Selection leads to remarkable variability in the outcomes of hybridization across replicate hybrid zones

S. Eryn McFarlane[1], Joshua P. Jahner[1], Dorothea Lindtke[2], C. Alex Buerkle[1], Elizabeth G. Mandeville[3]

[1] Department of Botany, University of Wyoming, Laramie, Wyoming, USA

[2] Agroscope, Animal GenoPhenomics, Posieux, Switzerland

[3] Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

ORCID

S. Eryn McFarlane	https://orcid.org/0000-0002-0706-458X
Joshua P. Jahner	https://orcid.org/0000-0001-8121-6783
Dorothea Lindtke	https://orcid.org/0000-0003-1339-1410
C. Alex Buerkle	https://orcid.org/0000-0003-4222-8858
Elizabeth G. Mandeville	https://orcid.org/0000-0002-9461-9140

Correspondence:
S. Eryn McFarlane, Department of Botany, University of Wyoming, Laramie, WY, USA.
Email: eryn.mcfarlane@gmail.com

Keywords: *ancestry, conservation, drift, incompatibility, introgression, secondary contact, speciation*

Running head: Variable outcomes of hybridization

1 Abstract

Hybrid zones have been held up as an opportunity to see speciation in action, as windows into the evolutionary process. Cases of repeated hybridization provide opportunities for comparative analysis. Within this context, there is an assumption that if the same genetic incompatibilities are maintaining reproductive isolation across all instances of secondary contact, those incompatibilities should be identifiable by consistent patterns across hybrid zones. This is in contrast with changes in allele frequencies due to genetic drift, which will be idiosyncratic for each hybrid zone. To test this assumption, we simulated 20 replicates of each of 24 hybrid zone scenarios with varied genetic incompatibilities, rates of migration, and selection. We found remarkable variability in the outcomes of hybridization in replicate hybrid zones, particularly with Dobzhansky-Muller incompatibilities and strong selection. Specifically, we found substantial differences among replicates in the overall genomic composition of individuals, including admixture proportions, inter-specific ancestry complement in the diploid genome, and number of ancestry junctions. Additionally, we found substantial variation among replicates at focal loci, including a locus that contributed to incompatibility, a locus linked to a locus under selection, and a locus not under selection. From this, we conclude that processes other than selection are responsible for some consistent outcomes of hybridization, whereas selection on incompatibilities can lead to genomically widespread and highly variable outcomes among replicates. We use these results to highlight the challenge of mapping between pattern and process in hybrid zones and call attention to how selection against incompatibilities will commonly lead to unpredictable, variable outcomes. Our hope is that this study informs future research on replicate hybrid zones and encourages further development of statistical techniques, theoretical models, and exploration of additional axes of variation to understand the maintenance of reproductive isolation, while recognizing the difficulties of confidently mapping process to pattern.

2 Introduction

Following and during speciation, the local or global persistence of novel species will commonly be tested through range shifts and secondary contact between diverged lineages. Species will be more likely to persist in the face of potential interspecific gene flow if hybrid fitness is low due to incompatible alleles or genotypes (Coyne & Orr 2004). For some groups, we have good descriptions of the genetics of incompatibilities and evolutionary models for their origin (Fishman & Sweigart 2018, Larson *et al.* 2018, Maheshwari & Barbash 2011, Orr 1995, Powell *et al.* 2020). Likewise, we have well-developed expressions for equilibrium clinal variation following secondary contact (Barton 1979, 1983, Barton & Hewitt 1985, Barton & Gale 1993, Endler 1977). Together these have formed an expectation that we can use empirical observations of gene exchange and ancestry variation in hybrid zones to identify the genetic loci associated with reproductive isolation and variation in fitness between hybrids and parental type individuals. For example, geographic and genomic cline methods identify loci with deviations in allelic ancestry or state relative to geographic or genomic expectations of either no variation between loci (Gompert & Buerkle 2012, Szymura & Barton 1986). There is growing interest in using contrasts of multiple hybrid zones to build our understanding of their consistent and idiosyncratic features (e.g. Simon *et al.* 2021, Westram *et al.* 2021). This mirrors diverse research in evolutionary biology that uses replicates or multiple instances of evolution to reveal parallel or concordant evolution and to implicate predictable and potentially deterministic features of evolution (Elmer *et al.* 2014, Ferris *et al.* 2021, Tenaillon *et al.* 2016).

In hybrid zones, outcomes of secondary contact and the composition of hybrid populations will be driven by drift, gene flow, selection, and attributes of the genome such as linkage and recombination rates. It is an open question as to what level of consistency among replicate hybrid populations we should expect with different levels of selection: that is, would high consistency among replicates indicate strong selection and effective reproductive isolation, or would consistency among replicates instead be a hallmark of neutral, clinal variation? We might expect consistency to be indicative of natural selection, particularly if we are thinking of directional selection (Pardo-Diaz *et al.* 2012, Whitney *et al.* 2006). However, the fitnesses of hybrid genotypes that contribute to reproductive isolation do not involve directional selection, but instead underdominance (low heterozygote fitness) at individual loci, or multilocus, epistatic incompatibilities (Bateson 1909, Dobzhansky 1937, Muller 1942). Rather than deterministic and consistent shifts in allele frequency, the evolutionary dynamics associated with these fitness matrices include unstable equilibria and frequency dependence that could lead to high variance in outcomes, despite strong selection.

In a world in which the evolutionary processes governing hybridization generate predictable outcomes, one could map between empirical observations and evolutionary processes, and hybrid zone characteristics (e.g. genetic composition and ancestry of individuals, number of hybrids, time since contact) would allow inference of the different demographic conditions, including selection and gene flow. Moreover, replicate hybrid zones could be used to understand how these evolutionary processes vary across space, time, and ecological variation (Buerkle & Rieseberg 2001, Harrison & Larson 2016, Westram *et al.* 2021). In the context of identifying loci important to determining hybridization outcomes, consistency or parallelism of loci with exceptional differentiation or introgression between populations (via F_{ST}, and geographic or genomic clines) have been suggested as evidence for either loci associated with adaptive introgression or genomic regions associated with reproductive isolation (Langdon *et al.* 2022, Matute *et al.* 2020, Rieseberg *et al.* 1999, Simon *et al.* 2020, 2021, Westram *et al.* 2021). But the question remains: if hybrid zones are indeed windows into evolution (Harrison 1990), do different scenes viewed through the window imply differences in process? If instead the outcomes of hybridization are highly idiosyncratic or capricious (*sensu* Lewontin 1966), potentially due to stochasticity following initial contact, the ability to link patterns to processes will be limited no matter how many replicate hybrid zones are sampled.

To resolve expectations for consistency and contingency in evolutionary replicates, we simulated replicate outcomes of hybridization, varying selection, migration, and genetic architecture of incompatibilities. It should be noted that we have simulated only simple genetic architectures, specifically two or four locus incompatibilities, rather than tackling polygenic selection. These simulations were originally performed in association with Lindtke & Buerkle (2015), but were not previously used to quantify consistency in outcomes of hybridization. Lindtke & Buerkle (2015) demonstrated that different genetic architectures causing reproductive isolation resulted in variation in introgression and the effectiveness of reproductive isolation. In contrast, we examined variation among replicates to 1) quantify the degree of variability across replicate hybrid zones with identical evolutionary conditions (genomic architecture and fitness consequences of incompatibilities, population size, migration); 2) determine how evolutionary parameters affect consistency across replicate hybrid zones; and 3) contrast variability of genotypes at specific loci, particularly those underlying the incompatibilities between species. We characterize the outcomes of hybridization across different simple genetic architectures and strengths of migration and selection. We use these results to provide a cautionary perspective on the use of replicate hybrid zone studies for inferring the processes that regulate the outcomes of hybridization. We also evaluate the implications of these results for managing hybrid individuals in conservation or restoration contexts. Finally, we make some suggestions about how to move forward with analyses of replicate hybrid zones in the future.

