Reproductive isolation via polygenic local adaptation in sub-divided populations: effect of linkage disequilibria and drift

Himani Sachdeva

Department of Mathematics, University of Vienna, Vienna 1090, Austria.

Abstract

This paper considers how local adaptation and reproductive isolation between hybridizing populations is influenced by linkage disequilibria (LD) between multiple divergently selected loci in scenarios where both gene flow and genetic drift degrade local adaptation. It shows that the combined effects of multi-locus LD and genetic drift on allele frequencies at selected loci and on heterozygosity at neutral loci are predicted accurately by incorporating (deterministic) effective migration rates into the diffusion approximation (for selected loci) and into the structured coalescent (for neutral loci). Theoretical approximations are tested against individual-based simulations and used to investigate the conditions for the maintenance of local adaptation on an island subject to one-way migration from a differently adapted mainland, and in an infinite-island population with two different habitats under divergent selection. The analysis clarifies the conditions under which LD between sets of locally deleterious alleles allows these to be collectively eliminated despite drift, causing sharper and (under certain conditions) shifted migration thresholds for loss of adaptation. Local adaptation also has counter-intuitive effects on neutral (relative) divergence: F_{ST} is highest for a pair of subpopulations belonging to the same (rare) habitat, despite the lack of reproductive isolation between them.

Keywords: linkage disequilibria, genetic drift, effective migration rates, local adaptation, speciation, reproductive isolation

Environmental adaptation in natural populations typically involves selection that varies over space and time, and acts on many genetic loci (Buckler et al. 2009; Soria-Carrasco et al. 2014; Lamichhaney et al. 2012). The ability of populations to adapt to their local environment thus depends on the ease with which alleles favoured in one region can establish and/or be maintained across multiple loci, despite gene flow from other regions (where different alleles may be locally advantageous). A key question is: to what extent does selection act against *combinations* of deleterious alleles, as opposed to individual alleles (Barton 1983; Feder et al. 2012)? More generally, when do statistical associations or linkage disequilibria (LD) between sets of locally adaptive alleles— which may be tightly clustered or distributed across the genome— protect such alleles from swamping, especially in marginal habitats, which are often subject to maladaptive, predominantly asymmetric gene flow from differentially adapted core populations?

The buildup or maintenance of LD between divergently selected alleles within (partially) hybridizing populations may be viewed as an example of the broader process of 'coupling' between barrier loci that reduces inter-population genetic exchange (Barton and De Cara 2009; Butlin and Smadja 2018). In the absence of major transitions (e.g., ploidy changes), the coupling of barrier loci across the genome is believed to be crucial for the emergence of genomewide (as opposed to localized) barriers to gene flow and the completion of reproductive isolation (RI) and speciation (Kulmuni et al. 2020).

Much of previous theoretical work on the role of LD during local adaptation has focused on the deterministic analysis of *two* loci subject to divergent selection across interconnected subpopulations or in continuous space (Li and Nei 1974; Slatkin 1975; Bürger and Akerman 2011; Geroldinger and Bürger 2015). Such models often allow for explicit analytical solutions under fairly general conditions; these are, however, difficult to generalize to greater numbers of loci, which is essential for understanding the evolution of RI.

Barton (1983) considered migration-selection balance at arbitrary numbers of divergently selected, equal-effect loci on a linear genome, and showed that in large populations, where the effects of drift can be neglected, the strength of coupling or LD between a set of introgressing deleterious alleles is governed by the selection density, i.e., the ratio of the total selection (against the entire set of alleles) to the total recombination rate (across the map length spanned by the alleles). When total selection is stronger than total recombination, and immigration sufficiently weak that deleterious alleles segregate at low frequencies, sets of such alleles are eliminated cooperatively by selection much faster than they are broken apart by recombination: then, the frequency of any allele at migration-selection balance is influenced more by indirect selection due to LD with other deleterious alleles than direct selection due to its own deleterious effect. Conversely, when recombination is much faster than selection, allele frequencies evolve essentially independently across loci, primarily under selection on their own effects.

Further, selection against immigrants and their descendants also impedes the exchange of neutral alleles between populations, if neutral alleles are in strong LD with alleles that are divergently selected across populations. This reduction in neutral exchange can be quantified in terms of the *effective* migration rate at a neutral locus for discrete demes connected via migration (Bengtsson 1985), or in terms of the strength of a 'barrier' to gene flow for spatially continuous populations subject to heterogeneous selection (Nagylaki 1976). Barton and Bengtsson (1986) calculated effective migration rates and barrier strengths in large populations (i.e., neglecting drift) for a variety of multilocus configurations and spatial geometries, and showed that when barrier loci are linked across a linear genome, then the effective migration rate of a neutral allele at an arbitrary genomic location (which is unlinked to any barrier locus) is strongly reduced only when selection density, i.e, net selection against introgressing deleterious alleles per unit map length is high. With barrier loci that are *unlinked* to each other, effective migration rates at neutral markers are reduced relative to the actual migration rate by a factor that (to a very good approximation) only depends on the relative fitness of immigrants (Bengtsson 1985).

This relatively simple picture of the effects of LD must be refined if populations are small and drift at least comparable to per locus selection. Not only can drift, in conjunction with maladaptive gene flow, then directly impair the efficacy of selection against individual deleterious alleles, it might also generate negative LD between sets of deleterious alleles via Hill-Robertson interference (Hill and Robertson 1966): this will counteract positive LD due to migration, thus also compromising the efficacy of indirect, LD-mediated selection which underlies the cooperative elimination of groups of locally deleterious alleles in large populations.

Understanding the combined effects of LD and genetic drift on local adaptation also becomes important in view of the fact that extended populations are often patchworks of smaller, interconnected subpopulations. When density regulation occurs primarily at the scale of subpopulations, the rate of drift is governed by local sizes rather than the size of the population as a whole. Additionally, if traits subject to divergent selection across subpopulations are polygenic, then individual loci contributing to trait variation may have rather weak selective effects, so that typical values of Ns are quite small, resulting in local adaptation via multiple transient allele frequency shifts (Yeaman 2015).

A mathematical understanding of the effects of genetic drift on multiple loci under selection remains largely elusive, despite the centrality of such an understanding to fundamental

evolutionary questions related to the limits to natural selection, the evolution of sex and recombination, and the maintenance of genetic variation. The diffusion approximation, which provides a powerful framework for analysing the combined effects of drift, selection, mutation, subdivision and other deviations from panmixia on allele frequencies (Wright 1937), is only practically useful for single loci or, at best, multiple loci under linkage equilibrium (i.e., neglecting LD). Thus, theoretical work concerning the effects of drift on multi-locus evolution (that goes beyond simulations) is based almost entirely on various kinds of heuristics and scaling arguments (Weissman and Barton 2012; Good et al. 2014; Roze 2021).

This paper considers local adaptation in the context of a large population consisting of many interconnected, small subpopulations occupying different habitats, where fitness is influenced by multiple loci with habitat-dependent selective effects. I illustrate how the gross effects of LD between selected loci on local adaptation (at any individual selected locus) in the presence of drift are accurately predicted by the simple heuristic of incorporating appropriately defined effective migration rates (for deleterious alleles) into the diffusion approximation. Analogously, the effects of LD between selected loci and neutral markers on neutral genetic diversity within any subpopulation is accurately captured by incorporating effective migration rates (for neutral alleles) into the structured coalescent (see also Petry (1983) and Charlesworth et al. (1997) who used this approach to approximate how adaptive divergence at a single selected locus affects diversity at linked neutral markers).

I then use theoretical approximations based on these heuristics to explore the conditions for local adaptation on an island subject to maladaptive gene flow from a large and perfectly adapted mainland, as well as in an infinite-island population with two habitats under divergent selection, where the degree of local adaptation in either habitat is influenced by the other. An important focus is to understand how local adaptation in a rare or marginal habitat (which encompasses a small fraction of all subpopulations in the population) is influenced by factors such as the number of loci under divergent selection (which influences the extent of LD), the selective effects of divergently selected loci, the size of subpopulations, the number of migrants exchanged, and the relative abundances of the two habitats. Finally, I explore how neutral diversity in either habitat is influenced by LD between neutral alleles and selective alleles under conditions of local adaptation, and how this translates into expectations for various F_{ST} measures. Theoretical approximations are compared against individual-based simulations throughout, in order to test their validity.

