Parental Population Range Expansion Before Secondary Contact Promotes Heterosis

Ailene MacPherson[1,2,∗,†]

Silu Wang[1,3,†]

Ryo Yamaguchi[4,5]

Loren H. Rieseberg[1]

Sarah P. Otto[1]

1. University of British Columbia, Vancouver, BC V6T 1Z4, Canada;

2. Current Address: University of Toronto, Toronto, Ontario M5S 1A5, Canada;

3. Current Address: University of California and Berkeley, Berkeley, CA 94720;

4. Tokyo Metropolitan University, Hachioji, Tokyo, 192-0397, Japan;

5. Hokkaido University, Sapporo, Hokkaido, 060-8588, Japan;

∗ Corresponding author; e-mail: ailene.macpherson@utoronto.ca.

† Contributed equally to manuscript.

Manuscript elements: Figure 1-5, Table 1, Appendix and Supplementary Figures Figure S1-S3 and Table S1, Dryad submission.

Keywords: Hybrid zone, allele surfing, range expansion, genetic incompatibilities.

Manuscript type: Article.

Prepared using the suggested LaTeX template for *Am. Nat.*

Abstract

Population genomic analysis of hybrid zones is instrumental to our understanding of the evolution of reproductive isolation. Many temperate hybrid zones are formed by the secondary contact between two parental populations that have undergone post-glacial range expansion. Here we show that explicitly accounting for historical parental isolation followed by range expansion prior to secondary contact is fundamental for explaining genetic and fitness patterns in these hybrid zones. Specifically, ancestral population expansion can result in allele surfing where neutral or slightly deleterious mutations drift to high frequency at the expansion front. If these surfed deleterious alleles are recessive, they can contribute to substantial heterosis in hybrids produced at secondary contact, counteracting negative effects of Bateson-Dobzhansky-Muller incompatibilities (BDMIs) hence weakening reproductive isolation. When BDMIs are linked to such recessive deleterious alleles the fitness benefit of introgression at these loci can facilitate introgression at the BDMIs. The extent to which this occurs depends on the strength of selection against the linked deleterious alleles and the distribution of recombination across the chromosome. Finally, surfing of neutral loci can alter the expected pattern of population ancestry, thus accounting for historical population expansion is necessary to develop accurate null genomic models of secondary-contact hybrid zones.

Population expansions often lead to secondary contact and hybridization with closely related species (Arntzen et al., 2017; Duvernell et al., 2019; Szymura, 1976; Weir and Schluter, 2004). This is most commonly reported for population expansions out of separate glacial refugia, which has led to many of the hybrid zones we observe today across both aquatic and terrestrial habitats (April et al., 2013; Avise et al., 1998; Duvernell et al., 2019; Gao et al., 2015; Haffer, 1969; Hewitt, 1999; Licona-Vera et al., 2018; Taberlet et al., 1998; Weir and Schluter, 2004). The well-documented association between invasive species and hybridization provides another example in which the rapid expansion of previously isolated species has led to widespread genetic mixing (Edmonds et al., 2004).

Range expansion can have significant evolutionary and ecological consequences (Edmonds et al., 2004; Excoffier et al., 2009). One important genetic consequence of range expansion is an increase in the fixation rate of alleles, including deleterious ones, at the range edge due to an increase in genetic drift. This phenomena, termed "gene (allele) surfing" by Klopfstein et al. (2006), was first characterized by Edmonds (2004). While all alleles, beneficial, neutral, and deleterious can surf, surfing of deleterious alleles can lead to a substantial reduction in population mean fitness at the range edge, termed "expansion load" (Peischl et al., 2013), and may even limit the extent of range expansion (Peischl and Excoffier, 2015).

Here we investigate the effect of post-expansion allele surfing on hybridization at secondary contact. Upon hybridization, deleterious alleles that have increased in frequency on the edge of one parental range can be masked by their beneficial counterparts present in the other parental population. Such masking may result in substantial heterosis (hybrid fitness advantage over the parental lineages) and alter the introgression dynamics that ultimately determines whether the diverged lineages could become independent evolutionary trajectories or not. Here we explore the effect of parental population range expansion before secondary contact on the fitness regime in hybrids both initially (among F1 crosses) and on the long-term consequences for introgression.

There are a range of possible evolutionary outcomes following secondary contact of closely related lineages: (1) fusion of the hybridizing lineages, (2) emergence of a hybrid species or,

(3) the maintenance and strengthening of the species boundary. Between these three distinct outcomes is the formation of a hybrid zone, a geographical region where divergent lineages hybridize (Barton, 2001; Barton and Hewitt, 1985; Endler, 1977; May et al., 1975; Slatkin, 1973). Among hybrid zones, there is extensive variation in the direction and strength of selection as well as the rate at which reproductive isolation evolves within them (Barton and Hewitt, 1985; Roux et al., 2016; Servedio and Noor, 2003). Models of hybrid zones typically assume that gene flow and selection are at equilibrium (Barton and Gale, 1993; Barton and Hewitt, 1989; Endler, 1977; May et al., 1975; Slatkin, 1973). A history of parental population expansion prior to secondary contact violates such assumptions. Hence, understanding the effect of allele surfing on hybridization will fill a gap in our understanding of speciation.

Theoretical Background

Models of Secondary Contact

The tension-zone model (Barton, 1979; Barton and Hewitt, 1989) explicitly examines the impact of population structure on secondary-contact. Ancestral populations are initially separated by a geographical barrier. Secondary-contact of the parental populations following the removal of this barrier results in the formation of genetic clines due to the balance of migration and selection. The shape and dynamics of the resulting clines hence depend on the nature of selection and introgression at each genetic locus. Motivated by analysis of transects of hybrid zones and the observation of concordant genetic clines along these transects (Barton and Hewitt, 1985), the classic tension-zone model assumes populations are arrayed along a linear habitat. Here we consider a variant of this model where parental populations are initially isolated at opposite ends of this linear habitat separated by a wide geographic barrier and must undergo range expansion prior to coming into secondary contact.

The selection regimes that are often considered within hybrid populations include the null case of neutral loci and universal selection against deleterious alleles. In addition, there is se-

lection specifically against hybrid genotypes as a result of underdominance or negative epistasis in the form of Bateson-Dobzhansky-Muller incompatibilities (BDMIs). In the model presented below we will focus on the dynamics at neutral, universally deleterious, and pairs of epistatic BDMI loci. Our model will contrast two different types of BDMI's, a dominant BDMI for which all recombinant genotypes are less fit, affecting both F1 and later hybrids, and a recessive BDMI for which breakdown is only experienced in F2, backcross, and later crosses (Turelli and Orr, 2000).

Range Expansion and Allele Surfing

Our expectations for the effect of range expansion, model design, and parameter choices are grounded in the results of Peischl and colleagues (Peischl et al., 2013; Peischl and Excoffier, 2015; Peischl et al., 2015). Using a combination of individual-based simulations and analytical results, they modeled the decline in population mean fitness on the edge of an expanding range as a result of the accumulation of deleterious mutations by drift. While we do not analyze the dynamics of edge mean fitness directly, this quantity is directly related to heterosis observed among F1 hybrids formed by crossing individuals at the edge of two parental ranges. Peischl and Excoffier (2015), as in our model below, consider deleterious (partially) recessive mutations initially at mutation-selection balance in the range core. As expansion progresses, population mean fitness initially declines rapidly as a result of the fixation of segregating variation. This rapid decline is followed by a period of gradual decline as a result of the continued fixation of new mutations (Peischl and Excoffier, 2015). Analogously we expect F1 heterosis to initially increase rapidly as a function of the distance travelled by parental populations prior to secondary contact, followed by a continued increase in heterosis at a reduced rate.

Peischel and Excoffier (2015) found that the decline in mean fitness on the range edge is predominantly the result of alleles with small to intermediate effects for which $Ns \leq 2$. In addition to the strength of selection (s) and the population size (N), the accumulation of mutations is also impacted by selective interference between non-recombining loci and hence by the pattern

of recombination and segregation across the genome (Peischl et al., 2015). Specifically, they compare expansion dynamics of a completely asexual population to the case of 20 freely-segregating but otherwise non-recombining chromosomes. While segregation has only mild effects on mean fitness on the range edge itself, it reduces selective interference among chromosomes and allows beneficial core alleles to spread more readily toward the range edge, hence narrowing the region of reduced fitness on the range edge.

Introgression and Heterosis

Introgression in the tension-zone model (Barton, 1979; Barton and Hewitt, 1989) is characterized by clines in allele frequency across the transect. Relatively insensitive to the nature of selection against hybrids these clines are expected to be sigmoidal in shape at selection-migration balance with allele frequency at location x given by:

$$p(x) = \frac{1}{1 + e^{-4d(x-c)}}.$$

Parameterized in this manner, the transition of the allele frequency from fixed in one parental population to absent in the other population is gradual and centered around the location c. The slope of the cline at c is given by d from which the width of the cline is defined as $W = 1/d$. The transient dynamics of allele frequency at neutral and selected loci are also expected to be sigmoidal with a time dependent center and width. In contrast to selected loci, the neutral alleles will continue to introgress eventually equilibrating at an expected frequency of 0.5 across the range.