3 Methods

3.1 Simulation framework

Following Lindtke & Buerkle (2015) and Gompert *et al.* (2012), we simulated diploid whole genomes (2N = 20) using the software `dfuse` (`https://www.uwyo.edu/buerkle/software/dfuse/`, `https://doi.org/10.5061/dryad.0506g`). A total of 510 loci were simulated per individual, with 51 equally spaced loci per chromosome. Individuals were arranged in a chain of 11 demes, with potentially admixed individuals occupying demes 2–10 (150 individuals per deme), and two differentiated parental species anchoring the chain in demes 1 and 11. Simulations of 11 demes allowed for population structure, with individuals in the spatially central deme unlikely to have direct parentage involving the two species (F1 or backcrossed hybrids). We also explored additional, complementary analyses for a 3 deme case (as in Lindtke & Buerkle 2015) to verify results were robust to different deme structures.

To determine the degree of variability among replicate hybrid zones across a broad range of evolutionary conditions, we considered 20 independent replicates of 24 different scenarios that varied by migration rate, strength of selection, type of incompatibility, and the presence or absence of environmental selection. Migration rate was set as $m/2$, where m was either 0.01 or 0.2 and could occur in a stepwise fashion between demes (except from deme 2 to 1 and from deme 10 to 11, where migration was not allowed). Viability selection (c) against individuals carrying mismatched alleles in either an inter-genomic (hereafter 'DMI', i.e. a Dobzhansky-Muller incompatibility) or intra-genomic incompatibility (hereafter 'pathway') was either 0, 0.2, or 0.9. For all DMI scenarios, two loci on the first chromosome (1.4, 1.6) were specified as the causal variants,

whereas two loci on both the first and second chromosomes (1.2, 1.4, 2.2, 2.4) were specified for all pathway scenarios (see Fig. 1 for genotypic fitness values). Finally, scenarios were run in the presence or absence of environmental selection, which was calculated based on the mismatch between an individual's genotype at locus 1.4 (aa = 0; ab = 0.5; bb = 1) and its deme's environmental value (1–4 = 0; 5–7 = 0.5; 8–11 = 1). It is worth noting that while individuals with an aa genotype had the highest environmental-based fitness in demes 1-4, total fitness was influenced by both environment and the underlying genetic architecture.

Individual genotypes, admixture proportion (q), inter-specific ancestry (Q_{12}), and the number of junctions between ancestry blocks were all recorded for each individual every 10 generations (from 10–100). The latter three statistics describe the ancestry of hybrid individuals: the admixture proportion (q) is the proportion of ancestry in an individual from one of the two parental species, the inter-specific ancestry (Q_{12}) is the combination of ancestry for pairs of alleles across all loci in diploid individuals, (Fig. S1; Gompert *et al.* 2014, Lindtke *et al.* 2014, Shastry *et al.* 2021), and the number of junctions describes the extent of recombination between ancestry from the parental genomes. With more generations of hybridization and recombination, inter-specific ancestry (Q_{12}) declines and the number of junctions increases. A more comprehensive description of the simulations can be found in Lindtke & Buerkle (2015).

3.2 Quantifying variability across replicate hybrid zones

Our primary goal was to determine under which scenarios replicate hybrid zones are consistent with one another. For all analyses, we considered only the central deme (i.e., deme 6) and focused on the outcomes from generations 10 and 100. As our primary metric of variability, we calculated Fisher's F statistics from analysis of variance (ANOVA) in `R v4.0.5` (R Core Team 2021) to quantify the ratio of among versus within replicate variance for a given simulation scenario. Larger F statistics correspond to a greater variance among replicates relative to variation within replicates. We also report the R^2, the proportion of variance explained by replicate.

We first focused on the variability of individual level parameters by calculating ANOVA F statistics based on q, Q_{12}, and the number of transitions (junctions) in ancestry along chromosomes. Next, we asked to what extent specific genetic loci respond similarly across replicates. To do this, we calculated ANOVA F statistics based on the genotypes of three specific loci: a locus directly under selection (the fourth locus on the first chromosome: 1.4); a locus on the same chromosome as selected loci (the tenth locus on the first chromosome: 1.10); and a physically unlinked locus (the fourth locus on the third chromosome 3.4). In addition to ANOVA on genotypes, we performed ANOVA on minor ancestry. We set the response variable as either the locus specific ancestry (0, 0.5 or 1), if mean q for the replicate was < 0.5, or, if $q > 0.5$ then we had the response variable as 1-genotype. This ensures that a minor ancestry score of 1 meant an individual had two copies of the rarer ancestry at the replicate level. As a complementary approach to assess genotypic variability among individuals and replicates, we used principal component analysis (PCA) in `R` based on raw genotypes from all 510 loci. For each replicate, we calculated a 90% confidence ellipse of the first two PCs using the *ordiellipse* function in `vegan v2.6-2` (Oksanen *et al.* 2022). Finally, we examined the variability in the genetic composition of hybrids using using multinomial regression that reflects the probabilities of homozygous and heterozygous genotypes given an individual's genome-wide admixture proportion (Gompert & Buerkle 2010).

4 Results

We examined 24 scenarios that varied with respect to the degree of gene flow ($m = 0.01$ versus 0.2), strength of selection ($c = 0$ versus 0.2 versus 0.9), genetic architecture (DMI versus pathway), and the presence or absence of environmental selection. In general, proportionally more variation was associated with differences among replicates (i.e. had larger ANOVA F statistics) when selection was high, gene flow was low, and when incompatibilities were due to a DMI (Fig. 2). When holding gene flow, selection, and genetic architecture constant, the results were qualitatively identical with or without the inclusion of environmental selection (Fig. 2). At generation 10, there was significant variation explained by replicate in number of junctions in all 24 scenarios (F range: 3.26–146, R^2 range: 0.020 - 0.482), in admixture proportion (q) for all 24 scenarios (F range: 2.16–1148, R^2 range: 0.014 - 0.880), and in inter-specific ancestry (Q_{12}) for 12 scenarios (F range: 0.72–153, R^2 range: 0.005 - 0.494) (Table S1). For the three deme structure, we also found substantial among

replicate variation in q when selection was high and migration was low (Fig. S2), but the high selection, high migration scenario was more consistent than in the 11 deme structure. At generation 100, replicates differed in the number of junctions for all 24 scenarios (F range: 17.76–1183, R^2 range: 0.079 - 0.883), in q for all 24 scenarios (F range: 7.24–3094, R^2 range: 0.044 - 0.952), and in Q_{12} for 19 scenarios (F range: 0.94–433, R^2 range: 0.006 - 0.734) (Table S1). For those scenarios with the largest F statistics for q and Q_{12}, extreme variation was not isolated to the chromosome with incompatible loci, but instead extended throughout the genome (Fig. 3; note that Q_{12} was less variable in the three deme structure, Fig. S3). Accordingly, in these scenarios the genetic composition of hybrids varied greatly across replicates for key characteristics such as of the proportion of F2 hybrids and the direction of backcrossing (Fig. 4), although there was less variation among replicates in the three deme case (Fig. S4). Across scenarios, log-transformed F statistics for q and Q_{12} were highly correlated with one another in both generation 10 (r = 0.874; Fig. S5) and generation 100 (r = 0.933; Fig. S6). Further, a replicate's median q, and to a lesser extent Q_{12}, in generation 10 were often correlated with its median value in generation 100, especially when selection was strong ($c = 0.9$) (Figs. S7, S8).