Models and Approximations

Mainland-island model. Consider an island with one-way migration from a mainland, where mainland and island populations are subject to divergent selection at L unlinked loci. Loci are biallelic, with alternative alleles favoured on the mainland and island at each locus, independent of the state of other loci. Selection is multiplicative across loci, i.e., there is no epistasis. Further, effect sizes are assumed to be equal, so that the fitness of any individual on the island depends only on the number y of locally deleterious alleles it carries, and is given by e^{-sy}, where s is the selective effect per deleterious allele.

For simplicity, we can also assume that there is no polymorphism in the mainland population at any of the divergently selected loci, i.e., each locus is fixed for the locally advantageous allele (this assumption is not crucial and is adopted merely to avoid proliferation of parameters). Polymorphism is maintained on the island despite drift and continual gene flow from the mainland, provided mutation rates are above some threshold value.

There is a fixed number N of haploid individuals on the island. The lifecycle on the island is as follows: in each generation, a fraction m of individuals on the island are replaced by

migrants from the mainland. Individuals then undergo mutation, with the rate of mutation between alternative alleles being μ per locus per individual. Following mutation, the next generation of individuals on the island is formed by sampling $2N$ parents (with replacement) with probabilities proportional to their relative fitness. Finally, the $2N$ parents are paired randomly and recombinant haploid offspring created via free recombination between pairs.

Infinite-island model. Consider a population with n_D islands, where each island has a fixed number N of individuals. In theoretical analyses, it is convenient to take the $n_D \to \infty$ limit, i.e., assume an infinite number of islands. Individual fitness on any island depends on the habitat (or local environment) of that island. For simplicity, we will consider only two habitats; a fraction ρ of islands support one habitat, and the remaining fraction $1-\rho$ the second. We will assume $\rho<1/2$, so that the first habitat is 'rare' (i.e., it characterizes a minority of islands in the population) and the second 'common'.

As before, fitness is influenced by L unlinked biallelic loci, with alternative alleles favoured in the two habitats at each of the L loci. The relative fitness of an individual carrying i *locally* deleterious alleles in a deme belonging to habitat k is $e^{-s_k i}$, where s_k is the selective effect per locally deleterious allele in habitat k. We will use the subscripts r and c to denote the rare and common habitats. For simplicity, only the case with $s_r=s_c=s$ (symmetric selection) is considered in this paper, but none of the approximations described below depend on this.

With an infinite number of demes, the mutation rate can be set to zero, since mutation is not essential for the maintenance of polymorphism as long as ρ is not vanishingly small, and because we are primarily interested in the effect of gene flow on local adaptation. The extension to non-zero mutation rates is straightforward.

In each generation, a fraction m of individuals in each deme are replaced by migrants from a common pool, formed by uniformly drawing individuals from across all demes. Following migration, the next generation of individuals is formed by sampling $2N$ parents within each deme with probabilities proportional to local relative fitness. N haploid offspring are then created in each deme by free recombination between parental pairs. For the infinite-island model, we will also follow neutral markers that are unlinked to any selected locus and to each other; this allows us to investigate how the extent of adaptive divergence between habitats influences neutral gene flow and genome-wide RI.

Since the main goal is to clarify how LD and drift jointly influence local adaptation in a structured population, other kinds of complexity are neglected by making various simplifying assumptions: organisms are assumed to be haploid (thus neglecting the effects of dominance); loci are assumed to be unlinked and selective effects taken to be equal across all loci. More crucially, the model assumes an extreme form of divergent selection wherein any selected allele has opposite effects in the two habitats (irrespective of the state of other loci). Finally, there is no explicit space and no isolation-by-distance.

In the following, I first briefly discuss the diffusion approximation (Wright 1937), which accounts for the effects of genetic drift on allele frequencies in a subdivided population but neglects LD. I then discuss deterministic analyses that account for LD but neglect drift (Barton 1983; Barton and Bengtsson 1986); it is useful to represent the effects of LD in such analyses by an effective migration rate for the selected or neutral allele (Bengtsson 1985; Petry 1983). I then describe how effective migration rates can be incorporated within the diffusion framework, leading to approximations that yield highly accurate predictions for the combined effects of LD and drift on allele frequencies.

Diffusion approximation (assuming LE). When net selection against maladapted genotypes is weak relative to recombination, LD between selected variants can be neglected and loci

assumed to evolve independently, i.e., under linkage equilibrium (LE). Then, as long as $1/N$, s, m, and μ are sufficiently small, the probability distribution $\psi[p]$ of the allele frequency p at any locus under mutation-selection-migration-drift equilibrium is accurately predicted by the diffusion approximation (Wright 1937). For definiteness, we will use p to denote the frequency of the allele that is locally favoured in the common habitat and thus locally disadvantageous in the rare habitat (or on the island in case of mainland-island migration).

For an island subject to one-way migration from the mainland (which is assumed to be fixed for the alternative allele), the equilibrium allele frequency distribution can be expressed in terms of the scaled parameters Ns, $N\mu$ and Nm, and is given by:

$$\psi_{island}[p] \propto p^{2Nm+2N\mu-1}(1-p)^{2N\mu-1}e^{-2Nsp} \tag{1}$$

Integrating over the normalised distribution gives the expected allele frequency $\mathbb{E}[p]$ on the island.

For the infinite-island model, one can write an implicit expression for the allele frequency distribution $\psi_i[p|\overline{p}]$ at any locus for a deme within habitat i, conditional on $\overline{p}$, the allele frequency at that locus in the migrant pool, which (at equilibrium) is simply equal to the average allele frequency across the entire population. Then, $\psi_i[p|\overline{p}]$ is given by:

$$\psi_i[p|\overline{p}] \propto p^{2Nm\overline{p}-1}(1-p)^{2Nm(1-\overline{p})-1}e^{-2Ns_i[X_ip+(1-X_i)(1-p)]} \tag{2}$$

The subscript i can take on values r and c, corresponding to the rare and common habitats respectively. We have X_r=1 and X_c=0. One can now calculate the expected allele frequency $\mathbb{E}[p_i|\overline{p}]$ in either habitat as a function of $\overline{p}$ by integrating over the frequency distribution above, and then find $\overline{p}$ by numerically solving $\overline{p}=(1-\rho)\mathbb{E}[p_c|\overline{p}]+\rho\mathbb{E}[p_r|\overline{p}]$ (see also Szép et al., 2021).

Effective migration rates (neglecting drift). The LE analysis above is expected to apply only when recombination is much faster than all other processes. In particular, this requires Ls, which is the net selective disadvantage of maladapted genotypes (that are perfectly adapted to the alternative habitat), to be weaker than recombination, i.e., $Ls \ll 1/2$. However, if selection against maladapted genotypes is strong, i.e., $Ls \gtrsim 1$, sets of incoming alleles will be eliminated cooperatively before recombination can break them apart, causing allele frequencies across different loci to evolve in a coupled manner. Thus, in this regime, we must account for the effects of LD on allele frequencies by explicitly modeling multilocus dynamics (Barton 1983).

If selection per locus and migration are much stronger than drift, i.e., $1/N \ll m \sim s \ll Ls \sim 1$, then genotype frequencies essentially evolve deterministically. Further, since most incoming alleles are rapidly eliminated (for large Ls), the frequency of deleterious genotypes is low, and the probability of mating between two individuals, both carrying deleterious alleles is negligible. Under these conditions, the equilibrium frequencies $\{P_y\}$, associated with genotypes carrying $y=1,2,\ldots L$ deleterious alleles, follow a set of coupled *linear* equations:

$$P_y=\left[mw_L\binom{L}{y}\left(\frac{1}{2}\right)^L+\sum_{k=y}^{L}P_kw_k\binom{k}{y}\left(\frac{1}{2}\right)^k\right] \quad \text{where} \quad w_i=2e^{-is} \tag{3}$$

Here, $w_k=2e^{-ks}$ is the expected number of offspring of an individual carrying k deleterious alleles, and $\binom{k}{y}\left(\frac{1}{2}\right)^k$ the probability that such an individual transmits exactly y of these deleterious alleles to an offspring. Note that for s=0, any individual has an average of 2 offspring and transmits (on average) half of its genome to an offspring. Equation (3) can be solved (Appendix A), and used to calculate the (deterministic) average deleterious allele frequency

$p_{det}=\left(\sum_{y=1}^{L} y\, P_y\right)/L$:

$$p_{det}=m\,\frac{e^{-sL}}{(2-e^{-s})^{L-1}}\sum_{k=0}^{\infty} e^{-s\,k}\left(1+\frac{e^{-sk}}{2^k}-\frac{e^{-s(k+1)}}{2^k}\right)^{L-1} \tag{4}$$

