text{power}}{2d+1}\) (central columns) and \(\frac{\text{power}}{5xthr_1}\) % or a proportion of false positive in the outliers above 5% were excluded.

To further evaluate the method, we performed a cross-validation analysis, using 25% of the simulations (randomly chosen) from scenarios with epistatic selection in which the incompatibility loci were detectable (see Methods). We report which combination of \(d, thr_1, thr_2\) was the best for each of the 100 iterations with \(\frac{\text{power}}{2d+1}\) (Table 3) as the optimizing metric (see Table SI for results using power as the optimization metric). The mode of the distribution corresponds to the best combination \(d = 9, thr_1 = 900, thr_2 = 60\), translating into a resolution of 9.5Mb. The second most common combination, \(\{d = 9, thr_1 = 900, thr_2 = 80\}\), is really similar, therefore confirming our result. When we did not penalize for the size of the window \(d\), the best combination was observed in 61 out of 100 iterations and given by \(\{d = 16, thr_1 = 400, thr_2 = 40\}\) (Table SI). Since the proportion of false positives in outliers is one of the criteria we consider, the effects of \(thr_1\) and \(thr_2\) are not straightforward. Indeed, while the false positive rate will decrease (or remain unchanged in the worst-case scenario) when the filters become more stringent, the proportion of false positives among outliers may increase, due to the possible exclusion of true positive outliers. Therefore, despite being really similar, we cannot merge the cases where \(\{d = 9, thr_1 = 900, thr_2 = 60\}\) and \(\{d = 9, thr_1 = 900, thr_2 = 80\}\) by simply choosing the more stringent option. Therefore, we considered combinations that provided not only the best optimization metrics, but also combinations found within 5% of this value (i.e. a list of better choices, instead of simply best choices). Under these conditions, the \(\{d = 9, thr_1 = 900, thr_2 = 80\}\) combination, that is the second best option, is found as the better option in 81 out of 100 iterations (Table 3). We therefore present in Table 4 the power, false positive rate and proportion of false positives calculated for the remaining simulations (see Methods) and averaged for the 81 cases where \(\{d = 9, thr_1 = 900, thr_2 = 80\}\) was the best set of criteria. Similarly, when using power as the optimizing metric, the better option corresponds to \(\{d = 16, thr_1 = 450, thr_2 = 40\}\).
Table 3: Distribution of combination of \{d, thr_1, thr_2\} using $\text{power}_{2d+1}$ as the optimizing metric. Left column corresponds to the number of a time a combination is the best one, the right one to the number of time a combination is within 5% of the best combination. Only combination that appeared twice as the best combination, or 5 times within 5% of the best combination.

| d   | thr_1 | thr_2 | Count (best) | Count (top5%) |
|-----|-------|-------|--------------|---------------|
| 9   | 900   | 60    | 43           | 56            |
| 9   | 900   | 80    | 28           | 81            |
| 9   | 900   | 40    | 9            | 10            |
| 9   | 900   | 100   | 8            | 32            |
| 9   | 850   | 100   | 4            | 5             |
| 9   | 950   | 40    | 2            | 4             |
| 9   | 900   | 0     | 0            | 10            |
| 9   | 900   | 20    | 0            | 10            |

As displayed in Table 4, using the common set of criteria determined above, we can detect incompatibility loci over a broad range of scenarios with a resolution of 9.5Mb. As long as epistatic selection is not too weak, we are able to detect the loci with power near 50%, while maintaining a false positive rate below 5% (corrected for multiple testing, see Methods), for both recessive incompatibilities and codominant incompatibilities. The proportion of false positives in outliers fluctuates more but remains below 5% in all but 2 scenarios: (1) strong codominant epistatic selection and (2) antagonistic single-locus and codominant epistatic selection (Table 4A). For both scenarios, this excess of false positive in outliers is mainly due to the distortion generated by selection extending beyond what is defined as “true positive” (Fig. S8 and S16). This “too large” distortion is either due to the strength of epistatic selection itself or the antagonist interaction between single-locus and epistatic selection maintaining one of the two incompatible allele at high frequency (0.77 after generation 1000). Our approach fails to detect the incompatibility loci when migration is high.

Single-locus selection is similarly detectable (Table 4B). About 65% of loci under strong single-locus selection are detected, a rate similar to the strong epistatic selection scenario. Weaker single-locus selection is detected only about 35% of the time.

