

1 **Quantitative analyses of coupling in hybrid zones**

2 Thomas J Firneno Jr¹, Georgy Semenov², Erik B. Dopman³, Scott A. Taylor², Erica L. Larson¹,
3 Zachariah Gompert⁴

4 ¹*Department of Biology, University of Denver, Denver, Colorado, 80208, USA*

5 ²*Department of Ecology and Evolutionary Biology, University of Colorado Boulder,*
6 *Boulder Colorado, 80211, USA*

7 ³*Department of Biology, Tufts University, Medford, Massachusetts 02155, USA*

8 ⁴*Department of Biology, Utah State University, Logan, UT, 84321, USA*

9 Thomas J Firneno Jr , ORCID ID: 0000-0002-4975-2794, thomas.firneno@du.edu

10 Georgy Semenov, ORCID: 0000-0002-7218-7885, georgy.semenov@colorado.edu

11 Erik B. Dopman, ORCID: 0000-0002-8633-5527, erik.dopman@tufts.edu

12 Scott A. Taylor, ORCID: 0000-0001-9580-9125, scott.a.taylor@colorado.edu

13 Erica L. Larson, ORCID ID: 0000-0003-3006-645X, erica.larson@du.edu

14 Zachariah Gompert, ORCID: 0000-0003-2248-2488, zach.gompert@usu.edu

15 Corresponding Author:

16 Zachariah Gompert, zach.gompert@usu.edu

Abstract

In hybrid zones, whether barrier loci experience selection mostly independently or as a unit depends on the ratio of selection to recombination as captured by the coupling coefficient. Theory predicts a sharper transition between an uncoupled and coupled system when more loci affect hybrid fitness. However, the extent of coupling in hybrid zones has rarely been quantified. Here, we use simulations to characterize the relationship between the coupling coefficient and variance in clines across genetic loci. We then re-analyze 25 hybrid zone data sets and find that cline variances and estimated coupling coefficients form a smooth continuum from high variance and weak coupling to low variance and strong coupling. Our results are consistent with low rates of hybridization and a strong genome-wide barrier to gene flow when the coupling coefficient is much greater than 1, but also suggest that this boundary might be approached gradually and at a near constant rate over time.

30 Introduction

31 Species are often separated by genetic and phenotypic discontinuities that evolved and are
32 maintained by natural selection. Whether such discontinuities generally arise from selection on
33 one, a few, or many genes is largely unknown (Nosil and Schluter 2011; Nosil *et al.* 2021). For
34 example, theory suggests that speciation with gene flow occurs more readily when selection is
35 concentrated on a few large effect genes (Yeaman and Whitlock 2011), and there is some
36 empirical evidence consistent with this prediction (Kozak *et al.* 2019; Unbehend *et al.* 2021).
37 However, empirical studies also show that speciation can involve divergence at many genes
38 (Michel *et al.* 2010; Martin *et al.* 2019; Kautt *et al.* 2020), although this does not necessarily
39 imply that many genes causally drive speciation. When multiple genes contribute to speciation,
40 linkage disequilibrium (LD) among these genes and other regions of the genome increases the
41 total selection experienced by each locus (Barton 1983; Barton and de Cara 2009). In essence,
42 LD causes selection on causal loci (i.e., barrier loci) to spill over leading to indirect selection at
43 other loci. Barton (1983) referred to this process as “coupling”, the usage of which is consistent
44 with the definition of coupling described in Perspective 2 of Dopman *et al.* 2023, but is
45 somewhat different from the definition from Ritchie and Butlin 2023 that emphasizes
46 associations between genes affecting signals and preference. Whether selection affects
47 individual genes or larger genomic regions (e.g., chromosomes or the entire genome) depends
48 on coupling and this in turn influences whether genes or genomes diverge during speciation
49 (Barton and Bengtsson 1986; Wu 2001). The transition to a coupled system where LD causes a
50 genome-wide barrier to gene flow might be important for completing the speciation process
51 (Barton and de Cara 2009; Nosil *et al.* 2021).

52 Hybrid zones are powerful systems for studying speciation in general (Peñalba *et al.* 2023), and
53 specifically for assessing whether selection occurs on barrier loci independently or as a unit
54 creating a genome-wide barrier to gene flow (Abbott *et al.* 2013; Harrison and Larson 2014;
55 Gompert *et al.* 2017). In hybrid zones, spatial structure and selection (s) against hybrids
56 maintains LD among barrier loci, whereas recombination (r) in hybrids breaks down these
57 associations. The balance between these processes determines the extent to which barrier loci
58 operate independently or as a unit and this is captured by Barton’s coupling coefficient $\theta = s / r$
59 (Barton 1983). Specifically, Barton (1983) considered a model of simple underdominance, such
60 that an individual heterozygous for n barrier loci has a fitness of $(1 - s)^n$; r denotes the average
61 recombination rate between neighboring underdominant loci, such that the total map length with
62 L loci is $R = (L - 1)/r$. When θ is greater than 1, the system is coupled and the total selection
63 experienced by each locus (denoted s^*) approaches sL (at least up to $sL = 1$, at which point
64 hybrid fitness is 0), that is the combined effects of all barrier loci (Barton 1983) (Figure 1A).
65 Whenever coupling occurs, it is s^* (direct and indirect selection combined) rather than s (direct
66 selection due to the causal effect of a locus on fitness) that determines the shape of a cline,
67 particularly the slope near the center of the cline (Barton 1983).

68 Barton (1983) noted a sharp transition between uncoupled and coupled systems in analytical
69 models at $\theta = 1$ when holding R and the total selection constant but increasing the number of
70 barrier loci L . More recently, results from individual-based simulations have documented a

similarly sharp transition in time when populations diverge with gene flow (Flaxman *et al.* 2014; Nosil *et al.* 2017). Specifically, divergence remains low until a sufficient number of adaptive mutations (i.e., barrier loci) build-up at which point the system enters a positive feedback loop where divergent selection increases genetic differentiation. This increases LD for the system as a whole, in turn further increasing the selection experienced by each locus. These sharp transitions in space and time suggest that LD among barrier loci can lead to rapid transitions from early to late stages of speciation. However, other theoretical results suggest a smoother transition from uncoupled to coupled systems (i.e., Kruuk *et al.* 1999) and provide limited evidence for the critical transition in terms of LD among barrier loci at $\theta = 1$ perhaps because of a slow approach to equilibrium expectations (Baird 1995).

Recent work has placed an increasing emphasis on coupling (Bierne *et al.* 2011; Abbott *et al.* 2013; Kunerth *et al.* 2022), and several hybrid zone studies have tested for or analyzed coupling (Vines *et al.* 2016; Ryan *et al.* 2017; Cruzan *et al.* 2021). Nonetheless, we know relatively little about coupling in most hybrid zones. One reason for our limited knowledge is that coupling is not directly measured in the most commonly estimated cline summaries, such as cline concordance and coincidence, and these estimates provide only a binary classification of clines. We aim to overcome these limitations by first using simulations to determine the relationship between the coupling coefficient (θ) and the variance in clines across loci. Specifically, we test the predictions that: (i) cline variances decline with increased values of θ and (ii) that the decline in cline variances is nonlinear with a notable transition to lower variances at $\theta = 1$. We then re-analyze genomic data from 25 hybrid zones to infer cline variances and convert these into quantitative estimates of coupling. This allows us to ask whether the hybrid zones we consider exhibit only strong or weak coupling, as predicted by tipping point models, or form more of a connected continuum with a notable zone of intermediate systems (i.e., a grayzone of speciation; Roux *et al.* 2016). We conclude by assessing the consistency of coupling across replicate hybrid zone transects and differential coupling of autosomes versus sex chromosomes based on a subset of these data sets.

Main text

Measuring cline variance

We used two complementary approaches to quantify the variance in clines across a set of loci and evaluate whether the variances relate to the coupling coefficient (θ) in a linear or non-linear manner (Figure 1). First, we used geographic clines in allele frequencies. One difficulty with the geographic approach is that expected cline shape differs for cases with and without coupling. Uncoupled, single locus clines are expected to be sigmoidal, whereas coupled, multilocus clines can be better approximated by a steeper sigmoidal function in the center of the cline and shallower exponential decay in the tails (Barton 1983; Szymura and Barton 1986). This is because the shape of the central portion of the cline reflects the total selection (s^*) on a locus, including that caused by LD with barrier loci, whereas the shape of the tails of the cline should mostly reflect direct selection on the locus (or on tightly linked loci in LD outside of the hybrid zone). Thus, to provide a common framework for quantifying cline variance using a single

model, we focus on the center of each cline – the region governed by total selection (s^*), which is equivalent to s in the uncoupled, single locus model. We fit a linear model for the logit allele frequency, $\log(p/(1-p))$. The slope in this model is expected to be four times the allele frequency gradient (i.e., the reciprocal of cline width) at the center of the cline (i.e., where $p = 0.5$) (Barton and Hewitt 1989; Kruuk *et al.* 1999). The variance (or rather the standard deviation, SD) in slope among loci then serves as our metric of cline variance in this model.

