HHS Public Access

Author manuscript

Mol Ecol Resour. Author manuscript; available in PMC 2024 November 25.

Published in final edited form as:

Mol Ecol Resour. 2021 October ; 21(7): 2278–2287. doi:10.1111/1755-0998.13428.

A narrow window for geographic cline analysis using genomic data: Effects of age, drift, and migration on error rates

Gaston I. Jofre[1,2,3], **Gil G. Rosenthal**[1,2]

[1]Department of Biology, Texas A&M University, TAMU, College Station, TX, USA

[2]Centro de Investigaciones Cientıficas de las Huastecas “Aguazarca”, Calnali Hidalgo, Mexico

[3]Department of Biology, University of North Carolina, Chapel Hill, NC, USA

Abstract

The use of genomic and phenotypic data to scan for outliers is a mainstay for studies of hybridization and speciation. Geographic cline analysis of natural hybrid zones is widely used to identify putative signatures of selection by detecting deviations from baseline patterns of introgression. As with other outlier-based approaches, demographic histories can make neutral regions appear to be under selection and vice versa. In this study, we use a forward-time individual-based simulation approach to evaluate the robustness of geographic cline analysis under different evolutionary scenarios. We modelled multiple stepping-stone hybrid zones with distinct age, deme sizes, and migration rates, and evolving under different types of selection. We found that drift distorts cline shapes and increases false positive rates for signatures of selection. This effect increases with hybrid zone age, particularly if migration between demes is low. Drift can also distort the signature of deleterious effects of hybridization, with genetic incompatibilities and particularly underdominance prone to spurious typing as adaptive introgression. Our results suggest that geographic clines are most useful for outlier analysis in young hybrid zones with large populations of hybrid individuals. Current approaches may overestimate adaptive introgression and underestimate selection against maladaptive genotypes.

Keywords

conservation genetics; ecological genomics; outlier analysis; spatial analysis

1 | INTRODUCTION

Hybridization occurs when two genetically distinct individuals reproduce and generate viable offspring of mixed ancestry (Abbott et al., 2013; Barton & Hewitt, 1985). After a

Correspondence Gaston I. Jofre, Department of Biology, University of North Carolina, Chapel Hill, NC, USA. gastonjr@email.unc.edu.

AUTHOR CONTRIBUTIONS

Gaston I. Jofre was responsible for the conceptualization, simulations, data modification, analysis, writing-original draft, project administration, writing review, and editing. Gil G. Rosenthal was responsible for the conceptualization, project administration, writing review and editing.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

hybridization event, selection against introgression maintains species boundaries (Coughlan & Matute, 2020; Mallet, 2005). Alternatively, alleles favoured by selection can introgress from one parent population to another across a hybrid zone (Martin & Jiggins, 2017; Muirhead & Presgraves, 2016). Such histories of selection can leave signatures in the spatial distribution of allele frequencies along a hybrid zone relative to neutral expectations.

Cline theory has long been used to detect phenotypic introgression across wild hybrid zones and draw inferences about selection (e.g., asymmetric sexual selection, Baldassarre et al., 2014; Parsons et al., 1993). More recently, cline theory has been widely employed to the same end to detect outlier regions of the genome more permeable or resistant to introgression (Lipshutz et al., 2019; Ryan et al., 2017; Stankowski et al., 2017). Cline theory addresses the changes in ancestry-allele frequencies from genomic markers across a spatial gradient (Figure 1a), which may covary with environmental parameters (Barton & Hewitt, 1985; Endler, 1977; Haldane, 1948).

At its simplest, the analysis uses two parameters, centre and width. The centre (Figure 1a, vertical line) is the point where the slope of the fitted sigmoidal curve is the steepest (Barton & Gale, 1993; Szymura & Barton, 1986), and is the geographic location at which mean ancestry-allele frequencies are exactly intermediate. The width of the cline (Figure 1a, horizontal arrow) is the inverse of the maximum slope, and is inversely related to the rate of change of ancestry-allele frequencies along the spatial gradient (Barton & Gale, 1993).

In the absence of selection, the allele frequencies of neutral markers change progressively across the cline (Figure 1b; Barton & Hewitt, 1989). Their widths increase with the rate of migration from parental populations (Barton & Hewitt, 1985; Slatkin, 1973). Cline centres from neutral markers are expected to coincide with each other in the spatial gradient (Gay et al., 2008). Cline analysis seeks outliers that deviate from the null expectation of no selection (Barton & Hewitt, 1985; Gay et al., 2008; Szymura & Barton, 1986), indicating candidate targets of selection (e.g., Fitzpatrick et al., 2010; Payseur, 2010; Stankowski et al., 2017). Here, we focus on comparing allele-frequency clines for individual markers relative to the cline for genome-wide ancestry (Figure 1b). Candidates for adaptive introgression show signatures of directional selection, shifting cline centres away from the ancestry average (Figure 1b; Barton & Hewitt, 1985; Fitzpatrick et al., 2010). By contrast, markers involved in reproductive isolation have reduced introgression and produce steep clines with narrow widths and drastic changes in allele frequencies across short distances (Figure 1b; Barton & Hewitt, 1989). Selection against migrants from the cline edges maintains the cline centres at the middle of the hybrid zone. The cline width is approximately proportionate to selection against migrants; as selection increases the cline becomes steeper (Barton & Gale, 1993; Payseur, 2010; Sciuchetti et al., 2018).

While cline analysis offers an appealing tool for detecting candidate genomic regions under selection, the interaction of migration and drift can generate spurious outliers in the absence of selection. Narrow clines, like those expected from underdominance, can emerge simply from small deme sizes and low migration rates (Figure 1b; Felsenstein, 1975; Slatkin & Maruyama, 1975). By contrast, migration can shift cline widths and centres not due to

adaptive introgression but simply due to the homogenizing effects of gene flow (Barton & Hewitt, 1985; Slatkin, 1973).

In this study, we apply individual-based forward-time simulations of stepping-stone hybrid zones (e.g., Hvala et al., 2018; Polechova & Barton, 2011; Sciuchetti et al., 2018) to evaluate the effects of a range of demographic parameters on cline structure and on the distribution of spurious signatures of selection. Specifically, we aim to calculate the rate of erroneous outlier detection in geographic clines across different demographic histories of migration, drift, and selection.