When both epistatic selection and single-locus selection (antagonistically) affect the two loci, power decreases from 61% to 49% in the codominant case and from 44% to 20% in the recessive one (Table 4A), compared to epistatic selection acting in isolation. Using a common combination of \{d, thr_1, thr_2\} and the $\text{power}_{2d+1}$ optimization metric does not reduce our ability to detect this scenario in the codominant case (Table 2) but it does for the recessive case.

When using power as the optimizing metric, the best combination of criteria is \{d = 16, thr_1 = 400, thr_2 = 40\} and the power achieved is far greater than when using $\text{power}_{2d+1}$ as the optimizing metric, reaching above 95% in some cases (Table S2). This increase comes at the cost of lower resolution (16.5Mb vs. 9.5Mb). Only the scenarios with strong migration and default epistatic selection perform poorly. In the absence of epistatic selection, the focal loci are detected in 3.25% of the cases, which is less than expected assuming the position of outliers is independently and uniformly distributed along the chromosome (6.8%). A total of 172 outliers are found in 130 of the 400 cases, but in 79 of the 100 replicates. The distribution of outliers detected across the 100 replicates for each of the time samples can be found in the Supplement (Fig. S20).
A/  
| Scenario                | Prop. false pos. in outliers | Power | False pos. rate |
|-------------------------|------------------------------|-------|-----------------|
| Default                 | 0.017                        | 0.613 | $4.1 \times 10^{-5}$ |
| Default rec.            | 0.018                        | 0.436 | $3.1 \times 10^{-5}$ |
| High ep.                | 0.062                        | 0.696 | $1.8 \times 10^{-4}$ |
| High ep. rec.           | 0.032                        | 0.633 | $7.4 \times 10^{-5}$ |
| Low mig.                | 0.042                        | 0.734 | $1.4 \times 10^{-4}$ |
| Low mig. rec.          | 0.036                        | 0.659 | $9.9 \times 10^{-5}$ |
| SL-sel and ep.         | 0.145                        | 0.485 | $4.2 \times 10^{-4}$ |
| SL-sel and ep. rec.    | 0.021                        | 0.202 | $2.9 \times 10^{-5}$ |

B/  
| Scenario                | Prop. false pos. in outliers | Power | False pos. rate |
|-------------------------|------------------------------|-------|-----------------|
| Neutral                 | 0.013                        | 0.350 | $1.30 \times 10^{-5}$ |
| SL-sel.                 | 0.11                         | 0.671 | $1.73 \times 10^{-5}$ |
| High SL-sel.            | 0.178                        | 0.124 | $2.60 \times 10^{-6}$ |
| Low ep.                 | $2.50 \times 10^{-3}$       | 0.068 | $7.80 \times 10^{-6}$ |
| Low ep. rec.            | $6.25 \times 10^{-3}$       | 0.0125| $2.60 \times 10^{-6}$ |
| High mig.               | 0.087                        | 0.134 | $1.17 \times 10^{-4}$ |
| High mig. high sel.     | 0.033                        | 5.88$ \times 10^{-2}$ | $3.38 \times 10^{-5}$ |

Table 4: Ability to detect the focal loci using the “optimal” combination of criteria: $\{d = 9, thr_1 = 900, thr_2 = 80\}$, as determined by the cross-validation analysis for the various scenarios. Here, the correction for multiple testing is a factor $9/100$, meaning that the false positive rate should be below $4.5 \times 10^{-3}$. Table A displays scenarios used during the cross-validations and therefore corresponds to the average over the 81 cases where $\{9,900,80\}$ was the best combination or within a 5% distance of it. Table B corresponds to the other scenarios.

To gauge the applicability of our method to hybrid zones with different ages, we examined performance at the four time points at which we sampled the hybrid population. For most scenarios, power is similar across time points (Fig. [S21][S24]). This result indicates that our method is robust to variation in hybrid zone age and it can be applied without assumptions about this parameter.