Our second approach involved fitting clines in the probability of locus-specific ancestry along a genome-wide admixture gradient (i.e., as a function of a hybrid index; Szymura and Barton 1986; Gompert and Buerkle 2009, 2011). Here, we use the logit-logistic genomic cline function proposed by Fitzpatrick (2013), where the probability a gene copy at locus i for individual j was inherited from species 2 is $\phi_{ij} = (h_j^{v_i}) / (h_j^{v_i} + (1-h_j^{v_i}) * e^{u_i})$. In this equation, h_j denotes the hybrid index (proportion of the genome inherited from a species 2), v_i measures the cline slope (gradient) for locus i relative to the average ($v = 1$) and u_i defines the cline center for locus i relative to the genome average and to v_i (Fitzpatrick 2013). Consequently, information on ancestry at a locus depends on the genotype, an individual's hybrid index (based on all loci) and the genomic cline parameters (based on all of the analyzed hybrids) (Gompert and Buerkle 2011). We use a reparameterization following Bailey (2022) with $\text{logit}(c_i) = u_i/v_i$ to define the more intuitive cline center parameter (c_i) that indicates the hybrid index value at which $\phi_{ij} = 0.5$ (i.e., the probability of ancestry from both species is equal). We measure the variance in clines here as the standard deviation (SD) in $\log(v_i)$ and $\text{logit}(c_i)$, both of which have expected means of 0 (i.e., the values averaged across all loci used to define the hybrid index should be 0).

Each of these models has benefits and drawbacks. Geographic clines can only capture hybrid zone dynamics when there is a geographic gradient of gene flow and the scale of cline parameters are dependent on the geographic scale of dispersal, which varies among organisms and is often poorly known. Genomic clines are always relative to genome-average admixture (hybrid index) and thus are not in geographic units, but instead measure the change in the ancestry probability per unit change in hybrid index. This makes the variance in genomic clines easier to compare across hybrid zones, however genomic clines cannot capture the absolute extent of introgression and fail when hybrids are rare (i.e., when the hybrid zone lacks an admixture gradient in h). For this reason, we consider both geographic and genomic clines in simulations (where the scale of dispersal is set), but consider only genomic clines in our analyses of empirical data sets.

We take a hierarchical Bayesian approach to fitting geographic and genomic clines. This is important because it allows us to explicitly estimate the cline variances as parameters in our models. Thus, we assume that the slopes (β) for geographic clines and $\log(v)$ and $\text{logit}(c)$ for genomic clines at each locus represent independent draws from a higher level (normal) distribution with a mean and standard deviation. We are most interested in the standard deviation which describes the variability in clines across the genome. For the geographic cline analysis, we estimate both the mean and standard deviation. We placed a weakly informative normal prior (mean = 0, SD = 5) on the mean (μ_β) and a weakly informative gamma prior (shape = 0.1, rate = 0.01) on the standard deviation (σ_β). In the case of the genomic clines we expect

means of 0 and set them as such (i.e., we use soft centering but do not impose a sum-to-zero constraint), but we estimate the standard deviations (σ_v and σ_c) by setting weakly informative normal priors (mean = 0, SD = 1) on both. We fit these models using Hamiltonian Monte Carlo (HMC), which is an algorithm that allows for more efficient exploration of and sampling from complex posterior probability distributions (Neal 2011). This allows us to better estimate these standard deviation parameters, which can exhibit poorer mixing using alternative Markov chain Monte Carlo methods (ZG personal observation). Model fitting was done using Stan via the R (versions 4.1 and 4.2) interface, *rstan* (version 2.21.7).

Simulations connecting cline variance to coupling

We simulated hybrid zones with known coupling coefficients and analyzed the simulated data using the models described in the preceding section to measure the association between coupling and the variance in cline parameters. We were especially interested in whether there was a sharp decrease in cline variance at a coupling coefficient of $\theta = 1$. Hybrid zones were simulated using the *dfuse* software described by Lindtke and Buerkle (2015). This software runs individual-based simulations of secondary contact using a stepping stone model and tracks ancestry junctions (Baird 1995; Buerkle and Rieseberg 2008). We modified the existing software to include a model of multiplicative underdominance equivalent to that considered by Barton (1983), that is where the fitness of a hybrid heterozygous at n barrier loci is $(1 - s)^n$. All simulations included 110 demes, each with an adult carrying capacity of 50, arrayed in a one-dimensional stepping-stone model. We set the migration rate between neighboring demes to either 0.1 or 0.2. This set of conditions was chosen to approximate a pair of hybridizing species distributed continuously in space using a large number of small, well-connected demes. We simulated hermaphroditic, diploid organisms with a single 1 Morgan chromosome (thus there was one expected recombination per meiosis). The simulations assumed random mating within demes with viability selection on progeny. Each generation began with a reproduction phase, which involved creating offspring until either the progeny carrying capacity was reached (100 individuals or twice the adult carrying capacity) or until all of maternal gametes from the adults from the previous generation were exhausted (five per individual). This was followed by progeny dispersal and mortality selection where the probability of survival was given by hybrid fitness, that is $(1 - s)^n$. Surviving progeny (now adults) were then culled randomly to the adult carrying capacity if more than 50 progeny survived (see Lindtke and Buerkle 2015 for a full description of the simulation model).

We conducted simulations with coupling coefficients $\theta = 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1, 1.1, 1.5$, and 2, and with the number of underdominant barrier loci set to $L = 2, 10, 100, 200, 500$ or 1000. These were spaced equally across the chromosome, with an average recombination rate between neighboring loci of $r = 1/(L - 1)$. We then used the relationship $\theta = s/r$ to calculate the appropriate per locus selection coefficient to achieve the desired coupling coefficient (a few combinations of L and θ would have required $s > 1$ and fitness < 0 ; these combinations were dropped). We ran the simulations for 2000 generations but examined output every 500 generations to verify that the clines appeared stable (a lack of change could reflect an

approximate equilibrium outcome or very gradual approach to a not yet reached equilibrium, see Baird 1995). In total, 1140 hybrid zone data sets were simulated. In each case, genotypes at 51 diagnostic (fixed differences between species) biallelic markers spaced evenly along the chromosome (every 2 cM) were output for analysis.

To run the geographic cline analyses, we defined the center of each simulated hybrid zone as the deme where the mean allele frequency was closest to 0.5 (demes near the outer edges of the simulated hybrid zone, i.e., demes 1 to 20 and 90 to 110, were excluded from consideration). We then focused our analyses on the 11 demes centered on this central deme (i.e., the central deme plus five on each side). We fit our hierarchical Bayesian geographic cline models to the logit allele frequencies from these 11 sites. Our focus on 11 sites represents a compromise between avoiding the exponential decay portion of multilocus clines, retaining a sufficient number of demes for analysis, and not unduly constraining the possible values for the slope. This differs from the threshold of $\text{logit}(p)$ between -2 and 2 used by Kruuk et al. (1999), which would not have been as practical across the range of simulated hybrid zones. We fit each cline model using HMC with 20 chains each comprising 1200 iterations and a 1000 iteration warmup. We computed the Gelman-Rubin convergence diagnostic to verify adequate HMC mixing and likely convergence of the HMC algorithm to the posterior distribution.

Across the 1140 simulated hybrid zone data sets, the mean cline slope (μ_β , measured in units of change in logit allele frequency per deme) ranged from -1.65 to -0.11 (mean = -0.88) and the SD in slopes (σ_β) ranged from 0.0027 to 0.67 (mean 0.26). Together the mean and SD in cline slopes explained 85.2% of the variation in the simulated coupling coefficients (standardized regression coefficients from simple linear regression with an interaction term: $|\mu_\beta| = 0.47$, s.e. = 0.008; $\sigma_\beta = -0.027$, s.e. 0.008; $|\mu_\beta|:\sigma_\beta = -0.084$, s.e. = 0.010; model P-value < 0.0001) (Figure 2A). Because the scale of variability in clines (σ_β) depends on the average slope, we also considered the coefficient of variation (CV), that is $\sigma_\beta/|\mu_\beta|$. We found that the CV was negatively associated with θ and by itself explained 74.0% of the variation in this parameter for the simulated data sets. Importantly, we found that μ_β and the CV varied smoothly with θ , without an obvious, abrupt transition in these high-level cline parameters at $\theta = 1$. With that said, there was evidence that μ_β and CV approached an asymptote around $\theta = 1$ (Figure 2B,D; Table S1). A similarly non-linear relationship was detected between σ_β and θ , such that the SD in clines increased and then decreased as a function of the coupling coefficient with the transition occurring around $\theta = 1$ (Figure 2C; Table S1). This transition was especially pronounced when viewed on a log-log scale, where there was evidence of a bimodal distribution of σ_β and possible bistability around $\theta = 1$ as is sometimes observed at tipping points (Figure S1) (e.g., Nosil et al. 2017).

To fit genomic clines, we first designated the subset of demes with mean allele frequencies between 0.1 and 0.9 to constitute the hybrid zone (as opposed to non-hybrid parental populations). Such demes need not be contiguous and do not necessarily include only or even mostly hybrid individuals, but intermediate allele frequencies at the diagnostics marker loci suggest at least the possibility for hybrids. We then sampled up to 300 individuals as putative hybrids from these demes. We only fit models when one or more demes met this criterion. We

used the known hybrid indexes and parental allele frequencies from *dfuse* and fit the hierarchical Bayesian genomic cline models described above using HMC. We ran four independent HMC chains per data set with 2000 total iterations and a 1000 iteration burnin for each chain. As with the geographic cline analysis, we used the Gelman-Rubin convergence diagnostic to verify good HMC performance.