Both the magnitude and heterogeneity of recombination rate are known to shape the extent of introgression across the genome. Introgression is usually suppressed in regions of low recombination rate due to genetic linkage to loci that reduce hybrid fitness (Aeschbacher et al., 2017; Janoušek et al., 2015; Juric et al., 2016; Kirkpatrick, 2017; Martin et al., 2019; Navarro and Barton,

2003; Noor et al., 2001; Rieseberg, 2001; Schumer et al., 2017).

Methods

Allele surfing and range expansion are complex eco-evolutionary processes (Peischl and Excoffier, 2015; Peischl et al., 2015). We begin here by describing the genetics, life cycle, and geography of our eco-evolutionary model of secondary contact. Specifically, we consider secondary contact between two populations that were initially isolated in two distant refugia (of $n_{core} = 5$ demes each) followed by a subsequent period of range expansion before finally coming into secondary contact (Figure 1).

The life cycle of the species is characterized by discrete non-overlapping generations with three stages: population census, reproduction, and migration. Selection occurs during reproduction such that an individual with genotype i reproduces at a density-dependent rate, producing a Poisson distributed number of gametes with mean $2\left(1+\rho\left(1-\frac{N_T}{K}\right)\right)V_i$, which combine randomly to produce diploid offspring. Here $\rho = 0.01$ is the intrinsic growth rate, N_T is the total number of individuals in the local population, and $K = 200$ is a constant determining the strength of density-dependence in each deme. Finally, V_i is a measure of the viability of genotype i.

An individual's genotype consists of $n_N = 80$ neutral loci, $n_D = 100$ 'selected' loci experiencing purifying selection, and $n_B = 4$ two-locus BDMI pairs, and $n_{FD} = 12$ neutral loci that are artificially forced to exhibit fixed differences between parental populations as in the classic tension-zone model for a total of $n_{Loci} = n_N + n_D + 2n_B + n_{FD} = 200$ loci. All loci are assumed to be biallelic and are arranged randomly along the genome. Selected loci are either fully recessive $h_D = 0$ or partially recessive $h_D = 0.25$. We examine three values of directional selection $s_D = 0.001, s_D = 0.003$, and $s_D = 0.005$, which correspond to values of Ks, a quantity equivalent to Ns when $V_i = 1$, of 0.2, 0.6 and 1 respectively. Hence the strength of selection is weak to intermediate across the parameter range explored such that all selected loci are expected to experience extensive surfing during range expansion as predicted by Peischl and Excoffier (2015).

For reference purposes we also examine the special case of neutral evolution ($s_D = 0.0$).

In addition to the selected loci, we consider epistatic selection between pairs of BDMI loci. Following the model and notation of Turelli and Orr (2000), for each pair of BDMI loci, B and C, we assume the parental populations are fixed for either the bbCC or BBcc genotype. The BDMI between the pair of loci B and C is characterized by three specific incompatibilities: H_0 the fitness of the double heterozygous genotypes (BbCc), H_1 the fitness of the heterozygous-homozygous genotypes (bbCc, Bbcc), and H_2 the fitness of the double homozygous genotype (bbcc). Breakdown in F1 hybrids is determined by the value of H_0 whereas F2 and backcross hybrids are affected by all three forms of incompatibilities. We consider two different types of BDMI's: "dominant BDMIs" for which $H_0 = H_1 = H_2 = 1 - s_B$ and "recessive BDMIs" where $H_0 = H_1 = 1$ and $H_2 = 1 - s_B$. We consider two strengths of epistatic selection at the BDMI loci: weakly incompatible case ($s_B = 0.01$) and strongly incompatible case ($s_B = 0.05$). See Table 1 for a summary of model parameters.

Viability across loci is multiplicative with a baseline viability of V_0 such that the viability of genotype i is given by the product:

$$V_i = V_0 * \underbrace{V_B}_{\text{BDMI}} * \underbrace{\prod_{j=1}^{n_N+n_{FD}} 1}_{\text{neutral}} * \underbrace{\prod_{j=1}^{n_D} \begin{cases} 1 & A_jA_j \\ 1 - h_D s_D & A_ja_j \\ 1 - s_D & a_ja_j \end{cases}}_{\text{selected}}$$

$$V_B = \prod_{j=1}^{n_B} \begin{cases} 1 & B_jB_j/c_jc_j \\ 1 & b_jb_j/C_jC_j \\ 1 - s_B & \text{otherwise} \end{cases} \quad \text{Dominant BDMI} \tag{1}$$

$$V_B = \prod_{j=1}^{n_B} \begin{cases} 1 - s_B & b_jb_j/c_jc_j \\ 1 & \text{otherwise} \end{cases} \quad \text{Recessive BDMI}$$

As there are strong eco-evolutionary feedbacks in the model and we wanted to limit the likelihood of extinction of core populations during burn-in, we set V_0 to ensure that the core populations would have an expected size of K at mutation-selection balance. Specifically, we set:

$$V_0 = \frac{1}{(1+s_D)^{n_D \int_{\frac{1}{2K}}^{1-\frac{1}{2K}} [x^2 \phi(x, s_D, h_D, \mu, n_{core} * K)] dx}} \tag{2}$$

where $\phi(x, s, h, \mu, N_e)$ is Wright's distribution for mutation-selection-drift equilibrium. We initialize the simulations assuming $N_e = n_{core} * K$ as an approximation for the effective population size in the range core. As shown in Figure S1, burn-in simulations indicate that this definition of V_0 provides an equilibrium deme size near K.

The absolute fitness, W_i, and relative fitness, w_i, of an individual with genotype i is determined by its viability in the following manner:

$$\begin{aligned} W_i &= \left(1 + \rho\left(1 - \frac{N_T}{K}\right)\right) V_i \\ w_i &= V_i / \bar{V} \end{aligned}$$

where $\bar{V}$ is the mean population viability. Even though the absolute fitness, W_i, depends on the local population size, N_T, an individual's relative fitness, w_i, does not. We will use V_i as a proxy for fitness throughout, as it is proportional to absolute fitness and determines relative fitness.

Parents of genotype i produce a random Poisson-distributed number of gametes with mean $2W_i$. Given the previously identified importance of recombination to both the dynamics of expansion (Peischl et al., 2015) and genome-wide patterns of introgression (Noor and Bennett, 2009), we consider five different recombination scenarios (Figure 1). In the first two cases the recombination rate is uniform across the genome with either free ($r = 0.5$) or constrained ($r = 0.005$) recombination between any two consecutive loci. In the third case the genome is divided into four segments of equal length with recombination occurring freely, $r = 0.5$, between them at three recombination "hot-spots" and no recombination within each segment. This third scenario

resembles the genomic structure used by Peischel et al. (2015) and can represent either highly heterogeneous recombination on a single chromosome or independent segregation among four otherwise non-recombining chromosomes. The resulting genome-wide average recombination rate in this case is given by $\bar{r} = 0.0075$. In the fourth and fifth scenarios we model the evolution of a single chromosome where the recombination rate varies continuously across the genome as given by the average pattern observed across a wide range of plant and animal taxa respectively, (see Haenel et al. (2018) *Figure 1, Table 1*). The genome-wide average recombination rate in these latter two scenarios was $\bar{r} = 0.005$ to facilitate comparison to the uniformly constrained and chromosomal segregation cases. Given this average rate, there is an expected ≈ 1 cross-over events per genome per generation, on par with expectations for many diploid organisms (Stapley et al., 2017).

In addition to recombination, gametes carry novel mutations. The per-generation per-site mutation rate, with both forward and backward mutations allowed, was set to $\mu = 10^{-4}$ at both the neutral and selected loci. To simplify the tracking of reproductive isolation we do not allow mutation at the BDMI loci. To provide an analogous neutral comparison, mutation is also suppressed at the 12 "fixed difference" loci. Following meiosis, mating is random with gametes combining randomly to produce diploid offspring. These offspring then migrate at a rate $m = 0.01$ to neighbouring demes as described in more detail below. Following migration, each population is censused, completing the generation.

We consider the demographic model depicted in Figure 1, modelling first evolution in allopatry of populations A and B, followed by their expansion, and finally introgression between the two populations initially occupying opposite ends of a finite linear "stepping-stone" habitat. The model was coded in C++ and proceeds through the following stages *(program and simulation results are deposited on Dryad)*.

Initialization: The simulations begin by initializing the habitat with the left-most n_{core} demes occupied by population A and the right most n_{core} populations occupied by population B. Individuals migrate among demes such that they move to the left (right) neighbouring deme with

probability $m/2$ ($m = 0.01$ is used throughout) with reflective population boundaries. These demes will be referred to as the core of population A and B respectively. We initialize each core deme with K individuals, drawing the initial allele frequencies at the selected loci using Wright's distribution (see Equation (2)). This initialization is an approximation to the steady state, ensuring that the burn-in phase efficiently reaches equilibrium.

Burn-in: For the core populations to reach the true eco-evolutionary equilibrium, we begin by simulating evolution for $10,000$ burn-in generations. To test the quality of the burn-in we evaluate the dynamics of the mean viability $\bar{V}$ and the population size of each deme (see Figure S1).