**Stronger epistatic selection is detected further away**

When epistatic selection is strong, the focal outlier is usually detected further away from the positions of incompatibility loci (Fig. 4A; this is also true for single-locus selection Fig. [S25]). In the absence of selection, the SFS is bimodal, with one mode corresponding to rare alleles and the second mode near 0.5 due to fixed differences between the two parental populations (Fig. 1D). Selection distorts the local SFS (Fig. 2A, S3-S17; alleles with frequencies at or near 0.5 are more strongly affected because many of them are in tight linkage disequilibrium with incompatibility alleles. About half of the mutations at frequency 0.5 originated from population 1 and are in strong linkage disequilibrium with allele $A$ (or $b$), while the other half originated from population 2 and are in strong association with allele $a$ (or $B$). Due to epistatic selection, allele $A$ decreases in frequency, and so do the neutral alleles in linkage disequilibrium with $A$. Conversely, allele $a$ and linked neutral alleles increase in frequency. This dynamic splits in two parts the mode of the SFS near 0.5. When selection is strong, neutral alleles in linkage disequilibrium with $A$ are pushed to a lower frequency, making it harder to distinguish the SFS under strong epistatic...
selection from the baseline SFS. Therefore, the signal is recovered further away. Put another way, incompatibility alleles decrease in frequency at a faster rate compared to recombination when selection is strong. As a result, linkage disequilibrium between incompatibility alleles and neutral alleles extends further and so does the distortion in the SFS.

We recover a similar pattern for migration (Fig. 4B). With stronger migration, the detected outliers are on average further away from the positions of incompatibility loci because a high proportion of incompatibility alleles have been recently introduced into the hybrid population and linked neutral alleles have had less time to recombine away. However, migration and epistatic selection seem to interact non-linearly. Indeed, for a migration rate of $m = 0.05$, incompatibility loci under strong epistatic selection are detected closer to their true positions (Fig. S20) than are incompatibility loci under weak epistatic selection. Finally, in the presence of antagonistic single-locus selection, the detected outliers are on average further away from the positions of incompatibility loci (Fig. S21).

**Discussion**

Although hybrid incompatibilities are important contributors to reproductive isolation and speciation, they remain challenging to identify. We have shown that epistatic selection against incompatibility loci produces a localized distortion in the site frequency spectrum in hybrid populations. Using this distortion, we can identify chromosomal regions that contain incompatibility loci under a variety of scenarios.

Our approach features several advantages. First, the method is applicable to any pair of nascent species that form a natural hybrid zone; breeding organisms in the laboratory is not required. Second, our strategy can detect incompatibility loci that affect any aspect of hybrid
viability or hybrid fertility without specifying the phenotype of interest in advance. In addition, the method performs relatively well across a wide range of epistatic selective strength. Third, using the genome-wide SFS to predict the local SFS accounts for effects of demographic history, obviating the need to reconstruct the history of hybridization (which can be complex). Fourth, many of the known examples of incompatible genes involve strongly diverged species (Powell et al., 2020), presenting a risk that they arose after the speciation process was complete. The method we described relies on individuals from natural hybrid zones. Genetic incompatibilities detected are therefore likely to be involved in the speciation process. Finally, our approach incorporates the full SFS and is specifically designed for the analysis of genome sequence data. Below, we discuss the application and limits of the method.

One of the biggest challenges in the detection of hybrid incompatibilities is disentangling epistatic selection from single-locus selection. Although disrupted interactions between alleles at two loci reducing hybrid fitness and a beneficial allele at a single-locus increasing fitness are biologically distinct phenomena, they leave similar signatures in the SFS. Selection against an incompatibility (derived) allele increases the frequency of the non-incompatibility (ancestral) allele in a manner that resembles a selective sweep involving a single-locus. Epistatic selection is less effective than single-locus selection because an incompatibility allele is only disfavored when it is combined with the incompatibility allele at the second locus (this difference is strongest when epistatic interactions are fully recessive). As a result, we might expect epistatic selection to be detected only when it is strong. In contrast, epistatic selection leading to a fitness reduction of 10% in double heterozygotes is easily detectable using our method. Cases with even weaker epistatic selection can be detected with limited power, depending in part on the balance between resolution and power that is chosen (Tables 4 and S2). Nevertheless, when applied in isolation, our method is not able to exclude the possibility that single-locus selection explains detected outlier loci.