2 | MATERIALS AND METHODS

2.1 | Simulation and modelling

To model how hybrid zone properties affected outlier detection, we used Admix'em (Cui et al. (2016); github: https://github.com/melop/admixem). This program generates realistic simulations of admixed diploid populations (Massey, 2017; Schumer & Brandvain, 2016; Schumer et al., 2018). For all simulations, we used a one-dimensional stepping-stone model with 19 discrete hybrid demes and two infinitely large source populations, each at one extreme of the cline (Figure 2; Table S1). This mimics the biological inspiration for our model, natural hybrid zones across an elevation gradient in montane streams, with small pools separated by waterfalls and flumes. Transformed one dimensional transects are required for geographic cline analysis (Derryberry et al., 2014; Szymura & Barton, 1986). When natural hybrid zones are distributed along two dimensions, researchers often fit demes to a one dimensional transect (e.g., Gay et al., 2008; Stankowski et al., 2017; Szymura & Barton, 1986). Thus, we used the stepping-stone model, for simplicity, as it is an approximation of many natural hybrid zones (Hvala et al., 2018), and because is widely used in the literature (Feldman & Christiansen, 1974; Gavrilets, 1997; Sciuchetti et al., 2018; Slatkin & Maruyama, 1975). We restricted the number of hybrid demes to 19 (similar to Felsenstein, 1975), to model clines with boundaries away from their cline centres, where one or the other allele will be close to fixation (see Slatkin & Maruyama, 1975), while keeping adequate computational speed.

Simulations began at generation zero, with only source populations 1 and 2. We made the simplifying assumptions that populations were completely isolated, with complete lineage sorting, and without shared polymorphisms (see: Hvala et al. [2018]; Sciuchetti et al. [2018] for stepping-stone models; and Schumer et al. [2015]; Blanckaert & Bank [2018] for hybrid swarm models). Between generation zero and generation 1, each source population sent a proportion of migrants to populate each deme. The relative proportion of migrants from source population 1 versus source population 2 varied gradually along the 19 hybrid demes in 5% increments: from 95% population 1 and 5% population 2, to 90% and 10%, through to 5% and 95%. As a result, at the end of generation 1 each admixed deme had an ancestry proportion p from source population 2 increasing from $p = .05$ to $p = .95$ along the cline. For example, Deme 10 had p of 0.5 and was located in the centre of the stepping-stone cline, representing the initial point of contact. Starting between generation 1 and generation 2, each source population, and each of the 19 hybrid demes, sent a proportion of migrants only to adjacent demes (see Figure 2).

The genome of each individual consisted of 10 diploid chromosomes, each with one randomly-drawn recombination event per arm per meiosis. This recombination rate is similar to that suggested by Dumont and Payseur (2008). We tracked ancestry in 11 biallelic markers randomly distributed in each chromosome, 110 markers in total. We allowed random mating for 1000 generations and sampled individuals in the hundredth (for young hybrid zones), and thousandth generation (for old hybrid zones). Generations did not overlap; at the end of each generation parental individuals were removed after mating.

We varied the effect of drift by modifying the deme size and migration rate. We restricted the deme sizes to 100, 500, and 1,000 individuals. While limited by computational resources, our deme sizes are biologically realistic, notably for species of conservation concern. Mark-recapture studies on the biological inspiration for this model, swordtail fish in mountain streams, show hybrid deme sizes of <1000 adult individuals (Culumber et al., 2014). The Grevy's zebra in northern Kenya and Ethiopia has a global population size of 3000 individuals, has formed a hybrid zone with the plains zebra in an area of sympatry (Cordingley et al., 2009), only 25 individuals have been identified as hybrids (Schieltz & Rubenstein, 2015). While we cannot simulate hybrid zones on the scale of the broad mouse hybrid zone in central Europe (Janousek et al., 2012), our deme sizes are at the upper range of previous literature on cline theory (Table S2). Migration rates were symmetrical and restricted to 0.1 and 0.01 to account for low and high migration rates on natural hybrid zones. In a natural swordtail fish migration rate towards the hybrid zone is ~0.06 ± 0.02 (Schumer et al., 2017). In fire-bellied newts, migration rates towards the hybrid zone can be as low as 0.05 ± 0.02 and as high as 0.1 ± 0.02 (Tominaga et al., 2018). These parameter values were also intended to complement previous models of cline analysis (Table S2). We performed 100 replicated simulations with each parameter combination, for each mode of selection (see Figure 2).

Mode of selection in hybrid zones:

i. No selection:

We ran the simulations as described above with the null expectation of no selection with respect to ancestry.

ii. Directional and underdominance selection:

To determine the effects of selection on hybrids, we modelled the effects of directional selection (with incomplete dominance) and underdominance. We modelled directional selection (Figure S1A) as one allele having lower fitness than the other allele along the spatial gradient. We modelled directional selection on one marker at the middle of chromosomes three, denoted as 3.5, and another on chromosome seven, denoted as 7.5. We modelled underdominance on markers in the middle of chromosomes five and nine, denoted as 5.5 and 9.5 as heterozygous disadvantage along the spatial gradient (Figure S1B). We assumed additive fitness, with relative fitness as $f = 1 - s_{3.5} - s_{5.5} - s_{7.5} - s_{9.5}$, where s is the selection coefficient at each of the four focal markers (see selection coefficients in Figure S1A and B).

iii. Bateson–Dobzhansky–Muller incompatibility:

Unlike underdominant selection, Bateson–Dobzhansky–Muller (BDM) incompatibilities (Dobzhansky, 1937; Muller, 1940) may be a more common cause of reduced hybrid fitness (Coyne & Orr, 2004). We modelled the effects of epistatic interactions by simulating a BDM interaction between two autosomal markers, one in the middle of chromosomes three (3.5) and another on chromosome seven (7.5). We assumed codominance in the incompatible alleles and applied additive fitness, where relative fitness is $f = 1 - s_{3.5} - s_{7.5}$ where s is the selection coefficient at each of the two focal markers (see selection coefficients in Figure S1C).

2.2 | Geographic cline analysis

Our simulated data set consisted of 12 types of hybrid zones without selection, 12 with directional selection and underdominance, and 12 with BDM interactions. Each type of hybrid zone had a different combination of N, m, and Generations, and includes 100 replications (a total of 3600 hybrid zones). To study the effects of our four variables (N, m, Generations, mode of selection) on the detection of false positive and false negatives using geographic clines, we sampled 30 random individuals from of the 19 hybrid demes and source populations per hybrid zone. Admix'em generated files with ancestries at each marker from each individual. We then used custom scripts (https://github.com/gjofre/simulation_clines) to modify these ancestry files into average allele frequencies per marker per deme, and average genome wide hybrid index per deme; necessary values for geographic cline analysis. The Metropolis–Hastings algorithm in the package hzar (Derryberry et al., 2014) in R (R Core Team, 2020), was used to fit geographic clines to allele frequencies for each of the 110 markers genotyped. To fit a cline, we estimated the following five parameters in hzar: cline centre c, cline width w, the ends of the cline $P_{\min}$ and $P_{\max}$, the slope of the cline tails $\tau_{P1}$and τ_{P2}, and the size of the cline tails $\delta_{P1}$and δ_{P2} (see Figure S2; also see Szymura & Barton, 1986). For computational speed, we fitted three models that estimate different cline parameter combinations. Model I estimated only centre c and width w from the data, assumed no tails ($\tau = 1$ and $\delta = 0$), and included fixed ends ($P_{\min}$ fixed to 1 and $P_{\max}$ fixed to zero). Model II estimated c, w, and $P_{\min}$ and $P_{\max}$ from the data, and assumed no tails ($\tau = 1$ and $\delta = 0$). Model III was the same as Model II, and with tail estimates τ and δ allowed to vary from the data. In hzar we estimated each model parameter using three independent chain runs using 500,000 MCMC steps after a burnin of 100,000 steps. The model with the lowest corrected Akaike information criterion (AIC: Akaike, 1998) was chosen for each marker. From the selected model we then used the average and the 95% confidence intervals calculated by hzar for each cline parameter for outlier detection.

