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

Gaston Jofre[1] and Gil Rosenthal[1]

[1]Texas A&M University System

September 11, 2020

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 measured the effects of drift on genetic differentiation, and on false positive and false negatives detection using geographic clines. We modeled multiple stepping stone hybrid zones with distinct age, deme sizes, and migration rates, and evolving under different types of selection. We found that in young hybrid zones, drift increases overall genomic divergence, distorts cline shapes and increases both false positive and false negative rates. In old hybrid zones, genomic divergence and cline distortion are higher. Our results suggest that geographic clines are most useful for outlier analysis in young hybrid zones with large populations of hybrid individuals.

divergence and cline distortion are higher. Our results suggest that geographic clines are most useful for outlier analysis in young hybrid zones with large populations of hybrid individuals.

Keywords: outlier analysis, migration, drift, ecological genomics, tension zone

Introduction

Hybridization occurs when two genetically distinct individuals reproduce and generate viable offspring of mixed ancestry (Barton and Hewitt 1985; Abbott et al. 2013). After a hybridization event, selection against introgression maintains species boundaries (Mallet 2005; Coughlan and Matute 2020). Alternatively, alleles favored by selection can introgress differentially across a hybrid zone (Muirhead and Presgraves 2016; Martin and Jiggins 2017). These different types of selection can leave signatures that can be identified using outlier identification methods.

Cline theory has long been used to detect phenotypic introgression across wild hybrid zones and draw inferences about selection (e.g asymmetric sexual selection, Barton & Szymura 1989, Parsons 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 (Ryan et al. 2017; Stankowski et al. 2017; Lipshutz et al. 2019). Cline theory addresses changes in ancestry-allele frequencies across a spatial gradient (Figure 1), which may covary with environmental parameters (Haldane 1948; Endler 1977; Barton and Hewitt 1985). Geographic cline analysis compares cline positions and shapes to detect loci potentially under selection (Szymura and Barton 1986). At its simplest, the analysis uses two parameters, the center of the cline, which indicates the direction of selection acting on each allele, and the width of the cline, which indicates the maximum rate of change in allele frequencies (Barton and Gale 1993). Neutral alleles are expected to pass easily across the cline (neutral diffusion), and changes in their frequencies across a spatial gradient represents the null expectation of no selection (Barton and Hewitt 1985; Szymura and Barton 1986; Gay et al. 2008). Cline shapes that deviate from the null are considered outliers (Payseur 2010; Stankowski et al. 2017). Specifically, outlier loci that are involved in reproductive isolation between parental species have reduced widths (Payseur 2010; Sciuchetti et al. 2018); and outlier loci that indicate high rate of introgression can have centers shifted towards either end of the cline (Barton and Hewitt 1985).

Theoretical models show that genetic drift can distort cline shape (see Figure 1), reducing the width and shifting the center in either direction (Felsenstein 1975; Slatkin and Maruyama 1975; Polechova and Barton 2011; Sciuchetti et al. 2018). Felsenstein (1975) and Slatkin and Maruyama (1975) showed that drift causes a reduction in the expected width, and a shift in the expected center. Polechova and Barton (2011) and Sciuchetti et al. (2018) modeled selection against hybrid genotypes and show that the deviation in cline center and reduction in width of a given locus can be the result of drift alone instead of a given model of selection. However, Polechova and Barton (2011) and Sciuchetti et al. (2018) used one deme size. Thus, the effects of different deme sizes, migration rates, and models of selection on hybrid zone cline shapes remains unexplored.

Migration and drift can thus generate spurious outliers not caused by selection. For example, small deme sizes and low migration rates increase inbreeding (Hossjer et al. 2016), causing local fixation along the hybrid zone. In this study, we evaluate how often genetic drift causes neutral genomic regions to have cline shapes characteristic of regions under selection, and how often drift modifies the cline shapes of regions under selection. Specifically, we aim to calculate the rate of false positive detection and the rate of false negative detection in geographic clines using simulations in multiple hybrid zones with different demographic histories. We model multiple stepping stone hybrid zones under different evolutionary scenarios.