The total number of hybrids, defined here as individuals with hybrid indexes between 0.1 and 0.9, in the simulated hybrid zones declined with increasing coupling coefficients (linear regression $r^2 = 0.71$, $P < 0.0001$) (Figure S2). The SDs for genomic cline center (σ_c) and slope (σ_v) together explained 52.3% of the variation in the simulated coupling coefficients (standardized regression coefficients from linear regression: $\sigma_c = -0.0167$, s.e. = 0.013; $\sigma_v = -0.24$, s.e. = 0.013; $\sigma_c:\sigma_v = 0.097$, s.e. = 0.0090; model P-value < 0.0001) (Figure 3A). These results consider only simulated data sets with at least 10 hybrids (defined as individuals with hybrid indexes between 0.1 and 0.9); similar results were obtained when analyzing data sets that included 50 or more hybrids (see Figure S3). Using polynomial regression, we found some evidence of a non-linear relationship between the coupling coefficient (θ) and these cline hyperparameters (σ_c and σ_v) (Figure 3; Table S2), but we did not detect a sharp transition for either parameter around $\theta = 1$. However, we were only able to analyze a modest proportion of data sets with $\theta > 1$ (e.g., 34% with $\theta = 1.5$ and 2% with $\theta = 2$, compared to 83% with $\theta = 1$ and all simulations with $\theta \leq 0.5$) as these produced few actual hybrids, often too few to have a reasonable hybrid index axis for cline fitting (see Figure S2). Thus, the lack of hybrids when coupling coefficients notably exceed 1 makes detecting such a sharp transition difficult but at the same time represents a transition to strong reproductive isolation. In other words, while we cannot estimate the variance in clines without hybrids, we can conceptualize such hybrid zones as having a cline variance of 0.

Overall, our geographic and genomic analyses of simulated hybrid zones revealed a strong relationship between the coupling coefficients (θ) and the cline hyperparameters (σ_β and μ_β for geographic clines and σ_c and σ_v for genomic clines), with mixed evidence of a smooth versus more abrupt transition in these parameters as a function of θ . Importantly, even if patterns for clines vary continuously with θ , feedbacks in nature could cause most hybrid zones to fall into a low coupling state with high cline variance (and wide mean geographic clines) or a high coupling state with low cline variance (and steep mean geographic clines). It is to this topic of natural hybrid zones that we now turn. Although our analyses of simulations suggested a stronger relationship between geographic cline parameters and coupling than between genomic cline parameters and coupling, this was for a set of simulations conducted on a similar scale of dispersal. This will not be true for diverse natural systems making geographic comparisons difficult. We thus focus exclusively on genomic clines for natural hybrid zones. This also allows us to analyze mosaic hybrid zones, (e.g., Harrison 1986; Bierne *et al.* 2003), and other hybrid zones lacking a major geographic axis.

Testing robustness of our estimates of coupling from simulated data

Our goal was to use the relationship between cline parameters and θ documented above to

estimate θ in empirical data sets. Before doing so, we conducted several tests of the robustness of our results based on simulations. These are detailed in the online Supplementary Material, but we highlight the key results here. First, we compared the performance of the linear regression models used to estimate θ to a non-linear regression approach, specifically the random forest machine learning model. The out-of-bag (i.e., for simulations not used to fit the model) percent variance explained from a random forest model relating σ_c and σ_v to θ was similar to the explanatory power of a linear regression model (random forest out-of-bag variance explained = 58.0% compared to linear model $r^2 = 0.523$), suggesting robustness of our analytical approach. We also found that estimates of the cline parameter SDs were reasonably robust to alternative sampling schemes within the simulated hybrid zones (see the Supplementary Material and Figure S4). Likewise, our results were qualitatively consistent for different genetic map sizes. Specifically, across a set of simulations that included 0.5, 1 or 2 Morgan chromosomes, σ_c and σ_v explained 37% of the variation in the simulated values of θ (linear regression, model $P < 0.0001$) (see the Supplementary Material and Figure S5 for details). Nonetheless, differences in map size quantitatively affected patterns of hybridization and introgression such that the number of hybrids and the SD in clines declined more rapidly with increased θ when recombination occurred less frequently (Figure S5A). Consequently, a regression model incorporating an interaction between map size and the cline SDs increased our ability to explain variation in θ from $r^2 = 0.371$ to $r^2 = 0.534$.

Lastly, we conducted a full additional set of simulations with a lower rate of gene flow ($m = 0.05$). We found that the relationship between cline parameters and θ was weaker in these simulations (linear model $r^2 = 0.415$, random forest out-of-bag variance explained = 50.6%) (Figure S6), and that estimates of coupling coefficients for these simulations based on models fit from the original simulations with $m = 0.1$ or 0.2 were less accurate (estimated values of θ based on the linear regression model fit for $m = 0.1$ and 0.2 explained 22.5% of the variation in the true, simulated values of θ for $m = 0.05$). Importantly, random forest regression models fit for $m = 0.1$ and 0.2 failed to predict θ values for $m = 0.05$ (out-of-bag variance explained ~0%) (in contrast both linear regression and random forest performed well in terms of predicting $m = 0.1$ results from models fit based on $m = 0.2$ and vice versa, see Figure S7). Because of this, we focus mostly on the linear regression model for predicting θ with empirical hybrid zones. We also caution that our estimates of θ for the empirical data sets should be taken as estimates given the specific simulation conditions and linear model used here (we discuss ways that this could be improved in the Concluding remarks section). However, our estimates of cline SDs for the empirical hybrid zones are themselves informative about the variability of clines across the genome and thus the extent that evolutionary processes are operating on all loci similarly, which is central to understanding hybrid zone dynamics and speciation regardless of the connection to a theoretical coupling coefficient.

Empirical data sets

We fit genomic cline models for 25 empirical hybrid zones (see Table 1 for details) to determine whether these data sets smoothly spanned the continuum from weak to strong coupling (or

likewise, high to low cline variances). We chose these data sets to span taxonomic diversity and only included hybrid zones with genome-wide SNP data. This is not meant to be an exhaustive set of hybrid zones, but we hope it is representative of the genomic hybrid zone data sets that exist in the literature. Our use of these data sets to estimate θ assumes that each is a hybrid zone maintained by selection against hybrids (e.g., a tension zone) or at least that hybrids are not favored within the hybrid zone (as they would be in models of bounded hybrid superiority) (Moore 1977). Past theory and reviews suggest that most hybrid zones are tension zones or at least exhibit clines similar to expectations from tension zones (Barton and Hewitt 1985). Moreover, by using genomic rather than geographic clines our analyses can extend to cases where hybridization does not occur in a simple geographic context (e.g., to mosaic hybrid zones). Thus, we expect this assumption to be at least reasonable for most of the data sets, and even if a few do not fit the tension zone model very well, this should not qualitatively alter our core conclusions. Details of the data sets and data processing are provided in the online Supplementary Material.

For each data set, we analyzed only ancestry informative loci (here defined as SNPs with an allele frequency difference between the parental taxa > 0.3), and we limited our analysis to 1000 randomly sampled loci (mean number of loci = 583, SD = 389, minimum = 25, maximum = 1000). We chose a minimum allele frequency difference cutoff of 0.3 to ensure loci carried sufficient information about genetic ancestry to provide meaningful estimates of genomic clines (clines in ancestry), while also not excluding data sets with lower levels of genetic differentiation (see, e.g., Gompert et al. 2012). Still, this does not ensure an identical level of ancestry informativeness of loci across data sets. We limited the number of loci to 1000 as our goal was to efficiently estimate the variance in clines (even a sample size of 25 provides a relatively precise estimate of a variance) not to describe detailed patterns of introgression across the genome. Additionally, using no more than 1000 loci minimizes the effect of LD caused by tight physical linkage on our results (as opposed to LD more generally, which is a signal of interest). Cline variances were inferred using the Bayesian genomic cline model described above with 8 HMC chains each comprising a 1000 iteration warmup and 1500 iterations for sampling. We then estimated coupling coefficients based on the values of σ_c and σ_v and the parameterized linear model or random forest regression model from the simulated data sets (the parameterized models were based on the core simulations with $m = 0.1$ or 0.2).

For the 25 empirical hybrid zones, estimated SDs in genomic cline center (σ_c) and slope (σ_v) ranged from 0.29 to 1.73 (mean = 0.83) and 0.12 to 0.56 (mean = 0.29), respectively, and thus broadly overlapped with estimates from our simulated hybrid zones (Figures 4 and 5, see Figure S8 for the distribution of hybrid indexes in each hybrid zone). Using the linear model fit on the simulated data sets, estimates of the coupling coefficient (θ) for the empirical hybrid zone data sets ranged from 0.08 to 1.30 (mean = 0.55). We found no evidence of a gap in inferred coupling coefficients, rather we documented a smooth continuum from low to high coupling across these 25 data sets (Figure 5). Moreover, similar results were obtained when estimating θ from our fit random forest regression model; this was true both with respect to the distribution of coupling coefficients across the data sets (Figure S9) and the specific estimates of θ for each data set (Pearson correlation = 0.82, 95% CIs = 0.63-0.92) (Figure S10). As noted in the

preceding section, both the linear regression and random forest models were parameterized from a specific set of simulations and the relationship between the cline variances and θ could (and likely does) vary under different conditions. However, we also documented a smooth continuum in the cline SDs (σ_c and σ_v) across the 25 hybrid zone data sets. This suggests that even if these metrics do not relate to the theoretical parameter θ in the manner suggested here, we still find a smooth continuum in the degree to which loci introgress independently (uncoupled) or not (coupled) across these hybrid zones, and it is this that is most relevant for our understanding of speciation.