Each of the three focal loci (1.4: locus under selection; 1.10: locus physically linked to causal locus; 3.4: physically unlinked locus) had different proportions of variation explained by replicate, depending on the simulation scenario. Replicate was a significant predictor of variation in all scenarios for each of the 3 loci (Table S2). The ratio of variance explained by replicate in locus 1.4 ranged from F = 2.7 to 233, where there was little variation explained due to replicate in the DMI scenario with high migration and moderate selection ($m = 0.2$, $c = 0.2$, F = 2.7, P < 0.001, $R^2 = 0.017$). In contrast, in the DMI scenario with low migration and high selection, there was huge proportion of variation in genotype explained by replicates (F = 233.1, P < 0.001, $R^2 = 0.598$). Similar patterns across scenarios were found for the other two loci, suggesting that the patterns extended across the genome. We found that using minor ancestry instead of genotype recovered a similar pattern, where replicate explained substantial variation in nearly all scenarios (Table S3). The PCA based on the genotypes from all 510 loci failed to differentiate the replicates from generation 10, except for the scenarios involving a DMI with $m = 0.01$ and $c = 0.9$ (Fig. S9). At generation 100, however, PCA was able to distinguish among replicates for all low gene flow ($m = 0.01$) scenarios, but replicates were qualitatively identical for all high gene flow ($m = 0.2$) scenarios (Fig. S10). Finally, there was substantial variation in the multinomial regressions of ancestry each of the three focal loci in both the 11 deme scenario (Figs. 5, S11, S12, S13) and the three deme scenario (Fig. S14).

5 Discussion

For many of the evolutionary scenarios we modeled, there was remarkable variation in the outcomes of hybridization across replicates (Fig. 2) despite their identical evolutionary and demographic conditions. Notably, both genome-wide ancestry composition of individuals and genotypes at individual loci varied extensively across replicates. Strong selection on the same loci led to increased variation among replicates, in contrast to naïve expectations that selection would lead to more consistent outcomes. Moreover this result extended across the entire genome with pronounced intra- and inter-chromosomal linkage disequilibrium (Fig. 3). Even loci that were not on the same chromosome as the causal loci for incompatibilities were affected by the strong viability selection. This lack of a direct causal link between selection strength and outcomes of hybridization suggests low identifiability of evolutionary processes in replicate hybrid zones. Similar outcomes do not indicate strong selection, nor do similar outcomes exclude the possibility of strong selection acting in the system. Indeed, we found the least variability explained by among replicate differences when selection was weak, suggesting that consistency could indicate a lack of reproductive isolation. Variability in outcomes could arise from many causes. Our results highlight a prominent role for stochasticity in the composition of hybrid zones (Blount *et al.* 2008), particularly when selection is strong, meaning that chance events in early generations of hybridization lead to replicate hybrid zones proceeding in dramatically different directions. This suggests that a "re-run of the tapes" could yield drastically different outcomes (Gould 1989). Below, we discuss these results in the context of both speciation research and conservation genetics.

5.1 Replicate hybrid zones in speciation genetics

When the geographic ranges of species collide in secondary contact, outcomes of hybridization have been expected to be similar when mechanisms of reproductive isolation are the same. Further, previous work has shown that selection on individual loci can lead these loci to deviate in ancestry from neutral loci elsewhere in the genome (Gompert & Buerkle 2011), though attributing deviation at exceptional loci to drift or selection is fraught (Gompert & Buerkle 2011, McFarlane *et al.* 2021). These assumptions underlie the basic premise of replicate hybrid zone research: similarities across replicates should be indicative of shared isolating barriers, independent of the effects of genetic drift at each individual population, thereby giving insight into the nature of reproductive isolation between a pair of species (Langdon *et al.* 2022, Matute *et al.* 2020, Rieseberg *et al.* 1999, Simon *et al.* 2020, 2021, Westram *et al.* 2021). Despite this expectation, many of our simulations highlight remarkable variation in individual-level parameters (Figs. 2, 4), genotypes (Fig. 3), and introgression (Fig. 5) across replicates. For example, more than 80% of the variation in q is explained by among replicate variation when migration is rare and selection against DMIs are strong (Table S1). When reproductive isolation is strong, one might anticipate consistency, where strong selection might shore up reproductive isolation. But if there are variable outcomes, then reproductive isolation could collapse in some places but not others, under identical conditions. These unstable equilibria resulting from frequency dependent selection and underdominance (in the context of multi-locus epistasis) in our simulations suggest that a substantial amount of stochasticity could result when reproductive isolation is tested in secondary contact.

In contrast to the prediction that strong selection should lead to consistency across replicates, we found that Dobzhansky-Muller incompatibilities combined with strong selection lead to remarkable variation among replicates. Since underdominance created by DMIs can lead to alternative homozygous genotypic states with full fitness, alternative alleles can rise to high frequency in our different simulation replicates (Fig. 1). This was reflected by a bimodal distribution of q scores among replicates particularly when selection on DMIs was high (Fig. 2). We found that the stronger the barriers to gene flow at specific loci, the more variable were the outcomes under identical conditions, and the more variation was explained by among replicate differences, and thus, the more unstable the allele frequencies are across replicates. Further, any drift will lead to alternative homozygotes being favoured due to frequency dependent selection (Ayala & Campbell 1974, Heino *et al.* 1998, Gompert *et al.* 2012). This means that small perturbations to identical systems due to stochasticity can result in drastically different outcomes (Fig. 1). The unstable equilibrium dynamics predominate in the first few generations after secondary contact, but eventually steady migration will bring hybrid zones back to migration-drift balance. In the meantime, selection for either parental homozygote has set the trajectory of the system towards later generations. This is why late generations are well predicted by early generations (Figs. S7, S8). This is consistent with previous work that has demonstrated the sensitivity of evolutionary trajectories to starting conditions (Simões *et al.* 2008, Losos 2011) and the historical sequence of events (i.e. historicity; Lewontin 1966); while the replicates in our simulations have identical conditions, stochasticity and drift lead to differences in early generations that are maintained.

Our conception of speciation typically includes a set of phenotypes and underlying genotypes that contribute to reproductive isolation between diverged taxa. If, as we have assumed as an idealized context here, each species is homogenous, without population structure, and reproductive isolation arises as the result of strongly decreased fitness of genotypes at a simple set of loci, repeated instances of secondary contact would nevertheless be expected to lead to different genetic compositions of hybrid individuals, and different opportunities for gene flow between parental taxa through hybrids. In other words, if taxa hybridize and make fertile and viable progeny in multiple instances of secondary contact, a single, consistent set of loci contributing to reproductive isolation would lead to variable evolutionary dynamics of hybrids and gene flow. That variability will make it more challenging to map the genes associated with reproductive isolation, even if there is a single, consistent genetic architecture. Indeed, we have previously noted that genetic drift can lead to patterns in the genome that can mimic the predicted outcomes of selection in secondary contact (Gompert & Buerkle 2011, McFarlane *et al.* 2021). But beyond the challenge of mapping pattern to process, the variation in the effectiveness of reproductive isolation even under (unrealistically) constant conditions suggests that the nature of isolation and gene flow between species is more complex than we have typically appreciated, or than would be suggested by a simple interpretation of 'speciation genetics'. For these reasons, we suggest that understanding the variation in effectiveness of incompatibilities, where secondary contact occurs in more than one location (or time point), is an interesting avenue for future research.