One can define an effective migration rate $m_e(s,L)$ for a selected allele as that rate of migration which would cause the allele frequency at a *single* locus under migration-selection balance (with selective disadvantage s for the deleterious allele) to be equal to p_{det}, the average deleterious frequency that emerges in the multi-locus model, where immigrant genotypes carrying L such deleterious alleles are introduced at rate m per generation. Then, we have (Appendix A):

$$\frac{m_e(s,L)}{m}=\frac{e^{-s(L-1)}-e^{-sL}}{(2-e^{-s})^{L-1}}\sum_{k=0}^{\infty} e^{-s\,k}\left(1+\frac{e^{-sk}}{2^k}-\frac{e^{-s(k+1)}}{2^k}\right)^{L-1} \tag{5}$$

Following Barton and Bengtsson (1986), we can also define an effective migration rate $m_e^*(s,L)$ associated with neutral alleles which are injected into a recipient population (at rate m per generation) on an immigrant genetic background with L unlinked deleterious alleles of effect s. Then $m_e^*(s,L)/m$ is the probability that such a neutral allele eventually escapes via one or more recombination events onto a background with no deleterious alleles. For an unlinked neutral allele, this is simply (Bengtsson, 1985; see also Appendix A):

$$\frac{m_e^*(s,L)}{m}=\left(\frac{e^{-s}}{2-e^{-s}}\right)^{L} \tag{6}$$

The ratio $m_e^*(s,L)/m$, which represents the reduction in the effective migration rate of a neutral allele relative to the raw migration rate, converges to the average reproductive value (RV) of maladapted migrant individuals, as $m\to 0$ (Kobayashi et al. 2008). Here, the average RV of a migrant refers to its long-term (relative) genetic contribution to the population into which it migrates (Fisher 1930).

In the so-called infinitesimal limit: $s\to 0$, $L\to\infty$, with sL constant, i.e., assuming that a given total selective disadvantage sL is due to larger and larger numbers of loci of weaker effects, we find that both (scaled) effective migration rates, $m_e(s,L)/m$ and $m_e^*(s,L)/m$, approach e^{-2sL} (Appendix A; see also fig. S1(a) in Appendix C). Thus, in the infinitesimal limit, the effective migration rate of any allele (neutral or selected) is reduced relative to the raw migration rate by a factor equal to the RV of maladapted migrant individuals, reflecting their (lower) long-term genetic contribution.

That the average RV of a migrant is exactly e^{-2sL} in the infinitesimal limit (under rare migration) can be seen as follows: a migrant has relative fitness equal to e^{-sL}; thus, the average fitness of its immediate progeny, i.e., of an F_1 individual will be $\left(\frac{1+e^{-s}}{2}\right)^L$ (obtained by averaging $e^{-s\,X}$ over the probability of transmitting $X=0,1,\ldots L$ deleterious alleles to the offspring). Expanding in powers of s with sL constant, this is: $e^{-sL/2}(1+s^2L/8+\ldots)$, which reduces to $e^{-sL/2}$ in the infinitesimal limit. Assuming that all mating events involve only backcrossing (as expected under rare migration), second and third generation descendants would have average fitness $e^{-sL/4}$ and $e^{-sL/8}$ respectively in the infinitesimal limit, and so on. Thus, the average RV of a migrant (in the infinitesimal limit) is given by $e^{-sL}\,e^{-sL/2}\,e^{-sL/4}\,e^{-sL/8}\ldots=e^{-sL\sum_{j=0}^{\infty}1/2^j}=e^{-2sL}$. In effect, this can be obtained by simply assuming that each generation of backcrossing exactly halves the contribution of the ancestral migrant genome to the succeeding generation.

Conceptually similar arguments have also been used to derive expressions for the effective size of a population by relating it to the long-term genetic contribution of individuals in the population (Robertson 1961).

Introducing effective migration rates into the diffusion approximation. In subdivided or peripheral populations, where individual subpopulations are small, drift may be comparable to selection per deleterious allele, but much weaker than selection against immigrant genotypes (that carry many such alleles). This corresponds to a regime, where: $1/N \sim m \sim s \ll Ls \sim 1$. In this case, the evolutionary dynamics of genotypes with $\mathcal{O}(L)$ alleles will be essentially deterministic, while the dynamics of genotypes with fewer, i.e., $\mathcal{O}(1)$ deleterious alleles will be significantly perturbed by drift.

Any approximate description of the resultant allele frequency distributions must account for both these features: first, that alleles embedded in immigrant or early-backcross genotypes (that carry a large number of deleterious alleles) experience strong negative selection due to statistical association (LD) with such alleles, but are largely unaffected by drift; and second, that alleles embedded in genotypes that carry a small number of deleterious alleles are significantly affected by drift but not by LD. If we further assume that most alleles are present in genotypes with one or a few deleterious alleles and thus unaffected by LD, then the equilibrium allele frequency distribution must be close to the single locus distribution under drift-migration-selection-mutation balance, but with a *reduced* rate of migration: such a reduction would reflect selection against immigrant and early-backcross genotypes, which causes sets of deleterious alleles to be eliminated together.

To make these arguments more concrete, consider first the mainland-island case. If the rate of drift $1/N$ is much higher than the (effective) rate of migration and mutation, allele frequency distributions will be U-shaped, and the island population near fixation for one or other allele at each locus. If the expected deleterious allele frequency per locus under migration-selection-mutation-drift balance is $\mathbb{E}[p]$, then U-shaped distributions imply that the island is near fixation (on average) for the locally deleterious allele at $X \approx L\mathbb{E}[p]$ loci and for the locally beneficial allele at $L-X$ loci. Thus, island and mainland populations are near fixation for divergently selected alleles at $L-X \approx L(1-\mathbb{E}[p])$ loci, so that immigrant genotypes from the mainland experience a relative selective disadvantage on the island that is $\sim sL(1-\mathbb{E}[p])$ (on average). This suggests that the relevant reduced migration rate governing allele frequency distributions (and consequently the expected frequency $\mathbb{E}[p]$) can be approximated by $m_e(\, s, L(1-\mathbb{E}[p])\,)$. In making this approximation, we are neglecting the fact that the number of nearly-fixed differences between populations itself follows a distribution. Thus the selective disadvantage of immigrant genotypes on the island is $sL(1-\mathbb{E}[p])$ only on *average*, and has variance $\propto s^2 L$: this however, is quite small if individual loci are weakly selected (i.e., in the limit $s \to 0$ with sL constant) and can thus be neglected to a first approximation.

We use $\psi_{island}\,[\,p\,|\,\mathbb{E}[p]\,]$ to denote the allele frequency distribution on the island, conditional on the expected allele frequency $\mathbb{E}[p]$. This is simply given by eq. (1), but with the raw migration rate m replaced by the effective migration rate $m_e(\, s, L(1-\mathbb{E}[p])\,)$. We can then obtain an implicit equation for $\mathbb{E}[p]$ by using the fact that: $\mathbb{E}[p] = \int dp\, \psi_{island}\,[\,p\,|\,\mathbb{E}[p]\,]$. Then we have:

$$\mathbb{E}[p] = 2N(\mu+m_e)\frac{{}_1\tilde{F}_1\,(1+2Nm_e+2N\mu,\, 1+2Nm_e+4N\mu,\, -2Ns)}{{}_1\tilde{F}_1\,(2Nm_e+2N\mu,\, 2Nm_e+4N\mu,\, -2Ns)}, \qquad m_e = m_e(\, s, L(1-\mathbb{E}[p])\,) \tag{7}$$

where ${}_1\tilde{F}_1$ is the confluent hypergeometric function of the first kind. Equation (7) can be solved numerically to obtain $\mathbb{E}[p]$.

Now consider the case of the infinite-island population where a fraction ρ of islands belong to one habitat, and the remaining fraction to the other, with alternative alleles favoured in the two habitats at each of the L divergently selected loci. If the two habitats are simultaneously locally adapted, then individuals migrating between demes within the same habitat will have higher RV, i.e., contribute more genetic material to future generations, than individuals migrating between demes belonging to different habitats. Thus, deleterious alleles can be associated with effective migration rates that now depend on both the habitat from which the allele originates as well as the habitat into which it immigrates.

