Parental population range expansion before secondary contact promotes heterosis

Ailene MacPherson[1,h,a,c,g], Silu Wang[d,h,g], Ryo Yamaguchi[e,f,g], Loren H. Riesesberg[b,c,g], Sarah P. Otto[a,c,g]

[a] *Department of Zoology, University of British Columbia, Vancouver, Canada*
[b] *Department of Botany, University of British Columbia, Vancouver, Canada*
[c] *UBC Biodiversity Research Centre*
[d] *Integrative Biology, University of California Berkley*
[e] *Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, Tokyo, Japan*
[f] *Department of Advanced Transdisciplinary Sciences, Hokkaido University, Sapporo, Hokkaido, Japan*
[g] *Declarations of interest: none*
[h] *Contributed equally to manuscript*

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 had 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, neutral or slightly deleterious mutations drift high frequency at the front of the expansion. If these surfed deleterious alleles are recessive, they can contribute to substantial heterosis in hybrids produced at secondary contact, counteracting negative-epistatic interactions between BDMI loci and hence can deteriorate reproductive isolation. Similarly, surfing at neutral loci can alter the expected pattern of population ancestry and suggests that accounting for historical population expansion is necessary to develop accurate null genomic models in secondary-contact hybrid zones. Furthermore, this process should be incorporated in macroevolutionary models of divergence as well, since such heterosis facilitated by parental-range expansion could dampen genomic divergence established in the past.

Keywords: allele surfing, genomic clines, BDMI

Introduction

Hybrid zones are natural laboratories for understanding the genetics and evolution of reproductive isolation (Barton and Hewitt, 1985), as there is extensive variation in the direction and strength of selection among hybrid zones as well as the rate at which reproductive isolation evolves within them. One of the most abundant types of hybrid zones is that formed by secondary contact between divergent parental lineages (Barton and Hewitt, 1985). These hybrid zones are ideal for studying speciation genetics because they vary

Email address: amacp@zoology.ubc.ca (Ailene MacPherson)

in their fitness regimes (the relative fitness of the parental populations to that of the hybrids) as well as in the dynamics of reproductive isolation.

One rarely considered factor that may explain variation in outcomes upon secondary contact is the extent of range expansion of the parental populations prior to secondary contact. Shaped by geological and climatic changes, most extant secondary contact hybrid zones have undergone histories of extensive parental population range expansion before secondary contact (Arntzen et al., 2017; Szymura, 1976). For example, many were influenced by glacial cycles (Haffer, 1969; Avise et al., 1998; Bernatchez and Wilson, 1998; April et al., 2013), with parental populations diverging as they were confined to refugia and expanding as glaciers receded before finally coming into secondary contact. These range expansions can have significant genetic consequences (Edmonds et al., 2004; Excoffier et al., 2009) and leave distinct spatial genetic signatures in secondary-contact hybrid zones (Bertl et al., 2018). A history of range expansion prior to introgression may reshape hybrid zone dynamics and the genomic differentiation between parental populations. This may, in turn, alter the speciation trajectory. Despite their potential effects, the consequences of parental population range expansion is rarely considered in the speciation literature. Here we explore the effect of parental population range expansion before secondary contact on the fitness regimes in the hybrid zones, and the consequences that this has on the dynamics of introgression and reproductive isolation.

First identified by Edmonds et al. (2004) and coined "gene (allele) surfing" by Klopfstein et al. (2006), 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. A result of the repeated population bottlenecks and founder effects at the range edge, allele surfing is a complex eco-evolutionary process shaped by both the dynamics of population growth and density-dependence as well as selection, migration, and drift (Klopfstein et al., 2006). 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). This full eco-evolutionary perspective contrasts with a neutral coalescent approach which makes explicit assumptions about population demography (Bertl et al., 2018; Austerlitz et al., 1997).

Here we ask whether range expansion and the surfing of deleterious alleles influences the fitness regimes and dynamics of hybrid zones. In particular, we explore how the expansion of divergent parental populations allows the fixation of recessive deleterious mutations at their range edge prior to secondary contact. Upon hybridization, the negative fitness effects of the deleterious mutations that are fixed in one parental population can be masked by wild-type alleles from the other parental population. On a genome-wide scale, this masking can lead to substantial heterosis in F1 hybrids, facilitating introgression. The effect of allele surfing on fitness regimes at secondary contact may in turn alter the speciation trajectory of differentiated lineages (e.g., fusion, hybrid speciation, or completion of reproductive isolation).

Methods

Allele surfing and range expansion is a complex eco-evolutionary process, (Peischl and Excoffier, 2015; Peischl et al., 2015). We begin here by describing the genetics, life cycle, and fitness landscape of our eco-evolutionary model of secondary contact. Specifically, we consider secondary contact between two populations that were initially isolated in two distant refugia 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 life-history 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 $2\left(1+\rho\left(1-\frac{N_T}{K}\right)\right)V_i$. Here ρ is the intrinsic growth rate, N_T is the total number of individuals in the local population, and K is a constant determining the strength of density-dependence. Finally, V_i is a measure of the viability of genotype i.