2.3 | Outlier detection

A marker was considered an outlier if the 95% confidence intervals from either centre or width (provided by the lowest AIC cline model), did not overlap with the confidence intervals from the genome-wide average cline model with the lowest AIC. Centres and widths of any two clines were considered coincident (same centre) or concordant (same

width) if their parameters overlapped with each other's 95% confidence intervals. This allowed us to categorize neutral regions as true negatives or false positives, and regions under selection as true positives, false negatives, or spurious centre outliers (Figure S3; Table S3).

We categorized markers without selection as true negatives if both cline centre and width were coincident and concordant with the hybrid index cline. We categorized them as false positives if their clines were either noncoincident or discordant compared to the genome-wide cline (Figure S3A). In markers under any mode of selection, we categorized false negatives as clines if both cline centre and width were coincident and concordant with the hybrid index cline. We categorized markers under directional selection as true positives if they displayed the cline centre shifted towards the side of the geographic gradient with higher fitness (Figure S3B). We categorized markers under underdominant selection and in BDM incompatibilities as true positives if their centre was coincident with the hybrid index cline, and if their width was significantly lower than the width of the hybrid index cline. Finally, we categorized markers under underdominant selection and in BDM incompatibilities as spurious centre outliers if their width was lower than the width of the whole genome hybrid index cline and if their centre was not coincident with the hybrid index cline (Figure S3C).

We computed an error matrix for each hybrid zone (see Table S3). For each replicate in our data set, we calculated the average false positive rate (FPR) as the proportion of neutral markers that were flagged as outliers. To calculate the false negative rate (FNR) we omitted the data from the markers with no selection, and calculated the proportion of markers under each mode of selection that were detected as false negatives. For markers under underdominance and involved in a BDM incompatibility, we calculated the spurious centre rate (SCR), as the proportion of markers that were detected as spurious centre outliers.

2.4 | Statistical analyses

To assess which of our four variables (N, m, generations, and mode of selection) had a higher impact on the distribution of FPR, FNR, and SCR, we fitted three linear models using the lme package in R. We first log converted the different deme sizes (N = 100, 500, 1000), migration rates (m = 0.01, 0.1), and generations (100, 1000). Then we included N, m, generations, and mode of selection as fixed effects. We also included the pairwise interactions between these variables.

Our models had the form of:

$$\begin{aligned} X = {} & N + \log m + \text{logGeneration} + \text{Selection mode} \\ & + \log N : \log m + \log N : \text{logGeneration} + \log N : \text{Selection mode} \\ & + \log m : \text{logGeneration} + \log m : \text{Selection mode} \\ & + \text{logGeneration} : \text{Selection mode} \end{aligned}$$

where $X$was FPR, FNR, or SCR as the response of each model. For the FNR linear model, we omitted the data from the markers with no selection, since there are no true positives or false negatives. For the SCR model we included only the data from markers with underdominance and BDMI (see Figure S3). We then performed post hoc comparisons

of estimated means of FPR, FNR, and SCR, in each hybrid zone model parameter with a type III ANOVA test. To evaluate the interaction effects from our independent variables on our response variables we compared the sum of squares from the ANOVA test, and the coefficient estimates from each variable, and interaction between variables in each model.

3 | RESULTS

3.1 | False positives

As predicted, false positives increased with hybrid zone age and decreased with deme size and migration rate (Figure 3; Figure S4). Our linear model revealed that deme size had the strongest effect with FPR, followed by hybrid zone age and migration rates (Table 1, Table S4, Figure S5). The mode of selection had no effect on mean FPR values. Mean FPR from neutral markers did not significantly differ in our hybrid zones regardless of the mode of selection acting upon (Table 1; Figure S4). Markers under selection did not influence the neutral markers sampled in our analysis.

Hybrid zones with small deme sizes generated the most clines with centres drastically shifted, showing spurious patterns of introgression (Figure S6A; Video S1). In the worst-case scenario, old hybrid zones with small deme sizes ($N = 100$) and low migration rates ($m = 0.01$) had an average FPR of 0.98. While small hybrid zones produced spurious outliers due to drift, there were strong main and interaction effects of migration rate and age that kept false positive rates high even in large populations if hybridization was old and migration was low. Under the most favourable conditions (young hybrid zone with high migration and large deme size) the false positive rate averaged only 1.35% across replicates (Figure 3).

3.2 | False negatives

The mode of selection had the strongest effect on FNR (Table 1; Table S4; Figure S5). In markers with underdominance and in BDM incompatibilities, FNR was zero across all parameters (Figure S4). Only markers under directional selection were likely to be false negatives, and only in young hybrid zones with small deme sizes (Figure S6B). Migration rates slightly increased FNR, and after 1000 generations of admixture all markers were categorized as true outliers. In 3% of our replicates the centres were even shifted to the opposite side of the gradient (red arrows in Figures S7A and B, Video S2).

3.3 | Spurious centres

Under most parameter combinations, markers with underdominance or involved in BDM incompatibilities were more likely to be spurious centres than true positives (Figures S6C and D). Our linear model showed that deme size had a strong decreasing effect on SCR (Figure 4; Table S4; Figure S5). Small deme sizes had the highest rates of spurious outliers, especially in old hybrid zones. Migration rate had a decreasing effect on SCR, but only in young hybrid zones.

For BDM incompatibilities, the average SCR was lower than it was for underdominant selection (Figure 4; Figure S4; Video S3 and S4), with centre shifts accumulating with

hybrid zone age (Figure 4, Figures S6C and D). Large deme sizes ($N = 1000$) sharply decreased SCR. These results suggest that markers involved in BDM incompatibilities will generate fewer markers with spurious centre outliers compared to markers under underdominant selection.

4 | DISCUSSION

Geographic cline theory (Barton & Hewitt, 1985) makes powerful predictions about how allele frequencies should change across a spatial gradient given a range of selection dynamics. Numerous studies have used whole-genome data with geographic cline analysis to detect outliers from average ancestry patterns (Payseur, 2010). However, we show that drift can distort cline shapes from hybrid zones, consistent with previous studies (Felsenstein, 1975; Nagylaki & Lucier, 1980; Polechova & Barton, 2011; Sciuchetti et al., 2018; Slatkin & Maruyama, 1975). Our simulations suggest that under a broad range of parameter combinations, relying on cline outliers may cause researchers to overestimate the importance of adaptive introgression and underestimate the deleterious genetic effects of hybridization.