Let $\overline{p}_r$ and $\overline{p}_c$ denote the expected frequencies in the rare and common habitat respectively (of the allele which is favoured in the common habitat). If drift is strong or at least comparable to other evolutionary processes, then any deme is close to fixation for one or other allele at each locus. In this scenario, immigrants will typically carry alleles that are positively selected vis-a-vis the resident allele at some loci and negatively selected at other loci. In principle, one can calculate the deterministic introgression dynamics of such a mosaic genome, containing both positively and negatively selected alleles (Sachdeva and Barton 2018). However, as a first approximation, we will assume that all that matters is the net selective disadvantage of such a genome, which is, on average, proportional to the excess number of locally deleterious alleles that it carries, relative to a typical resident. This is $\sim sL(\overline{p}_c - \overline{p}_r)$ for individuals migrating *between* habitats, and zero for migrants *within* the same habitat.

As before, we can write down the distribution $\psi_i[\,p\,|\,\overline{p}_r, \overline{p}_c\,]$ of allele frequencies on any island within habitat i, conditional on $\overline{p}_r$ and $\overline{p}_c$, the average frequencies across the rare and common habitats respectively. Integrating over these yields $\mathbb{E}_i[p\,|\,\overline{p}_r, \overline{p}_c]$, the expected frequency within any deme in habitat i, conditional on $\overline{p}_r$ and $\overline{p}_c$. Finally, by using the fact that $\overline{p}_i = \mathbb{E}_i[p|\overline{p}_r, \overline{p}_c]$ for all i (at equilibrium), we arrive at the following coupled equations for $\overline{p}_r$ and $\overline{p}_c$:

$$\begin{aligned}\overline{p}_r &= \mathbb{E}_r[p\,|\,\overline{p}_r\,,\,\overline{p}_c] \\ &= 2N\left(m_e^{(rr)}\overline{p}_r + m_e^{(cr)}\overline{p}_c\right)\frac{{}_1\tilde{F}_1\left[1{+}2N(\,m_e^{(rr)}\overline{p}_r + m_e^{(cr)}\overline{p}_c),\ 1{+}2N(\,m_e^{(rr)} + m_e^{(cr)}),\ 2Ns\right]}{{}_1\tilde{F}_1\left[2N(\,m_e^{(rr)}\overline{p}_r + m_e^{(cr)}\overline{p}_c),\ 2N(\,m_e^{(rr)} + m_e^{(cr)}),\ 2Ns\right]}\end{aligned} \tag{8a}$$

$$\begin{aligned}\overline{p}_c &= \mathbb{E}_c[p\,|\,\overline{p}_r\,,\,\overline{p}_c] \\ &= 2N\left(m_e^{(rc)}\overline{p}_r + m_e^{(cc)}\overline{p}_c\right)\frac{{}_1\tilde{F}_1\left[1{+}2N(\,m_e^{(rc)}\overline{p}_r + m_e^{(cc)}\overline{p}_c),\ 1{+}2N(m_e^{(rc)} + m_e^{(cc)}),\ -2Ns\right]}{{}_1\tilde{F}_1\left[2N(m_e^{(rc)}\overline{p}_r + m_e^{(cc)}\overline{p}_c),\ 2N(m_e^{(rc)} + m_e^{(cc)}),\ -2Ns\right]}\end{aligned} \tag{8b}$$

$$m_e^{(rr)} = \rho\, m \qquad m_e^{(cc)} = (1-\rho)\, m \qquad m_e^{(rc)} = \rho\, m_e(s, L(\overline{p}_r - \overline{p}_c)) \qquad m_e^{(cr)} = (1-\rho)\, m_e(s, L(\overline{p}_r - \overline{p}_c)) \tag{8c}$$

Note that in this case, the allele frequency distributions and the expected frequencies depend on an effective migration *matrix* with elements $\{m_e^{(ij)}\}$, which denote the effective rate of migration from any deme in habitat i to a given deme in habitat j. For $i{=}j$, i.e., migration within the same habitat, $m_e^{(ij)}$ is assumed to be equal to the raw migration rate multiplied by the fraction of islands in the habitat of origin. For $i{\neq}j$, i.e., migration between habitats, $m_e^{(ij)}$ is approximated by the effective migration rate $m_e(s, L(\overline{p}_r - \overline{p}_c))$ times the fraction of islands in the habitat of origin (eq. (8c)). Equation (8) can be solved numerically to obtain $\{\overline{p}_r, \overline{p}_c\}$— the expected frequencies in the two habitats.

Barriers to gene flow and neutral divergence in the infinite-island population. When both habitats are locally adapted, F_{ST} at any arbitrary genomic location will be elevated due to a reduction in the effective rate of immigration into any deme, which in turn, is a consequence of the low RV of migrants from dissimilar, differently adapted habitats. Moreover, this elevation will be higher for demes in the rare as opposed to the common habitat, due to the fact that most immigrants into the former originate from the dissimilar habitat, resulting in a more severe reduction in effective immigration. Let $F_{ST}^{(r)}$ and $F_{ST}^{(c)}$ denote divergence at an unlinked neutral marker for a deme belonging to the rare and common habitats. Here, divergence is measured relative to genetic diversity at the level of the entire population. Following Slatkin (1991), we can express the expected value of various F-statistics in terms of expected pairwise coalescence times: $F_{ST}^{(r)}{=}1{-}T_r/T_{tot}$ and $F_{ST}^{(c)}{=}1{-}T_c/T_{tot}$, where T_r, T_c and T_{tot} denote the average coalescence time of two lineages, both sampled from a single deme belonging to the rare habitat (T_r), or both from a single deme within the common habitat (T_c), or each lineage sampled independently from across the entire population (T_{tot}).

Often, in practice, it is only possible to estimate neutral divergence between *pairs* of subpopulations, here labeled i and j. In this case, the expected value of $F_{ST}^{(i,j)}$ is: $1{-}\frac{(T_{ii}+T_{jj})/2}{(T_{ii}+T_{jj}+2T_{ij})/4}$, where T_{ii}, T_{jj} and T_{ij} are, respectively, the expected coalescence times for pairs of lineages both sampled from deme i, or from deme j, or one from i and one from j. As before, we must separately consider the cases where both sampled demes belong to the rare habitat or to the common habitat or one to the rare and the other to the common habitat: this results in three divergence measures $F_{ST}^{(rr)}$, $F_{ST}^{(cc)}$ and $F_{ST}^{(rc)}$

Let m_{rc} and m_{rr} denote the probability per unit time for a lineage in a given deme in the rare habitat to trace back to demes within the rare and common habitats respectively. We can define analogous backward migration rates m_{cc} and m_{cr} for lineages that originate in demes belonging to the common habitat. We can then obtain the various expected coalescence times using the structured coalescent (Appendix B), and express these in terms of the population-size-scaled migration rates $M_{rr}{=}Nm_{rr}$, $M_{rc}{=}Nm_{rc}$, $M_{cc}{=}Nm_{cc}$, $M_{cr}{=}Nm_{cr}$. As before, the key approximation is to assume that if the two habitats are differently adapted, then the backward rate for migration between habitats is attenuated by a factor that depends on the total selective disadvantage of immigrant genomes (see also Feder and Nosil (2010) for a similar approximation). More specifically, we will assume that M_{rc} and M_{cr} are proportional to $m_e^*[s, L(\overline{p}_r{-}\overline{p}_c)]$ (see eq. (9f)), where m_e^* is given by eq. (6), and $\overline{p}_r$ and $\overline{p}_c$ are obtained using eq. (8). Thus, we have:

$$F_{ST}^{(r)}{=}\frac{1}{1{+}2M_{rc}{+}2M_{rr}} \tag{9a}$$

$$F_{ST}^{(c)}{=}\frac{1}{1{+}2M_{cr}{+}2M_{cc}} \tag{9b}$$

$$F_{ST}^{(rr)}{=}\frac{1}{1{+}4M_{rc}{+}4M_{rr}} \tag{9c}$$

$$F_{ST}^{(rc)}{=}\frac{1{+}(M_{rc}{+}M_{rr}){+}(M_{cr}{+}M_{cc})}{1{+}3(M_{rc}{+}M_{rr}){+}3(M_{cr}{+}M_{cc}){+}8(M_{rc}{+}M_{rr})(M_{cr}{+}M_{cc})} \tag{9d}$$

$$F_{ST}^{(cc)}{=}\frac{1}{1{+}4M_{cr}{+}4M_{cc}} \tag{9e}$$

$$\begin{aligned} M_{rr}{=}\rho\, Nm \qquad & M_{rc}{=}(1{-}\rho)\, Nm_e^*[s, L(\overline{p}_r{-}\overline{p}_c)] \\ M_{cr}{=}\rho\, Nm_e^*[s, L(\overline{p}_r{-}\overline{p}_c)] \qquad & M_{cc}{=}(1{-}\rho)\, Nm \end{aligned} \tag{9f}$$

Note that eq. (9f) is similar to eq. (8c), but with the effective migration rate $m_e(s, L(\overline{p}_r{-}\overline{p}_c))$ for selected alleles (eq. (5)) replaced by the corresponding rate $m_e^*(s, L(\overline{p}_r{-}\overline{p}_c))$ for neutral alleles

(eq. (6)). Technically, the migration rates in eqs. (8) and (9) are different, being forward and backward migration rates respectively; however, these are equal under conservative migration.