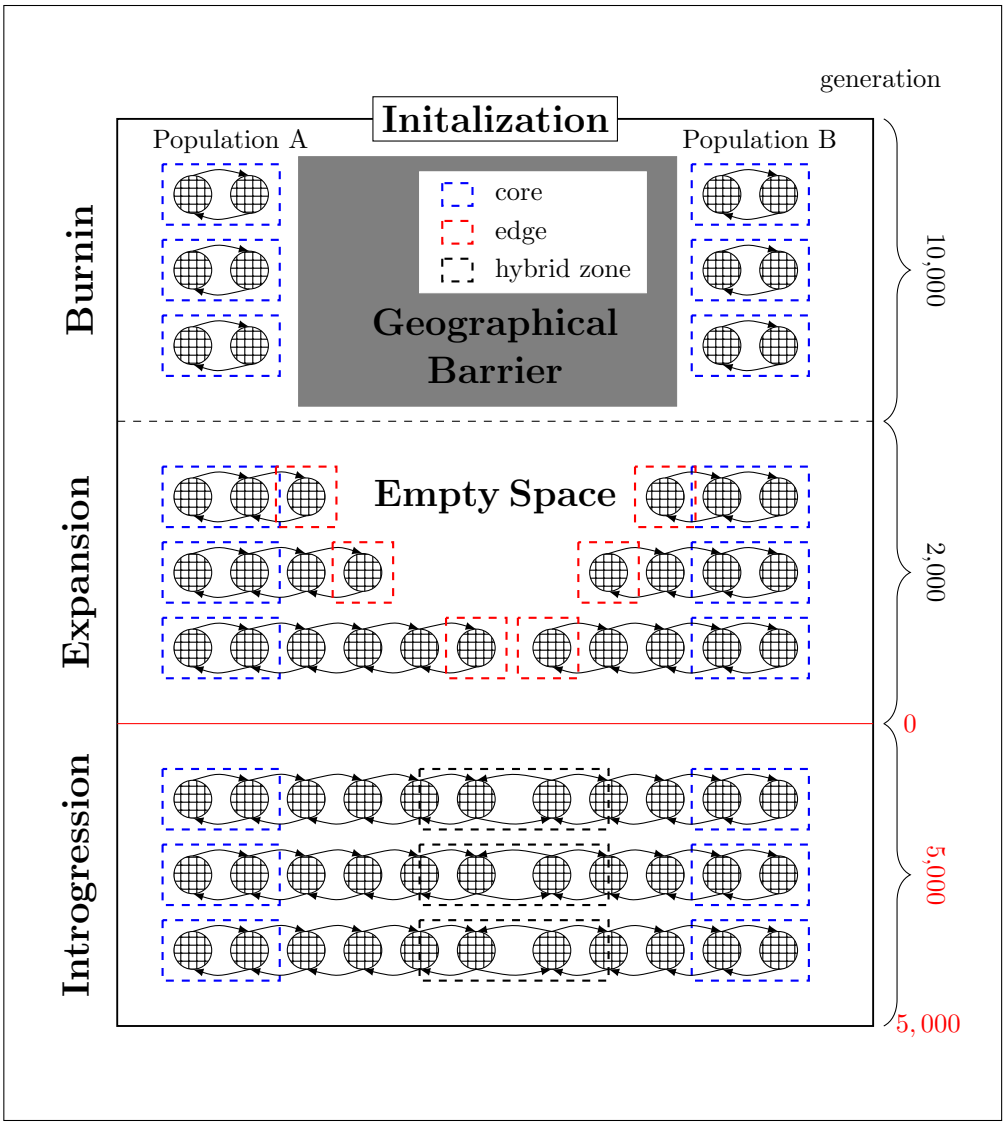

Figure 1: **Simulation schematic.** Simulations consist of four parts: initialization, burn-in (evolution in allopatry), range expansion, and introgression. The absolute number of generations is shown on the right-hand side.

An individual's genotype consists of $n_{neut} = 50$ neutral loci, $n_{del} = 100$ deleterious "background" loci, and $n_{BDMI} = 4$ two-locus BDMI pairs for a total of $n_{Loci} = n_{neut} + n_{del} + 2n_{BDMI} = 158$ loci, each of which is assumed to be biallelic. Deleterious loci are fully recessive, $h = 0$, with both heterozygotes and derived homozygotes having fitness $1 - s_{del}$. In the main text we consider a "dominant BDMI" where the fitness effects at the jth pair of BDMI loci, locus B_j and C_j, are defined such that individuals with genotypes B_jB_j/c_jc_j and b_jb_j/C_jC_j have fitness of 1 and a fitness of $1 - s_{BDMI}$ otherwise, where allele B_j is fixed in population A and C_j is fixed in population B. Throughout we fix $s_{BDMI} = 0.01$. Our results for such a dominant BDMI do not differ significantly from those obtained for a "recessive BDMI" (see supplementary material). Viability across loci is multiplicative with a baseline viability of V_0 such that the viability of

genotype i is given by the product:

$$V_i = V_0 * \prod_{j=1}^{n_{nut}} 1 * \prod_{j=1}^{n_{del}} \begin{cases} 1 & A_jA_j \\ 1 - hs_{del} & A_ja_j \\ 1 - s_{del} & a_ja_j \end{cases} * \prod_{j=1}^{n_{BDMI}} \begin{cases} 1 & B_jB_j/c_jc_j \\ 1 & b_jb_j/C_jC_j \\ 1 - s_{BDMI} & \text{otherwise} \end{cases} \quad (1)$$

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 equal size of K, at mutation-selection balance. Specifically, we set:

$$V_0 = \frac{1}{(1+s_{del})^{n_{del}\int_{\frac{1}{2K}}^{1-\frac{1}{2K}}[x^2\phi(x,s_{del},h_{del}=0,\mu,K)]dx}} \quad (2)$$

where $\phi(x, s, h, \mu, K)$ is Wright's distribution for mutation-selection-drift equilibrium given a population size of K. We excluded the rare cases when extinction still occurred during the burn-in. To quantify the effect of the BDMI we calculate the ratio V/V_b where V_b is defined as the first three terms of equation (1).

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

$$\begin{aligned} W_i &= \left(1 + \rho\left(1 - \frac{N_T}{K}\right)\right) V_i \\ w_i &= V_i/\bar{V} \end{aligned} \quad (3)$$

where $\bar{V}$ is the mean population viability. Note that this definition of fitness is density-independent. Even though the absolute fitness, W_i depends on the local population size N_T, an individuals relative fitness w_i does not. Due to the dependence of W_i on population size we will use V as a proxy for fitness throughout as it proportional to absolute fitness. Parents of genotype i produce a random Poisson distributed number of gametes with mean $2W_i$. The recombination rate between loci is constant across the genome and occurs at rate r. We will explore the case of both free recombination $r = 0.05$, and low recombination $r = 0.01$. During population burn-in and range expansion mutation occurs during gamete production. The per-generation, per-site mutation rate at the background loci occurs at rate $\mu = 10^{-4}$. To simplify the tracking of reproductive isolation we do not allow mutation at the BDMI loci. Gametes combine randomly to produce diploid offspring, which then migrate at a rate m to neighbouring demes as described in more detail below. Following migration, populations are censused, completing the generation.