Our analyses above consider a single hybrid zone for each species, but the evolutionary outcomes of secondary contact can vary among populations due to genomic and ecological context (Gompert et al. 2017; Mandeville et al. 2017). Thus, to assess this possibility with respect to coupling and thereby further evaluate the robustness of our general results to details of the systems considered, we compared estimates of coupling across transects for a subset of species. Specifically, the datasets we obtained included five species with two hybrid zones or transects, and one species (*Mus*) with three transects, resulting in eight pairs of transects (only one transect from each was included in the core analyses above). Using these additional transects, we found moderate consistency in estimates of θ for the pairs (Pearson correlation = 0.67, 95% CI = -0.17-0.95, $P = 0.101$, similar results were obtained for estimates based on random forest regression, $r = 0.75$, 95% CI = -0.01-0.96, $P = 0.053$) (Figure S10; Table S3).

Lastly, given the widespread interest in the role of sex chromosomes in speciation (Haldane 1922; Coyne and Orr 1989) and observation that sex chromosomes often have steeper clines across hybrid zones (e.g., Tucker et al. 1992; Carling and Brumfield 2008; Hooper et al. 2019), we asked whether coupling was stronger (higher estimates of θ) for sex chromosomes than autosomes. For this, we ran additional analyses separately estimating cline SDs and θ for eight hybrid zone data sets (including replicate transects) that we were especially familiar with and where there was clear information about which SNPs were on the X or Z sex chromosome vs autosomes (see Table S4). Importantly, differences in recombination rates or effective population size for sex chromosomes relative to autosomes, in addition to differences in the number of barrier loci, could contribute to differences in estimates of θ . Across these data sets, estimates of θ for autosomes and sex chromosomes were positively correlated (Pearson correlation = 0.65, 95% CI = -0.09-0.93, $P = 0.078$). Moreover, in seven of the eight datasets θ was higher for the X/Z chromosome than for the autosomes (mean difference in θ across data sets = 0.12) (Table S4). The one exception was for the *Gryllus* Connecticut transect where θ was notably higher for the autosomes (0.86) than for the X chromosome (0.39).

Concluding remarks

We documented a smooth continuum of coupling across 25 natural hybrid zones, ranging from very weak coupling to near-complete coupling suggestive of a strong genome-wide barrier to gene flow ($\theta > 1$) (Figure 5B). As such, we found no evidence of a tipping point or positive feedback loop near $\theta = 1$ that would rapidly drive systems to higher levels of coupling resulting

in a dearth of systems with $\theta \sim 1$ (analogous to the gap in genetic differentiation documented in sympatric *Timema* stick insects, see Riesch *et al.* 2017). Several factors likely contributed to our finding of a smooth continuum of inferred coupling coefficients. First, if species mostly diverge in allopatry with limited gene flow, there would be limited opportunity for the feedback of increased coupling to reduce effective migration and further increase LD among barrier loci. This is consistent with findings from theoretical work by Flaxman *et al.* (2014) that found evidence of a sudden transition to genome-wide congealing (analogous to coupling) when gene flow was strong relative to selection, but more gradual divergence otherwise (also see Sinitambirivoutin *et al.* 2023). However, this feedback could still operate upon secondary contact and thus in hybrid zones if species come into contact before speciation is complete (see, e.g., Flaxman *et al.* 2014). Second, different organisms vary in ecology and demographic histories, including the time since secondary contact, and these differences could act to smooth out the empirical distributions of cline SDs and estimates of coupling across systems (this is part of a general difficulty in treating pairs of species as a reconstructed “chrono-sequence”, see, e.g., Nosil *et al.* 2017; Bolnick *et al.* 2023). Third, different sets of loci, such as barrier loci versus neutral loci, can become coupled at different rates (Barton 1983; Barton and Bengtsson 1986; Schilling *et al.* 2018). Our analyses necessarily average over such variation, which could contribute to the continuum of coupling documented here. And finally, some predictions for a sudden transition in coupling in terms of parameter space may not reflect the temporal dynamics by which reproductive isolation evolves. For example, we may not see the expected transition from uncoupled to coupled dynamics at $\theta = 1$ when holding R and total selection constant but increasing L (Barton 1983) if speciation does not progress with total selection held constant.

Despite the smooth continuum we detected in terms of cline SDs and estimated coupling coefficients, we did observe patterns that suggest a transition in hybrid zone dynamics around $\theta = 1$. Specifically, in simulations, hybrids become increasingly rare around $\theta = 1$ and few of our empirical hybrid zones had estimates of θ notably larger than 1. Thus, consistent with past work (e.g., Barton 1983) our results suggest that by $\theta = 1$ the overall barrier to gene flow across much of the genome is quite strong. The small number of empirical hybrid zones with θ larger than 1 suggests that species pairs with such high levels of coupling rarely have patterns that evolutionary biologists would classify as hybrid zones. In other words, $\theta = 1$ might roughly approximate a genome-wide species boundary even if the approach to this boundary occurs gradually and at a near constant rate with increases in θ (e.g., with an increase in the number of barrier loci and thus a decrease in r).

Our results are relevant for at least two other classic issues in the study of hybrid zones and speciation. First, considerable work has been done on the consistency or variability of overall and locus-specific patterns of introgression across replicate hybrid zones or hybrid zone transects (Buerkle and Rieseberg 2001; Nolte *et al.* 2009; Janoušek *et al.* 2012; Larson *et al.* 2014; Schaefer *et al.* 2016; Mandeville *et al.* 2017). This is relevant as it bears on the degree to which reproductive isolation and hybridization outcomes are contingent on the ecological context of secondary contact. Such analyses have rarely, if ever, considered cline SDs and coupling explicitly when comparing hybrid zones. We found relatively high levels of consistency even in systems where locus-specific patterns of introgression have been shown to be less

consistent, that is where patterns of introgression for individual loci have varied substantially across transects (e.g., Mus, Teeter *et al.* 2010). Thus, our results hint at a greater consistency in the overall barrier to gene flow, as captured by cline SDs and estimates of θ , than in locus-specific patterns. Second, there is considerable evidence that sex chromosomes (X or Z) contribute disproportionately to reproductive isolation and exhibit reduced introgression in hybrid zones (Tucker *et al.* 1992; Carling and Brumfield 2008; Janoušek *et al.* 2012; Hooper *et al.* 2019). Our results suggest that, consistent with these other patterns, sex chromosomes show higher levels of coupling (lower cline SDs) than autosomes, which could be at least partially responsible for the lower overall rates of introgression (Muirhead and Presgraves 2015).

Our simulations and analyses documented a link between genome-wide variation in clines and a theoretical quantity relevant for understanding the extent to which genes or the entire genome experience a barrier to gene flow (the coupling coefficient, θ). Using simulations, we showed that this relationship is somewhat consistent under different conditions, but does vary to an extent and is likely to vary even more under conditions that diverge more from those we considered, such as in mosaic hybrid zones lacking a strong spatial axis. Our focus here was on cline SDs, but additional information, such as the prevalence of hybrids or the distribution of hybrids and introgression in space, could provide additional information about coupling. Using such additional metrics perhaps combined with more tailored simulations based on individual hybrid zones could provide a powerful framework to infer coupling coefficients using, for example, approximate Bayesian computation or neural networks.

More generally, our results demonstrate the relevance of quantifying cline variances rather than focusing solely on patterns of introgression for individual loci in analyses of hybrid zones. Under some conditions, especially strong coupling in the later stages of speciation, cline variances could be more informative about the process of speciation than patterns of introgression for individual loci. In contrast, when coupling is weak, loci resistant to introgression would be more likely to reside in genomic regions causally connected to reproductive isolation (Gompert *et al.* 2012). Additional understanding could come from quantifying cline variances and coupling at different genomic scales. For example, here we documented differences in variances and coupling for autosomes versus sex chromosomes. Finer scale analyses looking at coupling along chromosomes (e.g., in megabase windows) or within versus outside of large structural variants could provide additional insights on how the ratio of selection to recombination varies across the genome and thus on the genetics of speciation. In light of our findings, we think further empirical analyses of cline variances and coupling focused on the transition between weak and strong coupling, ideally within specific taxonomic groups, are critical for advancing understanding of the dynamics of speciation.

Acknowledgements

The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. The National Science Foundation provided support for this project to ZG (NSF DEB 1844941), TJF (NSF DBI 2208825), ELL (NSF DEB 2012041), SAT

475 (NSF IOS-1754909; DEB-192889). Silhouettes for figure 4 were obtained from phylopic.org. We
476 thank our colleagues that generated each of the 25 empirical hybrid zone data sets for making
477 these data publicly available or providing them to us upon our request.

478 **Data and code availability**

479 Simulated hybrid zone data sets and our formatted input files for the empirical hybrid zones will
480 be archived on Dryad (DOI pending). Computer scripts used for simulations, data processing
481 and analyses are available from GitHub, <https://github.com/zgompert/ClineCoupling>.

482 **Author contributions**

483 All authors conceived of and designed the study. TJF compiled the hybrid zone data, with help
484 from GS, SAT, ELL and ZG. ZG conducted the hybrid zone simulations and analyzed the data.
485 TJF and ZG wrote an initial draft of the chapter. All authors contributed substantially to editing
486 and revising the chapter.