Individual-based simulations. Individual-based simulations are used to check the prediction accuracy of the various approximations introduced above. For the mainland-island model, a single focal deme with N haploid individuals carrying L selected unlinked loci is simulated. In each generation, a Poisson number of individuals with mean Nm is replaced by mainland individuals, who carry the locally deleterious allele at each of their L loci. Mutation is then introduced by flipping the state of each locus of each individual independently with probability μ. Selection is implemented by multinomial sampling of $2N$ parents (with replacement) from the existing N individuals by choosing sampling weights to be equal to (relative) individual fitness. N offspring in the next generation are then produced by randomly pairing parents and creating offspring genotypes by independently choosing the allelic state at each locus to be the same as that of either parent with equal probability.

The simulation procedure for the island model with n_D demes is essentially the same (excluding mutation), except that migration is implemented by first removing a Poisson-distributed number of individuals (with mean Nm) from each deme to create a common migrant pool; individuals from this pool are then randomly redistributed back into the n_D demes, such that the size of each deme remains constant at N. Simulations are initialised such that there is polymorphism at each locus within each population. For this model, an additional set of L_1 neutral bi-allelic markers are simulated; these are unlinked to each other and to selected variants, and are initialized with allele frequency 0.5. Once the population has equilibrated, F_{ST} is calculated as $1-H_{within}/H_{tot}$, where H_{within} is the heterozygosity within demes, averaged over (relevant) demes and over all neutral markers, and H_{tot} is the diversity across the (relevant) combined population, averaged over all neutral markers.

Data availability. A MATHEMATICA notebook for the numerical approximations and FORTRAN code for individual-based simulations is available at `xxx`.

Results

We will first analyse the mainland-island model, focusing on how local adaptation on the island depends on Ns (the strength of selection relative to drift), Nm (the expected number of migrants per generation), and Ls (the maximum possible maladaptation load). We will then analyse local adaptation in the infinite-island setting, focusing on how Ns, Nm, Ls, and ρ (which governs the relative abundances of the two habitats) influence the expected frequency of locally beneficial alleles and the expected heterozygosity of selectively neutral alleles in the rare and common habitats. Throughout, theoretical predictions (eqs. (7)-(9) using eq. (5)-(6)) will be compared against individual-based simulations.

Mainland-island model

Figure 1(a) (main plot) shows the expected equilibrium frequency of the locally favoured allele on the island as a function of Nm, for various numbers L of divergently selected loci (various colors), for selective effect s=0.02 per deleterious allele, Ns=2 and $N\mu$=0.01. As expected, the locally favoured allele becomes less abundant with increasing migration from the mainland, for all values of L. This decrease in frequency with Nm is relatively smooth for small L; for larger L, we observe a sharp threshold for the loss of local adaptation, such that there is only a mild decrease in frequency of the locally favoured allele with increasing migration when Nm is lower

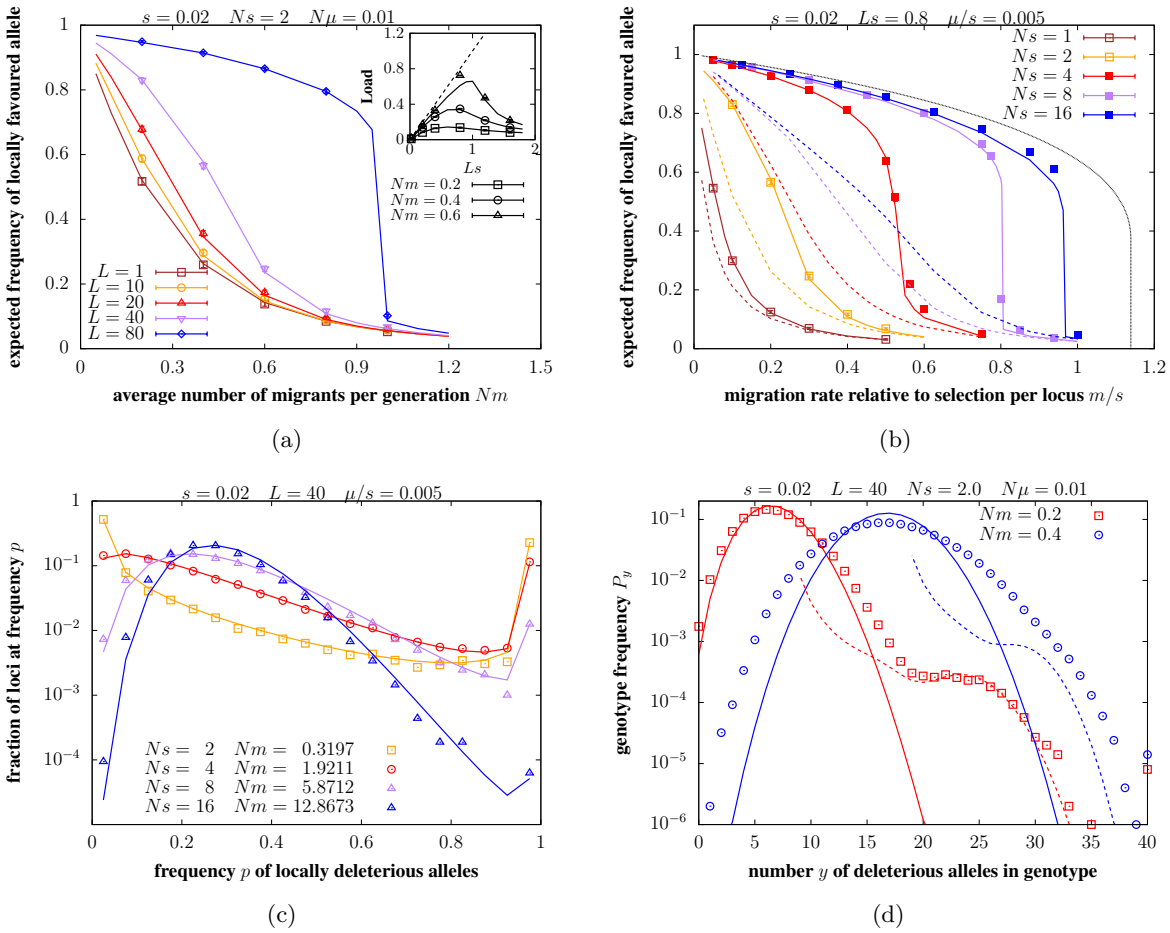


Figure 1: Local adaptation under mainland-island migration. (a) Expected frequency $1-\mathbb{E}[p]$ of the locally favoured allele on the island vs. Nm, the average number of migrants per generation, for various L (different colours) for Ns=2, s=0.02, and $N\mu$=0.01. Inset: Expected load $Ls\mathbb{E}[p]$ vs. Ls (which is varied by varying L for s=0.02) for various migration levels Nm; the maximum possible load Ls (dashed line) is also shown for reference. (b) Expected frequency of the locally favoured allele vs. Nm for various values of Ns (different colors) for μ=10^{-4}, s=0.02 and L=40. Results of individual-based simulations are depicted by symbols in both (a) and (b); solid lines show theoretical predictions that account for both LD and drift (obtained from eq. (7) together with eq. (5)); dashed lines in (b) show LE predictions that only account for drift (obtained from eq. (1)); the dotted black line in (b) shows the deterministic prediction in the $L\to\infty$ limit (with Ls=0.8) that only accounts for LD (see text). (c) Distribution of allele frequencies, i.e., the probability that the frequency of the locally deleterious allele at a locus lies in the interval $[p, p+\Delta p]$ vs. p (for Δp=0.05). The different colors correspond to different population sizes N=100, 200, 400, 800 (so that Ns=2, 4, 8, 16 for s=0.02), with Nm chosen in each case such that the expected frequency of the locally deleterious allele is $\mathbb{E}[p]$=0.3. Theoretical predictions (lines) are in agreement with the results of individual-based simulations (symbols); however, some deviation can be observed for larger values of Nm. (d) Genotype frequencies P_y, which represent the probability that a randomly chosen genotype in the population carries y deleterious alleles, vs. y, for two different values of Nm, for Ns=2, s=0.02. Symbols depict results of individual-based simulations; solid lines show deterministic predictions and dashed lines predictions under LE (see text for more details about the two kinds of predictions). Other parameters for (c) and (d) are: L=40 and $N\mu$=0.01.

than this threshold. Accordingly, for low Nm, the expected frequency of the locally favoured allele increases with increasing L: this is consistent with the idea that when large numbers of loci are under divergent selection, *net* selection against migrants and their immediate descendants is so strong that most locally deleterious alleles are collectively eliminated before they can escape via backcrossing onto fitter backgrounds.