Materials and Methods

Simulation and modeling

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 (Schumer and Brandvain 2016; Massey 2017; Schumer et al. 2018). For all simulations, we modeled admixture in a cline with 19 demes and two infinitely large source populations, each at one extreme of the cline (Figure 2; Supporting Information Table 1). We restricted the number of hybrid demes to 19 (similar to Felsenstein 1975), to model clines with boundaries away from their cline centers, where one or the other allele will be close to fixation (see Slatkin and Maruyama 1975), and while keeping an adequate computation speed . We assumed that at generation 1 each admixed deme has an ancestry proportion p from population 2 increasing from p =0.05 to p =0.95 along the cline (see Figure 2). The genomes consisted of ten 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 markers randomly distributed in each chromosome, 110 markers in total. We allowed random mating for 1,000 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. Migration rate was symmetrical and restricted to 0.1 and 0.01. We performed 100 replicated simulations with each parameter combination, for each type of selection (see Figure 2).

Models of selection in hybrid zones:

A) No selection:

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

B) Directional and underdominance selection

To determine the effects of selection on hybrids, we modeled the effects of directional selection (with incomplete dominance) and underdominance. We modeled directional selection (Supporting Information Figure S1, A) as one allele having lower fitness than the other allele along the spatial gradient. We modeled directional selection on one locus at the middle of chromosomes three, denoted as 3.5, and another on chromosome seven, denoted as 7.5. We modeled underdominance (Supporting Information Figure S1, B), as heterozygous disadvantage along the spatial gradient. We modeled underdominance on loci in the middle of chromosomes five and nine, denoted as 5.5 and 9.5. Although underdominant selection models are unrealistic in nature (Sedghifar et al. 2016), they portray a simple scenario in which hybrids are unfit. We assumed additive fitness, where relative fitness is $f = 1 - x_{3.5} - x_{5.5} - x_{7.5} - x_{9.5}$, where x is a function of an individual's genotype at the specific focal loci.

C) 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 and Orr 2004) . We modeled the effects of epistatic interactions by simulating a BDM interaction between two autosomal loci, 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 $f = 1 - x_{3.5} - x_{7.5}$ (see selection coefficients in Supporting Information Figure S1, C).

Patterns of genomic differentiation using F_{ST}

To compare variation in genomic differentiation between young (100 generations) and old hybrid zones (1000 generations), and between models of selection, we calculated Wright's fixation index F_{ST}. Drift generates differentiation among demes, increasing F_{ST}, while migration does the reverse (Efremov 2004). To estimate the degree in genomic differentiation, we calculated per-locus F_{ST} among all hybrid populations following Weir and Cockerham (1984) and using the R package *pegas* (Paradis 2010). We calculated the mean, median, range, and interquartile range of the F_{ST} values from 100 loci with no selection, under directional selection, under underdominant selection, and in BDM interaction.

Geographic cline analysis

To study the effects of migration and drift on cline shape outliers, we sampled 30 random individuals from each deme in generations 100 and 1,000. Admix'em generates an output file with ancestries from the defined markers in the input files. We then used custom scripts (https://github.com/gjofre/simulation_clines) to generate the input files necessary for geographic cline analysis. The Metropolis–Hastings algorithm in the R package *hzar* (Derryberry et al. 2014), 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 center †, cline width⊒, the ends of the cline $P_{\min}$ and $P_{\max}$, the rate at which the cline tail decaysθ_{P1} and θ_{P2}, and the size of the cline tails $\mathcal{B}_{P1}$ and $\mathcal{B}_{P2}$. For computational speed, we fitted three models with different cline parameter combinations. Model I estimated only center †and width ⊒ from the data, assumed no tails ($\vartheta = 1$and $\mathcal{B} = 0$), and included fixed ends ($P_{\min}$ and $P_{\max}$ fixed to zero and one). Model II estimated †, ⊒, and $P_{\min}$ and$P_{\max}$ from the data, and assumed no tails. Model III was the same as Model II, and with tail estimates ϑ and $\mathcal{B}$allowed to vary. Each model parameter was estimated using three independent chain runs using 500,000 MCMC steps after a burn-in of 100,000 steps. The model with the lowest corrected Akaike information criterion (AIC, Akaike (1998)) was chosen for each marker. We then used the average estimates of all the parameters to estimate a genome-wide average cline for each replicate.