We suggest several potential solutions to the challenge of distinguishing epistatic selection from single-locus selection. First, clines in allele frequency or genotype frequency at SNPs near detected outliers can be examined in a second stage of analysis. An explicit epistatic model can be fit using genomic clines (Gompert and Buerkle, 2009, 2011). If samples from additional localities in the hybrid zone can be collected, geographic clines can be analyzed for evidence of adaptation of the form that would be predicted under single-locus selection (Endler, 1977). Second, longitudinal data from hybrid zones have the potential to detect differences in the trajectory of alleles under epistatic selection and single-locus selection. Allele trajectories have started to be exploited to detect selection (Foll et al., 2015) and to understand the evolution of quantitative traits (Franssen et al., 2017). Finally, we emphasize that we would expect many of the loci identified by our approach to be incompatibilities, at least when considering hybrid zones between lineages with reproductive barriers. The number of hybrid incompatibilities is expected to grow faster than linearly with divergence time (Orr, 1995) and this result has been recovered in multiple taxa: including Solanum (Moyle and Nakazato, 2010) Guerrero et al., 2017, Drosophila (Matute et al., 2010) or Mus musculus (Wang et al., 2015).

Our approach offers the potential to identify both loci involved in a hybrid incompatibility. However, because each genomic region is evaluated independently, the current method does not reveal whether detected loci interact epistatically. Schumer and Brandvain (2016) demonstrated that pairs of incompatibility loci tend to be found in the extreme tail of the distribution.
of admixture disequilibrium, providing a possible way forward. Unfortunately, our prelimi-
nary investigation of admixture disequilibrium revealed that this signature is not maintained
in older hybrid zones (Fig. S3-S17, S28). We suspect this disparity reflects our decision to
model (Bateson-Dobzhansky-Muller) incompatibilities with asymmetric fitness effects (whereas
(Schumer and Brandvain 2016) focused on incompatibilities with symmetric fitness effects) and
to consider scenarios with weaker selection, in addition to the limited time frame during which
epistatic selection distorts admixture disequilibrium. In another attempt to detect interactions
among selected loci, Beissinger et al. (2016) computed the $D'_S^2$ statistic (Ohta, 1982a,b), a mea-
sure that partitions linkage disequilibrium between and within populations. Although a certain
number of candidate genes were identified, the authors pointed out that single-locus selection
could not be excluded as an alternative explanation since selective sweeps can generate similar
patterns.

The need to include enough SNPs to formulate a reasonable site frequency spectrum and
the extra steps taken to minimize the detection of false positives jointly reduce the genomic
resolution of our method. An incompatibility locus can be associated with a region of about
9-15Mb, depending on whether we optimize for power or resolution. Nevertheless, most outliers
are found much closer to the incompatibility locus itself (Fig. 4). For example, in the default case
we examined, outliers were located within 1.4Mb of the incompatibility locus on average. The
same genome sequence data required for our method can be used to further refine the location
of incompatibility loci. We expect incompatibility alleles to be enriched for fixed differences
outside the hybrid zone; scanning parental populations for fixed differences or large frequency
differences could be a reasonable filter. Genes with relatively high levels of non-synonymous
divergence between parental populations could be prioritized. Searching outlier regions for
alleles that show narrow frequency clines could also localize incompatibility loci. Finally, in
species with annotated genomes or their relatives, knowledge of gene function could be used to
identify candidate genes. For example, genes that reduce fertility when mutated could be good
targets for further examination. Alternatively, our method could be used to suggest new classes
of genes involved in reproductive isolation by searching for pathways that are enriched across
outliers. Confirmation that a gene participates in an incompatibility will ultimately require
experimental validation (see for example (Bozdag et al. 2021)), which is becoming possible for
an increasing variety of species through advances in genome editing.

Several biologically realistic scenarios missing from our simulation model deserve consider-
ation. In addition to epistatic selection against hybrid incompatibilities, reduced introgression
can result from single-locus selection against deleterious mutations that accumulate in source
populations with small effective sizes (Harris and Nielsen 2016). Although our simulations as-
sumed that incompatibility alleles were fixed in source populations, incompatibilities are often
polymorphic in nature (Cutter 2012). More complex models of migration including structured
source populations, asymmetrical migration, and migration out of hybrid populations are likely
features of real hybrid zones. Finally, both hybrid populations and source populations will
routinely depart from mutation-drift equilibrium due to demographic factors. Examining the
extent to which each of these scenarios produces local distortions in the site frequency spectrum
of hybrid populations is a logical next step.