Our results suggest that drift makes geographic cline analysis inappropriate for any hybrid zones with small deme sizes, as well as for older hybrid zones with low migration rates. We show that, depending on the demography (deme sizes, migration rates and age) of a hybrid zone, drift increases spurious signatures of adaptive introgression. In the absence of selection, drift can generate spurious outliers with strong patterns of reduced or increased introgression that could be taken as evidence of selection.

Hybrid zones with small deme sizes and low migration rates show high false positive rates, suggesting limitations for cline analyses. Similar to the results by Polechova and Barton (2011) and Sciuchetti et al. (2018) we found that drift increases risk of fixation within a deme, and shifts the centres away from the expected hybrid index cline (Figure 3). Increasing the population size does decrease the effects of drift, generating more concordant and coincident clines, and significantly reducing the error rate. However, this reduction in FPR is not sufficient in old hybrid zones with low migration rate, where even with large population sizes, drift can maintain high false positive rates.

Similarly, even old hybrid zones accumulate spurious centre outliers even in the face of strong selection against heterozygotes. Since heterozygous genotypes have lower fitness, drift can increase the rate of fixation, and shift the centre to either side of the gradient (e.g., Figure 4; Figures S8 and S9), which could also be misinterpreted as directional selection. This effect is particularly stronger for markers under underdominance selection versus markers involved in BDM incompatibilities.

In 3% of our replicates with small deme sizes, genomic regions under directional selection, instead of having centres shifted towards the adaptive side of the geographical gradient, showed centres shifted towards the opposite (less fit) side of the gradient (Figure S7A and B). Asymmetric introgression of multiple candidate markers, with most cline centres displaced in a consistent direction relative to baseline, strengthens the argument for adaptive

introgression, as long as studies control for spurious sources of linkage disequilibrium. For example, Yang et al. (2018) found that out of 117 candidate color loci in Italian wall lizards, 83% introgressed in the same direction as the representative phenotype. Seven candidate markers, however, show clines shifted southwest from the hybrid index cline, possibly a byproduct of drift.

Between false positives on neutral markers, and false negatives on underdominant and BDMI markers, spatial analyses of hybrid zones may be markedly overestimating the importance of adaptive introgression, or directional selection on an allele from one parent species. Putative adaptive introgression has been invoked as evidence of asymmetric sexual selection (e.g., Baldassarre et al., 2014; Parsons et al., 1993; Semenov et al., 2021) and of resilience to climate change (e.g., Hamilton & Miller, 2016; Owens & Samuk, 2020). However, our simulations indicate that apparent adaptive introgression may instead reflect drift. Further, while underdominance has generally been regarded as a minor contributor to reproductive isolation (e.g., Sedghifar et al., 2016), heterozygote fitness can be markedly reduced, for example due to chromosomal rearrangements (Stathos & Fishman, 2014). Our results suggest that underdominance may often go undetected or misattributed.

Geographic cline analysis may be most appropriate in recently formed hybrid zones with large populations of hybridizing individuals. Knowing the demographic history of the model organism a priori can reduce the risk of misidentifying neutral regions as candidate regions under selection in natural populations. Geographic cline analysis is most helpful with validation of the same outliers across multiple independent hybrid zones. For example, previous studies have used overlapping data across independent hybrid zones to draw conclusions about diagnostic markers. In the European mouse hybrid zone, from 366 outliers with patterns of reduced introgression, 59 outliers overlapped between transects and contain genes involved in male reproduction (Janousek et al., 2012). In warblers, from 11 markers, Brelsford and Irwin (2009) identified two common outliers between transects which show patterns of reproductive isolation. In *Bufo* toads, from 179 outliers with restricted introgression, 26 overlapped between transects (van Riemsdijk et al., 2020). In swordtails, from 1087 outliers with an FPR of 5%, Schumer et al. (2014) identified 327 diagnostic markers involved in genetic incompatibilities, common between two independent hybrid zones. In individual transects, stochastic processes can have an influence in the data and generate outliers. In the absence of detailed demographic information, using multiple independent transects can minimize this issue.

Geographic cline analysis is one of a family of outlier detection methods used in evolutionary and ecological genomics, for example genomic clines (Fitzpatrick, 2013; Gompert & Buerkle, 2009, 2011; Szymura & Barton, 1986), ancestry tract lengths (Sedghifar et al., 2016), and ancestry junctions (Hvala et al., 2018). All of these methods are susceptible to increased error rates due to drift, in these cases by generating excess ancestry, increasing ancestry tract sizes, and reducing junction densities. Further, we should expect to see similar patterns when it comes to phenotypic introgression, particularly when interspecific differences are associated with a small number of markers. Without previous knowledge of demographic history, an option is to compare the detected patterns

of introgression with null scenarios that assume a very large population size or a very low population sizes (see Fitzpatrick, 2013; Gompert & Buerkle, 2011).

Asymmetric introgression of traits or candidate genes is often taken as prima facie evidence of selection. Our results suggest that drift can generate both false positives and negatives that look like adaptive introgression, rather than neutral movement of alleles or selection against recombinant genotypes. Hybridization and introgression are frequently favoured by selection, and many taxa show pervasive evidence of historical hybridization in the genome, but spatial analysis of hybrid zones may lead us to overestimate the importance of adaptive processes therein. Careful consideration of hybrid zone demographics, combined with independent replication or validation of outliers, can make geographic cline analysis a powerful framework for studying genome-wide patterns of introgression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Molly Schumer for sharing example configuration files needed for Admix'em. We are grateful to James Cai, Kirk Winemiller, Kira Delmore, Hannah Justen, and members of the Matute lab, particularly Daniel Matute, who provided insightful comments. We also thank four anonymous reviewers for providing helpful comments during the review process. G. I. J. R. was supported by a Rosemary Grant award from the Society for the Study of Evolution, a CONACyT international predoctoral fellowship, and NIH award R01GM121750 to D. Matute. G. G. R. and G. I. J. R. were supported by NSF awards 1354172 and 1755327. We are deeply thankful to Heath Blackmon for sharing his server for computational analyses. Additional analyses were conducted at the High Performance Research Computing center at Texas A&M University.

Funding information

National Institutes of Health, Grant/Award Number: R01GM121750; Consejo Nacional de Ciencia y Tecnología; National Science Foundation: NSF, Grant/Award Number: 1354172 and 1755327

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been made available in Dryad at https://doi.org/10.5061/dryad.0p2ngf1z9 (Jofre and Rosenthal, 2021).

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, … Zinner D (2013). Hybridization and speciation. Journal of Evolutionary Biology, 26, 229–246. 10.1111/j.1420-9101.2012.02599.x [PubMed: 23323997]

Akaike H. (1998). A new look at the statistical model identification. In Parzen E, Tanabe K, & Kitagawa G. (Eds.), Selected Papers of Hirotugu Akaike (pp. 215–222). Springer.