We consider the demographic model depicted in Figure 1, modelling evolution in allopatry, followed by expansion, and finally introgression between two populations (populations A and B) 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 will be deposited upon acceptance):

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 ($n_{core} = 5$ throughout). Individuals migrate among demes such that they move to the left (right) neighbouring deme with probability

$m/2$, assuming reflective population boundaries. These demes will be referred to as the core of population A and B respectively. We initialize each core deme with $K = 100$ individuals, drawing the initial allele frequencies at the background loci using Wright's distribution given an effective population size of K, mutation rate $\mu = 10^{-4}$, and selection coefficient s_{del}. This initialization is an approximation to the steady state, ensuring that the burn-in phase efficiently reaches equilibrium. This approximation does not, however, account for changes in the effective population size due to variation in V_i, migration among demes, and linkage disequilibrium built up by migration and drift.

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 convergence of the mean viability $\bar{V}$ in each deme and the population size of each deme, N_T (see Figure 2).

Expansion: The third stage of the simulations is characterized by the expansion of population A and B which remain allopatric. Expansion proceeds for 2000 generations. As the rate of expansion is determined implicitly with individuals allowed to migrate to the neighbouring demes at a rate $m/2$, a value of $m = 0.01$ was used throughout. As described above the geographic extent (x-axis distance) of expansion varies stochastically across simulations. Although artificial, constraining expansion to a fixed time rather than distance allows us to exchange parental populations increasing the power of our simulations.

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

$$\mathcal{L} = \frac{\bar{V}_{core} - \bar{V}_{edge}}{\bar{V}_{core}} \tag{4}$$

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:

$$\mathcal{H} = \bar{V}_{F1} - \bar{V}_{Par} \tag{5}$$

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. As the dynamics of introgression are our primary focus, we define time $t = 0$ as the time of secondary contact. Following secondary contact, introgression proceeds for five-thousand generations. Simulations proceed as described above (Figure 1) except that without mutations. 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 neutral loci to define a measure of population ancestry. In particular if we define the allele frequency at locus i in the core of population A (left most deme) as p_i^A and the allele frequency in the core of population B (right most deme) as p_i^B. Then we

can measure the ansestry score of deme d as the following weighted average.

$$A_d = \frac{\sum_{i=1}^{n_{Neut}} \frac{p_i^d - p_i^A}{p_i^B - p_i^A} \Delta_i}{\sum_{i=1}^{n_{Neut}} \Delta_i} \tag{6}$$

where $\Delta_i = |p_i^A - p_i^B|$ and p_i^d is the allele frequency at locus i in deme d. An ancestry score of 0 indicates ancestry A and 1 indicates ancestry B. As we use all loci to define ancestry not just those with fixed differences the ancestry score can exceed 1 or fall below 0.

Results

Evolution in allopatry: the eco-evolutionary equilibrium

After evolving in allopatry, we graphically confirmed that the simulations had reached eco-evolutionary equilibrium. Shown in Figure 2 (see Figure S1 for the case of low recombination) the equilibrium population size (Panel A) and equilibrium viability (Panel B) increase as a function of the strength of selection on the deleterious loci s_{del} due to the drift-related purging and migration-related increase in N_e relative to equation 2.

During isolation the two parental populations diverge via genetic drift, ultimately reaching an equilibrium determined by the balance of mutation, migration, selection, and drift. Of the 100 deleterious background loci, the distribution of the number of fixed differences between parental population pairs is shown in Panel C. As expected at mutation-selection-drift balance, the number of fixed difference decreases with increasing strength of selection. This eco-evolutionary equilibrium will play a role in the dynamics of range expansion, heterosis, and introgression to follow.

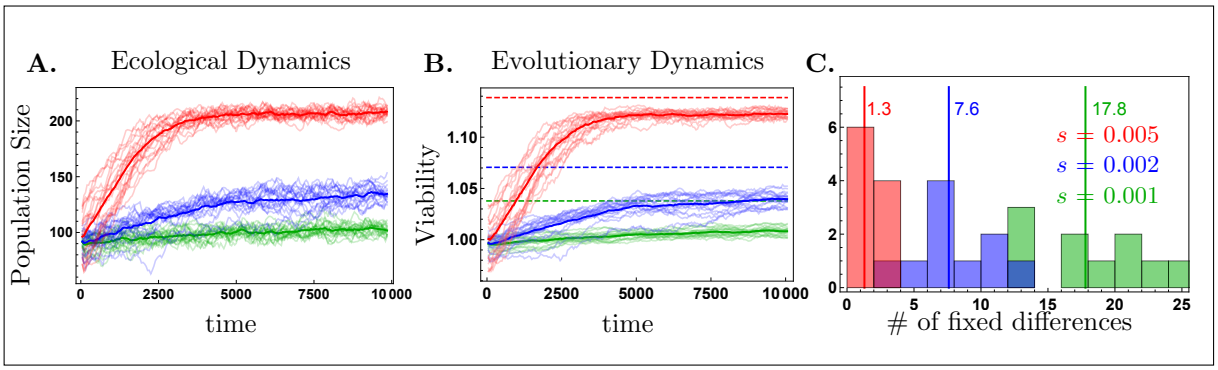

Figure 2: **Eco-evolutionary burn-in dynamics.** Panel A: The population size dynamics for each of the 20 replicate burn-ins across the 10000 burn-in generations for three strengths of selection red:$s = 0.005$, blue:$s = 0.002$, and green: $s = 0.001$ (dark curves show mean dynamics). Panel B: Dynamics of mean population fitness (viability) over the course of the burn-in. Dark curves show mean dynamics for a given strength of selection, dashed lines give the maximum possible viability, V_0, for a given strength of selection. Panel C: Distribution of the number of fixed differences between burn-in populations A and B. Vertical line gives the mean number of fixed differences.