Our measurement of method performance also assumed that sequences could be determined
with complete accuracy. Genotyping error could reduce power and increase the false-positive
rate, especially for low-coverage sequencing, which can lead to biased reconstructions of the site
frequency spectrum (Han et al., 2014). We recommend incorporating uncertainty in variant
calling when building the site frequency spectrum from sequencing reads (Nielsen et al., 2012)
before applying our approach.

Acknowledgments

We thank members of the Payseur lab for useful discussions. This work was supported
by the National Science Foundation (DEB 1353737 to B.A.P) and the National Institutes of
Health (R01GM120051 and R35GM139412 to B.A.P). This research was performed using the
compute resources and assistance of the UW-Madison Center For High Throughput Computing
(CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison,
the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin
Institutes for Discovery, and the National Science Foundation, and is an active member of
the Open Science Grid, which is supported by the National Science Foundation and the U.S.
Department of Energy’s Office of Science.

Data Availability

The data underlying this article are available on zenodo and can be accessed via a DOI link
(doi:10.5281/zenodo.4614847).

Author contributions

Conceptualization: Alexandre Blanckaert, Bret A. Payseur.
Investigation: Alexandre Blanckaert.
Software: Alexandre Blanckaert.
Supervision: Bret A. Payseur.
Visualization: Alexandre Blanckaert.
Writing: Alexandre Blanckaert, Bret A. Payseur.

References

Bank, C., R. Bürger, and J. Hermisson 2012. The limits to parapatric speciation: Dobzhansky–
Muller incompatibilities in a continent–island model. Genetics, 191(3):845–863.
Barton, N. and B. Bengtsson 1986. The barrier to genetic exchange between hybridising popu-
lations. Heredity, 57(3):357–376.
Barton, N. and G. Hewitt 1985. Analysis of hybrid zones. Annual review of Ecology and
Systematics, Pp. 113–148.
Bateson, W. 1909. Heredity and variation in modern lights. Darwin and modern science,
Pp. 85–101.
Beissinger, T., M. Gholami, M. Erbe, S. Weigend, A. Weigend, N. De Leon, D. Gianola, and H. Simianer 2016. Using the variability of linkage disequilibrium between subpopulations to infer sweeps and epistatic selection in a diverse panel of chickens. *Heredity*, 116(2):158–166.

Blanckaert, A., C. Bank, and J. Hermisson 2020. The limits to parapatric speciation 3: Evolution of strong reproductive isolation in presence of gene flow despite limited ecological differentiation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 375(1806).

Blanckaert, A. and J. Hermisson 2018. The limits to parapatric speciation II: Strengthening a preexisting genetic barrier to gene flow in parapatry. *Genetics*, 209(1):241–254.

Bozdag, O., J. Ono, J. Denton, E. Karakoc, N. Hunter, J. Leu, and D. Greig 2021. Breaking a species barrier by enabling hybrid recombination. *Current Biology*, 31(4):R180–R181.

Braverman, J., R. Hudson, N. Kaplan, C. Langley, and W. Stephan 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, 140(2):783–796.

Brideau, N., H. Flores, J. Wang, S. Maheshwari, X. Wang, and D. Barbash 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*, 314(5803):1292–1295.

Cutter, A. 2012. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in ecology & evolution*, 27(4):209–218.

Dobzhansky, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics*, 21(2):113.

Dobzhansky, T. 1937. Genetic nature of species differences. *The American Naturalist*, 71(735):404–420.

Endler, J. 1977. *Geographic variation, speciation, and clines*. Princeton University Press.

Foll, M., H. Shim, and J. Jensen 2015. WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular ecology resources*, 15(1):87–98.

Franssen, S., R. Kofler, and C. Schlötterer 2017. Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity*, 118(1):42–51.

Gompert, Z. and A. Buerkle 2009. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, 18(6):1207–1224.

Gompert, Z. and A. Buerkle 2011. Bayesian estimation of genomic clines. *Molecular Ecology*, 20(10):2111–2127.

Guerrero, R., C. Muir, S. Josway, and L. Moyle 2017. Pervasive antagonistic interactions among hybrid incompatibility loci. *PLoS Genetics*, 13(6):e1006817.

Haller, B. and P. Messer 2019. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular biology and evolution*, 36(3):632–637.
Han, E., J. Sinsheimer, and J. Novembre 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular biology and evolution*, 31(3):723–735.