Outlier detection analysis

We then scanned for outliers across the genome. A marker was considered an outlier if the 95% confidence intervals for the lowest AIC value of either ⊒ or † did not overlap with the confidence intervals of the lowest AIC genome-wide average cline model in hzar. Centers and widths of any two clines were considered coincident (same †) or concordant (same ⊒) 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 or false negatives (Supporting Information Table 2).

We categorized markers under directional selection as true positives if they displayed the cline center shifted towards the side of the geographic gradient with higher fitness and as false negatives otherwise. We categorized markers under underdominant selection and in BDM incompatibilities as true positives if their center coincided with the center of the whole genome hybrid index cline, and if their width was lower than the width of the whole genome hybrid index cline. Finally, we categorized markers without selection as false positives if they were non-coincident or discordant compared to the genome-wide cline.

We computed an error matrix for each hybrid zone (see Supporting Information Table S2). For each replicate, we calculated the average false positive rate (FPR) as the proportion of neutral loci that were flagged as outliers and the false negative rate (FNR) as the proportion of loci under each form of selection that were undetected (see SI table 2).

Statistical analyses

We fitted a linear model to observe which hybrid zone parameters (N , m , generation on admixture, and type of selection) had a higher effect on F_{ST} , FPR, and FNR. We first log converted the different deme sizes, migration rates, and generations of admixture. We then analyzed the effects of these parameters onF_{ST} , FPR, and FNR with the lm function in R. For the FNR linear model, we omitted the data from the hybrid zones with no selection, since there are no true positives and false negatives. We then performed post hoc comparisons of estimated means ofF_{ST} , FPR, and FNR in each hybrid zone model parameter.

Results

Effect of model parameters in genetic differentiation

To observe the effects of drift on genetic differentiation, we compared the effects of each model parameter on F_{ST} values across replicated hybrid zones. Underdominance and BDM incompatibilities created genetic differentiation (Figure 3; Supporting Information Table 3; Supporting Information Figure S2). Deme size, migration rate, and generation in admixture had also an effect on meanF_{ST}. Larger deme sizes and higher migration rates decreased genetic differentiation. More generations in admixture, by contrast, increased genetic differentiation.

4

We found high genetic differentiation without any need for selection. Average F_{ST} was highest in hybrid zones with small deme sizes, and having a higher migration rate decreased F_{ST} further (Supporting Information Figure S2 A-C). F_{ST} values peaked in old hybrid zones with low migration rates (m= 0.01). Drift can drastically inflate genetic differentiation in hybrid zones with small population sizes and generate patterns of selections in neutral genomic regions.

Under directional selection, hybrid zones with small deme sizes had higher and wider F_{ST} distribution than hybrid zones with bigger deme sizes (Supporting Information Figure S2 D-F). Old hybrid zones had higher and narrower F_{ST} distributions than young hybrid zones, but only when migration rates were high (m =0.1). Old hybrid zones with low migration rates (m =0.01), had drastically decreased average F_{ST} values. With reduced migration between adjacent demes, some markers fixed along the geographic gradient, resulting in low F_{ST} values. These results suggest that in old hybrid zones with reduced migration rate, markers can be under directional selection and yet show patterns of low differentiation.

In hybrid zones under underdominant selection, or with BDMIs, small deme sizes did not affect mean F_{ST} (Supporting Information Figure S2 G-M). Similar to the hybrid zones under directional selection, at 1000 generations of admixture the F_{ST} distribution became narrower in both hybrid zones with high (m =0.1) and low migration rates (m= 0.1). We observed again an effect of drift in old hybrid zones drastically reducing F_{ST} values. However, this effect happened only at a few loci and only in hybrid zones with small deme sizes. Selection against hybrid genotypes, either via underdominance or BDM incompatibilities, showed more consistent patterns of genomic differentiation.