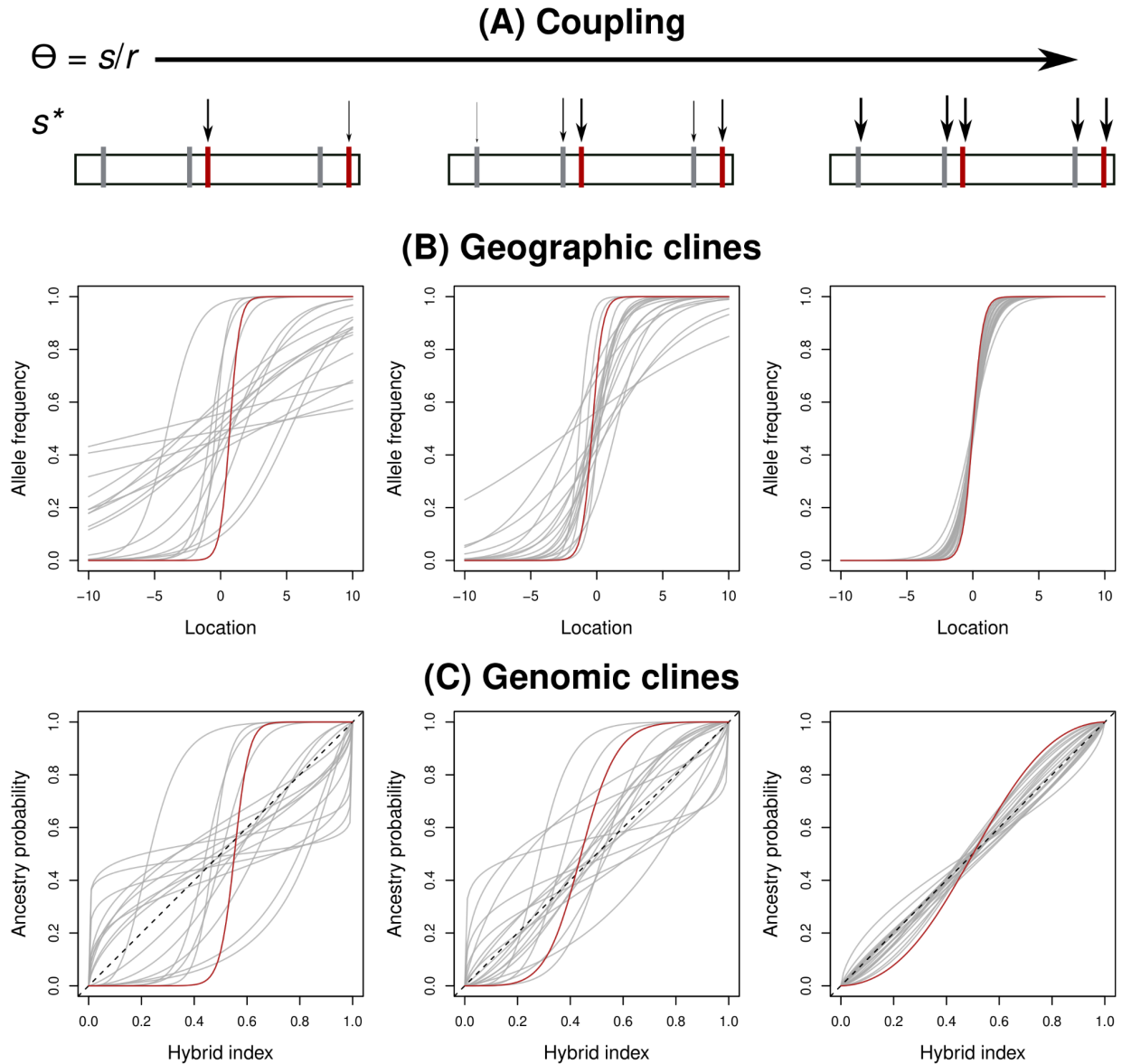


Figure 1. Coupling and its consequences. (A) The coupling coefficient $\theta = s / r$ determines the extent to which barrier loci operate independently or as a unit, depicted here with a few example loci. As coupling (θ) increases the total selection (s^* , represented by the size of the vertical arrows) experienced by barrier loci (red bars) and neutral loci (gray bars) increases because of increased linkage disequilibrium. With low coupling, geographic (B) and genomic (C) clines vary across the genome, whereas with high coupling geographic clines steepen and exhibit similar cline centers and widths and genomic clines converge to the genome-average admixture gradient (hybrid index). Geographic and genomic clines at barrier loci and neutral loci are shown in red and gray, respectively in (B) and (C). Dashed lines in (C) denote the genome-average admixture gradient.

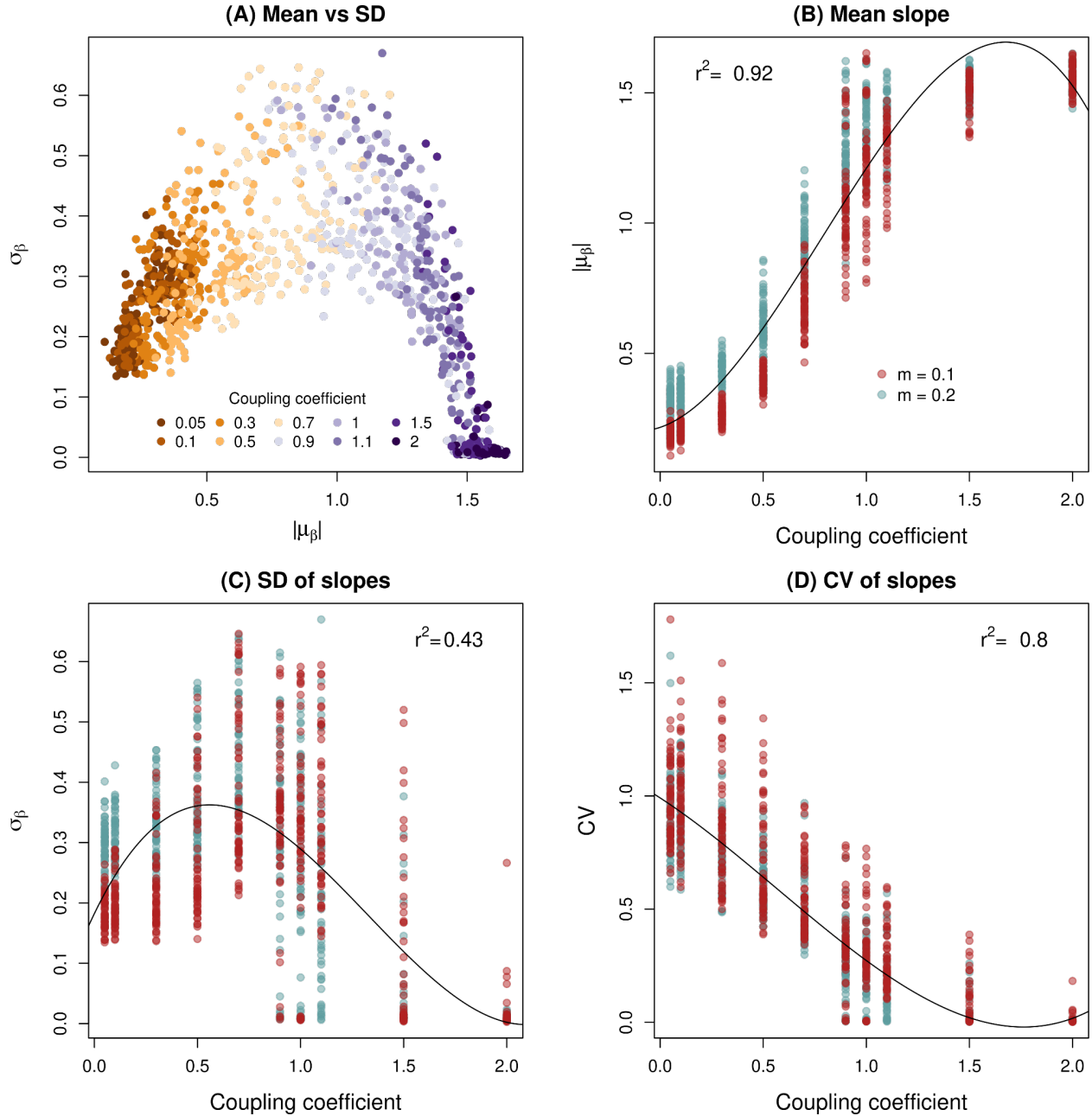


Figure 2. Relationship between geographic clines and coupling in simulated hybrid zones. Panel (A) shows estimates of mean cline slope (μ_β) and the standard deviation (SD) in slopes (σ_β) from the simulated hybrid zones. Points are colored by the known coupling coefficient. Panels (B-D) show the relationship between the coupling coefficient (θ) and μ_β (B), σ_β (C), and the coefficient of variation (CV) for cline slopes (D). Points denote results from individual simulations and are colored based on the migration rate (m) between neighboring demes. The best fit line from polynomial regression is shown along with the corresponding coefficient of determination (r^2).