5.2 Implications for conservation research

In the realm of conservation, where replicate hybrid zones are studied because of the potential negative consequences of hybridization between endemic and introduced species (Allendorf *et al.* 2001, Rhymer & Simberloff 1996, Todesco *et al.* 2016, Wolf *et al.* 2001), our findings are somewhat challenging. Given finite budgets for conservation, the ability to generalize from one setting with potentially hybridizing individuals to other (unsampled) settings would be extremely useful. However, except in specific circumstances (e.g. when selection against hybrids is known to be low), we anticipate substantial variation among hybrid zones even under identical demographic and evolutionary processes. This means that to assess the likely extent of hybridization in each population, for example to protect individuals in those settings with the least hybridization, each locality might need to be sampled. Since this pattern is driven by genetic drift, the variation among localities will be even higher with small effective population sizes, as would be expected if one species has been introduced to the range of another. The predictability of evolution is negatively (but not linearly) related to effective population size (Szendro *et al.* 2013), so when considering small, introduced or endangered populations, practitioners must assume that drift and the corresponding stochasticity in the composition of populations are even stronger. One optimistic point worth highlighting is the consistency found across time within replicate hybrid zones. If conditions do not change over time, the distribution of admixture proportions in generation 100 is highly correlated with the distribution in generation 10 (Figs. S7, S8). Practitioners will need to weigh the need for temporal replication to 'check in' relative to the need for spatial replication, based on our finding that every new hybrid zone could have a different distribution than those already sampled. However, one caveat is that this temporal stability is based on a system with constant conditions, which is fairly unrealistic.

5.3 Caveats

There are a few noteworthy differences between our simulations and some empirical studies of replicate hybrid zones and the conditions that might apply to instances of secondary contact. The first is that we have simulated equal numbers of individuals between admixing populations. Since we have the same N_e for each parental population, this leads to the split in high frequency alleles in the DMIs between replicates; either pair of alleles is equally fit, and equally likely, so which ever becomes more frequent due to drift is favored by selection (Fig. 1). Other simulations of hybrid zones have demonstrated that varying N_e does not necessarily lead to increased consistency (McFarlane *et al.* 2021), although it would be interesting to quantify how extreme ratios, as might be seen with introduced, hybridizing species would affect variation among replicates. We predict that this could lead to less variation among replicates, given that the ratio of individuals is far from 50/50. In empirical cases where secondary contact is due to introductions of few individuals from one species into a large population of a second species, the outcomes of hybridization (i.e. which pair of alleles go to fixation) should be more easily be predicted based on starting allele frequencies (Kimura 1962).

Some studies have used minor ancestry at haplotypes, rather than admixture proportion, to determine to what extent there are regions of the genome with depleted variation, suggesting incompatibilities (e.g. Langdon *et al.* 2022, Schumer *et al.* 2018). While we explored this for specific loci (Table S3), we are conscious that polarizing by minor parent assumes that backcrossing in either direction has the same consequences, regardless of the ancestry of the genetic background. Given that our simulations assume the same parental population size and equal contributions to admixture (i.e. there is no asymmetry), the population determined to be the minor ancestor varies between replicates (Fig. 1). When we compared minor ancestry at specific loci, we found a very low correlation between (arbitrarily chosen) replicate 1 and 2 and significant variation across replicates in nearly all scenarios (Table S3), suggesting that using minor ancestry rather than genotype (i.e. a unimodal vs bimodal distribution of overall variation) does not substantially decrease the variation between replicates.

A second key difference between our simulations and the expectations in empirical systems is the heterogeneous recombination landscape that has been demonstrated to affect speciation (Nachman & Payseur 2012). Our simulations have a homogenous recombination landscape based on a Poisson process (Lindtke & Buerkle 2015). This homogenous recombination means that recombination hotspots and cool spots, which could lead to purging of even beneficial alleles in empirical systems (Moran *et al.* 2021), have not been simulated here. This also means that we would not have found a correlation between recombination and minor ancestry as

in (Langdon *et al.* 2022), although we also did not find a correlation of minor ancestry between replicate populations. Future work could simulate such recombination landscapes and ask how this affects variation in outcomes across replicate hybrid zones.

A third caveat is that we have not simulated adaptive introgression, where an allele from one population introgresses through the hybrid zone into the other population due to directional selection, but we expect that some of the same problems we have demonstrated will apply as well, even if underdominance does not drive variation as we expect it does for DMIs. Some empirical studies have found evidence of adaptive introgression from introduced species, using replicate hybrid zones (e.g., Bay *et al.* 2019, Calfee *et al.* 2020), but few have explicitly ruled out genetic drift as a mechanism, and selection is regularly implicated even when hybrid zones do not show concordant patterns. Given that adaptive introgression is only expected to be possible when recombination is quite high, allowing a beneficial allele to be separated from a neutral or deleterious haplotype (Moran *et al.* 2021, Uecker *et al.* 2015), the *a priori* expectation must be that introgression is due to non-adaptive forces (*sensu* Barrett & Hoekstra 2011).

Finally, more generally, our results are illustrative of variability that arises in a particular set of circumstances and are not meant to be a perfect reflection of the most likely population structure and fitnesses in secondary contact for a pair of taxa in nature. While we have perturbed these simulations to examine robustness, such as considering both an 11 deme and 3 deme population structure (Lindtke & Buerkle 2015; Figs. S2, S3), simulations can never cover all possible biological variability. The expectations for any natural setting should be informed by knowledge of the system, which will include an appreciation of the complexities due to chance in finite populations, unstable equilibrium allele frequencies (and ancestries), and the dynamics of gene flow between populations with potentially unequal and variable size through time.

5.4 Strategies for analyses of replicate hybrid zones

Even when there is substantial variability across hybrid zones, there is still the opportunity to capitalize on replication when available, and this might be particularly fruitful when the genetic architecture of incompatibilities is expected to be polygenic or selection is less strong, as such cases have less variation explained by among replicate differences (Table S1). Given that polygenic architectures are likely responsible for many cases of reproductive isolation (Barton 2022), we are encouraged by our results. By accounting for this variability across replicates statistically (rather than assuming all replicates are comparable), one can examine both global and local patterns. For example, some studies have used replicate hybrid zones to ask how ecological variation affects distributions of admixture (e.g. Chatfield *et al.* 2010, Culumber *et al.* 2012, Mandeville *et al.* 2019, 2021, Velo-Antón *et al.* 2021), a question that assumes variation across replicates. Here, we have used ANOVA to estimate variance due to replicate, but we have not incorporated other variables (e.g. environmental variability) into our analyses. ANOVAs are excellent for our balanced simulations, but might not be the appropriate method for studies that want to account for variability among replicates and also ask how some other variable (e.g. human activity, temporal variation) affects hybridization outcomes. Some studies of replicate hybrid zones have used PCA or similar approaches to assess to what extent there were substantial differences among replicates, and then using pairwise correlations between studies, to determine to what extent there was concurrence between admixture distributions, allele frequencies, or cline parameters (e.g. Westram *et al.* 2021). We found that PCAs did not differentiate among replicates well in the 10th generation after admixture (Fig. S9), suggesting that PCAs are not sufficient to confirm similarity among replicates.