Expansion: The third stage of the simulations is characterized by the expansion of the parental populations A and B while remaining allopatric. Expansion proceeds for 2000 generations. Each generation individuals are allowed to migrate to the nearest neighbouring demes in a stepping stone manner. As described above the geographic extent (the absolute x-axis distance covered) over the course of the 2000 generations of expansion varies stochastically across simulations. Although artificial, constraining expansion to a fixed time rather than distance allows us to interchange parental populations increasing the power of our analyses. To facilitate analysis of geographical patterns after expansion we scale the x-axis such that the total range of populations A and B upon secondary contact is 100 (e.g., see Figure 4A-C).

To quantify the impact of allele surfing on population mean fitness, every 50 generations we measure expansion load, which is defined as the relative difference in mean fitness of range edge, the right (left) most deme, and the range core:

$$\mathcal{L} = \frac{\bar{V}_{core} - \bar{V}_{edge}}{\bar{V}_{core}} \tag{3}$$

Similarly, to quantify the effect of masking of deleterious recessive mutations in hybrids, every 50 generations we create an artificial population of 10 F1 hybrids between the range edge of parental populations A and B. We then calculate heterosis by comparing mean F1 hybrid

fitness to the mean fitness of their parents:

$$\mathcal{H} = \bar{V}_{F1} - \bar{V}_{Par} \tag{4}$$

Introgression: Following expansion we consider the dynamics of introgression between population A and B allowing all neighbouring populations to be connected by migration as shown in Figure 1. Following secondary contact, introgression proceeds for five-thousand generations. In this phase, we stop mutation ($\mu = 0$). Although artificial, this ensures that all results observed during introgression are a result of expansion history and not de novo mutation during introgression itself.

Following secondary contact, we can use the allele frequencies at the n_N neutral loci that are not constrained to be fixed to define a measure of population ancestry based on the hybrid index. Adapting the likelihood based measure developed by Buerkle (2005), let the allele frequency at neutral locus i in the core of population A (left-most deme) be r_i and the allele frequency in the core of population B (right-most deme) be s_i. Then the probability of observing focal allele given a hybrid index of h is given by:

$$\tilde{p}_i(h) = hr_i + (1-h)s_i \tag{5}$$

Given the observed allele frequency in deme d, p_d, we calculate the hybrid index of deme d by maximizing the likelihood:

$$\mathcal{L}(h_d|\vec{p}_d) = \prod_{l=1}^{L} \binom{2N_d}{2N_d p_{d,l}} \tilde{p}(h_d)^{2N_d p_{d,l}} \left(1 - \tilde{p}(h_d)\right)^{2N_d(1-p_{d,l})} \tag{6}$$

where N_d is the population size of deme d. As with the individual hybrid index developed by Buerkle (2005), h_d accounts for both fixed and non-fixed differences at neutral loci between core demes. As is common in empirical studies however, we restrict our inference to neutral loci exhibiting a difference in allele frequency between the core populations of $Abs[r_i - s_i] \geq 0.3$ (Corbett-Detig and Nielsen, 2017; Matute et al., 2020; Wang et al., 2021).

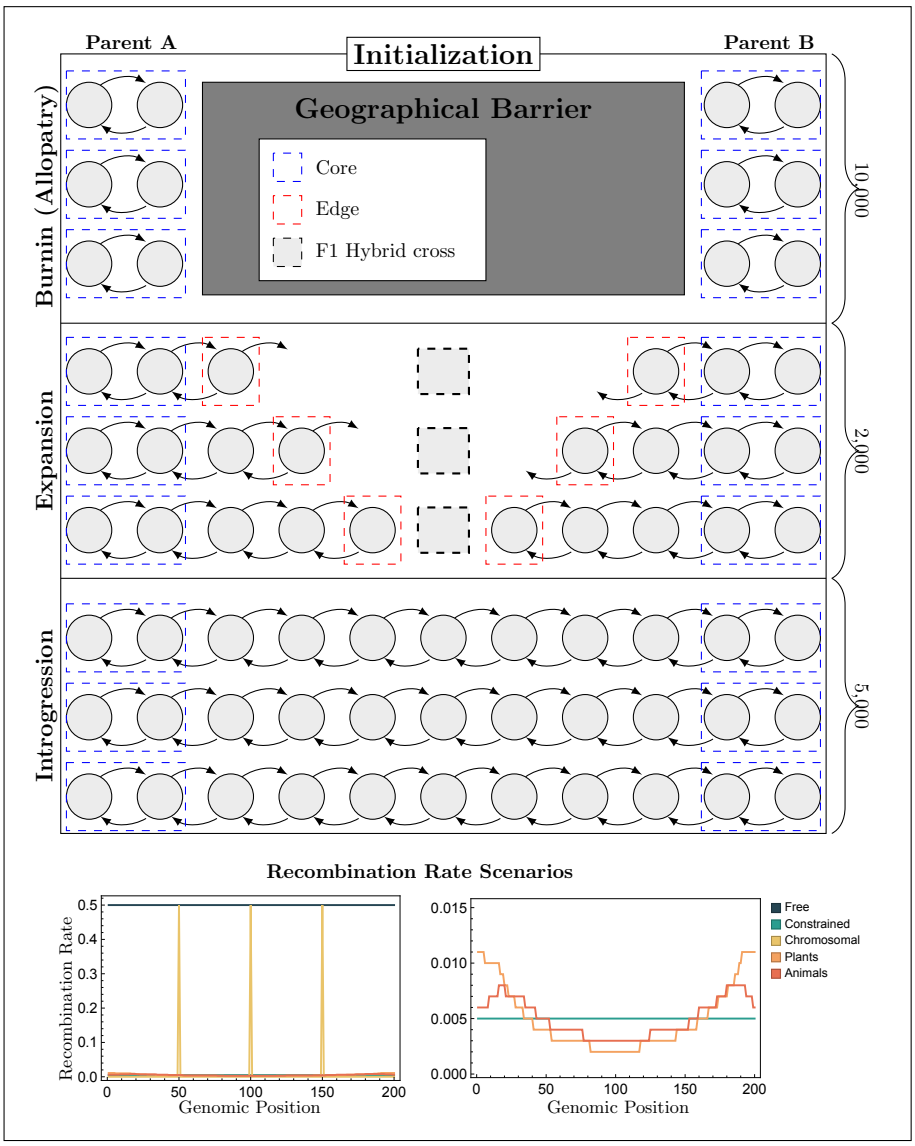

Figure 1: **Schematic of Model Design**. Top: Individual-based simulation model proceeds in three phases: 1) evolution in allopatry, 2) parental range expansion, and 3) secondary contact and introgression. Geography is modelled as a 1D stepping-stone habitat. Occupied demes (gray circles of which only a subset are shown) are connected via migration. Time is measured in units of generations of allopatry (10,000), generations of expansion (2,000), or generations of introgression (5,000) respectively. During range expansion parental edge populations are crossed every 50 generations to create F1 hybrids. Bottom left hand panel: The five recombination rate scenarios shown on the same axis as a function of genomic position from locus 1 to locus 200. Bottom Right: Variability in recombination rate under the "plants" and "animals" scenario compared to uniform constrained recombination.

Background Selection s_D	Background Dominance h_D	BDMI s_B (h_B)	Recombination $\bar{r}$
Neutral: 0.0	**Recessive: 0**	**No BDMI: 0 (NA)**	Free: 0.5
Weak: 0.001	Partially Recessive: 0.25	Recessive Weak: 0.01 (0)	**Constrained: 0.005**
Medium: 0.003		Dominant Weak: 0.01 (1)	Chromosomal: 0.0075
Strong: 0.005		Recessive Strong: 0.05 (0)	Plants: 0.005
		Dominant Strong: 0.05 (1)	Animals: 0.005

Table 1: **Parameter conditions used in individual-based simulation**. Individual based simulations were run in a full factorial design with 10 replicate simulations per treatment. The genomic location of neutral, background, and BDMI loci vary randomly across treatments but remained constant across the 10 replicates within each treatment. For all simulations the following parameter values remained fixed: $\mu = 0.0001$, $m = 0.01$, $K = 200$, $\rho = 0.1$, $n_D = 100$, $n_N = 80$, $n_{FD} = 12$, $n_B = 4$ for a total of 200 loci. Unless otherwise specified the parameters used are given in bold. Recombination rate scenarios are shown graphically in Figure 1.

Results

The core parental populations A and B reach eco-evolutionary equilibrium during the burn-in period (Figure S1). As expected during the subsequent range expansion phase, allele surfing greatly increased the frequency of deleterious mutations on the range edge and hence the number of fixed differences between the edges of population A and B (Figure 2A). The fixation of deleterious mutations on the range edge results in substantial expansion load as measured by Equation (3) (Figure S2). The selection and dominance coefficient of the deleterious alleles at the selected loci are important determinants of the extent of expansion load. In contrast, the dynamics of expansion load are insensitive to the recombination rate scenario, with nearly identical accumulation of load even in the case of complete linkage with chromosomal segregation.