Baldassarre DT, White TA, Karubian J, & Webster MS (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution, 68, 2644–2657. 10.1111/evo.12457 [PubMed: 24889818]

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

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

Barton NH, & Hewitt GM (1989). Adaptation, speciation and hybrid zones. Nature, 341, 497–503. [PubMed: 2677747]

Blanckaert A, & Bank C. (2018). In search of the Goldilocks zone for hybrid speciation. PLoS Genetics, 14, e1007613. 10.1371/journal.pgen.1007613 [PubMed: 30192761]

Brelsford A, & Irwin DE (2009). Incipient speciation despite little assortative mating: the yellow-rumped warbler hybrid zone. Evolution, 63, 3050–3060. 10.1111/j.1558-5646.2009.00777.x [PubMed: 19624726]

Cordingley JE, Sundaresan SR, Fischhoff IR, Shapiro B, Ruskey J, & Rubenstein DI (2009). Is the endangered Grevy's zebra threatened by hybridization? Animal Conservation, 12, 505–513. 10.1111/j.1469-1795.2009.00294.x

Coughlan JM, & Matute DR (2020). The importance of intrinsic postzygoticbarriers throughout the speciation process. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 375, 20190533.

Coyne JA, & Orr HA (2004). Speciation. Sinaeur Associates.

Cui RF, Schumer M, & Rosenthal GG (2016). Admix'em: a flexible framework for forward-time simulations of hybrid populations with selection and mate choice. Bioinformatics, 32, 1103–1105. 10.1093/bioinformatics/btv700 [PubMed: 26615212]

Culumber ZW, Ochoa OM, & Rosenthal GG (2014). Assortative mating and the maintenance of population structure in a natural hybrid zone. American Naturalist, 184, 225–232. 10.1086/677033

Derryberry EP, Derryberry GE, Maley JM, & Brumfield RT (2014). HZAR: hybrid zone analysis using an R software package. Molecular Ecology Resources, 14, 652–663. [PubMed: 24373504]

Dobzhansky T. (1937). Genetics and the origin of species. Columbia University Press.

Dumont BL, & Payseur BA (2008). Evolution of the genomic rate of recombination in mammals. Evolution, 62, 276–294. 10.1111/j.1558-5646.2007.00278.x [PubMed: 18067567]

Endler JA (1977). Geographic variation, speciation, and clines. Monographs in Population Biology, 10, 1–246. [PubMed: 409931]

Feldman MW, & Christiansen FB (1974). The effect of population subdivision on two loci without selection. Genetical Research, 24, 151–162. 10.1017/S0016672300015184 [PubMed: 4452480]

Felsenstein J. (1975). Genetic Drift in Clines Which Are Maintained by Migration and Natural-Selection. Genetics, 81, 191–207. 10.1093/genetics/81.1.191 [PubMed: 1205125]

Fitzpatrick BM (2013). Alternative forms for genomic clines. Ecology and Evolution, 3, 1951–1966. 10.1002/ece3.609 [PubMed: 23919142]

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, & Shaffer HB (2010). Rapid spread of invasive genes into a threatened native species. Proceedings of the National Academy of Sciences of the United States of America, 107, 3606–3610. 10.1073/pnas.0911802107 [PubMed: 20133596]

Gavrilets S. (1997). Hybrid zones with Dobzhansky-type epistatic selection. Evolution, 51, 1027–1035. 10.1111/j.1558-5646.1997.tb03949.x [PubMed: 28565489]

Gay L, Crochet PA, Bell DA, & Lenormand T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. Evolution, 62, 2789–2806. 10.1111/j.1558-5646.2008.00491.x [PubMed: 18752618]

Gompert Z, & Buerkle CA (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Molecular Ecology, 18, 1207–1224. 10.1111/j.1365-294X.2009.04098.x [PubMed: 19243513]

Gompert Z, & Buerkle CA (2011). Bayesian estimation of genomic clines. Molecular Ecology, 20, 2111–2127. 10.1111/j.1365-294X.2011.05074.x. [PubMed: 21453352]

Haldane JBS (1948). The Theory of a Cline. Journal of Genetics, 48, 277–284. 10.1007/BF02986626 [PubMed: 18905075]

Hamilton JA, & Miller JM (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conservation Biology, 30, 33–41. 10.1111/cobi.12574 [PubMed: 26096581]

Hvala JA, Frayer ME, & Payseur BA (2018). Signatures of hybridization and speciation in genomic patterns of ancestry. Evolution, 72, 1540–1552. 10.1111/evo.13509

Janousek V, Wang L, Luzynski K, Dufkova P, Vyskocilova MM, Nachman MW, Munclinger P, Macholan M, Pialek J, & Tucker PK (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus. Molecular Ecology, 21, 3032–3047. [PubMed: 22582810]

Jofre GI, & Rosenthal GG (2021). A narrow window for geographic cline analysis using genomic data: effects of age, size, and migration on error rates. Dryad Dataset. 10.5061/dryad.0p2ngf1z9

Lipshutz SE, Meier JI, Derryberry GE, Miller MJ, Seehausen O, & Derryberry EP (2019). Differential introgression of a female competitive trait in a hybrid zone between sex-role reversed species. Evolution, 73, 188–201. 10.1111/evo.13675 [PubMed: 30597557]

Mallet J. (2005). Hybridization as an invasion of the genome. Trends in Ecology & Evolution, 20, 229–237. 10.1016/j.tree.2005.02.010 [PubMed: 16701374]

Martin SH, & Jiggins CD (2017). Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development, 47, 69–74. 10.1016/j.gde.2017.08.007 [PubMed: 28923541]

Massey SE (2017). Strong Amerindian Mitonuclear Discordance in Puerto Rican Genomes Suggests Amerindian Mitochondrial Benefit. Annals of Human Genetics, 81(2), 59–77. 10.1111/ahg.12185 [PubMed: 28205222]

Muirhead CA, & Presgraves DC (2016). Hybrid Incompatibilities, Local Adaptation, and the Genomic Distribution of Natural Introgression between Species. American Naturalist, 187, 249–261. 10.1086/684583

Muller HJ (1940). Bearing of the Drosophila work on systematics. In Huxley J(Ed.), The new systematics (pp. 185–268). Clarendon Press.