Effect of model parameters on false positive detection

Deme size had the strongest effect on the FPR distribution, with the highest FPR (mean FPR 0.98) in hybrid zones with small deme sizes and low migration rates (Figure 3; Supporting Information Table 3; Supporting Information Figure S2). More generations in admixture also increased FPR. The mode of selection had no effect on mean FPR values (Figure 3; Supporting Information Table 3).

Hybrid zones with small deme sizes generated the most clines with their centers drastically shifted, showing spurious patterns of introgression (Figure 4, A & B ; Supporting Information Video 1), which highly inflated the FPR (Supporting Information Figure S3, A). Specifically, old hybrid zones with small deme sizes (N= 100) and low migration rates (m =0.01) had an average FPR of 0.98. Old hybrid zones with large deme sizes (N = 1000) had a drastic decrease in FPR, but only when they had migration rates of m = 0.1. Old hybrid zones with low migration rates still had average FPR values of 0.58. The best scenario was in young hybrid zones with large deme sizes and high migration rates, with mean FPR of less than 0.1.

These results suggest that when analyzing natural hybridization, young hybrid zones will have lower risk of generating false positives than old hybrid zones, especially with population sizes larger than N= 500. Under the most favorable conditions (young hybrid zone with high migration and large deme size) the false positive rate still averaged 1.35% across replicates (Figure. 4e). The risk of misidentification in hybrid zones with small population sizes is high in both young and old hybrid zones, regardless of migration rates. In old hybrid zones larger than N= 500, having migration rates larger than m= 0.1 will indeed decrease the risk of identifying false positives.

Effect of model parameters on false negative detection

Loci with underdominance and BDM incompatibilities had the highest risk of FNR across our simulations . Hybrid zones with bigger deme sizes showed lower average FNR, as did fewer generations of admixture. Migration rate had a minor effect, slightly increasing the overall FNR (Supporting Information Table 3).

Markers under directional selection showed low overall FNR which was elevated in young hybrid zones. Drift indeed distorted cline shape, but centers were rarely shifted to the opposite side of the gradient (Supporting Information Figure S4 A and B, Supporting Information Video 2). In all hybrid zones with population sizes larger than N= 500, and in all hybrid zones after 1000 generations, all markers under directional selection were categorized as true outliers (FNR = 0; Supporting Information Figure S4). Even when drift modifies

the cline shapes of markers under directional selection, the impact on FNR is low and only in young hybrid zones with small deme sizes.

By contrast, markers under underdominant selection showed increased FNR across all hybrid zones (Supporting Information Figure S3 C). In young hybrid zones, small deme sizes ($N = 100$) drastically increased FNR, regardless of migration rates. Old hybrid zones with small deme sizes (e.g. $N = 100$) showed the highest FNR, and even with larger deme sizes FNR was not substantially reduced. For example, in hybrid zones with deme sizes of $N = 1000$, regardless of migration rates, ~40% of the markers had drastic shifts in their cline centers (Supporting Information Figure S5). These results suggest that in small and young hybrid zones, and in old hybrid zones, drift can generate clines with patterns similar to markers under directional selection (e.g. Supporting Information Video 3).

In markers involved in BDM incompatibilities the average FNR was slightly lower relative to underdominant selection (Figure 3; Supporting Information Video 4). Young hybrid zones with large deme sizes had the lowest FNR, specifically with high migration rates (Supporting Information Figure S3 D, solid squares). On the other hand, old hybrid zones with high migration rates, showed the highest FNR when ($N = 100$ and 500). The cline centers shifted to a lesser degree, but the clines were outside the hybrid index confidence interval (Supporting Information Figure S6 A &C). In old hybrid zones with low migration rates, having large deme sizes decreased drastically the FNR. These results suggest that markers involved in BDM incompatibilities will generate fewer false negatives, compared to markers under underdominant selection.

Discussion

Geographic cline theory (Barton and 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, drift within demes can distort cline shapes from hybrid zones (Polechova and Barton 2011).