As the number of fixed differences between the range edges of populations A and B increases so too does the observed heterosis among F1 hybrids created in crosses of edge populations. The temporal dynamics of F1 heterosis (Figure 2B) follow our expectation based on the previously characterized dynamics of fitness on the edge of expanding population (Peischl and Excoffier, 2015). The extent of heterosis that results from the masking of (partially) recessive deleterious mutations that have surfed to fixation or near fixation on the range edge of one parental popula-

tion but not the other depends both on the strength of selection on the deleterious alleles, s_D, and the dominance of these alleles, h_D (see Figure 2B). As expected, dominant BDMIs decrease the fitness of F1 hybrids, yet heterosis from masking deleterious recessive mutations can partially or even completely compensate for the deleterious fitness effects of these incompatibilities (Figure 2C). Finally, heterosis from the masking of deleterious mutations is relatively insensitive to the recombination scenario (Figure 2D).

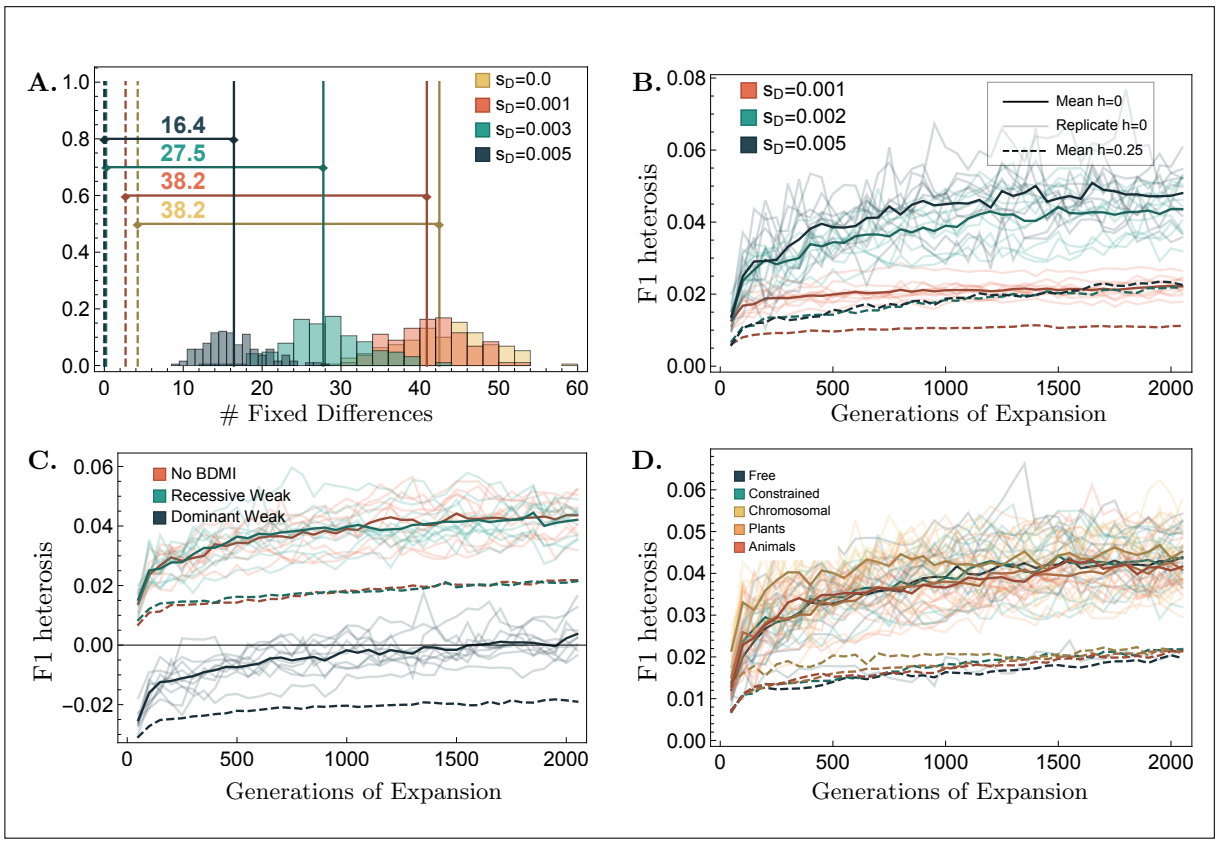

Figure 2: **Impact of allele surfing on F1 heterosis during range expansion**. Panel A: Number of selected loci fixed for different alleles between parental core populations at the end of evolution in allopatry (mean: dashed vertical line) verses parental edge populations at the end of expansion (histogram, mean: solid vertical line) for different strengths of selections. Panels C-D: F1 Heterosis as measured by Equation (4) in crosses between parental edge populations during range expansion. Dark solid curves give mean dynamics averaged across individual replicates (solid light curves) when deleterious loci are recessive $h_D = 0$. Dark dashed lines give mean dynamics for partially recessive case $h_D = 0.25$. Panel B: Effect of the strength of selection (case of $s_D = 0$ not shown) exhibits no heterosis as expected. Panel C: Effect of BDMI presence/absence and dominance shown here for the case of a weak BDMI $s_B = 0.01$. Panel D: Effect of recombination scenario. Unless specified otherwise parameters are given by boldface values in Table 1.

The dynamics of population mean fitness during range expansion and upon secondary contact are shown in Figure 3. As with F1 heterosis, upon secondary contact there is a substantial increase in fitness in the hybrid zone, which expands from the former range edge toward the core as wild-type alleles masking accumulated deleterious alleles introgress. While the presence

of BDMIs does not prevent the introgression of wild-type alleles, these incompatibilities result in a persistent reduction in population mean fitness at the point of secondary contact. Relative to dominant BDMIs, recessive incompatibilities result in a shallower and more diffuse reduction in fitness (Figure 3B and C).

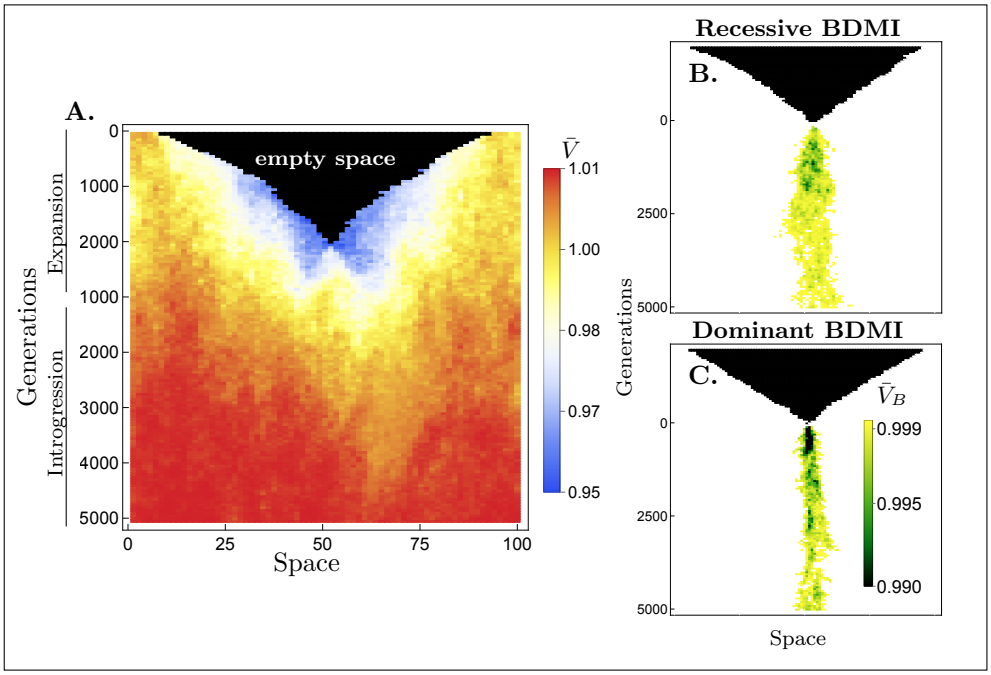

Figure 3: **Dynamics of mean fitness and the effect of the BDMIs upon secondary contact.** Panel A: exemplary dynamics of mean population viability $\bar{V}$ during range expansion and introgression in the absence of a BDMI, see boldface parameters in Table 1. Fitness effect of BDMI as measured by $\bar{V}_B$ in Equation (1) for a weak recessive (panel B) and weak dominant (panel C) BDMI.

Deleterious mutations fixed on the edge of one parental population and not the other will ultimately be removed by selection following secondary contact (recall that the simulation design assumes no mutations occur following secondary contact). While the deleterious alleles persist, however, these loci will exhibit a cline in allele frequency across the hybrid zone. Unlike the sigmoidal clines observed in a tension zone (Barton, 1979), however, the shape of the allele frequency cline is affected both by introgression and historical allele surfing. If we consider all loci for which the deleterious allele has a frequency $p^A_{edge} < 0.5$ at the range edge of parental

population A and $p^{B}_{edge} > 0.5$ at the edge of parental population B, then the cline in average allele frequency across these loci peaks just to the left of the contact zone (see Figure 4A).