Nagylaki T, & Lucier B. (1980). Numerical analysis of random drift in a cline. Genetics, 94, 497–517. 10.1093/genetics/94.2.497 [PubMed: 17249007]

Owens GL, & Samuk K. (2020). Adaptive introgression during environmental change can weaken reproductive isolation. Nature Climate Change, 10, 58–62. 10.1038/s41558-019-0628-0

Parsons TJ, Olson SL, & Braun MJ (1993). Unidirectional Spread of Secondary Sexual Plumage Traits across an Avian Hybrid Zone. Science, 260, 1643–1646. 10.1126/science.260.5114.1643 [PubMed: 17810207]

Payseur BA (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources, 10, 806–820. 10.1111/j.1755-0998.2010.02883.x [PubMed: 21565092]

Polechova J, & Barton N. (2011). Genetic Drift Widens the Expected Cline but Narrows the Expected Cline Width. Genetics, 189, 227–U905. 10.1534/genetics.111.129817 [PubMed: 21705747]

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Ryan SF, Fontaine MC, Scriber JM, Pfrender ME, O'Neil ST, & Hellmann JJ (2017). Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. Molecular Ecology, 26, 4725–4742. 10.1111/mec.14236 [PubMed: 28727195]

Schieltz JM, & Rubenstein DI (2015). Caught between two worlds: genes and environment influence behaviour of plains×Grevy's zebra hybrids in central Kenya. Animal Behavior, 106, 17–26. 10.1016/j.anbehav.2015.04.026

Schumer M, & Brandvain Y. (2016). Determining epistatic selection in admixed populations. Molecular Ecology, 25, 2577–2591. [PubMed: 27061282]

Schumer M, Cui R, Powell D, Dresner R, Rosenthal GG, & Andolfatto P. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. Elife, 3. 10.7554/eLife.02535

Schumer M, Cui R, Rosenthal GG, & Andolfatto P. (2015). Reproductive isolation of hybrid populations driven by genetic incompatibilities. PLoS Genetics, 11, e1005041. 10.1371/journal.pgen.1005041 [PubMed: 25768654]

Schumer M, Powell DL, Delclos PJ, Squire M, Cui RF, Andolfatto P, & Rosenthal GG (2017). Assortative mating and persistent reproductive isolation in hybrids. Proceedings of the National Academy of Sciences of the United States of America, 114, 10936–10941. 10.1073/pnas.1711238114 [PubMed: 28973863]

Schumer M, Xu CL, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, & Przeworski M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science, 360, 656–659. 10.1126/science.aar3684 [PubMed: 29674434]

Sciuchetti L, Dufresnes C, Cavoto E, Brelsford A, & Perrin N. (2018). Dobzhansky-Muller incompatibilities, dominance drive, and sex-chromosome introgression at secondary contact zones: A simulation study. Evolution, 72, 1350–1361. 10.1111/evo.13510

Sedghifar A, Brandvain Y, & Ralph P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular Ecology, 25, 2559–2576. [PubMed: 27148805]

Semenov GA, Linck E, Enbody ED, Harris RB, Khaydarov DR, Alstrom P, Andersson L, & Taylor SA (2021). Asymmetric introgression reveals the genetic architecture of a plumage trait. Nature Communications, 12, 1019. 10.1038/s41467-021-21340-y

Slatkin M. (1973). Gene flow and selection in a cline. Genetics, 75, 733–756. 10.1093/genetics/75.4.733 [PubMed: 4778791]

Slatkin M, & Maruyama T. (1975). Genetic drift in a cline. Genetics, 81, 209–222. 10.1093/genetics/81.1.209 [PubMed: 1205126]

Stankowski S, Sobel JM, & Streisfeld MA (2017). Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. Molecular Ecology, 26, 107–122. [PubMed: 27065228]

Stathos A., & Fishman L. (2014). Chromosomal rearrangements directly cause underdominant F1 pollen sterility in Mimulus lewisii-M imulus cardinalis hybrids. Evolution, 68, 3109–3119. [PubMed: 25125144]

Szymura JM, & Barton NH (1986). Genetic-Analysis of a Hybrid Zone between the Fire-B ellied Toads, Bombina-Bombina and Bombina-Variegata, near Cracow in Southern Poland. Evolution, 40, 1141–1159. [PubMed: 28563502]

Tominaga A, Matsui M, Yoshikawa N, Eto K, & Nishikawa K. (2018). Genomic Displacement and Shift of the Hybrid Zone in the Japanese Fire-Bellied Newt. Journal of Heredity, 109, 232–242. 10.1093/jhered/esx085 [PubMed: 29566204]

van Riemsdijk I, Arntzen JW, Bucciarelli G, McCartney-Melstad E, Rafajlović M, Scott PA, Toffelmier E, Shaffer HB, & Wielstra B. (2020). Spatial variation in introgression along a toad hybrid zone in France. bioRxiv, 746073. 10.1101/746073

Yang W, While GM, Laakkonen H, Sacchi R, Zuffi MAL, Scali S, Salvi D, & Uller T. (2018). Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard. Molecular Ecology, 27, 4213–4224. 10.1111/mec.14861 [PubMed: 30192998]

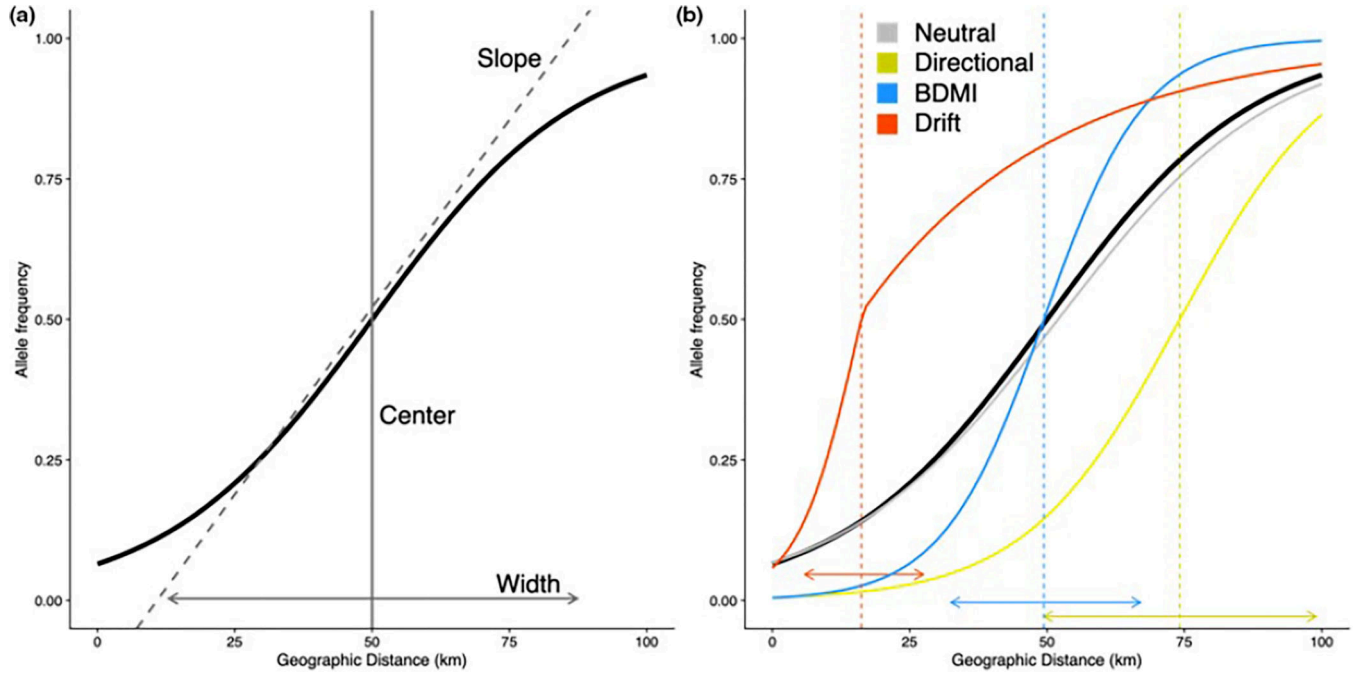
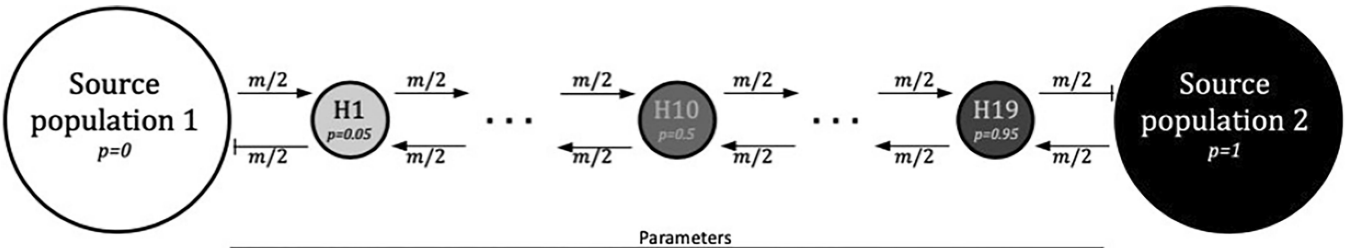