Expansion dynamics: expansion load, allele surfing, and F1 heterosis

Following burn-in, the parental populations expand across the initially empty linear stepping-stone habitat for a total of two-thousand generations. At generation 2000 they come into secondary contact and begin to introgress. As expected given the results of the eco-evolutionary range expansion of Pieschel et al. (2015), repeated founder effects at the range edge increase the rate of genetic drift allowing deleterious mutations to surf to high frequencies and even fixation. The increase in frequency of deleterious alleles at the background loci due to allele surfing leads to significant expansion load, as defined by equation 4 and as shown in Figure 4 (see Figure S2 for the case of low recombination or a recessive BDMI).

To understand the dynamics of heterosis, first consider the observed heterosis at time $t = 0$. Due to the BDMI loci the mean heterosis of the F1 hybrids is, as expected, less than 0. Shown in Figure 3 Panel A, as expansion proceeds the observed level of heterosis increases over time. The dynamics of heterosis depends on the balance of the number of fixed differences and selective consequence of masking of these recessive deleterious effects in F1 hybrids. At the end of the period of range expansion, the average number of newly fixed differences due to expansion varies little with the strength of selection (Figure 3B; $s = 0.001 : 23.1, s = 0.002 : 29.4$ and $s = 0.005 : 25.4$). Yet the selective benefit of masking these deleterious loci differs greatly leading to the substantial increase in heterosis under strong selection (red curve in Figure 3A).

As shown in panel C of Figure 3, in the absence of the BDMIs heterosis increases by a proportion $\frac{1}{(1-s_{BDMI})^4}$, as expected for the $n_{BDMI} = 4$ pairs of loci. Finally, despite its effects on allele surfing observed by Pieschel et al. (2015), we find that recombination has little effect on the dynamics of heterosis (Figure 3D). The absence of an effect of recombination here may be due to the fact recombination in our model is distributed uniformly across the genome rather than occurring a fixed number of places in the genome as in Pieschel et al. (2015).

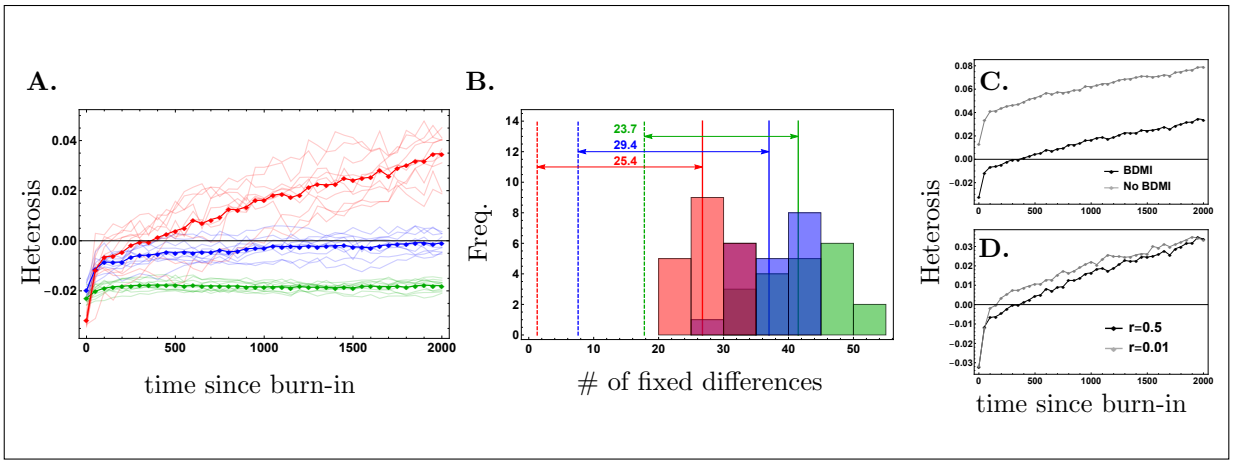

Figure 3: **Heterosis of F1 hybrids over the course of range expansion**. Panel A: Heterosis, see equation (5), over the course of expansion for three strengths of selection red: $s = 0.005$, blue: $s = 0.002$, and green: $s = 0.001$ with free recombination $r = 0.5$. Panel B: The number of fixed differences between edge demes after 2000 generations of range expansion with free recombination $r = 0.5$. Vertical lines give the mean initial (dashed) and final (solid) number of fixed differences (see Figure 2. Panel C: Difference in heterosis dynamics with (black) and without (gray) a BDMI for $s = 0.005$ and $r = 0.5$. Panel D: Effect of recombination on heterosis black:$r = 0.5$, dark gray:$r = 0.05$, and light gray: $r = 0.01$.

Introgression: Fitness, clinal dynamics, and population ancestry

Following the 2000 generations of range expansion, population A and population B come into secondary contact. We simulated the dynamics of introgression for the subsequent 5000 generations. Range expansion leads to a substantial decrease in population mean fitness and upon secondary contact masking of deleterious recessive alleles fixed on the range edge of one ancestral population but not the other leads to a substantial increase in fitness. As migration draws the introgressed alleles further into the range core, a region of relatively high fitness emerges (Figure 4A). A band of low fitness remains at the point of secondary contact due to the negative epistasis between the BDMI loci when there are fixed BDMIs (Figure S5 shows the effect of the BDMI loci on mean population viability).

Shaped by a combination of selection, migration, and drift, the spatial allele frequency dynamics after secondary contact are complex (Figure 5). This is particularly true for the allele frequency cline dynamics at the deleterious background loci. We will focus only on cases where the range edge of population A and population B are fixed for different alleles as a result of allele surfing , which occurs on average at 28.5% of loci for $s = 0.005$ (for clinal dynamics at all loci see figure S6). Range expansion leaves a distinct signal in the shape of these clines. Unlike clinal dynamics formed upon secondary contact in the absence of range expansion (Barton and Hewitt, 1985), parental populations are not initially uniformly fixed for different alleles. Instead the frequency of the derived allele increases steadily across the range as a result of surfing until reaching ultimate fixation near the point of secondary contact. In contrast, in the parental population fixed for the wild-type allele ($p = 0$) on the range edge, the frequency of the derived allele increases slightly from the edge to the core nearing the mutation-selection balance of $\frac{\mu}{s}$ in the core. As introgression proceeds

the initially fixed derived allele declines in frequency eventually approaching mutation-selection balance across the complete linear habitat. As illustrated in Figure 5, this can take a long time even when selection is strong. Hence genetic signals of range expansion persists for thousands of generations after secondary contact.