To quantify the transient dynamics of these allele frequency clines at each type of locus (e.g., neutral, selected, FD, BDMI) and their variation across parameter space we fit them phenomenologically with one of the following functions as determined by AIC based model selection.

$$\bar{p}_1(x) = \begin{cases} \frac{d}{b_A} Sech(a_A(x-c)) Tanh(b_A(x-c)) + e & if\ x < c \\ \frac{d}{b_B} Sech(a_B(x-c)) Tanh(b_B(x-c)) + e & if\ x > c \end{cases} \tag{7}$$

$$\bar{p}_2(x) = \frac{a}{1 + \exp\left[\frac{-4d}{a}(x-c)\right]} + e - \frac{a}{2}.$$

These model fits are defined such that the geographic center of the cline is given by c and height e, with the slope (derivative) of the cline at this point by d. Following the definition of Barton and Hewitt (1989), the width of the cline is given by $W = |\frac{1}{d}|$. For $\bar{p}_1(x)$ the value of a_A (a_B) determines the slope in allele frequency from the range core to its maximum (minimum) near the point of secondary contact as determined by the extent of allele surfing. The remaining parameters (b_A,b_B) determine the height and symmetry of the cline. By contrast, $\bar{p}_2(x)$ describes a sigmoidal function, as observed when parental population have fixed differences. The height of this symmetric sigmoid is given by a. The functions $\bar{p}_1$ versus $\bar{p}_2$ were developed to match the shape of the allele frequency clines in the case of the neutral and selected versus the fixed difference and BDMI clines respectively (see Figure 4A) although neither model is preferred exclusively in either of these cases across all parameter conditions.

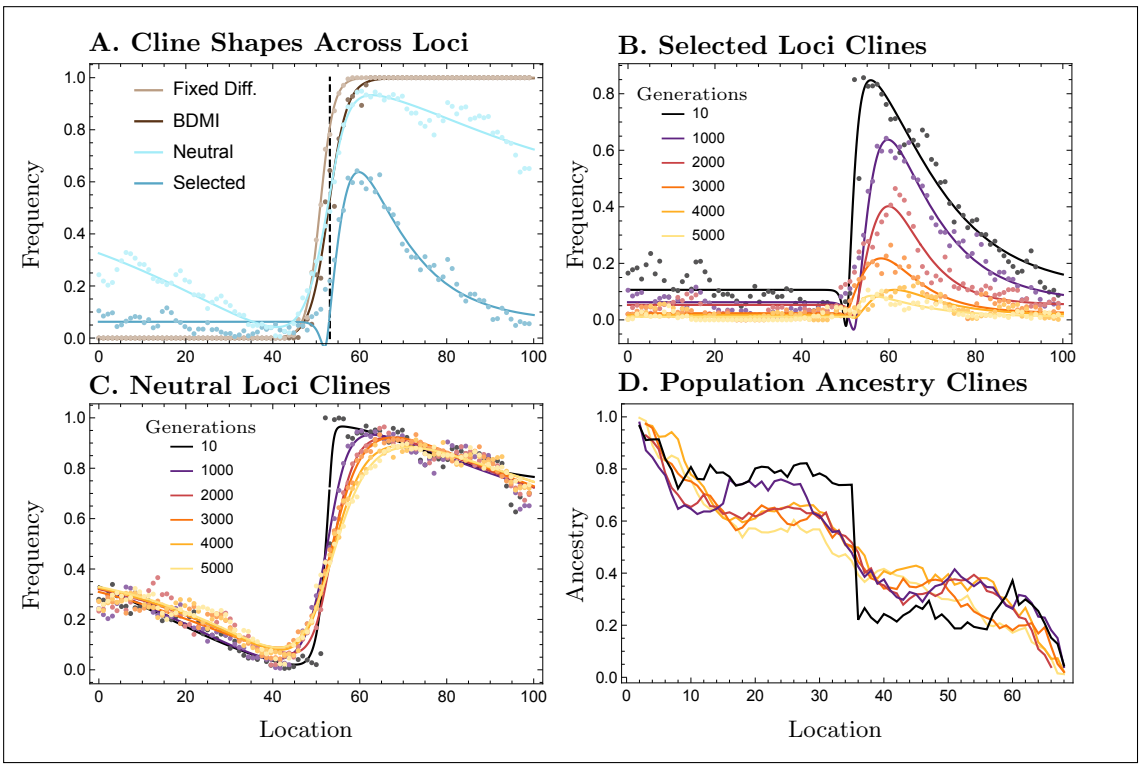

Figure 4: **Clinal dynamics and population ancestry.** Clines in average allele frequency are calculated by averaging across loci of a specific type (e.g., neutral, selected, fixed difference, BDMI) for which $p^A_{\text{edge}} < 0.5$ and $p^B_{\text{edge}} > 0.5$. Panel A: Shape of the allele frequency cline for each locus type 1000 generations after secondary contact. Brown versus blue clines exemplify difference in fit between Sech ($\bar{p}_1$) and sigmoidal ($\bar{p}_2$) curves given by Equation (7). Panel B: Temporal dynamics of clines at the selected loci throughout introgression. Panel C: Temporal dynamics of allele frequency clines at neutral loci following secondary contact. Panel D: Temporal dynamics of the cline in population ancestry, as defined by hybrid index in Equation (6). See Panel C for colour legend. Parameters are given by boldface values in Table 1.

We next focus on the dynamics of cline width at the selected and BDMI loci. As introgression proceeds, the width of the cline increases at both the selected (Figure 5A) and BDMI loci (Figure 5B), although less so for the the BDMI loci, as expected. Due to selection favouring the introgression of wild-type mutations, the cline width at the selected loci increases with the strength of selection s_D for the parameters explored. Interestingly, the cline width at the selected loci is less than that of the corresponding neutral loci. This is a result of the low frequency of the deleterious alleles in the parental range core and the subsequent spread of alleles from the range core

toward the hybrid zone following secondary contact. Essentially, selective recovery from deleterious allele surfing occurs from two directions, from secondary contact with the other range edge population and from the core of the same population. For further details see Appendix A.

The cline width at the BDMI loci decreases with the strength of selection against the BDMI incompatibilities s_B, with the width of BDMI clines less than than the clines at the neutral fixed difference loci. Cline width at the BDMI loci is also impacted by the strength of linked selection against deleterious alleles (s_D see Figure 5C), with strong linked selection facilitating introgression, at least over short time scales (0-1000 generations of introgression). Finally, the cline width at the BDMI loci is impacted by the distribution of recombination across the chromosome(s). The more uniform recombination is the more BDMI loci are able to introgress. In the case of chromosomal segregation, introgression at BDMI loci and hence cline widths are limited.