One could build more sophisticated hierarchical models beyond ANOVA, but the usual rules about sample size for hierarchical models will apply. For example, Calfee *et al.* (2020) fit a classic logistic cline model to inferred genome-wide individual ancestry proportions, which fit a cline specific to two replicate instances of the escape of *Apis mellifera scutellata* and subsequent hybridization with *Apis mellifera*. They were thus able to make inferences about the similarities and differences of these two hybrid zones, while accounting for between replicate variation. This means that different taxa and systems will be more fruitful for understanding causes of different outcomes of hybridization, such as multiply introduced species like trout (Mandeville *et al.* 2019), mussels (Simon *et al.* 2020), and organisms that are amenable to laboratory settings where there can be substantial replication (e.g. *Saccharomyces* sp.; (Dunn *et al.* 2013), *Tigriopus californicus*; (Pritchard & Edmands 2013)). There are opportunities to use replicate hybrid zones to shed light on both local and global patterns of hybridization and introgression, while being cautious to not assume that replicates will be

predictive of other replicates or unsampled sites.

One challenge when studying hybridization across spatial replicates is to think about how independent replicate sites are. At one end of the spectrum, 'replicates' may not be independent at all for study systems that have high levels of gene flow over the spatial scale of sampling. In contrast, some hybrid zones may be truly independent, such as instances of secondary contact isolated from one another in a mosaic hybrid zone or by a substantial distance. For example, bighorn sheep have been reintroduced to several isolated mountain ranges across North America using founders from multiple source ancestries (Jahner *et al.* 2019). More often, studies may involve multiple transects of secondary contact between diverged lineages, as in studies of house mice in Europe (Bímová *et al.* 2011, Payseur *et al.* 2004, Teeter *et al.* 2010) or *Gryllus* crickets in North America (Larson *et al.* 2014). Replicated hybrid zones may also have complex spatial structure, and replicates may be hierarchically nested, as is the case with replicating across fish hybrid zones in small streams within a larger river basin, and then also replicating across basins (e.g. *Catostomus* sucker hybridization or rainbow × cutthroat trout hybridization; Mandeville *et al.* 2017, Muhlfeld *et al.* 2017). Geographic distribution of hybridizing species, as well as the dispersal mode and capacity of study organisms, are likely to be quite important in defining independence of replicate sites, and may influence decisions about statistical treatment of comparisons. We emphasize that studies of replicate hybrid zones at all levels of independence are likely to be informative, but clearly defining the independence of replicates is an important goal.

6 Conclusion

This research was motivated by our own ambitions to use replicate hybrid zones to ask basic and applied questions in our study systems. Indeed, we have written about the promise of replicate hybrid zones in our own papers (Buerkle & Lexer 2008, Jahner *et al.* 2021, Lindtke *et al.* 2012, Mandeville *et al.* 2017, McFarlane *et al.* 2021), so we were fairly confident that we could link the outcomes of hybridization to the evolutionary processes that generated them, at least for some scenarios. As it turns out, early generation stochasticity can dramatically alter the outcomes of hybridization across hybrid zones, leading to extreme variability across replicates with identical starting conditions. Moreover, the degree of variability is amplified with increasing strength of selection on incompatibilities, suggesting that the 'simplest scenarios' will be the most difficult in which to tie pattern to process. Despite these results, we believe the future of replicate hybrid zone research remains promising, particularly if potential reasons for variation among hybrid zones are considered appropriately. We hope this study will encourage 1) the development of new statistical models that explicitly account for variability across hybrid zones and use this variability to better understand evolutionary and demographic causes of reproductive isolation; 2) future modeling exercises that examine the countless other important axes of variability that were not explored here; and 3) new empirical studies of replicate hybrid zones that acknowledge the potential lack of identifiability of evolutionary processes based on our observed outcomes of secondary contact.

7 Acknowledgements

We are grateful to Cassandre Pyne, Amanda Meuser, Amy Pitura, Teaghan Frauley, Chris Nice, and Zach Gompert for previous discussions of this manuscript. This research was supported by the Modelscape Consortium with funding from NSF (OIA-2019528). Analyses were supported by the University of Wyoming's Advanced Research Computing Center and its Teton Computing Environment, Intel x86_64 cluster (https://doi.org/10.15786/M2FY47).

8 Data accessibility

The original source code for the simulation software `dfuse` can be found at the Dryad Digital Repository associated with Lindtke & Buerkle (2015) (doi: 10.5061/dryad.0506g). All code and input files for analyses can be found at github `https://github.com/erynmcfarlane/predict_hybrids/tree/master/Remarkable_variation_submission`.

9 Author contributions

SEM and EGM had the initial idea to think carefully about the utility of replicate hybrid zones in evolution and conservation. DL ran all of the simulations. SEM, JPJ, and CAB went down a deep, statistical rabbit hole before ultimately deciding to use ANOVA to analyze everything. SEM wrote the initial draft of the manuscript, and all authors contributed to subsequent revisions.

References

Allendorf F, Leary R, Spruell P, Wenburg J (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Ayala FJ, Campbell CA (1974) Frequency-dependent selection. *Annual review of Ecology and systematics*, pp. 115–138.

Barrett RD, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341–359.

Barton NH (1983) Multilocus clines. *Evolution*, **37**, 454–471.

Barton NH (2022) The "new synthesis". *Proceedings of the National Academy of Sciences*, **119**, e2122147119.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 13–45, Oxford University Press, New York.

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

Bateson W (1909) Heredity and variation in modern lights. In: *Darwin and modern science* (ed. Seward AC), pp. 85–101, Cambridge University Press, Cambridge, UK.

Bay RA, Taylor EB, Schluter D (2019) Parallel introgression and selection on introduced alleles in a native species. *Molecular Ecology*, **28**, 2802–2813.

Bímová BV, Macholán M, Baird SJ, *et al.* (2011) Reinforcement selection acting on the european house mouse hybrid zone. *Molecular Ecology*, **20**, 2403–2424.

Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, **105**, 7899–7906.

Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution*, **23**, 686–694.

Buerkle CA, Rieseberg LH (2001) Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution*, **55**, 684–691.

Calfee E, Agra MN, Palacio MA, Ramírez SR, Coop G (2020) Selection and hybridization shaped the rapid spread of african honey bee ancestry in the americas. *PLoS genetics*, **16**, e1009038.

Chatfield MWH, Kozak K, Fitzpatrick B, Tucker P (2010) Patterns of differential introgression in a salamander hybrid zone: inferences from genetic data and ecological niche modelling. *Molecular Ecology*, **19**, 4265–4282.

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

Culumber Z, Shepard D, Coleman S, Rosenthal G, Tobler M (2012) Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (teleostei: Xiphophorus). *Journal of evolutionary biology*, **25**, 1800–1814.

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

Dunn B, Paulish T, Stanbery A, *et al.* (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS genetics*, **9**, e1003366.

Elmer KR, Fan S, Kusche H, *et al.* (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications*, **5**, 1–8.

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

Ferris KG, Chavez AS, Suzuki TA, *et al.* (2021) The genomics of rapid climatic adaptation and parallel evolution in North American house mice. *PLoS Genetics*, **17**, e1009495.

Fishman L, Sweigart AL (2018) When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annual Review of Plant Biology*, **69**, 707–731.

Gompert Z, Buerkle CA (2010) *introgress*: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

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

Gompert Z, Buerkle CA (2012) bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, **12**, 1168–1176.

Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**, 4555–4573.

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 439–450.

Gould SJ (1989) *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton & Company, Inc., New York, NY, USA.

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

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular ecology*, **25**, 2454–2466.

Heino M, Metz JA, Kaitala V (1998) The enigma of frequency-dependent selection. *Trends in Ecology & Evolution*, **13**, 367–370.

Jahner JP, Matocq MD, Malaney JL, *et al.* (2019) The genetic legacy of 50 years of desert bighorn sheep translocations. *Evolutionary Applications*, **12**, 198–213.

Jahner JP, Parchman TL, Matocq MD (2021) Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, **30**, 4245–4258.