In comparison to the dynamics at the background loci, the clinal dynamics at the BDMI loci are straightforward (see Figure S7). By design, each of the BDMI locus is initially fixed for one allele in population A and the alternative allele in population B creating a steep cline upon contact. Selecting against hybrid genotypes, this cline remains strong throughout introgression. This occurs despite the fact that the selection against hybrids at these loci is at least partially counteracted by linkage to wild-type background alleles that are favoured during introgression. This is most likely the result of the fact that each BDMI allele is statistically equally likely to be linked to the same number of masking wild-type background loci. The net balance of linkage to masking wild-type and derived background alleles is reiterated by the fact that the recombination rate has little effect on introgression at the BDMI loci, at least for the parameters investigated (see Figures S2, S3, and S4).

Like the allele frequency dynamics at the deleterious background loci, allele surfing leaves a distinct pattern in the shape of the allele frequency clines at the neutral loci. As expected at mutation-selection-drift balance (Wright's distribution) allele frequencies at the neutral loci in the core demes are distributed unimodally with a mean of 0.5. Allele surfing toward the range edge however creates a strongly bi-modal distribution of allele frequencies with many fixed in either population A or population B (see Figure 6 panel A and Figure S8). This change in the distribution of allele frequencies across the range leads to clines in the inferred population ancestry as defined by equation (6) even in the absence of any introgression (Figure 6).

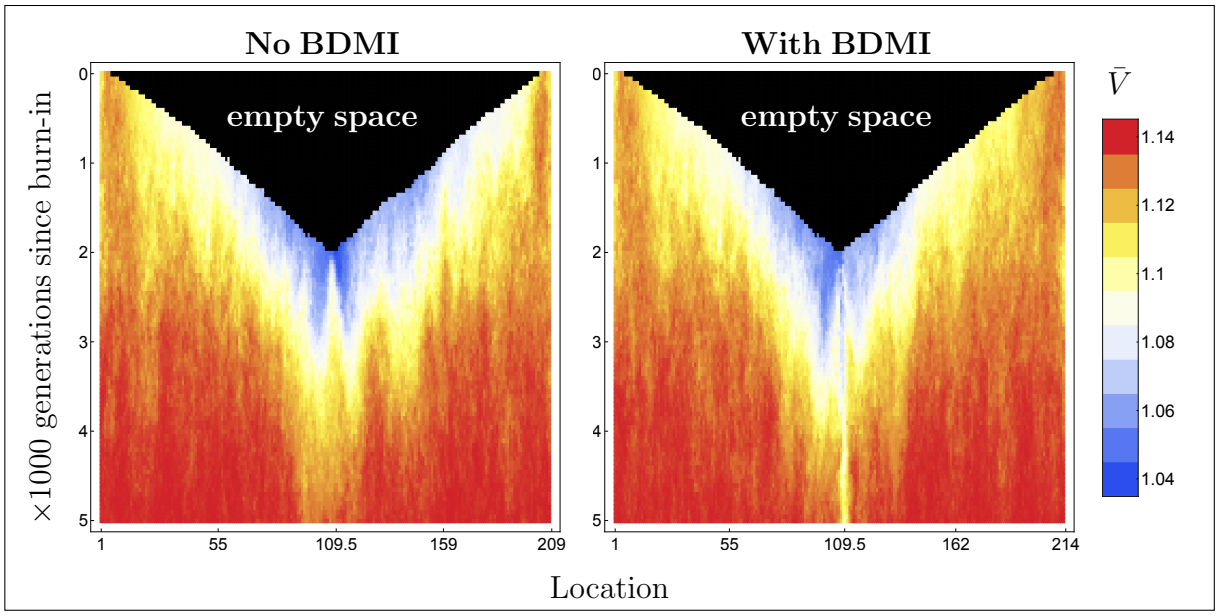

Figure 4: **Mean population viability during expansion and introgression**. Mean population viability, which is proportional to population mean fitness, over the 2000 generations of expansion and the first 3000 generations of introgression when ancestral populations are fixed for either the same (left panel) or different BDMI alleles (right panel). The x-axis represents geographic location occupied in the linear stepping stone habitat. Black area represents unoccupied space into which the parental populations can expand. Parameters: $s_{del} = 0.05, s_{BDMI} = 0.01$ and $n_{BDMI} = 4, r = 0.5$.