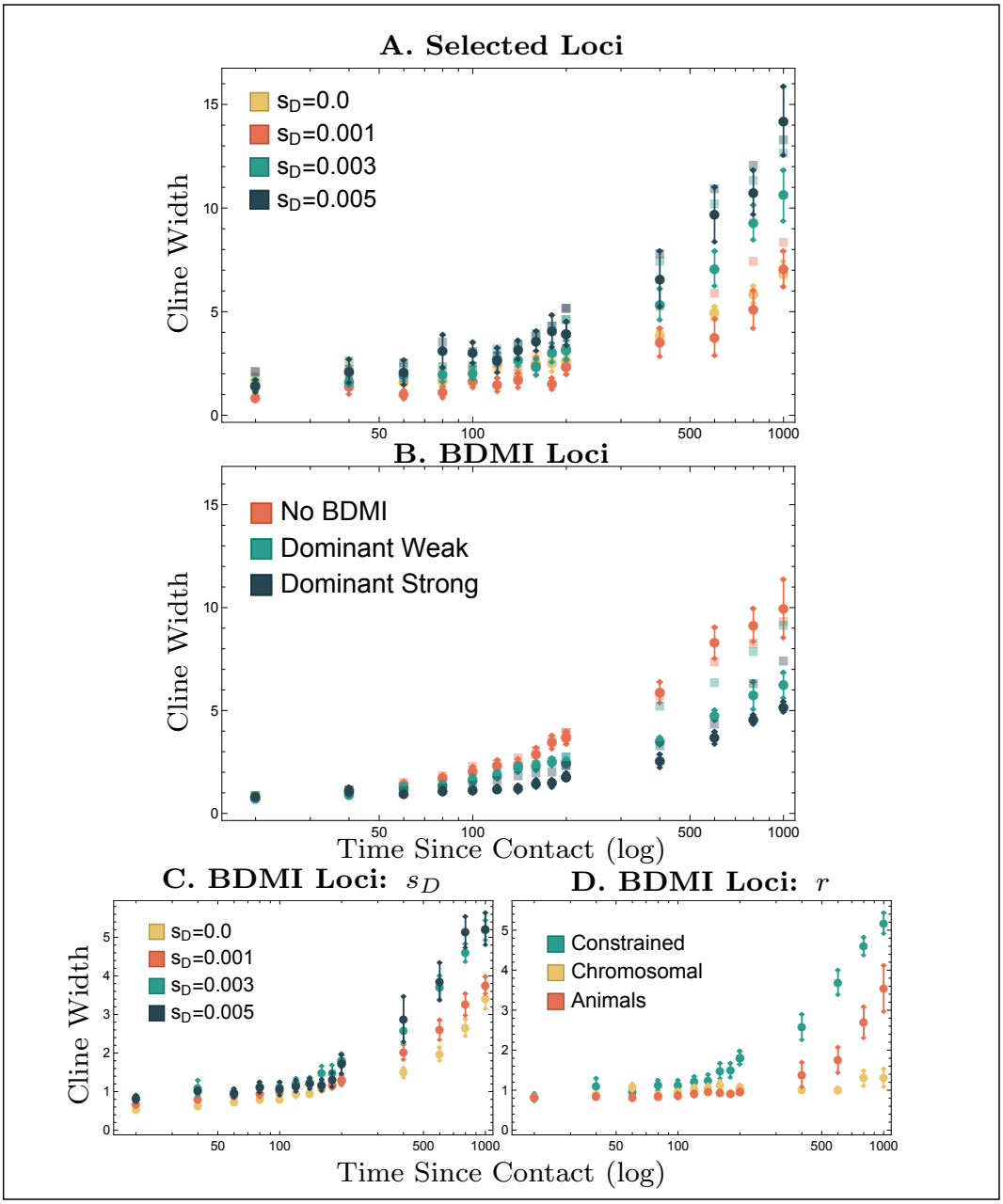

Figure 5: **Early temporal dynamics and determinants of cline width at selected and BDMI loci**. Panel A: temporal dynamics of cline width at selected loci. Effect of the strength of selection s_D on cline width of selected loci. Dark points give mean width across replicates with ± 1 SE whiskers. Opaque squares show mean cline width at neutral loci in the same genome. Panel B: Cline width at BDMI loci dark points show mean width with ± 1 SE whiskers. Opaque squares show mean for fixed difference loci in the same genome. Panel C: effect of the strength of selection on linked deleterious loci on cline width of BDMI loci. Panel D: effect of recombination scenario on cline width at BDMI loci (note plants case not shown). Unless specified otherwise parameters are given by boldface values in Table 1.

Finally, we characterize the population ancestry across the geographic range following secondary contact using the population-wide hybrid index at the n_N neutral loci that were not constrained to be fixed (Equation 6). As with the allele frequency clines at the neutral loci, the cline in population ancestry is not sigmoidal as would be expected in the tension-zone model. Affected by allele surfing during range expansion, the cline in population ancestry is gradual, showing a less marked shift in ancestry at the time of secondary contact (black line in Figure 4D). This gradual change in ancestry across the range persists over the course of introgression obscuring the location of initial contact, making the hybrid zone appear older than it is. These results suggest that incorporating historical range expansion is important for the accurate reconstruction of the biogeography of hybrid zones.

Discussion

Here we investigated the effect of parental population range expansion and allele surfing prior to secondary contact on hybridization and introgression. As found by previous models of range expansion (e.g., Peischl et al. (2015), Peischl and Excoffier (2015), and Peischl et al. (2013)), we find that the edge of two expanding parental populations can accumulate a significant number of deleterious mutations and hence expansion load. If these deleterious alleles are at least partially recessive, hybridization of range edge populations results in extensive heterosis, as wild-type alleles present in one population mask the effects of deleterious mutations accumulated in the other. This masking effect can compensate for the reduction in hybrid fitness due to genetic incompatibilities (e.g. BDMI) between parental populations facilitating introgression over generations after the initial contact. We further uncovered selective interference between the loci with surfed alleles and BDMI alleles in shaping introgression across species boundaries. The strength of the interference is dependent on the selection against surfed alleles and the distribution of the recombination rate across the genome.

In this model we have considered only a small subset of the possible selection regimes acting

in post-expansion secondary-contact hybrid zones. Namely, we have not explored the evolution of loci experiencing environment-dependent or underdominant selection. In addition, we have assumed that each parental population is fixed for a particular combination of compatible BDMI genotypes. In reality the genotypes at these loci are also subject to allele surfing during expansion, which could alter the prevalence of incompatible genotypes among hybrids and the permeability of these loci during intorgression.

Hybrid zones

Our results may help explain previously reported patterns seen in hybrids and hybrid zones. One commonly observed pattern is variation among cross combinations (both within and between species) in the level of heterosis among hybrid offspring (Lowry et al., 2008). This variation has been exploited by plant breeders to develop heterotic groups for hybrid crop production (Crow, 1998). Despite its importance, heterosis remains difficult to predict in natural populations (Pickup et al., 2013). Results from the present study imply that the history of the parental populations may be key. If divergence occurs in large adjacent populations, little heterosis is expected upon contact, whereas divergence in distant locations followed by range expansion with extreme drift, such as explored here, creates genomic conditions that facilitate heterosis and introgession via the masking of deleterious recessive alleles that counteract genetic incompatibilities.

Another partially unexplained phenomenon is the relatively high frequency of introgressed regions that appear to be positively selected in hybrid populations (Corbi et al., 2018; Hamilton et al., 2013; Rieseberg et al., 1999). This could be a byproduct of adaptation in finite populations, such that some favorable alleles were previously in only one of the parental populations (Barton, 2001) or the environment may be different in the hybrid zone centre, favoring a different combination of alleles. Alternatively, it seems likely that a substantial fraction of favorable introgressed regions are due to the masking of deleterious mutations, as shown here.

The spatial distribution of population ancestry is usually modeled as a sigmoidal clines, the expected shape assuming selection-dispersal balance in the tension zone model (Barton, 1979).

However, here we found distorted shapes of genomic clines (Figure 4D) as a result of allele surfing. This pattern might be undisclosed in many empirical system if the spatial genetic sampling within parental ranges was limited. For future investigations in hybrid zones, our results emphasize the importance of broad spatial sampling centered around the core of admixture in detecting signatures and effects of allele surfing in hybrid zone ancestry dynamics.

Genomic architecture of speciation

The effects of recombination rate heterogeneity in this expansion-hybridization context alters the conditions under which speciation can be impeded by introgression. Here we find elevated introgression at the BDMI loci due to hitchhiking with flanking loci, where surfed deleterious alleles were masked by heterospecific alleles. Interestingly, such interference-facilitated BDMI introgression is modulated by the distribution of recombination rates, holding the mean magnitude of recombination rate constant.

Our observation extends the existing understanding of the effect of recombination on introgression by accounting for the distribution (in addition to the magnitude) of recombination and the interference among neutral loci, BDMI loci, and loci with surfed deleterious alleles during parental range expansion. Both allele surfing and introgression are expected to be shaped by the magnitude and distribution of recombination events across the genome. In the expansion context, constrained recombination can reduce absolute fitness at the range edge suppressing population sizes and even resulting in substantial reductions in expansion rate (Peischl et al., 2015). In the introgression context, local recombination suppression due to chromosomal rearrangements can prevent introgression and facilitate differentiation (Navarro and Barton, 2003; Noor et al., 2001; Rieseberg, 2001).

In contrast to previous results (Peischl and Excoffier, 2015), we find that recombination has relatively little impact on the accumulation of expansion load (Figure S2) and thus F1 heterosis (Figure 2D), even in the case of chromosomal segregation. The contrast between these results may indicate that only small amounts of recombination are necessary to dramatically reduce selective

interference during range expansion. However for introgression, we find that genetic linkage and the distribution of recombination have large impacts on introgression, effects that can persist for hundreds of generations after the onset of hybridization (Figure 5). Specifically, introgression at neutral loci is facilitated by selection at linked selected loci. Even the introgression at BDMI loci is elevated under stronger selection at those deleterious loci (Figure 5). On the other hand, heterogeneity in the recombination rate suppresses introgression at BDMI loci helping maintain barriers to hybridization. Therefore, the expansion-introgression dynamics could be variable in organismal systems with different level of recombination heterogeneity in the genomes. These results highlight how the recombination landscape and range expansion interact to shape genomic ancestry at species boundaries.

Hybrid speciation

Our results also have implications for long term outcomes of natural hybridization. Most obviously, our simulations indicate that levels and heterogeneity of introgression may be greater than that predicted by standard hybrid zone models (Barton and Hewitt, 1985). Strong heterosis could contribute to the weakening of reproductive barriers, sustained hybrid swarms/species, or even the fusion of previously isolated populations. High level of heterosis would reduce the strength of reinforcement. This leads to the interesting prediction that reinforcement might be more likely when populations have diverged in adjacent areas rather than following extensive range expansion. This differs from the classic scenario that ignores allele surfing, in which reinforcement "completes speciation" following range expansion and contact between previously allopatric populations (Dobzhansky, 1937).

On the other hand, expansion-contact induced heterosis might sustain hybrid populations for extended period of time, allowing hybrid population to diverge from parental lineages under divergent selection (Buerkle et al., 2000). One compelling example resides in a North American wood warblers species complex, where hybrid populations existed long after post-expansion-hybridization between the parental species (Wang et al., 2021). Heterosis could sustained the

ancient hybrid population despite genetic incompatibility between the parental populations. The hybrid populations partitioned distinct climatic niche and diverged from parental lineages at genomic regions related to climatic adaptation (Wang et al., 2021). The expansion-hybridization resulted fusion, hybrid swarms, or hybrid species could be prevalent yet hidden in many extent populations with histories of expansion-admixture.