The inset of fig. 1(a) shows the expected maladaptation load $Ls\mathbb{E}[p]$ vs. Ls, where Ls is varied by varying L, while holding s=0.02 constant. Here, $\mathbb{E}[p]$ is the expected deleterious allele frequency, and Ls signifies the maximum possible load (shown by a dashed line in the plot): this corresponds to a situation with $\mathbb{E}[p]$=1, i.e., if the island is fixed for locally deleterious alleles at all divergently selected loci. Figure 1(a) (inset) shows that the expected load is maximum for intermediate Ls: this non-monotonic dependence emerges because for small L, the increase in the number of loci contributing to maladaptation far outweighs the milder decrease in maladaptation per locus (i.e., in $\mathbb{E}[p]$) with increasing L. However, for larger L, $\mathbb{E}[p]$ decreases more sharply than $\sim 1/L$ with L: this sharp decrease in maladaptation per locus (due to the co-operative elimination of deleterious alleles) compensates for the increase in L, causing total load to decline with L for large L.

Note that our focus on the composite parameter Ls (rather than on L and s separately) is motivated by the consideration that for highly polygenic architectures, i.e., for large L and small s, effective migration rates depend on L and s only via Ls. The sensitivity of allele frequencies to the exact genetic basis of load, i.e., to L and s for given Ls is explored in Appendix C (SI): selection against deleterious alleles is found to be most efficient in the infinitesimal limit, i.e., if the same total selective disadvantage is due to a very large number of loci of very weak effects. However, this dependence on L is significant only for $Ls \gtrsim 1$ (fig. S1(b), Appendix C, SI).

A striking feature of fig. 1(a) is the close agreement between the results of individual-based simulations (symbols) and theoretical predictions (lines) obtained by numerically iterating eq. (7) (using eq. (5)). This rather successful approximation of the effects of LD on allele frequencies in terms of effective migration rates suggests the following explanation for the emergence of sharp thresholds for loss of local adaptation at large Ls: a small increase in maladaptation at very many loci may, in aggregate, cause a substantial increase in the effective migration rate $m_e \approx m e^{-2Ls(1-\mathbb{E}[p])}$ associated with individual alleles. This, in turn, further increases swamping, pushing the frequency of the locally deleterious allele up, setting in motion a positive feedback, which culminates in collapse of local adaptation above a threshold migration rate.

Consider next how local adaptation is influenced by drift, i.e., by the the size of the island population, for a given genetic architecture of hybridisation load, i.e., for given L, s and μ. Figure 1(b) shows the expected frequency of the locally favoured allele vs. m/s, the strength of migration relative to the strength of selection per locus, for various Ns (various colors). As before, there is close agreement between individual-based simulations (symbols) and theoretical predictions (solid lines) across almost all parameter combinations, with some apparent (moderate) deviation only for $Nm \gtrsim 10$. Larger sizes (which translate into higher values of Ns) allow populations to sustain local adaptation despite higher levels of migration; thus, parameter combinations shown in fig. 1(b) correspond to a regime where drift substantially degrades local adaptation. This can also be seen by noting that expected allele frequencies approach the deterministic prediction p_{det} (black dotted line in fig. 1(b)) as Ns increases, but are still significantly lower than p_{det} even for the largest value of Ns. Here, p_{det} is obtained by solving for $p_{det}=(m/s)e^{-2Ls(1-p_{det})}$: note that this deterministic prediction accounts for the effects of LD between incoming unlinked deleterious alleles (which reduces the effective migration rate of any such allele), but not of drift.

Increasing Ns also results in sharper thresholds for loss of local adaptation: this is consistent with the fact that larger populations have higher mean fitness as locally deleterious alleles are

less likely to fix. Thus, immigrant genomes are at greater selective disadvantage with respect to the residents in larger populations (all else being equal), causing effective immigration rates of deleterious alleles to be lower then they would be in smaller populations. A related observation is that the expected frequency of the locally favoured allele (symbols/solid lines) is much higher than the corresponding LE prediction (dashed lines) for larger values of Ns. Recall that the LE predictions account for the effects of drift but not of LD, implying that LD between immigrant alleles has a more pronounced effect on allele frequency dynamics in larger populations.

At first glance, it is surprising that this simple heuristic of introducing effective migration rates into the diffusion approximation should predict the expected allele frequency so accurately across a wide range of parameters, including in large populations, where there may be substantial polymorphism at higher Nm, so that one or other allele is not necessarily close to fixation. This behoves us to ask: how sensitive are these approximations to assumptions about the (U-)shape of the underlying allele frequency distributions? We can investigate this by comparing allele frequency distributions in populations of different sizes (fig. 1(c)), choosing the migration level in each case to be such that the expected deleterious allele frequency is the same (here $\sim$0.3 based on the theoretical prediction).

As expected, the allele frequency distribution is U-shaped for smaller values of Nm, but becomes unimodal, i.e., is centered around $\mathbb{E}[p]\sim 0.3$ for larger Nm; accordingly, the probability of observing intermediate frequencies also increases as Nm increases. Interestingly, theoretical predictions for the allele frequency distribution (lines) are reasonably accurate even for larger values of Nm (purple and blue triangles in fig. 1(c), corresponding respectively to $Nm\sim 6$ and $Nm\sim 13$), for which allele frequency distributions deviate markedly from the canonical U-shape, and are characterised by high heterozygosity ($\mathbb{E}[2pq]\sim 0.355$ and 0.386 respectively). Thus, this rather crude approximation of multi-locus dynamics via a single effective migration rate that only depends on the expected number of genetic differences between the mainland and island appears to suffice even if there is substantial polymorphism at each locus, i.e., if these differences represent pairwise differences between individuals rather than fixed differences between populations.

Finally, we ask: do the approximations introduced here also predict the population distribution of multi-locus genotypes carrying different numbers $y=0, 1, 2, \ldots L$ of deleterious alleles? Figure 1(d) shows equilibrium genotype frequencies P_y vs. y for two values of Nm, as found in simulations (symbols), along with two kinds of analytical predictions (solid and dashed lines). The solid line represents the predicted genotype frequencies *under LE* given the expected deleterious allele frequency $\mathbb{E}[p]$ (which is determined as described above). Under LE, the allelic states of different loci in any genotype are independent Bernoulli random variables, and the probability of genotypes with exactly y deleterious alleles is simply $P_y^{(LE)}=\binom{L}{y}(\mathbb{E}[p])^y(1-\mathbb{E}[p])^{L-y}$.

We also compare against the *deterministic* prediction for P_y (dashed lines in fig. 1(d)) obtained by solving eq. (3) (see also eq. S1 in Appendix A): more specifically, the dashed lines show $P^{(det)}_{y-L\mathbb{E}[p]}$ vs. y, where the deterministic frequencies $P^{(det)}$ are calculated by assuming that the island is subject to divergent selection at $L(1-\mathbb{E}[p])$ loci. Note that the relevant deterministic frequencies are $P^{(det)}_{y-L\mathbb{E}[p]}$ and not $P^{(det)}_y$ as genotypes carrying y deleterious alleles have relative (Malthusian) fitness that is proportional to $y-L\mathbb{E}[p]$ (since the island is nearly fixed for $L\mathbb{E}[p]$ mainland (deleterious) alleles, by definition).