Harris, K. and R. Nielsen 2016. The genetic cost of Neanderthal introgression. *Genetics*, 203(2):881–891.

Lu, Y., A. Sandoval, S. Voss, Z. Lai, S. Kneitz, W. Boswell, M. Boswell, M. Savage, C. Walter, W. Warren, et al. 2020. Oncogenic allelic interaction in *Xiphophorus* highlights hybrid incompatibility. *Proceedings of the National Academy of Sciences*, 117(47):29786–29794.

Maheshwari, S. and D. Barbash 2011. The genetics of hybrid incompatibilities. *Annual review of genetics*, 45:331–355.

Martin, C. and P. Wainwright 2013. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, 339(6116):208–211.

Matute, D., I. Butler, D. Turissini, and J. Coyne 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, 329(5998):1518–1521.

Maynard Smith, J. and J. Haigh 1974. The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35.

Mayr, E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*, number 13. Harvard University Press.

Mihola, O., Z. Trachtulec, C. Vlcek, J. Schimenti, and J. Forejt 2009. A mouse speciation gene encodes a meiotic histone h3 methyltransferase. *Science*, 323(5912):373–375.

Moyle, L. and T. Nakazato 2010. Hybrid incompatibility “snowballs” between *Solanum* species. *Science*, 329(5998):1521–1523.

Muller, H. 1942. Isolating mechanisms, evolution and temperature. In *Biol. Symp*, volume 6, Pp. 71–125.

Nielsen, R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39:197–218.

Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PloS one*, 7(7):e37558.

Ohta, T. 1982a. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences*, 79(6):1940–1944.

Ohta, T. 1982b. Linkage disequilibrium with the island model. *Genetics*, 101(1):139–155.

Orr, H. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, 139(4):1805–1813.

Orr, H. and M. Turelli 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, 55(6):1085–1094.
Payseur, B. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, 10(5):806–820.

Payseur, B. and H. Hoekstra 2005. Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. *Genetics*, 171(4):1905–1916.

Payseur, B. and L. Rieseberg 2016. A genomic perspective on hybridization and speciation. *Molecular ecology*, 25(11):2337–2360.

Pool, J. 2015. The mosaic ancestry of the *Drosophila* genetic reference panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Molecular biology and evolution*, 32(12):3236–3251.

Porter, A., R. Wenger, H. Geiger, A. Scholl, and A. Shapiro 1997. The *Pontia Daplidice-Edusa* hybrid zone in northwestern Italy. *Evolution*, 51(5):1561–1573.

Powell, D., M. García-Olazábal, M. Keegan, P. Reilly, K. Du, A. Díaz-Loyo, S. Banerjee, D. Blakk, D. Reich, P. Andolfatto, et al. 2020. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science*, 368(6492):731–736.

Presgraves, D. 2010. The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11(3):175–180.

Rieseberg, L., J. Whitton, and K. Gardner 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152(2):713–727.

Satyaki, P., T. Cuykendall, K. Wei, N. Brideau, H. Kwak, S. Aruna, P. Ferree, S. Ji, and D. Barbash 2014. The *Hmr* and *Lhr* hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet*, 10(3):e1004240.

Schumer, M. and Y. Brandvain 2016. Determining epistatic selection in admixed populations. *Molecular Ecology*, 25(11):2577–2591.

Schumer, M., R. Cui, D. Powell, R. Dresner, G. Rosenthal, and P. Andolfatto 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife*, 3.

Simonsen, K., G. Churchill, and C. Aquadro 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, 141(1):413–429.

Szymura, J. and N. Barton 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, 40(6):1141–1159.

Teeter, K., B. Payseur, L. Harris, M. Bakewell, L. Thibodeau, J. O’Brien, J. Krenz, M. Sans-Fuentes, M. Nachman, and P. Tucker 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome research*, 18(1):67–76.

Turelli, M. and H. Orr 2000. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, 154(4):1663–79.
Wang, R., M. White, and B. Payseur 2015. The pace of hybrid incompatibility evolution in house mice. *Genetics*, 201(1):229–242.

Williamson, S., M. Hubisz, A. Clark, B. Payseur, C. Bustamante, and R. Nielsen 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet*, 3(6):e90.

Wright, S. 1938. The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences of the United States of America*, 24(7):253.