Kimura M (1962) On the probability of fixation of mutant genes in a population. *Genetics*, **47**, 713–719.

Langdon QK, Powell DL, Kim B, *et al.* (2022) Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLoS Genetics*, **18**, e1009914.

Larson EL, Vanderpool D, Sarver BA, *et al.* (2018) The evolution of polymorphic hybrid incompatibilities in house mice. *Genetics*, **209**, 845–859.

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. *Molecular ecology*, **23**, 1668–1678.

Lewontin RC (1966) Is nature probable or capricious? *BioScience*, **16**, 25–27.

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

Lindtke D, Buerkle CA, Barbará T, *et al.* (2012) Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology*, **21**, 5042–5058.

Lindtke D, Gompert Z, Lexer C, Buerkle CA (2014) Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular ecology*, **23**, 4316–4330.

Losos JB (2011) Convergence, adaptation, and constraint. *Evolution: International Journal of Organic Evolution*, **65**, 1827–1840.

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

Mandeville EG, Hall RO, Buerkle CA (2021) Ecological outcomes of hybridization vary extensively in catostomus fishes. *bioRxiv*.

Mandeville EG, Parchman TL, Thompson KG, *et al.* (2017) Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. *Evolution Letters*, **1**, 255–268.

Mandeville EG, Walters AW, Nordberg BJ, Higgins KH, Burckhardt JC, Wagner CE (2019) Variable hybridization outcomes in trout are predicted by historical fish stocking and environmental context. *Molecular ecology*, **28**, 3738–3755.

Matute DR, Comeault AA, Earley E, *et al.* (2020) Rapid and predictable evolution of admixed populations between two drosophila species pairs. *Genetics*, **214**, 211–230.

McFarlane SE, Senn HV, Smith SL, Pemberton JM (2021) Locus-specific introgression in young hybrid swarms: drift may dominate selection. *Molecular Ecology*, **30**, 2104–2115.

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M (2021) The genomic consequences of hybridization. *Elife*, **10**, e69016.

Muhlfeld CC, Kovach RP, Al-Chokhachy R, *et al.* (2017) Legacy introductions and climatic variation explain spatiotemporal patterns of invasive hybridization in a native trout. *Global Change Biology*, **23**, 4663–4674.

Muller HJ (1942) Isolating mechanisms, evolution, and temperature. *Biological Symposia*, **6**, 71–125.

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

Oksanen J, Simpson GL, Blanchet FG, *et al.* (2022) *vegan: Community Ecology Package. R package version 2.6-2*. https://CRAN.R-project.org/package=vegan.

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

Pardo-Diaz C, Salazar C, Baxter SW, *et al.* (2012) Adaptive introgression across species boundaries in heliconius butterflies. *PLoS genetics*, **8**, e1002752.

Payseur BA, Krenz JG, Nachman MW (2004) Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, **58**, 2064–2078.

Powell DL, García-Olazábal M, Keegan M, *et al.* (2020) Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. *Science*, **368**, 731–736.

Pritchard VL, Edmands S (2013) The genomic trajectory of hybrid swarms: outcomes of repeated crosses between populations of tigriopus californicus. *Evolution: International Journal of Organic Evolution*, **67**, 774–791.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

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

Schumer M, Xu C, Powell DL, *et al.* (2018) Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, **360**, 656–660.

Shastry V, Adams PE, Lindtke D, *et al.* (2021) Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Molecular Ecology Resources*, **21**, 1434–1451.

Simões P, Santos J, Fragata I, Mueller LD, Rose MR, Matos M (2008) How repeatable is adaptive evolution? the role of geographical origin and founder effects in laboratory adaptation. *Evolution: International Journal of Organic Evolution*, **62**, 1817–1829.

Simon A, Arbiol C, Nielsen EE, *et al.* (2020) Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evolutionary Applications*, **13**, 575–599.

Simon A, Fraïsse C, El Ayari T, *et al.* (2021) How do species barriers decay? concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, **34**, 208–223.

Szendro IG, Franke J, de Visser JAG, Krug J (2013) Predictability of evolution depends nonmonotonically on population size. *Proceedings of the National Academy of Sciences*, **110**, 571–576.

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

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mouse. *Evolution*, **64**, 472–485.

Tenaillon O, Barrick JE, Ribeck N, *et al.* (2016) Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, **536**, 165–170.

Todesco M, Pascual MA, Owens GL, *et al.* (2016) Hybridization and extinction. *Evolutionary applications*, **9**, 892–908.

Uecker H, Setter D, Hermisson J (2015) Adaptive gene introgression after secondary contact. *Journal of mathematical biology*, **70**, 1523–1580.

Velo-Antón G, Lourenço A, Galán P, Nicieza A, Tarroso P (2021) Landscape resistance constrains hybridization across contact zones in a reproductively and morphologically polymorphic salamander. *Scientific reports*, **11**, 1–16.

Westram AM, Faria R, Johannesson K, Butlin R (2021) Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, **30**, 3797–3814.

Whitney KD, Randell RA, Rieseberg LH (2006) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *American Naturalist*, **167**, 794–807.

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. *Conservation Biology*, **15**, 1039–1053.