At the lower migration level (red plot in fig. 1(d)), genotype frequencies P_y are close to the LE prediction (solid line) for small y, while they match the deterministic prediction (dashed line) for large y, with a crossover between the LE and deterministic regimes at intermediate y. This suggests that when migration is low and deleterious alleles correspondingly rare, the frequencies of genotypes with large numbers of deleterious alleles (large y) are governed by the

(deterministic) balance between migration, selection and recombination, with recombination only serving to break down highly deleterious genotypes into smaller, less deleterious fragments, but rarely bringing together such fragments to reconstitute the more deleterious genotypes (this is tantamount to assuming that $\{P_y\}$ satisfy *linear* coupled equations, as in eq. (3)). In contrast, both roles of recombination— the splitting of more deleterious genotypes to generate the focal genotype and the reconstitution of the focal genotype via recombination between less deleterious genotypes, appear to play a role in shaping the frequencies of genotypes with low numbers of deleterious alleles (small y): this is reflected in the fact that P_y for small y are close to the LE prediction, which assumes that genotypes are random assortments of independently segregating alleles. We note that P_y are slightly elevated above the LE prediction for very small y, reflecting positive selection on such genotypes due to their higher relative fitness.

At the higher migration level (blue plot), the distribution does not exhibit two distinct regimes. In particular, P_y for large y are much higher than the deterministic prediction, suggesting that when migration is high, and deleterious genotypes more common, mating events which bring together different sets of deleterious alleles to generate highly deleterious genotypes are more relevant (accounting for such events would require us to also include in eq. (3) terms of the kind $P_y P_{y'}$). We further note that the distribution of P_y for smaller y is significantly wider than the LE prediction in this case, suggesting significant selection not just on individual alleles but also sets of alleles.

The fact that allele frequency distributions are nevertheless accurately predicted by these approximations even at high Nm values suggests that the heuristic of effective migration rates captures the *gross* effects of LD on allele frequencies quite robustly, even when it does not provide a good handle on LD (or equivalently, on genotype frequencies) itself.

Infinite-island model

Let us now consider local adaptation across divergently selected habitats in the infinite-island setting, where a non-zero fraction ρ of islands belong to the rare habitat. In this case, both habitats influence each other: thus, we need to consider how allele frequencies in the different habitats co-evolve, instead of assuming the state of the more abundant habitat to be 'fixed' and independent of the rare habitat, as in the $\rho \to 0$ (mainland-island) limit.

Figure 2 shows the expected frequency of the locally beneficial in the rare habitat, as a function of Nm, the average number of migrants per island per generation, for different Ns (top vs. bottom rows), different ρ (right vs. left columns) and different L (different symbols in each subfigure). As before, the theoretical predictions for the infinite-island model (lines), obtained by numerically solving eq. (8) in conjunction with eq. (5), are in good agreement with the results of individual-based simulations (symbols) across all parameter combinations.

In the absence of mutation, there is a well-defined critical migration rate m_c, such that locally adaptive alleles segregate at non-zero frequencies in the rare habitat for $m<m_c$, but are eliminated for $m>m_c$. In other words, there is no adaptive divergence between habitats for $m>m_c$, i.e., alleles favoured in the common habitat fix across all islands, irrespective of habitat, at least under symmetric selection ($s_c=s_r=s$). Figure 2 shows that m_c increases with Ns and ρ: this simply means that local adaptation in the rare habitat can be sustained over a wider range of migration rates when individual subpopulations are larger (so that selection against deleterious alleles is more effective), and when the relative abundances of the two habitats are more similar (so that a smaller fraction of genotypes immigrating into the rare habitat are deleterious). For the parameter values considered in fig. 2, the critical threshold is approximately the same for L=10 and L=40, and thus is not very sensitive to Ls, the total selection difference between habitats. As discussed below, this lack of sensitivity is only observed at relatively low values of

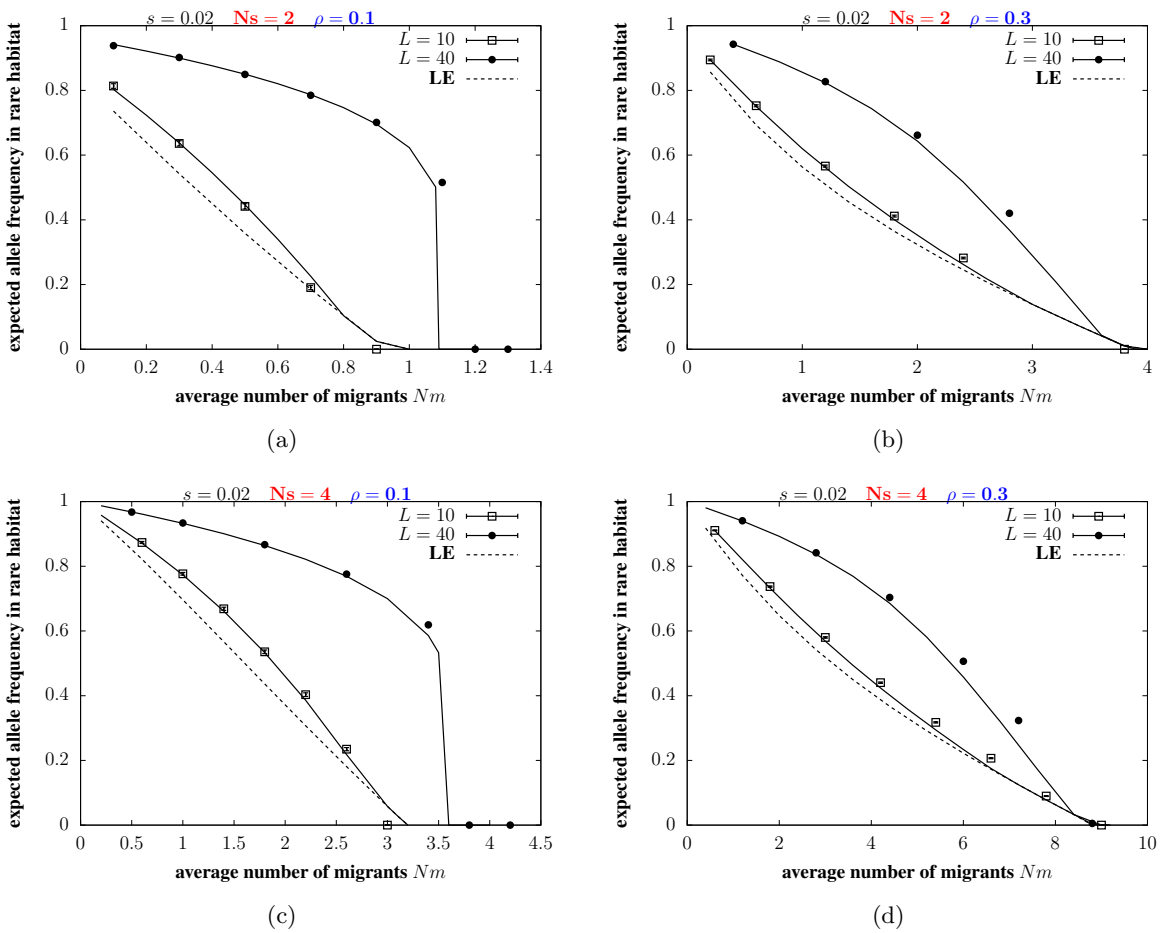


Figure 2: Local adaptation in the infinite-island model with two habitats. Plots show the expected allele frequency in the rare habitat, for alleles that are favoured locally in that habitat, vs. Nm for (a) Ns=2, ρ=0.1 (b) Ns=2, ρ=0.3 (c) Ns=4, ρ=0.1 (d) Ns=4, ρ=0.3, where ρ denotes the fraction of islands in the rare habitat. Empty squares depict simulation results for L=10 and filled circles for L=40 in each plot; solid lines depict theoretical predictions that account for both LD and drift (obtained from eq. (8) together with eq. (5)); dashed lines depict LE predictions that only account for drift (obtained from eq. (2)). Selective effect per deleterious allele is s=0.02 in all plots. The number of simulated demes is 500 in all individual-based simulations.

Ls, with m_c increasing sharply with Ls above a threshold $(Ls)_*$ (fig. 4).

Note that higher values of Ls (which correspond here to higher L) result in significantly sharper thresholds for loss of adaptation across all parameter combinations, even when the critical migration rate m_c (at which adaptation collapses) remains unchanged. As before, sharper thresholds at larger Ls reflect the fact that selection against deleterious alleles has a larger indirect (i.e., LD -mediated) component, or concomitantly, that reductions in frequencies of adaptive alleles at different loci are more strongly coupled, when Ls is larger.