FIGURE 1.

Hypothetical geographic clines from a hybrid zone between species 1 at distance zero and species 2 at distance 100 km. (a) Average whole genome cline indicating the main parameters. (b) Comparison between a cline produced by a marker under directional selection (yellow), a cline produced by a marker in a Bateson–Dobzhansky–Muller incompatibility-BDMI (blue), a cline produced by a neutral marker (grey), and a cline produced by a neutral marker under the effects of drift (red). Dashed lines denote their corresponding centres, and solid arrows denote their widths. The null expectation that clines should coincide with neutral markers is not met in the yellow, blue and red clines. Genetic drift has altered the allele frequencies in the red cline, reducing its width and shifting the centre, without the need of selection



Parameters

N	100				500				1000			
m	0.1		0.01		0.1		0.01		0.1		0.01	
Generations in admixture	100	1000	100	1000	100	1000	100	1000	100	1000	100	1000

FIGURE 2.

The stepping-stone model for all hybrid zones, and the parameters used. Two nonadmixed infinite size populations flank 19 admixed demes (i.e., H1, H2, …, H19). At generation 1 each admixed deme has an ancestry proportion p of individuals from population 2. Individuals symmetrically migrate with a migration rate of m to adjacent admixed demes, but not back into source populations. We modelled 12 different hybrid zones, each with a combination of deme size N (100, 500, 1,000), migration rate m (0.1, 0.01), and selection type (no Selection, directional, underdominance, and BDM incompatibility). We sampled 30 individuals and genotyped the same 110 markers on the 100th generation, and again in the 1000th generation. We replicated this process 100 times

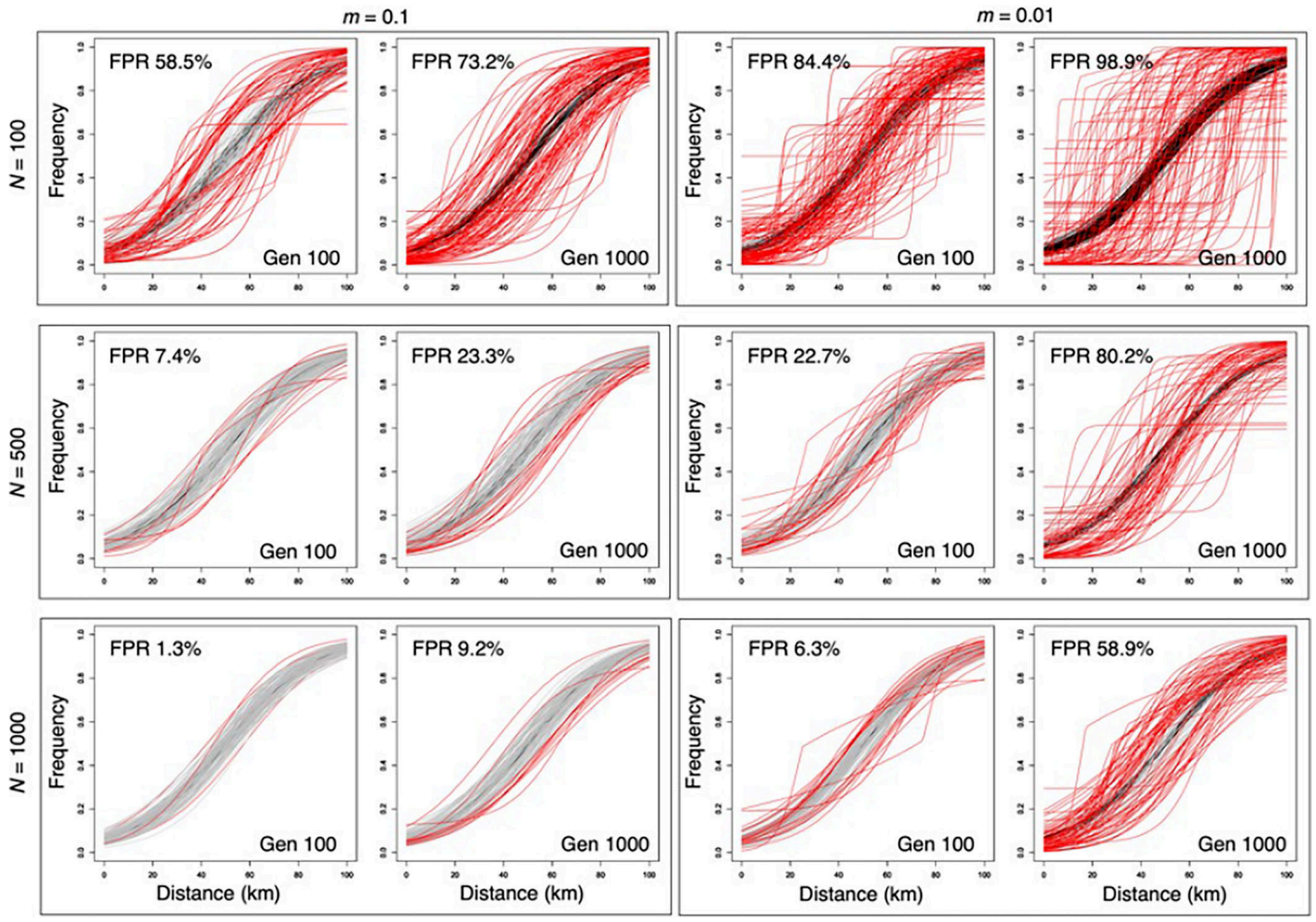


FIGURE 3.