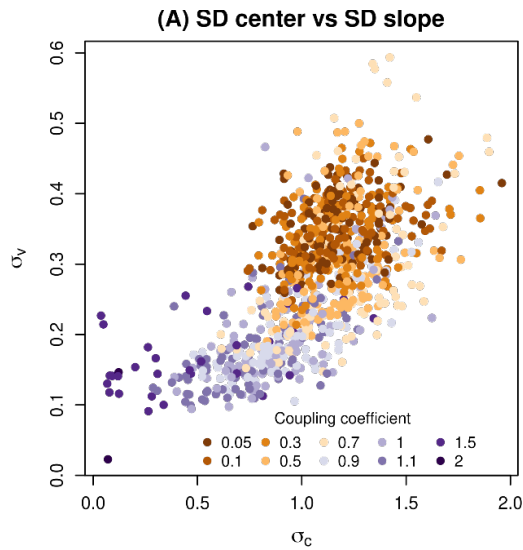
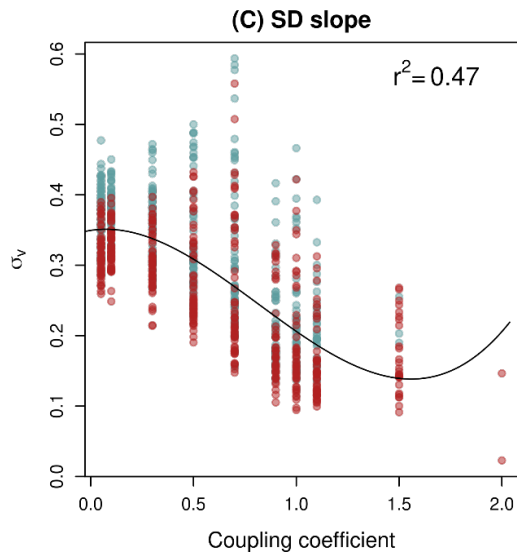
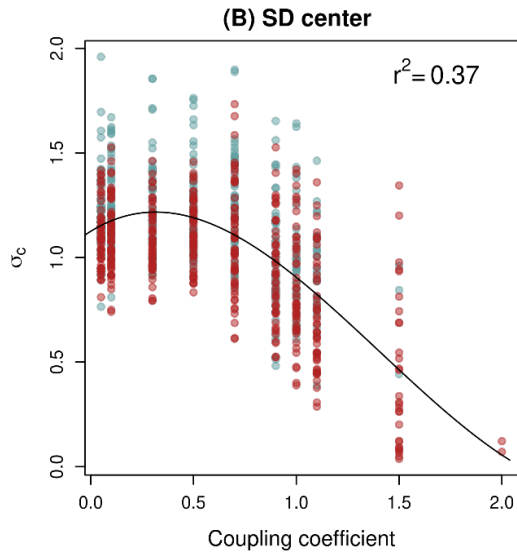


Figure 3. Relationship between genomic clines and coupling in simulated hybrid zones. Panel (A) shows estimates of the standard deviation (SD) for genomic cline center (σ_c) and slope (σ_v) from the simulated hybrid zones. Points are colored by the known coupling coefficient. Panels (B-C) show the relationship between the coupling coefficient (θ) and σ_c (B) or σ_v (C). Points denote results from individual simulations and are colored based on the migration rate (m) between neighboring demes (see Figure 2). The best fit line from polynomial regression is shown along with the corresponding coefficient of determination (r^2).



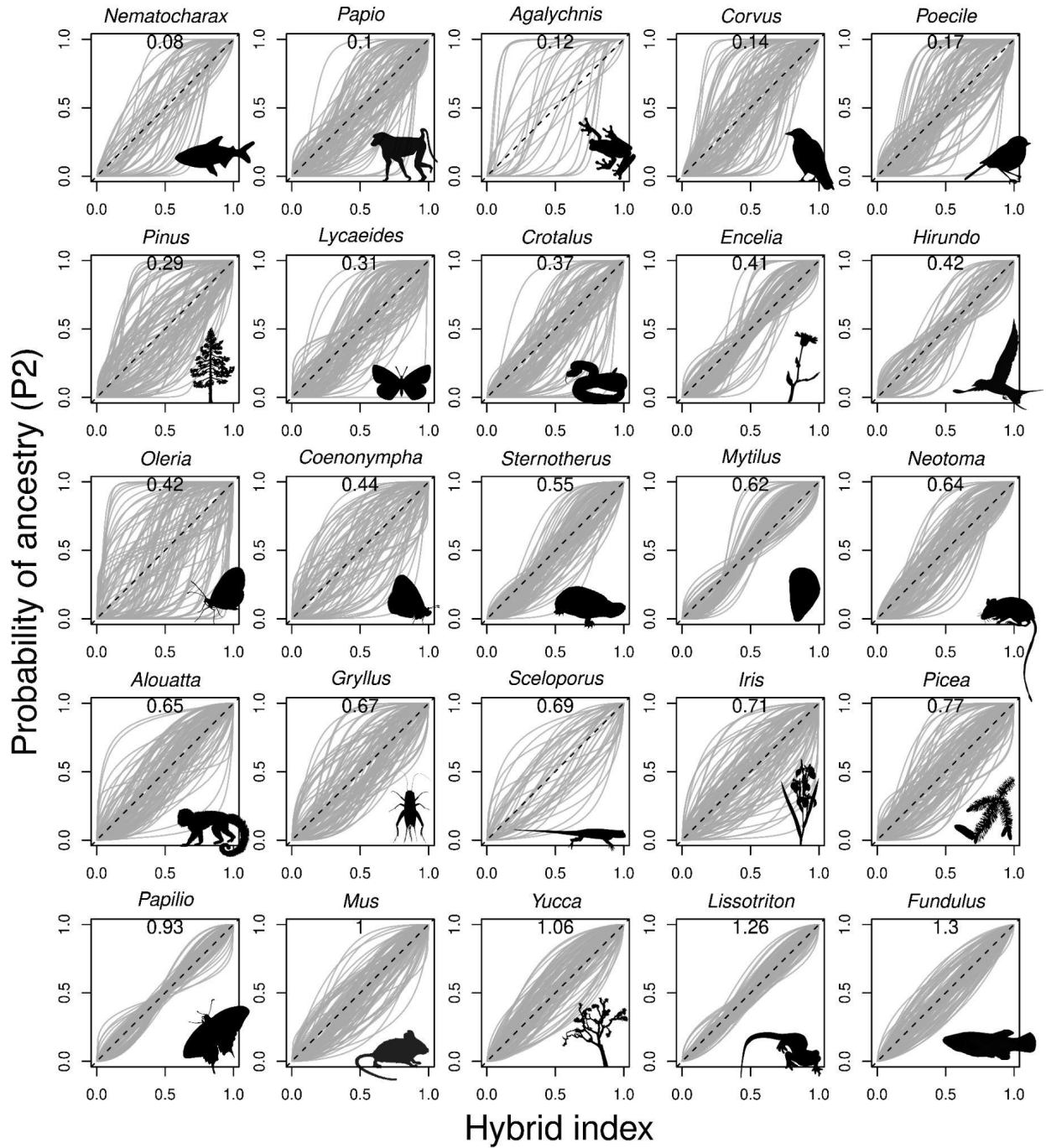


Figure 4. Summary of genomic clines empirical hybrid zones. The plots show estimated genomic clines (gray lines) for each of 25 hybrid zones. Each cline denotes the probability of ancestry for species 2 as a function of hybrid index (the total proportion of the genome inherited from species 2). Clines for 100 randomly chosen loci (or all loci if there were fewer than 100) are shown. The dashed one-to-one line denotes an ancestry probability equal to hybrid index. Our estimate of the coupling coefficient based on the variability among clines is reported.