Our results are consistent with theoretical (Felsenstein 1975; Slatkin and Maruyama 1975; Nagylaki and Lucier 1980) and empirical (Polechova and Barton 2011; Sciuchetti et al. 2018) studies showing that drift can distort spatial clines in hybrid zones. 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 false positive and false negative rates. 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. Conversely, drift can act to obscure the effects of selection, making it even less likely that outlier loci represent robust candidates.

Our comparisons show that the drift causes high variance in F_{ST} regardless of the type of selection acting upon hybrids, creating patterns of differentiation in neutral genomic regions similar to genomic regions under selection. Genomic differentiation in turn increases the rate of false positives when scanning the genome for signatures of selection. i This differentiation increases even further as the hybrid zone ages. Migration rates plays also an important role but, ultimately, having large deme sizes substantially decrease genomic differentiation and produce the least false positives and negatives.

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 centers away from the expected hybrid index cline (Figure 4). Increasing the population size does decrease the effects of drift, generating more concordant and coincident clines, and reducing significantly the rate of false positives. 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.

Small deme sizes can also distort our assessment of how selection is acting, even in cases in which we can correctly identify regions under selection. Hybrid zones with small deme sizes and restricted migration have

higher risk of distortion caused by drift, masking outliers as a different type of selection (e.g. adaptive introgression instead of an ecotone or selection against hybrids). Hybrid zones with underdominance and BDM incompatibilities, show clines with drastically shifted centers. Since heterozygous genotypes have lower fitness, drift can increase the rate of fixation, and shift the center to either side of the gradient (e.g. Supporting Information Figure S5 and S6; Supporting Information Videos 3 and 4), which could also be misinterpreted as directional selection. Additionally, in ~3% of our replicates with small deme sizes, genomic regions under directional selection, instead of having centers shifted towards the adaptive side of the geographical gradient, showed centers shifted towards the opposite (the less fit) side of the gradient.

Between false positives on neutral and false negatives on underdominant and BDMI loci, 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 (Parsons et al. 1993) and of resilience to climate change (Hamilton and Miller 2016). Our simulations indicate that apparent adaptive introgression may instead reflect drift or hybrid incompatibilities.

Our results suggest that geographic cline analysis is 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 (Schumer et al. 2014; Dufresnes et al. 2015; Janousek et al. 2015) or temporal analyses of selection on outlier candidates (Taylor et al. 2014; Ryan et al. 2018).

Geographic cline analysis is one of a family of outlier detection methods used in evolutionary and ecological genomics, for example genomic clines (Gompert and Buerkle 2011), 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 loci.

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 favored 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.

Acknowledgements

References

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. W. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Most, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert,

C. M. Smadja, R. Stelkens, J. M. Szymura, R. Vainola, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J Evolution Biol 26:229-246.

Akaike, H. 1998. A New Look at the Statistical Model Identification. Pp. 215-222 *in* E. Parzen, K. Tanabe, and G. Kitagawa, eds. Selected Papers of Hirotugu Akaike. Springer New York, New York, NY.

Barton, N. H. and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid Zones and the Evolutionary Process. Oxford University Press, Oxford, UK.

Barton, N. H. and G. M. Hewitt. 1985. Analysis of Hybrid Zones. Annu Rev Ecol Syst 16:113-148.

Coughlan, J. M. and D. R. Matute. 2020. The importance of intrinsic postzygoticbarriers throughout the speciation process. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 20190533.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sunderland, MA: Sinaeur Associates.

Cui, R. F., M. Schumer, and G. G. Rosenthal. 2016. Admix'em: a flexible framework for forward-time simulations of hybrid populations with selection and mate choice. Bioinformatics 32:1103-1105.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. hzar: hybrid zone analysis using an R software package. Molecular Ecology Resources 14:652-663.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press.

Dufresnes, C., A. Brelsford, J. Crnobrnja-Isailovic, N. Tzankov, P. L. Nicolas, and N. Perrin. 2015. Timeframe of speciation inferred from secondary contact zones in the European tree frog radiation (Hyla arborea group). Bmc Evol Biol 15.