10 Figures

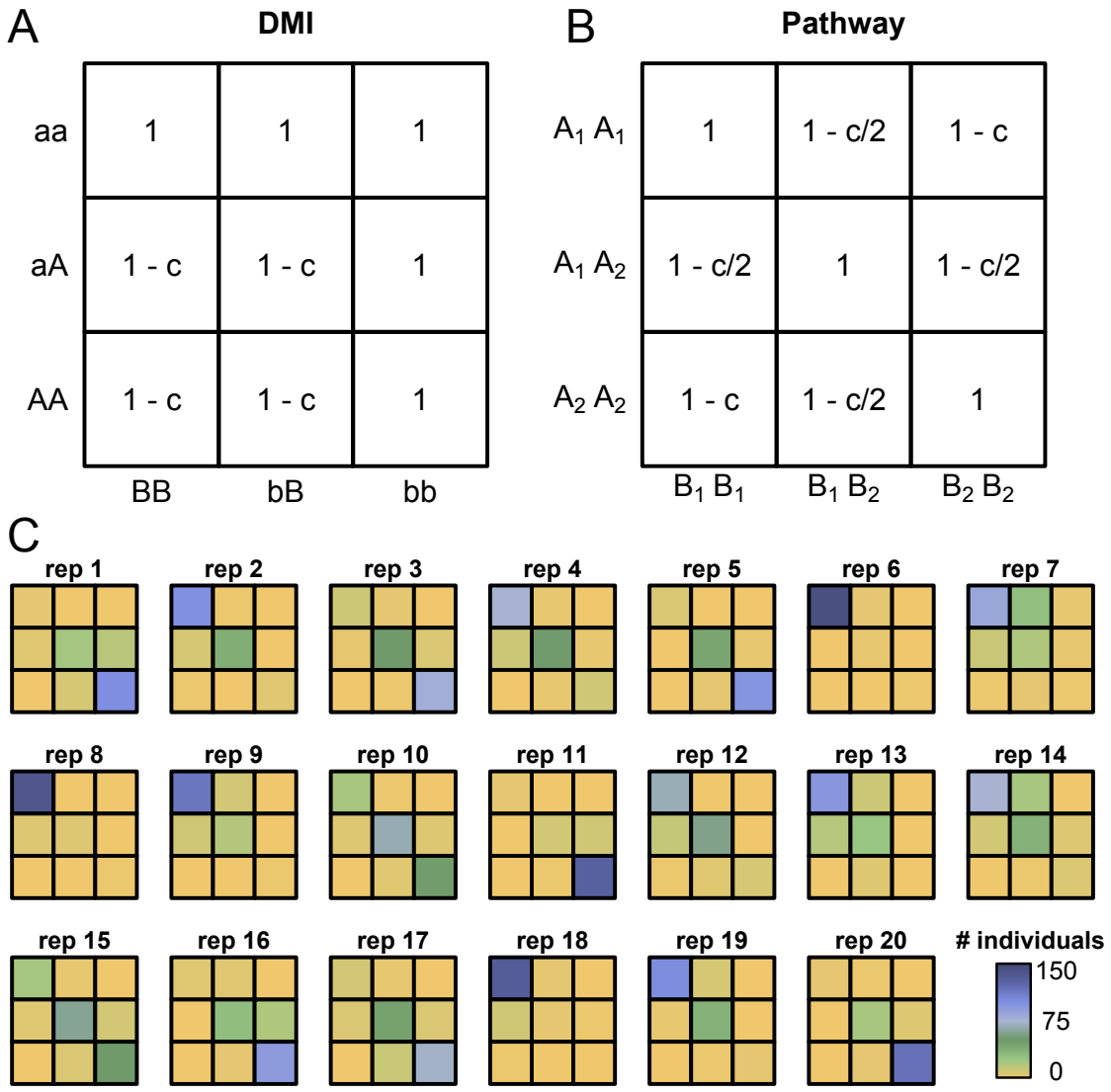

Figure 1: Genotype fitness matrices are shown for the (A) DMI and (B) pathway genetic architectures. Levels of selection (c) were set to either 0, 0.2, or 0.9. (C) The number of individuals from generation 10 having each genotypic combination for the two DMI loci (1.4 and 1.6) is displayed for each of the twenty replicates from deme 6 of the DMI, migration (m) = 0.01, selection (c) = 0.9 scenario.

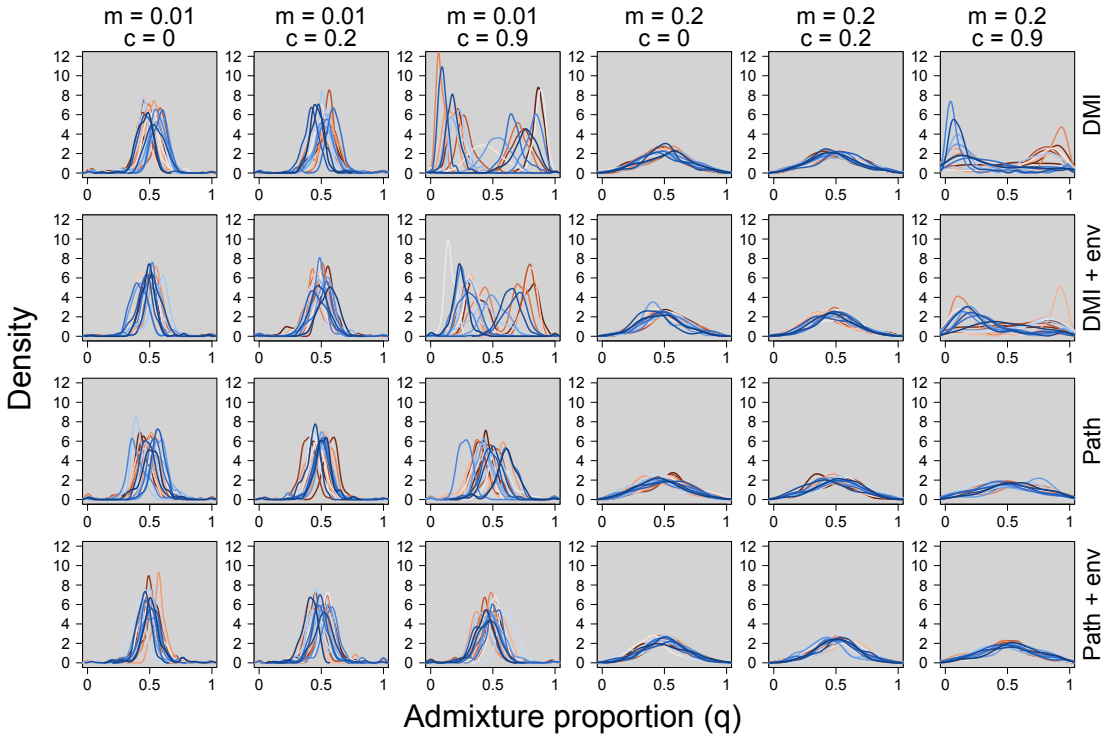

Figure 2: Selection leads to remarkable variation in the outcomes of hybridization across replicate hybrid zones. The distribution of inter-specific ancestry (q) from generation 10 is shown for 20 replicates from the 24 scenarios that varied by migration (m), selection (c), incompatibility (DMI versus pathway), and the presence or absence of environmental (env) factors. Colors correspond to different replicates and match those in other figures. See Fig. S6 for results from generation 100.