Distribution of false positives from markers under no selection. True negatives (grey), false positives (red), and genome hybrid index (black), obtained from 100 replicates in different demographic histories. Each cline was generated from an independent replicate

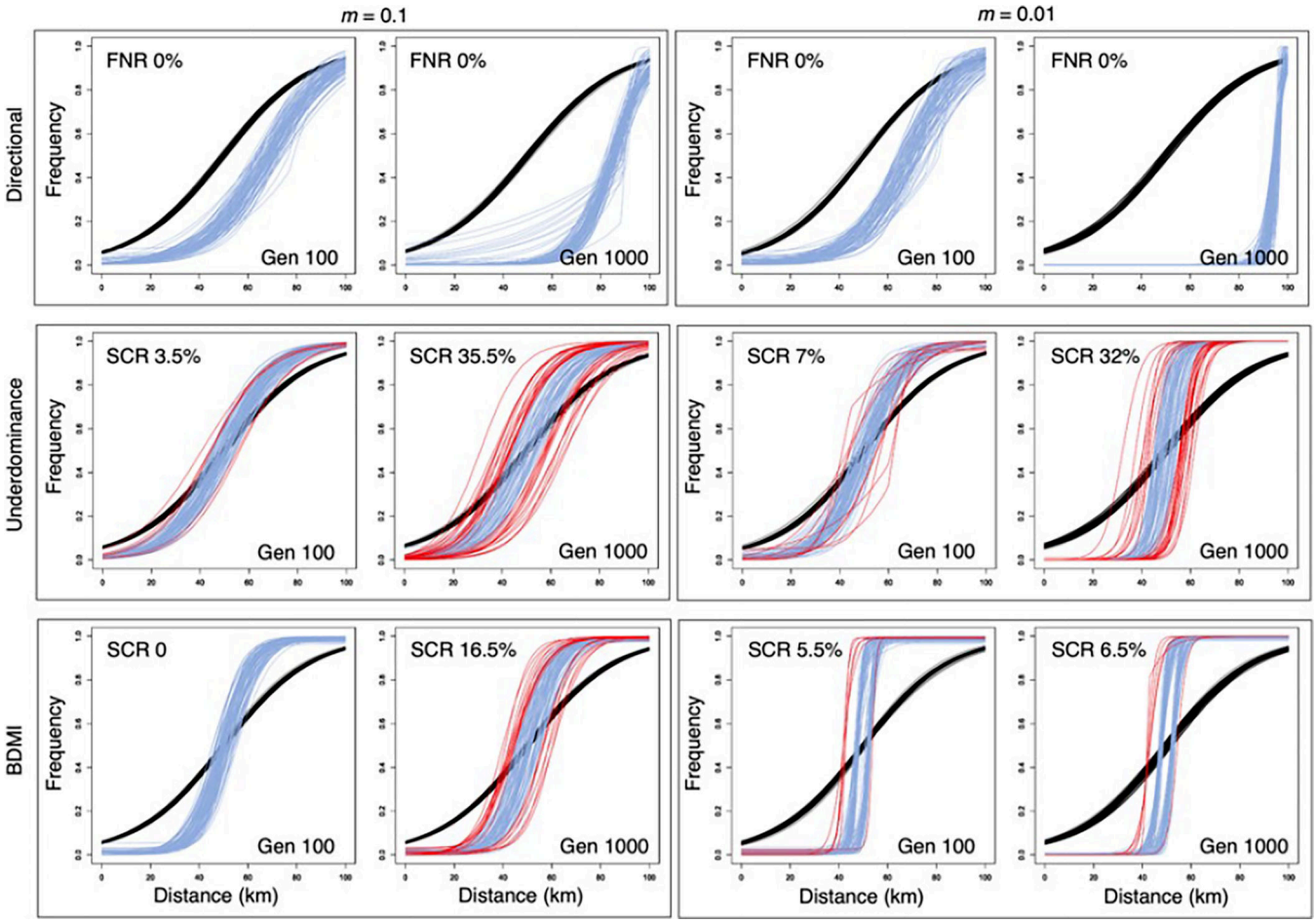


FIGURE 4.

Distribution of shifted cline centers in hybrid zones with large demes ($N = 1000$). True positives (blue), shifted cline centres (red), and genome-wide hybrid index (black), were obtained from 100 replicates in different migration rates and generations. Each cline was generated from an independent replicate. See the complete distribution of each mode of selection in Figures S7–S9

TABLE 1

Type ‘III’ ANOVA test sum of squares showing the interaction effect from our variables in our three linear models

	FPR				FNR				SCR			
Variables	**Sum Sq.**	***df***	***F* value**	**Pr(>*F*)**	**Sum Sq.**	***Df***	***F* value**	**Pr(>*F*)**	**Sum Sq**	***df***	***F* value**	**Pr(>*F*)**
Intercept	**27.466**	1	5130.27	**<2e−16**	**0.605**	1	159.59	**<2e−16**	**0.973**	1	8.01	**0.004688**
N	**26.695**	1	4986.38	**<2e−16**	**0.5151**	1	135.88	**<2e−16**	**6.201**	1	51.07	**<2e−16**
m	**4.998**	1	933.63	**<2e−16**	**0.0512**	1	13.51	**0.0002405**	**3.088**	1	25.43	**<2e−16**
Gen.	**8.435**	1	1575.65	**<2e−16**	**0.4384**	1	115.64	**<2e−16**	**1.421**	1	11.70	**0.0006361**
Mode of Sel.	0.026	3	1.64	0.178	**1.0478**	2	138.21	**<2e−16**	0.042	1	0.34	0.5574297
N & *m*	**0.373**	1	69.73	**<2e−16**	**0.0192**	1	5.06	**0.0244914**	0	1	0.00	0.9727065
m & Gen.	**24.589**	1	4592.86	**<2e−16**	0.0136	1	3.59	0.0581873	**8.05**	1	66.29	**<2e−16**
N & Gen.	**6.148**	1	1148.42	**<2e−16**	**0.3472**	1	91.58	**<2e−16**	**0.981**	1	8.08	**0.004526**
N & Mode of Sel.	0.031	3	1.91	0.126	**0.7485**	2	98.72	**<2e−16**	**0.584**	1	4.81	**0.0284437**
m & Mode of Sel.	0.002	3	0.13	0.945	**0.0356**	2	4.69	**0.0092452**	**1**	1	8.24	**0.0041377**
Gen. & Mode of Sel.	0.028	3	1.72	0.161	**0.4356**	2	57.45	**<2e−16**	0.375	1	3.09	0.0789969

Bold values denote significant interaction between our independent variables and our response variables FPR, FNR, and SCR.