Dumont, B. L. and B. A. Payseur. 2008. Evolution of the genomic rate of recombination in mammals. Evolution 62:276-294.

Efremov, V. V. 2004. The rate of approach to the equilibrium value of F(ST) in island and one-dimensional stepping-stone models of migration. Genetika 40:1268-1273.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Monographs in population biology 10:1-246.

Felsenstein, J. 1975. Genetic Drift in Clines Which Are Maintained by Migration and Natural-Selection. Genetics 81:191-207.

Gay, L., P. A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing Clines on Molecular and Phenotypic Traits in Hybrid Zones: A Window on Tension Zone Models. Evolution 62:2789-2806.

Gompert, Z. and C. A. Buerkle. 2011. Bayesian estimation of genomic clines. Molecular ecology 20:2111-2127.

Haldane, J. B. S. 1948. The Theory of a Cline. J Genet 48:277-284.

Hamilton, J. A. and J. M. Miller. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conservation Biology 30:33-41.

Hossjer, O., L. Laikre, and N. Ryman. 2016. Effective sizes and time to migration-drift equilibrium in geographically subdivided populations. Theor Popul Biol 112:139-156.

Hvala, J. A., M. E. Frayer, and B. A. Payseur. 2018. Signatures of hybridization and speciation in genomic patterns of ancestry. Evolution 72:1540-1552.

Janousek, V., P. Munclinger, L. Y. Wang, K. C. Teeter, and P. K. Tucker. 2015. Functional Organization of the Genome May Shape the Species Boundary in the House Mouse. Mol Biol Evol 32:1208-1220.

Lipshutz, S. E., J. I. Meier, G. E. Derryberry, M. J. Miller, O. Seehausen, and E. P. Derryberry. 2019. Differential introgression of a female competitive trait in a hybrid zone between sex-role reversed species. Evolution 73:188-201.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol 20:229-237.

Martin, S. H. and C. D. Jiggins. 2017. Interpreting the genomic landscape of introgression. Curr Opin Genet Dev 47:69-74.

Massey, S. E. 2017. Strong Amerindian Mitonuclear Discordance in Puerto Rican Genomes Suggests Amerindian Mitochondrial Benefit. 81:59-77.

Muirhead, C. A. and D. C. Presgraves. 2016. Hybrid Incompatibilities, Local Adaptation, and the Genomic Distribution of Natural Introgression between Species. Am Nat 187:249-261.

Muller, H. J. 1940. Bearing of the Drosophila work on systematics. The new systematics:185-268.

Nagylaki, T. and B. Lucier. 1980. Numerical analysis of random drift in a cline. Genetics 94:497-517.

Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419-420.

Parsons, T. J., S. L. Olson, and M. J. Braun. 1993. Unidirectional Spread of Secondary Sexual Plumage Traits across an Avian Hybrid Zone. Science 260:1643-1646.

Payseur, B. A. 2010. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources 10:806-820.

Polechova, J. and N. Barton. 2011. Genetic Drift Widens the Expected Cline but Narrows the Expected Cline Width. Genetics 189:227-U905.

Ryan, S. F., J. M. Deines, J. M. Scriber, M. E. Pfrender, S. E. Jones, S. J. Emrich, and J. J. Hellmann. 2018. Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. Proceedings of the National Academy of Sciences of the United States of America 115:E2284-E2291.

Ryan, S. F., M. C. Fontaine, J. M. Scriber, M. E. Pfrender, S. T. O'Neil, and J. J. Hellmann. 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.

Schumer, M. and Y. Brandvain. 2016. Determining epistatic selection in admixed populations. Molecular ecology 25:2577-2591.

Schumer, M., R. Cui, D. Powell, R. Dresner, G. Rosenthal, and P. Andolfatto. 2014. High-resolution Mapping Reveals Hundreds of Genetic Incompatibilities in Hybridizing Fish Species. Elife 3.

Schumer, M., C. L. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science 360:656-659.