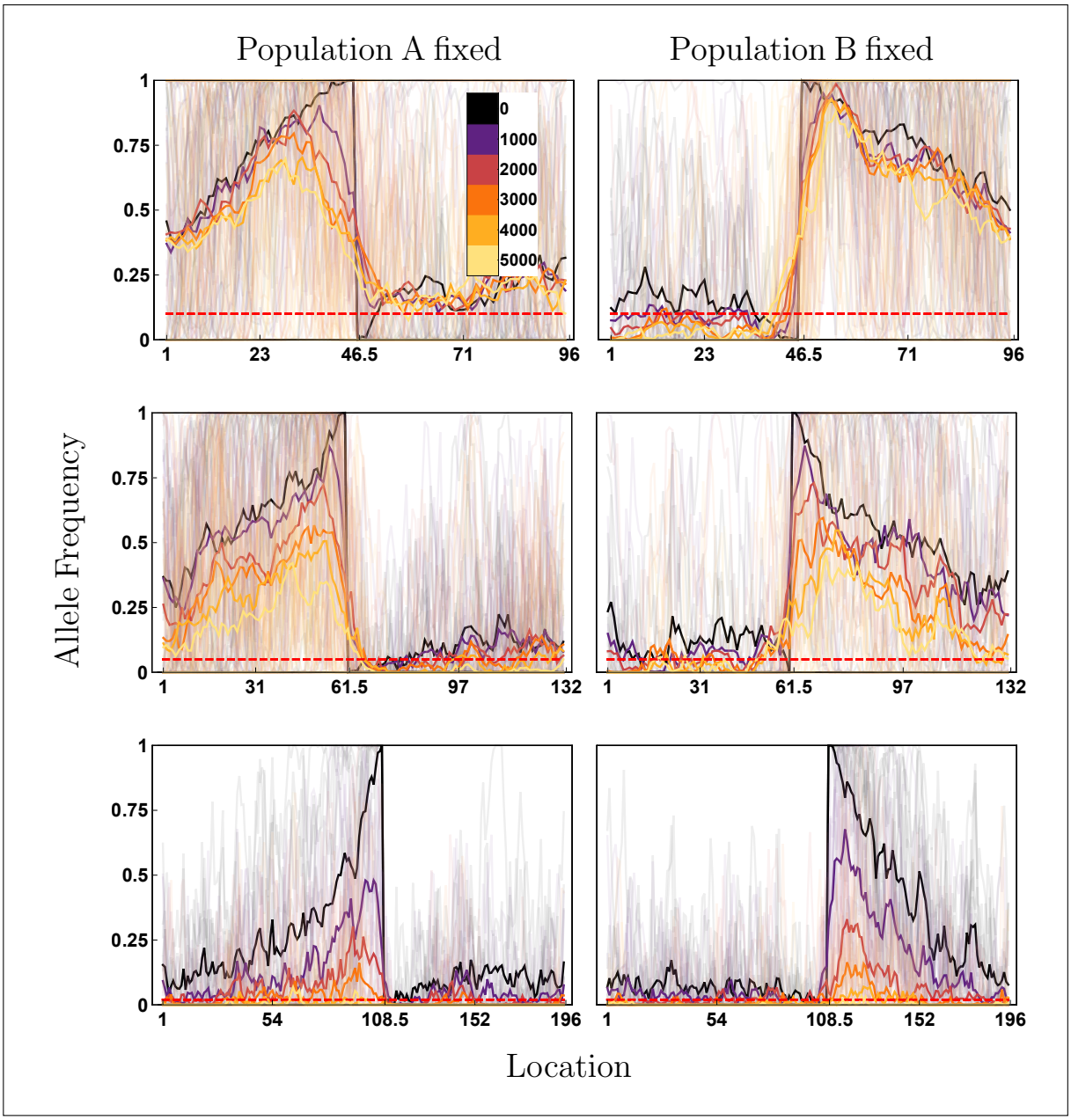

Figure 5: **Allele frequency profiles at background loci over the course of introgression.** Column A (B): Clinal dynamics for the subset of deleterious background loci fixed for the mutant allele in edge population A (B) and the wild-type allele in edge population B (A) shown for every 1000 generations (time measured in units of generations since end of expansion) over the course of introgression. Light lines give the allele frequency dynamics at individual loci as summarized by the dark lines showing the mean allele frequency dynamics. Row 1: $s = 0.001$, row 2: $s = 0.002$, row 3: $s = 0.005$. Dashed Red line indicates expected allele frequency under mutation-selection balance. Parameters: $r = 0.5$, dominant BDMI (see quation (1)) with $s_{BDMI} = 0.01$ and $n_{BDMI} = 4$.

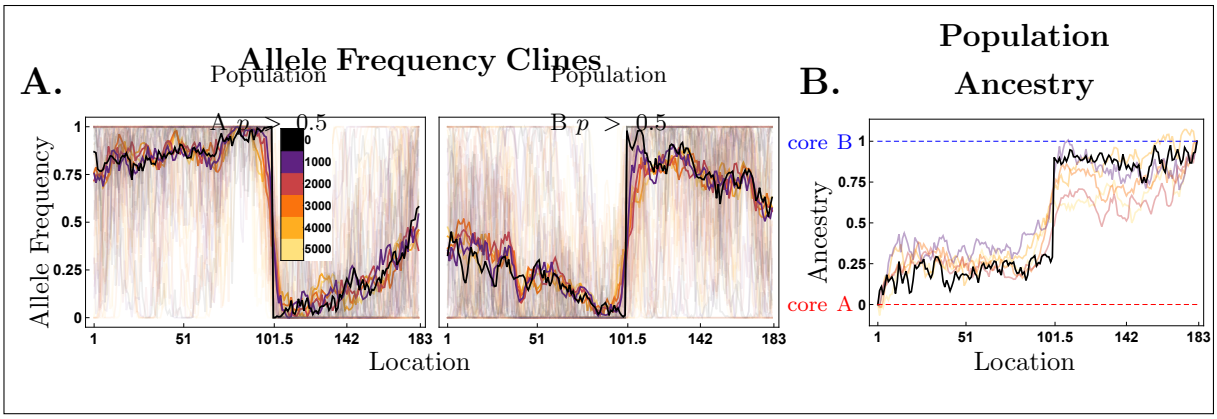

Figure 6: **Allele frequency clines at neutral loci and population ancestry**. Panel A: Left hand (Right hand) plot shows the allele frequency clines at neutral loci that have a derived allele frequency greater (less) than 0.5 in the core of population A. Dark lines give mean allele frequency dynamics light lines give dynamics at individual loci. Colours represent generations since secondary contact. Panel B: population ancestry inferred given the allele frequency at the neutral loci in the core demes of population A and B as defined by equation (6).