Genomic invasion

The heterotic effect of post-expansion hybridization can also facilitate introgression of native species genetic material between native and invasive species. In this case, if invasive species have undergone the most recent range expansion, deleterious alleles may have surfed to higher frequencies, facilitating introgression from stable native populations. Here we did not model invasion explicitly (asymmetry of expansion, in which the invasive species expanded its range while the native species did not). However, based on the expansion-hybridization induced heterosis and introgression our results predict, we expect masking of deleterious alleles in the invasive species following hybridization with its native congener. Such hybridization-induced heterosis could be the genomic mechanism underlying hybridization-accelerated anthropogenic invasions, including the process in which modern human replaced Neanderthals (Mallet, 2005; Mesgaran et al., 2016). Ironically, native species could potentially rescue invaders that are ecologically detrimental to them. Our result raises questions about the conservation of native populations in the face of anthropogenic invasion of closely related and rapidly-expanding invasive populations.

Acknowledgments

The Authors would like to thank Darren Irwin and Brett Payseur for their helpful feedback in the development of this project and to Matt Pennell for his helpful comments on the manuscript. AM was supported by a fellowship from the University of British Columbia and by the EEB department Postdoctoral Fellowship from the University of Toronto and support from the Natu-

ral Sciences and Engineering Research Council of Canada (PGS D 331015731 to SW and NSERC RGPIN-2016-03711 to SPO).

Literature Cited

Aeschbacher, S., J. P. Selby, J. H. Willis, and G. Coop. 2017. Population-genomic inference of the strength and timing of selection against gene flow. PNAS 114:7061–7066.

April, J., R. H. Hanner, A.-M. Dion-Côté, and L. Bernatchez. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Mol. Ecol. 22:409–422.

Arntzen, J. W., W. de Vries, D. Canestrelli, and I. Martínez-Solano. 2017. Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Mol. Ecol. 26:5663–5675.

Avise, J. C., D. Walker, and G. C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. Lond. B 265:1707–1712.

Barton, N. H. 1979. The dynamics of hybrid zones. Heredity 43:341–359.

———. 2001. The role of hybridization in evolution. Mol. Ecol. 10:551–568.

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

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

———. 1989. Adaptation, speciation and hybrid zones. Nature 341:497–503.

Buerkle, C. A. 2005. Maximum-likelihood estimation of a hybrid index based on molecular markers. Molecular Ecology Notes 5:684–687.

Buerkle, C. A., R. J. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. Heredity 84:441–451.

Corbett-Detig, R., and R. Nielsen. 2017. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. PLOS Genetics 13:e1006529.

Corbi, J., E. J. Baack, J. M. Dechaine, G. Seiler, and J. M. Burke. 2018. Genome-wide analysis of allele frequency change in sunflower crop–wild hybrid populations evolving under natural conditions. Mol. Ecol. 27:233–247.

Crow, J. F. 1998. 90 years ago: The beginning of hybrid maize. Genetics 148:923–928.

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

Duvernell, D. D., E. Westhafer, and J. F. Schaefer. 2019. Late Pleistocene range expansion of North American topminnows accompanied by admixture and introgression. J. Biogeogr. 46:2126–2140.

Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. PNAS 101:975–979.

Endler, J. A. 1977. Geographic Variation, Speciation, and Clines. Princeton University Press, Princeton, New Jersey.

Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic Consequences of Range Expansions. Annu. Rev. Ecol. Evol. Syst 40:481–501.

Gao, Y.-D., Y. Zhang, X.-F. Gao, and Z.-M. Zhu. 2015. Pleistocene glaciations, demographic expansion and subsequent isolation promoted morphological heterogeneity: A phylogeographic study of the alpine Rosa sericea complex (Rosaceae). Scientific Reports 5:11698.

Haenel, Q., T. G. Laurentino, M. Roesti, and D. Berner. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol. Ecol. 27:2477–2497.

Haffer, J. 1969. Speciation in Amazonian Forest Birds. Science 165:131–137.

Hamilton, J. A., C. Lexer, and S. N. Aitken. 2013. Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis x P. glauca) hybrid zone. New Phytol. 197:927–938.

Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68:87–112.

Janoušek, V., P. Munclinger, L. 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.

Juric, I., S. Aeschbacher, and G. Coop. 2016. The Strength of Selection against Neanderthal Introgression. PLOS Genetics 12:e1006340.

Kirkpatrick, M. 2017. The Evolution of Genome Structure by Natural and Sexual Selection. The Journal of Heredity 108:3–11.

Klopfstein, S., M. Currat, and L. Excoffier. 2006. The Fate of Mutations Surfing on the Wave of a Range Expansion. Mol. Biol. Evol. 23:482–490.

Licona-Vera, Y., J. F. Ornelas, S. Wethington, and K. B. Bryan. 2018. Pleistocene range expansions promote divergence with gene flow between migratory and sedentary populations of Calothorax hummingbirds. Biol. J. Linn. Soc. 124:645–667.

Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Philos. Trans. R. Soc. B 363:3009–3021.

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

Martin, S. H., J. W. Davey, C. Salazar, and C. D. Jiggins. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLOS Biology 17:e2006288.

Matute, D. R., A. A. Comeault, E. Earley, A. Serrato-Capuchina, D. Peede, A. Monroy-Eklund, W. Huang, C. D. Jones, T. F. C. Mackay, and J. A. Coyne. 2020. Rapid and Predictable Evolution of Admixed Populations Between Two Drosophila Species Pairs. Genetics 214:211–230.

May, R. M., J. A. Endler, and R. E. McMurtrie. 1975. Gene Frequency Clines in the Presence of Selection Opposed by Gene Flow. Am. Nat. 109:659–676.

Mesgaran, M. B., M. A. Lewis, P. K. Ades, K. Donohue, S. Ohadi, C. Li, and R. D. Cousens. 2016. Hybridization can facilitate species invasions, even without enhancing local adaptation. PNAS 113:10210–10214.

Navarro, A., and N. H. Barton. 2003. Accumulating Postzygotic Isolation Genes in Parapatry: A New Twist on Chromosomal Speciation. Evolution 57:447–459.

Noor, M. a. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. PNAS 98:12084–12088.

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier. 2013. On the accumulation of deleterious mutations during range expansions. Mol. Ecol. 22:5972–5982.

Peischl, S., and L. Excoffier. 2015. Expansion load: Recessive mutations and the role of standing genetic variation. Mol. Ecol. 24:2084–2094.

Peischl, S., M. Kirkpatrick, and L. Excoffier. 2015. Expansion Load and the Evolutionary Dynamics of a Species Range. Am. Nat. 185:E81–E93.

Pickup, M., D. L. Field, D. M. Rowell, and A. G. Young. 2013. Source population characteristics affect heterosis following genetic rescue of fragmented plant populations. Proc. R. Soc. Lond. B 280:20122058.

Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16:351–358.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid Zones and the Genetic Architecture of a Barrier to Gene Flow Between Two Sunflower Species. Genetics 152:713–727.

Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. PLOS Biology 14:e2000234.

Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2017. Natural selection interacts with the local recombination rate to shape the evolution of hybrid genomes. bioRxiv page 212407.

Servedio, M. R., and M. A. Noor. 2003. The Role of Reinforcement in Speciation: Theory and Data. Annu. Rev. Ecol. Evol. Syst. 34:339–364.

Slatkin, M. 1973. Gene Flow and Selection in a Cline. Genetics 75:733–756.

Stapley, J., P. G. D. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja. 2017. Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences 372:20160455.

Szymura, J. M. 1976. Hybridization between Discoglossid toads Bombina bombina and Bombina variegata in southern Poland as revealed by the electrophoretic technique. J. Zool. Syst. Evol. Res 14:227–236.

Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy, and J.-F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7:453–464.

Turelli, M., and H. A. Orr. 2000. Dominance, Epistasis and the Genetics of Postzygotic Isolation. Genetics 154:1663–1679.

Wang, S., M. J. Ore, E. K. Mikkelsen, J. Lee-Yaw, D. P. L. Toews, S. Rohwer, and D. Irwin. 2021. Signatures of mitonuclear coevolution in a warbler species complex. bioRxiv page 2020.04.06.028506.

Weir, J. T., and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B 271:1881–1887.

Appendix A: Ancestry of Selected Loci

The allele frequency clines at the selected loci are characterized by non-sigmoidal clines (see Figure 4B). Namely these clines are determined primarily by loci for which the allele frequency in both core populations, A and B, are low as expected at mutation-selection balance. This is in contrast to the allele frequency clines at neutral loci, as shown in Figure 4C, for which the allele frequency differences between the core A and B tend to be largest. The temporal dynamics of these two cline types differ as well. The dynamics of the neutral clines are characterized predominantly by changes in the cline width anchored around a constant center ({c,e}). In contrast, the selected loci clines exhibit dramatic changes in their width, center, and height over time (Figure 4).

Following secondary contact, the temporal dynamics at the selected loci are determined by both the introgression of beneficial alleles from the point of secondary contact and the spread of these same beneficial alleles from the range core. As with the neutral loci, initially the introgression of alleles from the point of secondary contact increases the cline width. As beneficial alleles from secondary contact and from the range core meet, the cline width stabilizes and the height and the center of the cline begins to fall. The net result is shown in Figure 5, with the selected clines exhibiting relatively narrow clines compared to the corresponding neutral loci.