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

Sedghifar, A., Y. Brandvain, and P. Ralph. 2016. Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular ecology 25:2559-2576.

Slatkin, M. and T. Maruyama. 1975. Genetic drift in a cline. Genetics 81:209-222.

Stankowski, S., J. M. Sobel, and M. A. Streisfeld. 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.

Szymura, J. M. and N. H. Barton. 1986. Genetic-Analysis of a Hybrid Zone between the Fire-Bellied Toads, Bombina-Bombina and Bombina-Variegata, near Cracow in Southern Poland. Evolution 40:1141-1159.

Taylor, S. A., T. A. White, W. M. Hochachka, V. Ferretti, R. L. Curry, and I. Lovette. 2014. Climate-Mediated Movement of an Avian Hybrid Zone. Curr Biol 24:671-676.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population-Structure. Evolution 38:1358-1370.

Author Contributions

G.I. Jofre conceptualization, simulations, data modification, analysis, writing-original draft, project administration, writing review, and editing. G.G. Rosenthal conceptualization, project administration, writing review and editing.

Data accessibility

The data that support the findings of this study are openly available in Dryad at https://doi.org/10.5061/dryad.0p2ngf1z9.
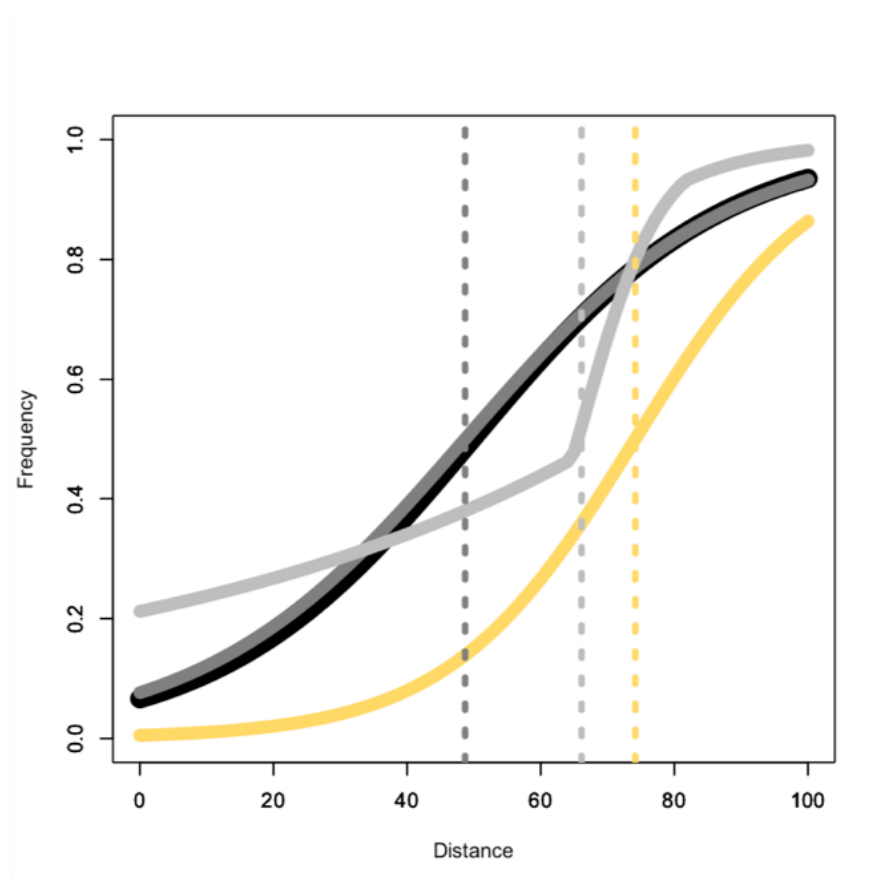
Figures and Tables

Figure 1. Hypothetical geographic clines from whole genome (black), neutral locus (gray), neutral locus distorted by drift (light gray), and locus under directional selection (yellow). The dotted lines denote their corresponding centers. Genetic drift has distorted the light gray cline, reduced its width and shifted the center.