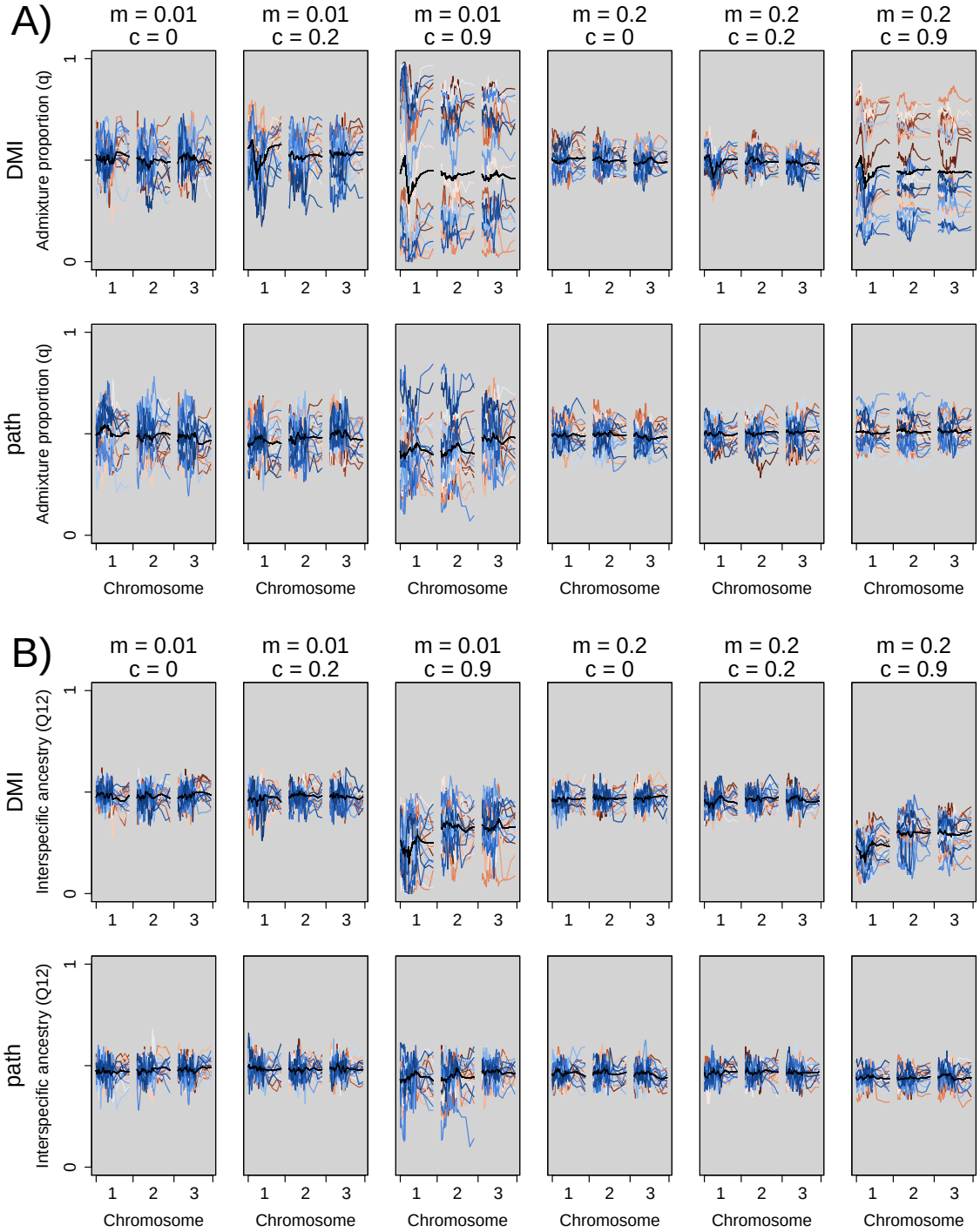

Figure 3: Locus-specific variation in average (among individuals in a replicate) admixture proportion (q, panel A) and inter-specific ancestry (Q_{12}, panel B) for replicates (different colours), six combinations (columns) of migration ($m = 0.01$ or 0.2) and selection ($c = 0$, 0.2 or 0.9), and two types of incompatibilities (DMI or pathways, in rows). Colors correspond to different replicates and match those in other figures. The black line is the average across all replicates, as in Lindtke & Buerkle (2015). The plots are for simulations after ten generations. Chromosome 1 has loci with DMI or pathway incompatibilities, chromosome 2 has pathway incompatibility loci, while chromosome 3 has only neutral loci, and is representative of the rest of the genome (chromosomes 4–10).

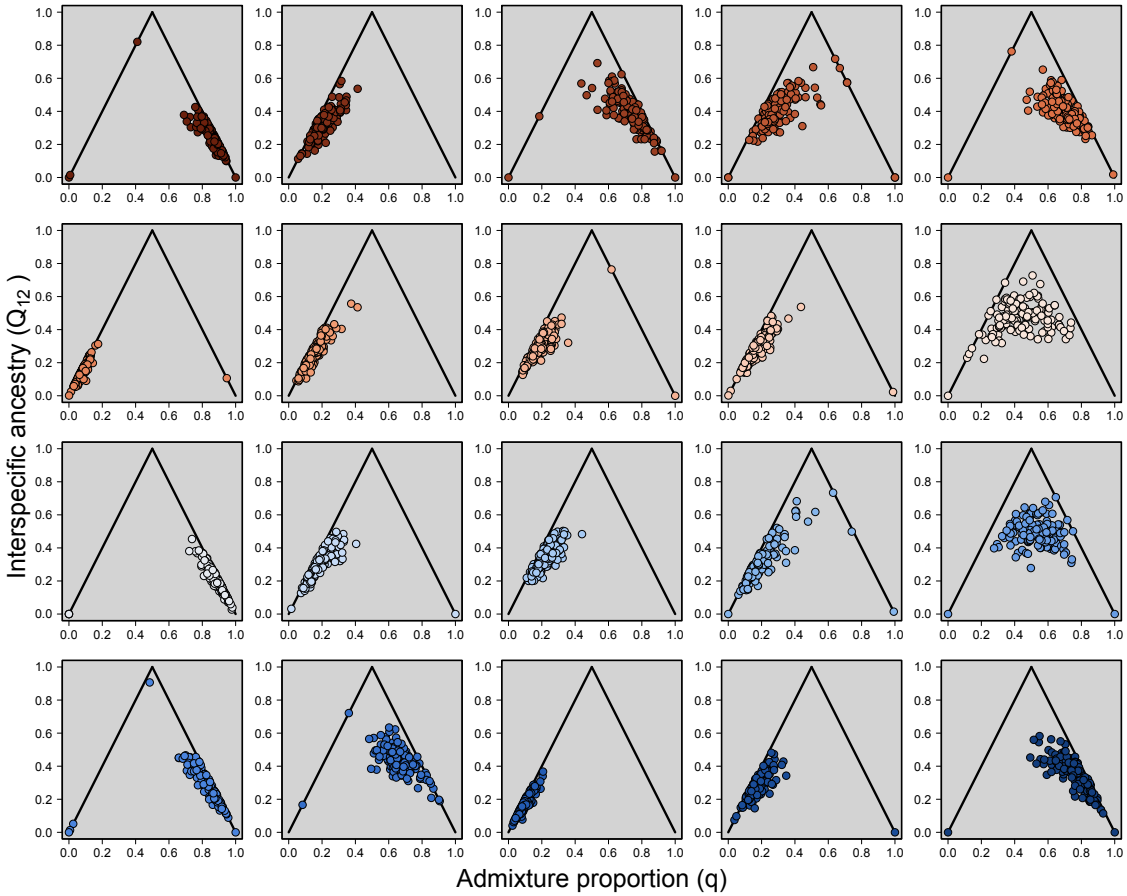

Figure 4: Individual ancestry varies greatly across the twenty replicates from the DMI, migration (m) = 0.01, selection (c) = 0.9 simulations (generation 10, deme 6). Paired estimates of admixture proportion (q) and inter-specific ancestry (Q_{12}) can be used to classify individuals as parentals ($q = 0$; $Q_{12} = 0$ or 1), F1 hybrids ($q = 0.5$; Q_{12}= 1), F2 hybrids (on average $q = 0.5$, $Q_{12} = 0.5$), or backcrosses (residing on the solid black lines) (Gompert *et al.* 2014, Lindtke *et al.* 2014, Shastry *et al.* 2021). Colors correspond to different replicates and match those in other figures.

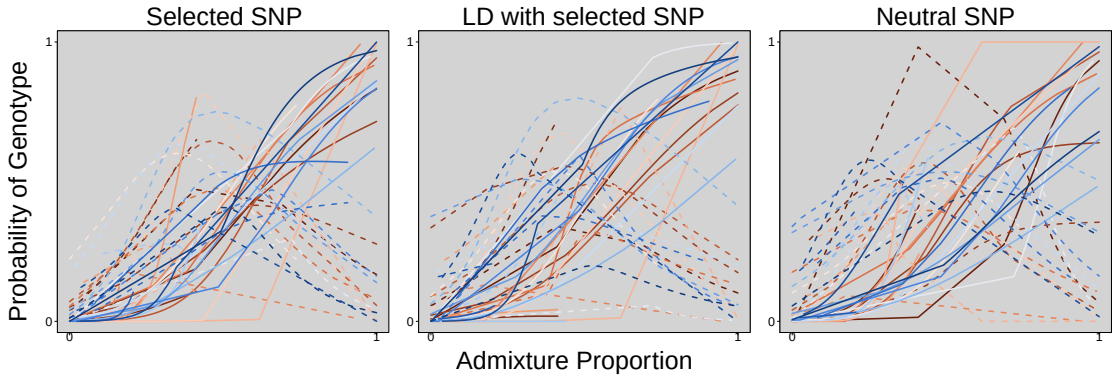

Figure 5: Probability of each AA (homozygous; solid line) and Aa (codominant heterozygote; dashed line) for each of three simulated SNPs in the DMI scenario where migration is low (m=0.01), and selection is high (c=0.9). Colors correspond to different replicates and match those in other figures. This multinomial regression shows the probability of either a homozygous or heterozygous genotype for a given admixture proportion q.