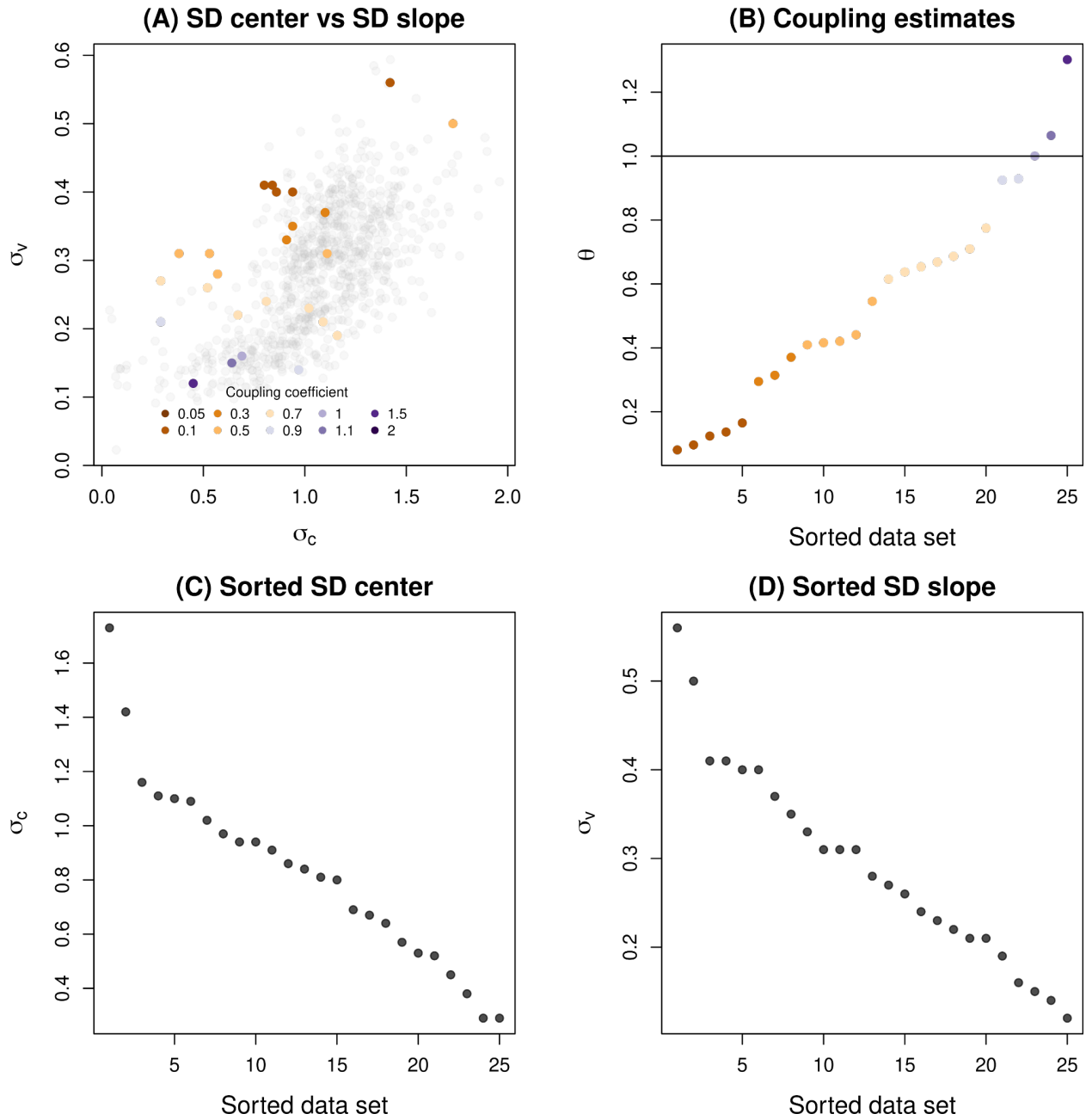


Figure 5. Cline variances and coupling in empirical hybrid zones. Panel (A) shows estimates of the standard deviation (SD) for genomic cline center (σ_c) and slope (σ_v) from the simulated (light gray points) and empirical (colored points) hybrid zones. Coupling coefficients for the 25 empirical hybrid zones were estimated from a linear regression model fit with the simulated hybrid zone data. Panel (B) plots the estimated coupling coefficients (θ) for the empirical hybrid zones sorted from smallest to largest. The horizontal line denotes $\theta = 1$. Panels (C) and (D) show estimates of the standard deviation (SD) for genomic cline center (σ_c) and slope (σ_v) for each of the 25 empirical hybrid zone data sets, here sorted from the largest (high variability among clines) to smallest (low variability among clines). These data suggest a continuum of estimated coupling coefficients (B) and cline parameters SDs (C and D).

535 **Table 1.** Empirical hybrid zone data sets.

Organism	Taxonomic Group	Divergence Time (MYA)	Number of loci	SD cline center (σ_c)	SD cline slope (σ_v)	θ
<i>Agalychnis</i>	Amphibian	-	25	1.42	0.56	0.12
<i>Alouatta</i>	Mammal	~ 3.0	1000	1.02	0.23	0.65
<i>Coenonympha</i>	Insect	< 0.02	81	1.11	0.31	0.44
<i>Corvus</i>	Bird	~ 0.443	588	0.86	0.40	0.14
<i>Crotalus</i>	Reptile	3.0 - 5.2	1000	0.91	0.33	0.37
<i>Encelia</i>	Plant	~ 1.4	1000	0.38	0.31	0.41
<i>Fundulus</i>	Fish	-	1000	0.45/0.33	0.12/0.08	1.30/1.58**
<i>Gryllus</i>	Insect	~ 0.2	110	0.81/0.48	0.24/0.15	0.67/1.15**
<i>Hirundo</i>	Bird	< 0.1	54	0.53	0.31	0.42
<i>Iris</i>	Plant	-	1000	1.16	0.19	0.71
<i>Lissotriton</i>	Amphibian	~ 1.0	737/730	0.35/0.97	0.14/0.14	1.26/0.92**
<i>Lycaeides</i>	Insect	~ 2.4	500	0.94	0.35	0.31

<i>Mus</i>	Mammal	~ 0.5	1000	0.69/0.70/1.01	0.16/0.14/0.23	1.00/1.07/0.66**
<i>Mytilus</i>	Bivalve	~ 3.5	419	0.29	0.27	0.62
<i>Nematocharax</i>	Fish	0.27 - 0.67	52	0.80	0.41	0.08
<i>Neotoma</i>	Mammal	~ 1.6	623	0.52	0.26	0.64
<i>Oleria</i>	Insect	-	1000	1.73	0.50	0.42
<i>Papilio</i>	Insect	0.5 - 0.6	164	0.29	0.21	0.93
<i>Papio</i>	Mammal	1.0 - 2.2	501	0.84	0.41	0.10
<i>Picea</i>	Plant	12.5 - 15.0	221	0.67	0.22	0.77
<i>Pinus</i>	Plant	~ 18.04	670	1.10	0.37	0.29
<i>Poecile</i>	Bird	~ 1.5	1000	0.94/0.82	0.40/0.38	0.17/0.19**
<i>Sceloporus</i>	Reptile	0.045 - 1.9	38	1.09	0.21	0.69
<i>Sternotherus</i>	Reptile	< 4.0	798	0.57	0.28	0.55
<i>Yucca</i>	Plant	0.1 - 0.2	1000	0.64	0.15	1.06

536 ** denotes that there were multiple transects that had coupling coefficients calculated for these organisms
537 SD = standard deviation
538 Θ = coupling coefficient

539 **References**

- 540 Abbott R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, *et al.*, 2013 Hybridization and
541 speciation. *J. Evol. Biol.* 26: 229–246.
- 542 Akopyan M., Z. Gompert, K. Klonoski, A. Vega, K. Kaiser, *et al.*, 2020 Genetic and phenotypic
543 evidence of a contact zone between divergent colour morphs of the iconic red-eyed
544 treefrog. *Mol. Ecol.* 29: 4442–4456.
- 545 Bailey R., 2022 Bayesian hybrid index and genomic cline estimation with the R package
546 gghybrid. Authorea Preprints. <https://doi.org/10.22541/au.164848698.82546348>
- 547 Baird S. J. E., 1995 A simulation study of multilocus clines. *Evolution* 49: 1038–1045.
- 548 Baiz M. D., P. K. Tucker, and L. Cortés-Ortiz, 2019 Multiple forms of selection shape
549 reproductive isolation in a primate hybrid zone. *Mol. Ecol.* 28: 1056–1069.
- 550 Barreto S. B., L. L. Knowles, P. R. A. de M. Affonso, and H. Batalha-Filho, 2020 Riverscape
551 properties contribute to the origin and structure of a hybrid zone in a Neotropical freshwater
552 fish. *J. Evol. Biol.* 33: 1530–1542.
- 553 Barton N. H., 1983 Multilocus clines. *Evolution* 37: 454–471.
- 554 Barton N. H., and G. M. Hewitt, 1985 Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* 16: 113–
555 148.
- 556 Barton N. and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising

560 Barton N. H., and M. A. R. de Cara, 2009 The evolution of strong reproductive isolation.
561 Evolution 63: 1171–1190.

562 Bierne N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, *et al.*, 2003 Introgression patterns in the
563 mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. Mol. Ecol. 12: 447–
564 461.

565 Bierne N., J. Welch, E. Loire, F. Bonhomme, and P. David, 2011 The coupling hypothesis: why
566 genome scans may fail to map local adaptation genes. Mol. Ecol. 20: 2044–2072.

567 Bolnick D. I., A. K. Hund, P. Nosil, F. Peng, M. Ravinet, *et al.*, 2023 A multivariate view of the
568 speciation continuum. Evolution 77: 318–328.

569 Buerkle C. A., and L. H. Rieseberg, 2001 Low intraspecific variation for genomic isolation
570 between hybridizing sunflower species. Evolution 55: 684–691.

571 Buerkle C. A., and L. H. Rieseberg, 2008 The rate of genome stabilization in homoploid hybrid
572 species. Evolution 62: 266–275.

573 Capblancq T., L. Després, and J. Mavárez, 2020 Genetic, morphological and ecological
574 variation across a sharp hybrid zone between two alpine butterfly species. Evol. Appl. 13:
575 1435–1450.

576 Carling M. D., and R. T. Brumfield, 2008 Haldane's rule in an avian system: using cline theory
577 and divergence population genetics to test for differential introgression of mitochondrial,
578 autosomal, and sex-linked loci across the *Passerina* hunting hybrid zone. Evolution 62:

582 Chiou K. L., C. M. Bergey, A. S. Burrell, T. R. Disotell, J. Rogers, *et al.*, 2021 Genome-wide
583 ancestry and introgression in a Zambian baboon hybrid zone. *Mol. Ecol.* 30: 1907–1920.

584 Coyne J. A., and A. H. Orr, 1989 Two rules of speciation, pp. 180–207 in *Speciation and its*
585 *Consequences*, edited by D. Otte & J. E. Sinauer Associates.

586 Cruzan M. B., P. G. Thompson, N. A. Diaz, E. C. Hendrickson, K. R. Gerloff, *et al.*, 2021 Weak
587 coupling among barrier loci and waves of neutral and adaptive introgression across an
588 expanding hybrid zone. *Evolution* 75: 3098–3114.

589 Dopman, E.B., K.L. Shaw, M.R. Servedio, R. Butlin, and C. M. Smadja. 2023. Coupling barriers
590 to gene exchange: Causes and consequences. *Cold Spring Harb Perspect Biol.* doi:
591 10.1101/cshperspect.a041432

592 DiVittorio C. T., S. Singhal, A. B. Roddy, F. Zapata, D. D. Ackerly, *et al.*, 2020 Natural selection
593 maintains species despite frequent hybridization in the desert shrub *Encelia*. *Proc. Natl.*
594 *Acad. Sci. U. S. A.* 117: 33373–33383.

595 Fitzpatrick B. M., 2013 Alternative forms for genomic clines. *Ecol. Evol.* 3: 1951–1966.

596 Flaxman S. M., A. C. Wacholder, J. L. Feder, and P. Nosil, 2014 Theoretical models of the
597 influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* 23: 4074–4088.