Hosted file

image2.emf available at https://authorea.com/users/357843/articles/480225-a-narrow-window-for-geographic-cline-analysis-using-genomic-data-effects-of-age-size-and-migration-on-error-rates

Figure 2. The stepping-stone model for all hybrid zones, and the parameters used. Two non-admixed 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 modeled 12 different hybrids zones, each with a combination of deme size N (100, 500,1000), migration rate m(0.1, 0.01), and selection type (No Selection, Directional and Underdominance, 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.

Hosted file

image3.emf available at https://authorea.com/users/357843/articles/480225-a-narrow-window-for-geographic-cline-analysis-using-genomic-data-effects-of-age-size-and-migration-on-error-rates

Figure 3. Model of selection, deme size (N) , migration rate (m), and generations in admixture modify the distribution of F_{ST} values, FPR, and FNR in different degrees. Mode of selection modifies only F_{ST} and FNR, with strong genomic differentiation even without any need for selection. Underdominant selection and BDM incompatibilities increased FNR. Deme size influenced F_{ST} , FPR, and FNR. Small deme sizes showed higher rates of differentiation and large deme sizes produce significantly less false positives and false negatives. Migration rate influenced only F_{ST} and FPR, with low migration rates increasing genomic differentiation and generating more false positives. Old hybrid zones increased F_{ST} values, FPR, and FNR, and specifically generate more false positives and false negatives. N.S. - No selection, Dir. – Directional selection, Und. -Underdominance, BDMI- Bateson–Dobzhansky–Muller incompatibility. See Supporting Information Figure S2 and S3 for a more detailed distribution of F_{ST} , FPR and FNR.
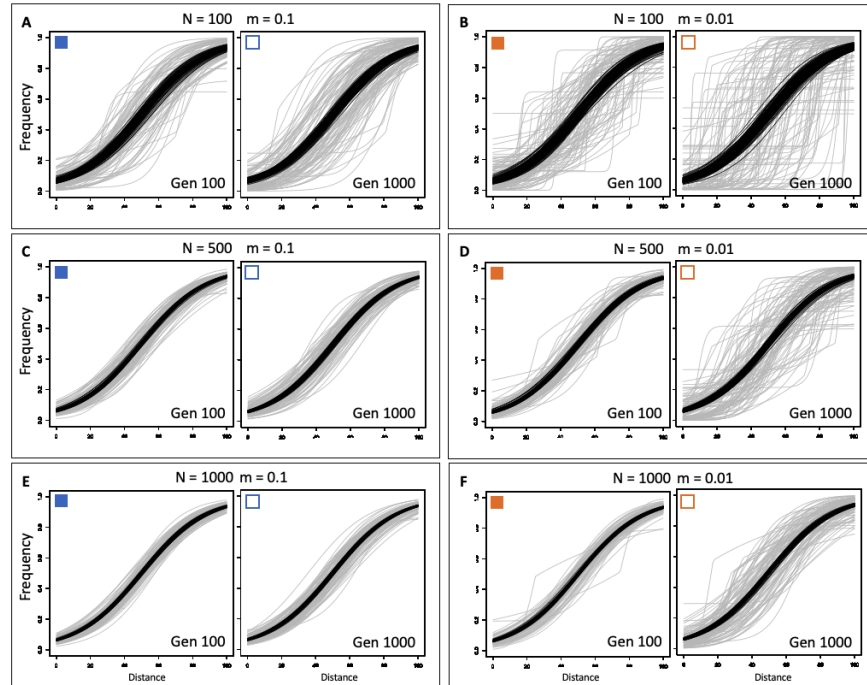
Figure 4. Clines from loci under no selection (grey) and average hybrid index clines (black), obtained from 100 replicates in different demographic histories. Each cline is from a locus from independent replicates. Blue squares are hybrid zones with $m= 0.1$. Orange squares are hybrid zones with $m= 0.01$. Solid squares are young hybrid zones after 100 generations of admixture. Open squares are old hybrid zones after 1000 generations of admixture.