Discussion

In this paper we provide proof-of-principle simulations to demonstrate that parental population range expansion can have important consequences on introgression in secondary contact hybrid zones. As suggested by prior theoretical literature (Edmonds et al., 2004; Peischl and Excoffier, 2015), allele surfing at deleterious loci can lead to substantial declines in population mean fitness at the range edge for these expanding populations. Allele surfing and the associated expansion load have been suggested to limit the rate and extent of range expansion and lead to spatial sorting (Peischl and Gilbert, 2020). Our results highlight the effect of allele surfing as a result of parental population expansion on the tendency of hybridization at secondary contact. In particular, we find that the masking of accumulated recessive deleterious alleles can favour hybridization, dramatically shape genetic clines across secondary-contact hybrid zones, and affect the expected patterns of population ancestry. Given the abundant evidence of parental range expansion before secondary contact (Hewitt, 1999; Taberlet, 1998) and its indispensable role in hybridization and spatial genetics, it has rarely been considered in speciation models. Only recently have the effects of historical range expansion on genomic signatures of in hybrid zones been considered (Bertl et al., 2018).

The model presented here makes several assumptions that have important consequences on the introgression dynamics. First, we consider only the effect of completely recessive ($h = 0$) deleterious alleles. While most deleterious alleles are, at least, partially recessive, the assumption that these alleles are recessive is necessary for them to contribute to heterosis. In contrast (co)dominant deleterious alleles, $h > 0.5$, will not contribute to heterosis and may have complex asymmetrical effects on the fitness regime of the hybrid zone. The effect of dominance is illustrated in Figures S3 and S4 which show the dynamics of heterosis in F1 hybrids when background loci are completely recessive $h = 0$, as explored above, and partially recessive

$h = 0.25$. In addition to focusing on recessive mutations our model assumes that mutant alleles at the background loci all have the same deleterious fitness effect, s_{del}. Quantifying the effect of allele surfing on hybrid zone dynamics would require the full distribution of selective and dominance effects of surfing alleles. A second implicit assumption of the model presented here is that background alleles effect individual viability (V). A comprehensive understanding of range expansion and introgression dynamics will require the consideration of mutational effects on multiple life-history traits and fitness components. Finally, our exploration of the effect of allele surfing and its effect on BDMIs and other forms of genetic incompatibilities is limited. In particular, we focus on the effect of allele surfing only at the background loci and hence do not allow for mutation at the BDMI loci. However, mutation and the possibility for segregating variation at the BDMI loci would allow them to surf as well, which may in turn influence their ultimate contribution to reproductive isolation upon secondary contact.

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 in 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 is key. If divergence occurs in the presence of gene flow, no heterosis is expected, whereas divergence under extreme drift, such as that seen here (where geographically isolated lineages expand and meet) creates genomic conditions that favor heterosis.

Another partially unexplained phenomenon is the relatively high frequency of introgressions that appear to be positively selected in hybrid populations (Rieseberg et al., 1999; Corbi et al., 2018; Hamilton et al., 2013). This could be a byproduct of adaptation in finite populations, such that some favourable alleles were previously in only one of the parental populations (Barton, 2001). Alternatively, the environment may be different in the hybrid zone centre, favouring a different set of alleles (Schilthuizen et al., 1999). However, it seems likely that a significant fraction of favourable introgressions are due to the masking of deleterious mutations, as shown here. Positively selected introgressions are frequently seen in progeny from controlled crosses as well (Rieseberg et al., 1996), consistent with this explanation.

Our results also have implications for longer 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 and potentially even to the fusion of previously isolated populations. High levels of heterosis would also reduce the strength of reinforcing selection. This leads to the interesting prediction that reinforcement might be more likely when populations have diverged in the presence of gene flow rather than following secondary contact between expanding populations. This differs from the classic scenario, in which reinforcement "completes speciation" following range expansion and contact between previously allopatric populations (Dobzhansky, 1982). On the other hand, heterosis might facilitate homoploid hybrid speciation,

assuming other ecological and evolutionary conditions favoured this outcome (Buerkle et al., 2000; Schumer et al., 2015). Future studies should explore the impact of heterosis on these processes.

Our results suggest that some genomic patterns of speciation may need to be reassessed. For example less genomic divergence has been interpreted as a single of rapid speciation. However, this could instead reflect of recent fusion between diverged lineages that has gone through expansion before secondary contact. For example, the latitudinal gradient of genetic divergence was interpreted as gradient of speciation rate (Weir and Schluter, 2004, 2007; Rabosky et al., 2018). However, if we take the parental expansion-facilitated heterosis into account, this difference in genomic divergence could be partly/mostly explained by recent glacial dynamics with differential expansion of parental populations. For instance, Townsend's and Hermit warblers were trapped in refugia during the last glaciation and expanded with glacial retraction until secondary contact (Weir and Schluter, 2004; Krosby and Rohwer, 2009). The coastal Townsend's warblers is a mix of the inland Townsend's and Hermit warblers, thus the current shallow genomic divergence is likely a result of historical and ongoing hybridization (Rohwer and Wood, 1998; Krosby and Rohwer, 2009; Wang et al., 2019a,b) facilitated, at lesat in part, by hterosis due to th maskingof deleterious alleles accumulated during range expansion.

References

April, J., Hanner, R.H., Dion-Côté, A.M., Bernatchez, L., 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Molecular Ecology 22, 409–422. doi:10.1111/mec.12116.

Arntzen, J.W., Vries, W.d., Canestrelli, D., Martínez-Solano, I., 2017. Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. Molecular Ecology 26, 5663–5675. doi:10.1111/mec.14273, tex.copyright: © 2017 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd.

Austerlitz, F., Jung-Muller, B., Godelle, B., Gouyon, P.H., 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. Theoretical Population Biology 51, 148–164. doi:10.1006/tpbi.1997.1302.

Avise, J.C., Walker, D., Johns, G.C., 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. Lond. B 265, 1707–1712. Tex.pmcid: PMC1689361.

Barton, N.H., 2001. The role of hybridization in evolution. Mol. Ecol. 10, 551–568. doi:10.1046/j.1365-294x.2001.01216.x.

Barton, N.H., Hewitt, G.M., 1985. Analysis of Hybrid Zones. Annu. Rev. Ecol. Evol. Syst 16, 113–148. doi:10.1146/annurev.es.16.110185.000553, publisher: Annual Reviews.

Bernatchez, L., Wilson, C.C., 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology 7, 431–452. doi:10.1046/j.1365-294x.1998.00319.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-294x.1998.00319.x.