To confirm this explanation of the clinal dynamics we ran a modified version of the simulations where each selected locus is replaced with a two-locus haplotype with complete linkage ($r = 0$). The first locus of these haplotypes is the original selected locus and the second locus is a neutral locus fixed for allele 1 across the range of population A and for allele 0 across the range of population B, as with the fixed difference loci above. In this way the ancestry of beneficial alleles, whether they are from the hybrid zone or spreading from the range core, can be determined while not modifying genomic linkage. Initially the sigmoidal shape and cline width in selected locus ancestry resembles that of the neutral fixed difference loci (Figure S3). On longer time scales (1000-5000 generations post after introgession) the width of the clines in selected locus

ancestry slightly exceeds that of the fixed difference loci due to the beneficial effects of the introgressing alleles. This effect is small and only visible on longer time scales due to the relatively weak selection at these loci ($s_D = 0.03$).

Appendix B: Supplementary Figures

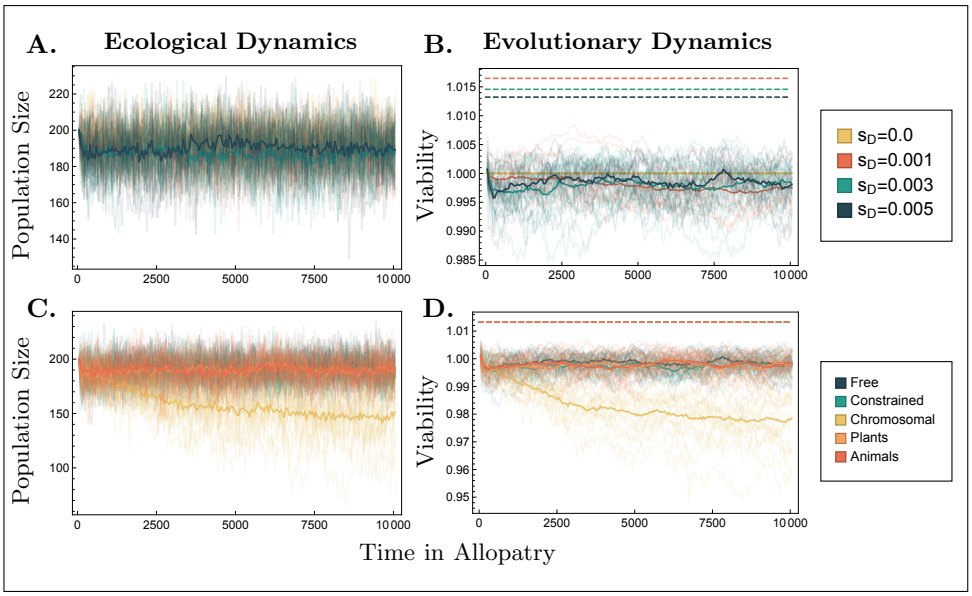

Figure S1: **Eco-evolutionary dynamics in Allopatry.** The population size (panels A and C) and mean population viability (panels B and D) of parental populations evolving in allopatry. Dashed lines in panels B and D give maximum possible viability. Top Row: Effect of selection. Bottom Row: Effect of recombination. Chromosomal scenario results in significantly reduced viability and subsequently population size due to the accumulation of mutation via Muller's ratchet. Unless specified otherwise, parameters are given by boldface values in Table 1

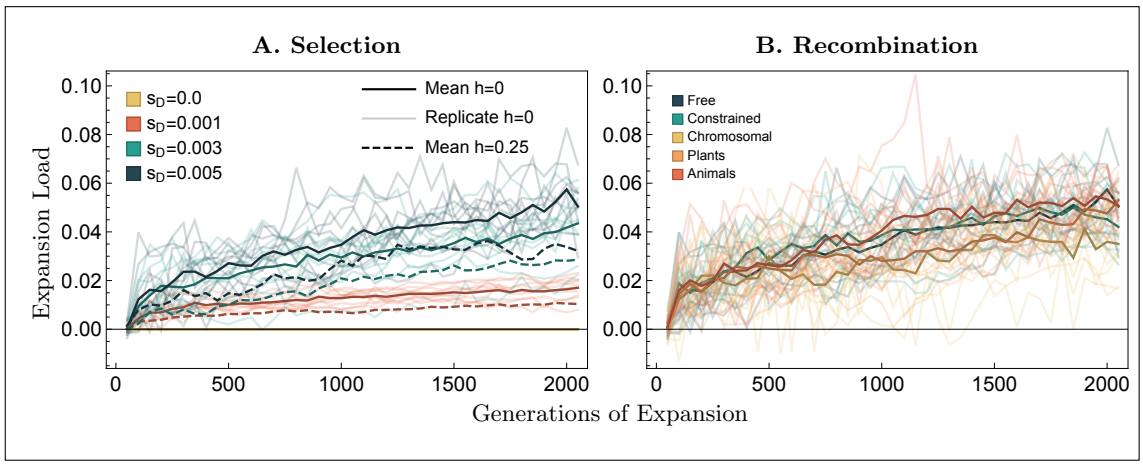

Figure S2: **Accumulation of load during range expansion.** Load as measured by Equation (3) over the course of range expansion. Panel A: Effect of the strength (color) and dominance (solid: $h_D = 0$, dashed: $h_D = 0.25$). Dark curves give mean dynamics whereas light curves give dynamics for individual replicates for (shown for $h = 0$ only). Panel B: Effect of recombination scenario on the accumulation of load during range expansion. Unless specified otherwise parameters are given by boldface values in Table 1.

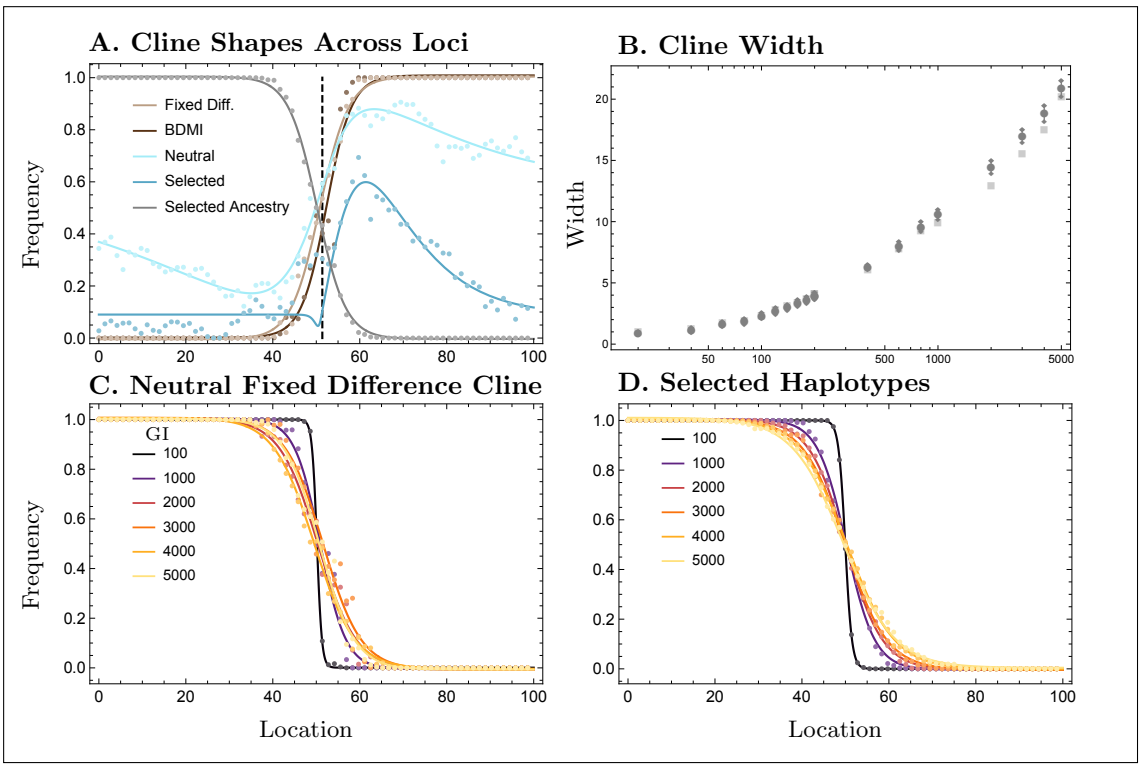

Figure S3: **Dynamics of selected loci clines with known ancestry.** Panel A: The cline shape for each locus after 1000 generations of introgression including the cline in the ancestry at the selected loci as determined by the linked neutral fixed difference (grey). Panel B: The width of the cline at the neutral fixed difference loci (light squares) versus the clines in selected locus ancestry (dark points). Panel C: Temporal dynamics in fixed difference clines fixed for the 1 allele across the range of parental population A and for the 0 allele across the range of parental population B. Panel D: Temporal dynamics in selected cline ancestry. Parameters are given by the bold face values in Table 1

.