598 Gauthier J., D. L. de Silva, Z. Gompert, A. Whibley, C. Houssin, *et al.*, 2020 Contrasting
599 genomic and phenotypic outcomes of hybridization between pairs of mimetic butterfly taxa
600 across a suture zone. *Mol. Ecol.* 29: 1328–1343.

604 2111–2127.

605 Gompert Z., T. L. Parchman, and C. A. Buerkle, 2012 Genomics of isolation in hybrids. *Philos.*
606 *Trans. R. Soc. Lond. B Biol. Sci.* 367: 439–450.

607 Gompert Z., E. G. Mandeville, and C. Alex Buerkle, 2017 Analysis of population genomic data
608 from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* 48: 207–229.

609 Haldane J. B. S., 1922 Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* 12: 101–
610 109.

611 Hamilton J. A., C. Lexer, and S. N. Aitken, 2013 Genomic and phenotypic architecture of a
612 spruce hybrid zone (*Picea sitchensis* × *P. glauca*). *Mol. Ecol.* 22: 827–841.

613 Harrison R. G., 1986 Pattern and process in a narrow hybrid zone. *Heredity* 56: 337–349.

614 Harrison R. G., and E. L. Larson, 2014 Hybridization, introgression, and the nature of species
615 boundaries. *J. Hered.* 105: 795–809.

616 Hooper D. M., S. C. Griffith, and T. D. Price, 2019 Sex chromosome inversions enforce
617 reproductive isolation across an avian hybrid zone. *Mol. Ecol.* 28: 1246–1262.

618 Jahner J. P., T. L. Parchman, and M. D. Matocq, 2021 Multigenerational backcrossing and
619 introgression between two woodrat species at an abrupt ecological transition. *Mol. Ecol.* 30:
620 4245–4258.

621 Janoušek V., L. Wang, K. Luzynski, P. Dufková, M. M. Vyskočilová, *et al.*, 2012 Genome-wide

625 Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588:
626 106–111.

627 Kozak G. M., C. B. Wadsworth, S. C. Kahne, S. M. Bogdanowicz, R. G. Harrison, *et al.*, 2019
628 Genomic basis of circannual rhythm in the European corn borer moth. *Curr. Biol.* 29: 3501–
629 3509.e5.

630 Kruuk L. E., S. J. Baird, K. S. Gale, and N. H. Barton, 1999 A comparison of multilocus clines
631 maintained by environmental adaptation or by selection against hybrids. *Genetics* 153:
632 1959–1971.

633 Kunerth H. D., S. M. Bogdanowicz, J. B. Searle, R. G. Harrison, B. S. Coates, *et al.*, 2022
634 Consequences of coupled barriers to gene flow for the build-up of genomic differentiation.
635 *Evolution* 76: 985–1002.

636 Larson E. L., T. A. White, C. L. Ross, and R. G. Harrison, 2014 Gene flow and the maintenance
637 of species boundaries. *Mol. Ecol.* 23: 1668–1678.

638 Leaché A. D., J. A. Grummer, R. B. Harris, and I. K. Breckheimer, 2017 Evidence for concerted
639 movement of nuclear and mitochondrial clines in a lizard hybrid zone. *Mol. Ecol.* 26: 2306–
640 2316.

641 Lindtke D., and C. A. Buerkle, 2015 The genetic architecture of hybrid incompatibilities and their
642 effect on barriers to introgression in secondary contact. *Evolution* 69: 1987–2004.

643 Mandeville E. G., T. J. Parchman, K. G. Thompson, B. J. Compton, K. B. Gelwicks, *et al.*, 2017

647 shapes barriers to introgression across butterfly genomes. PLoS Biol. 17: e2006288.

648 Menon M., J. C. Bagley, G. F. M. Page, A. V. Whipple, A. W. Schoettle, *et al.*, 2021 Adaptive
649 evolution in a conifer hybrid zone is driven by a mosaic of recently introgressed and
650 background genetic variants. Commun Biol 4: 160.

651 Michel A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, *et al.*, 2010 Widespread genomic
652 divergence during sympatric speciation. Proc. Natl. Acad. Sci. U. S. A. 107: 9724–9729.

653 Moore W. S., 1977 An Evaluation of Narrow Hybrid Zones in Vertebrates. Q. Rev. Biol. 52: 263–
654 277.

655 Neal R. M., 2011 Handbook of Markov Chain Monte Carlo, p. 50 in *Handbooks of Modern*
656 *Statistical Methods*, Chapman and Hall/CRC.

657 Nikolakis Z. L., D. R. Schield, A. K. Westfall, B. W. Perry, K. N. Ivey, *et al.*, 2022 Evidence that
658 genomic incompatibilities and other multilocus processes impact hybrid fitness in a
659 rattlesnake hybrid zone. Evolution 76: 2513–2530.

660 Nolte A. W., Z. Gompert, and C. A. Buerkle, 2009 Variable patterns of introgression in two
661 sculpin hybrid zones suggest that genomic isolation differs among populations. Mol. Ecol.
662 18: 2615–2627.

663 Nosil P., and D. Schluter, 2011 The genes underlying the process of speciation. Trends Ecol.
664 Evol. 26: 160–167.

669 Peñalba J.V., A. Runemark, J. Meier, P. Singh, G. Wogan, R. Sánchez-Guillén, J. Mallet, S.J.
670 Rometsch, M. Menon, O. Seehausen, J. Kulmuni, R.J. Pereira. The Role of Hybridization in
671 Species Formation. Cold Spring Harb Perspect Biol.

672 Riesch R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault, *et al.*, 2017 Transitions
673 between phases of genomic differentiation during stick-insect speciation. Nat Ecol Evol 1:
674 82.

675 Ritchie M.G. and R.K. Butlin. Genetic coupling in the genomic era. Cold Spring Harb Perspect
676 Biol.

677 Roux C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, *et al.*, 2016 Shedding light on the grey
678 zone of speciation along a continuum of genomic divergence. PLoS Biol. 14: e2000234.

679 Royer A. M., M. A. Streisfeld, and C. I. Smith, 2016 Population genomics of divergence within
680 an obligate pollination mutualism: Selection maintains differences between Joshua tree
681 species. Am. J. Bot. 103: 1730–1741.

682 Ryan S. F., M. C. Fontaine, J. M. Scriber, M. E. Pfrender, S. T. O'Neil, *et al.*, 2017 Patterns of
683 divergence across the geographic and genomic landscape of a butterfly hybrid zone
684 associated with a climatic gradient. Mol. Ecol. 26: 4725–4742.

685 Saarman N. P., and G. H. Pogson, 2015 Introgression between invasive and native blue
686 mussels (genus *Mytilus*) in the central California hybrid zone. Mol. Ecol. 24: 4723–4738.

687 Schaefer J. D., Duvernell, and D. C. Campbell, 2016 Hybridization and introgression in two

691 <https://doi.org/10.3390/genes9060274>

692 Scordato E. S. C., M. R. Wilkins, G. Semenov, A. S. Rubtsov, N. C. Kane, *et al.*, 2017 Genomic
693 variation across two barn swallow hybrid zones reveals traits associated with divergence in
694 sympatry and allopatry. *Mol. Ecol.* 26: 5676–5691.

695 Scott P. A., T. C. Glenn, and L. J. Rissler, 2019 Formation of a recent hybrid zone offers insight
696 into the geographic puzzle and maintenance of species boundaries in musk turtles. *Mol.*
697 *Ecol.* 28: 761–771.

698 Sinitambirivoutin M., P. Nosil, S. Flaxman, J. Feder, Z. Gompert, *et al.*, 2023 Early-warning
699 signals of impending speciation. *Evolution* 77: 1444–1457.

700 Slager D. L., K. L. Epperly, R. R. Ha, S. Rohwer, C. Wood, *et al.*, 2020 Cryptic and extensive
701 hybridization between ancient lineages of American crows. *Mol. Ecol.* 29: 956–969.

702 Sung C.-J., K. L. Bell, C. C. Nice, and N. H. Martin, 2018 Integrating Bayesian genomic cline
703 analyses and association mapping of morphological and ecological traits to dissect
704 reproductive isolation and introgression in a Louisiana Iris hybrid zone. *Mol. Ecol.* 27: 959–
705 978.

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

709 Teeter K. C., J. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman, *et al.*, 2010 The

713 FOR SEX CHROMOSOMES IN A HYBRID ZONE BETWEEN TWO SPECIES OF MICE.
714 Evolution 46: 1146–1163.

715 Unbehend M., G. M. Kozak, F. Koutroumpa, B. S. Coates, T. Dekker, *et al.*, 2021 bric à brac
716 controls sex pheromone choice by male European corn borer moths. Nat. Commun. 12:
717 2818.

718 Vines T. H., A. C. Dalziel, A. Y. K. Albert, T. Veen, P. M. Schulte, *et al.*, 2016 Cline coupling and
719 uncoupling in a stickleback hybrid zone. Evolution 70: 1023–1038.

720 Wu C.-I., 2001 The genic view of the process of speciation. Journal of Evolutionary Biology 14:
721 851–865.

722 Yeaman S., and M. C. Whitlock, 2011 The genetic architecture of adaptation under migration-
723 selection balance. Evolution 65: 1897–1911.

724 Zieliński P., K. Dudek, J. W. Arntzen, G. Palomar, M. Niedzicka, *et al.*, 2019 Differential
725 introgression across newt hybrid zones: Evidence from replicated transects. Mol. Ecol. 28:
726 4811–4824.

727 Muirhead CA, Presgraves DC. Hybrid incompatibilities, local adaptation, and the genomic
728 distribution of natural introgression between species. *American Naturalist* **187**: 249-261.