Bertl, J., Ringbauer, H., Blum, M.G., 2018. Can secondary contact following range expansion be distinguished from barriers to gene flow? PeerJ 6, e5325. doi:10.7717/peerj.5325.

Buerkle, C.A., Morris, R.J., Asmussen, M.A., Rieseberg, L.H., 2000. The likelihood of homoploid hybrid speciation. Heredity 84, 441–451. doi:10.1046/j.1365-2540.2000.00680.x.

Corbi, J., Baack, E.J., Dechaine, J.M., Seiler, G., Burke, J.M., 2018. Genome-wide analysis of allele frequency change in sunflower crop–wild hybrid populations evolving under natural conditions. Molecular Ecology 27, 233–247. doi:10.1111/mec.14202. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mec.14202.

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

Dobzhansky, T., 1982. Genetics and the Origin of Species. Columbia University Press.

Edmonds, C.A., Lillie, A.S., Cavalli-Sforza, L.L., 2004. Mutations arising in the wave front of an expanding population. PNAS 101, 975–979. doi:10.1073/pnas.0308064100. iSBN: 9780308064102 Publisher: National Academy of Sciences Section: Biological Sciences.

Excoffier, L., Foll, M., Petit, R.J., 2009. Genetic Consequences of Range Expansions. Annu. Rev. Ecol. Evol. Syst 40, 481–501. doi:10.1146/annurev.ecolsys.39.110707.173414. _eprint: https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Haffer, J., 1969. Speciation in Amazonian Forest Birds. Science 165, 131–137. doi:10.1126/science.165.3889.131. publisher: American Association for the Advancement of Science Section: Articles.

Hamilton, J.A., Lexer, C., Aitken, S.N., 2013. Differential introgression reveals candidate genes for selection across a spruce (Picea sitchensis x P. glauca) hybrid zone. The New Phytologist 197, 927–938. doi:10.1111/nph.12055.

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68, 87–112. doi:10.1006/bijl.1999.0332.

Klopfstein, S., Currat, M., Excoffier, L., 2006. The Fate of Mutations Surfing on the Wave of a Range Expansion. Mol. Biol. Evol. 23, 482–490. doi:10.1093/molbev/msj057.

Krosby, M., Rohwer, S., 2009. A 2000km genetic wake yields evidence for northern glacial refugia and hybrid zone movement in a pair of songbirds. Proc. R. Soc. Lond. B 276, 615–621. doi:10.1098/rspb.2008.1310. publisher: Royal Society.

Lowry, D.B., Modliszewski, J.L., Wright, K.M., Wu, C.A., Willis, J.H., 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Philos. Trans. R. Soc. B 363, 3009–3021. doi:10.1098/rstb.2008.0064. publisher: Royal Society.

Peischl, S., Dupanloup, I., Kirkpatrick, M., Excoffier, L., 2013. On the accumulation of deleterious mutations during range expansions. Molecular Ecology 22, 5972–5982. doi:10.1111/mec.12524. tex.copyright: © 2013 John Wiley & Sons Ltd.

Peischl, S., Excoffier, L., 2015. Expansion load: recessive mutations and the role of standing genetic variation. Molecular Ecology 24, 2084–2094. doi:10.1111/mec.13154.

Peischl, S., Gilbert, K.J., 2020. Evolution of Dispersal Can Rescue Populations from Expansion Load. Am. Nat 195, 349–360. doi:10.1086/705993.

Peischl, S., Kirkpatrick, M., Excoffier, L., 2015. Expansion Load and the Evolutionary Dynamics of a Species Range. Am. Nat 185, E81–E93. doi:10.1086/680220.

Pickup, M., Field, D.L., Rowell, D.M., Young, A.G., 2013. Source population characteristics affect heterosis following genetic rescue of fragmented plant populations. Proc. R. Soc. Lond. B 280, 20122058. doi:10.1098/rspb.2012.2058.

Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M., Alfaro, M.E., 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559, 392–395. doi:10.1038/s41586-018-0273-1.

Rieseberg, L.H., Arias, D.M., Ungerer, M.C., Linder, C.R., Sinervo, B., 1996. The effects of mating design on introgression between chromosomally divergent sunflower species. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 93, 633–644. doi:10.1007/BF00417959.

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

Rohwer, S., Wood, C., 1998. Three Hybrid Zones between Hermit and Townsend's Warblers in Washington and Oregon. The Auk 115, 284–310. doi:10.2307/4089188. publisher: American Ornithological Society.

Schilthuizen, M., Hoekstra, R.F., Gittenberger, E., 1999. Selective increase of a rare haplotype in a land snail hybrid zone. Proc. R. Soc. Lond. B 266, 2181–2185. doi:10.1098/rspb.1999.0906. publisher: Royal Society.

Schumer, M., Cui, R., Rosenthal, G.G., Andolfatto, P., 2015. Reproductive Isolation of Hybrid Populations Driven by Genetic Incompatibilities. PLOS Genetics 11, e1005041. doi:10.1371/journal.pgen.1005041. publisher: Public Library of Science.

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. doi:10.1111/j.1439-0469.1976.tb00938.x.

Taberlet, P., 1998. Biodiversity at the intraspecific level: The comparative phylogeographic approach. Journal of Biotechnology 64, 91–100. doi:10.1016/S0168-1656(98)00106-0.

Wang, S., Rohwer, S., Delmore, K., Irwin, D.E., 2019a. Cross-decades stability of an avian hybrid zone. J. Evol. Biol. 32, 1242–1251. doi:10.1111/jeb.13524.

Wang, S., Rohwer, S., de Zwaan, D.R., Toews, D.P.L., Lovette, I.J., Mackenzie, J., Irwin, D.E., 2019b. Selection on a pleiotropic color gene block underpins early differentiation between two warbler species. bioRxiv doi:10.1101/853390.

Weir, J.T., Schluter, D., 2004. Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B 271, 1881–1887. doi:10.1098/rspb.2004.2803.

Weir, J.T., Schluter, D., 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315, 1574–1576. doi:10.1126/science.1135590.