We also note that thresholds are sharper for smaller ρ. This can be rationalised as follows: sharp thresholds emerge when a small increase in migration causes a substantial enough drop in population fitness (or increase in load) to significantly increase the effective migration rate of deleterious alleles immigrating from the alternative habitat, setting in motion a positive feedback between declining population fitness and rising (effective) maladaptive immigration into the population. However, this kind of positive feedback only affects alleles migrating between differently adapted habitats, and not deleterious alleles that migrate within the same habitat (see also eq. (8c)). Thus, feedback effects are stronger and the threshold for loss of

local adaptation on an island within the rare habitat sharper if immigration into the island is predominantly from the alternative habitat (as is the case if ρ is small).

Barriers to gene flow and neutral divergence. Let us now consider how adaptive divergence at selected loci influences patterns of genetic diversity at unlinked neutral markers. In the absence of such divergence, F_{ST} depends only on Nm, the average number of migrants exchanged per generation between any island and the larger population in which it is embedded. Here, Nm may be viewed as a proxy for the physical subdivision of the population; under the neutral infinite-island model, such subdivision affects all subpopulations equally. However, when habitats are locally adapted, F_{ST} at neutral markers depends not only on the number of migrants exchanged but also on the RV of migrants, which is their long-term contribution to the neutral gene pool of the population into which they immigrate. As shown in this section (fig. 3), this dependency is captured accurately by expressing F_{ST} in terms of the elements of an effective migration rate matrix for neutral alleles (eq. (9)) and approximating effective migration rates as in eq. (6).

The focus on *unlinked* neutral markers is motivated by the consideration that reduced gene flow at such markers is more representative of the buildup of *genome-wide*, as opposed to localized, barriers to gene flow, and is thus a more appropriate measure of RI between habitats. However, as shown by Barton and Bengtsson (1986), even for neutral markers that are tightly linked to a particular barrier locus, it is unlinked barrier loci that contribute the most to the net reduction in effective migration, if hybrid dysfunction is due to very many loci of weak effects, spread across a sufficiently long genetic map.

Figure 3(a) shows average $F_{ST}^{(r)}$ and $F_{ST}^{(c)}$, i.e., the average F_{ST} for islands that belong respectively to the rare and common habitat, as a function of Nm. The expected F_{ST} in the absence of local adaptation, given by $1/(1+2Nm)$, is also shown (red line). As before, theoretical predictions (lines, obtained from eqs. (9a), (9b), (9f), in conjunction with eq. (6)) are in close agreement with simulations (symbols). We further note that adaptive divergence between habitats increases neutral F_{ST} in both habitats above the neutral expectation (red line), but that this increase is more pronounced for the rare habitat. This can be understood by noting that $F_{ST}^{(r)}$ and $F_{ST}^{(c)}$ measure respectively the extent to which neutral diversity within any deme belonging to the rare or common habitat is reduced relative to gene diversity at the level of the population as a whole. The neutral diversity within demes, in turn, reflects the effective rates of immigration into demes in either habitat. When both habitats are locally adapted, the majority (i.e., a fraction $1-\rho$) of immigrants into the rare habitat have very low RV, as they originate from the common habitat and thus carry genotypes that are locally deleterious. By contrast, only a minority (i.e., a fraction ρ) of immigrants into the rare habitat have low RV. This results in a stronger reduction in neutral diversity and concomitantly, a sharper increase in F_{ST} for islands in the rare habitat compared to the common habitat.

We also consider F_{ST} between pairs of islands embedded in the population. Figure 3(b) shows theoretical predictions (lines) and simulation results (symbols) for average $F_{ST}^{(rr)}$, $F_{ST}^{(rc)}$ and $F_{ST}^{(cc)}$, which correspond respectively to F_{ST} between a pair of islands where both belong to the rare habitat, where one belongs to the rare and the other to the common habitat, and where both belong to the common habitat. Note that $F_{ST}^{(rr)}>F_{ST}^{(rc)}>F_{ST}^{(cc)}$: thus, at low Nm, F_{ST} between two islands in the rare habitat is *greater* than F_{ST} between islands from different habitats, despite the fact that only the latter experience divergent selection and exhibit adaptive divergence. This somewhat paradoxical observation can be explained by noting that the rare habitat, though rare, will nevertheless be large enough to harbour high genetic diversity (at the level of the habitat as a whole), as long as it encompasses a finite fraction of all islands in an

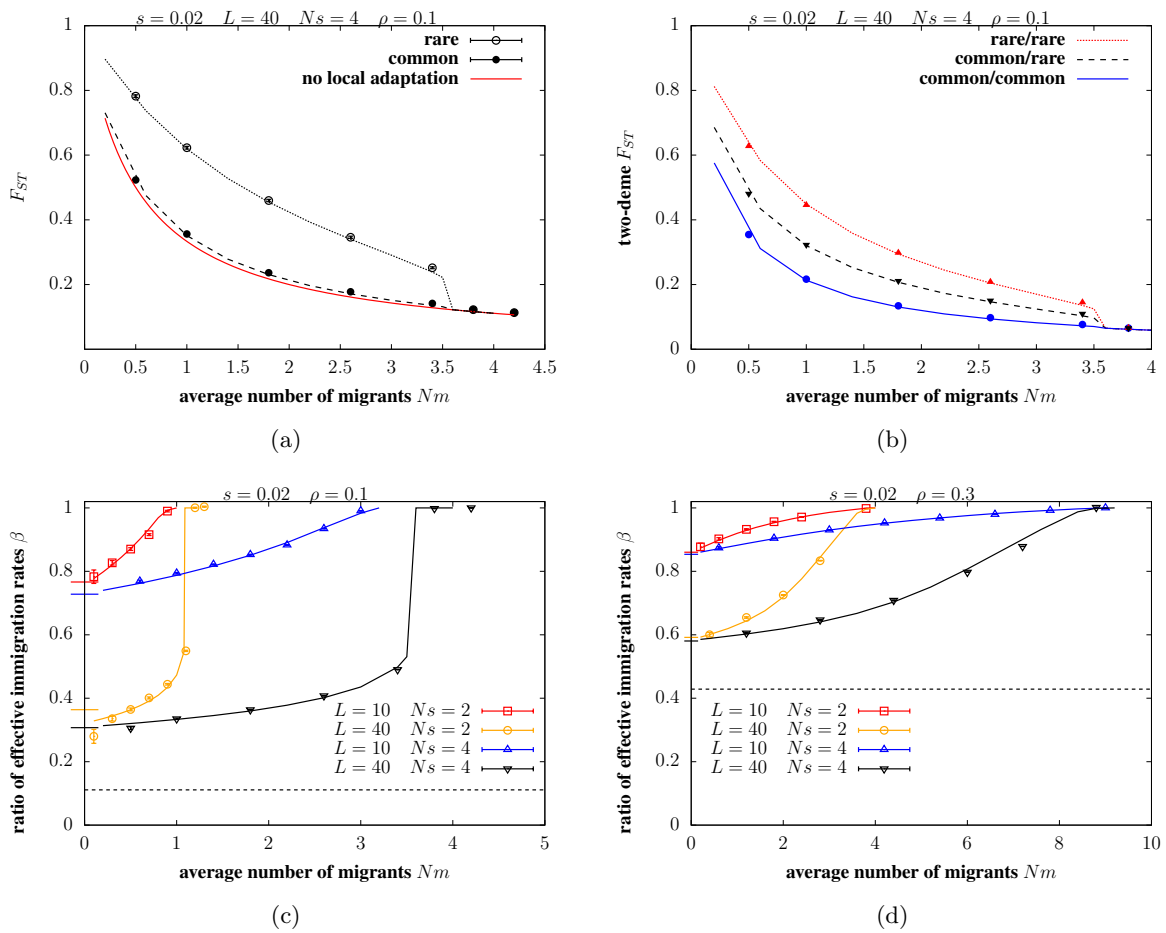


Figure 3: Neutral divergence in the infinite-island model. (a) F_{ST} in the rare and common habitats, i.e., $F_{ST}^{(r)}$ (empty circles) and $F_{ST}^{(c)}$ (filled circles) vs. Nm, for s=0.02, L=40, Ns=4, ρ=0.1. Symbols represent simulation results; black lines represent theoretical predictions (obtained from eq. (9a),(9b),(9f) together with eq. (6)); the red line represents F_{ST}=1/(1+2Nm): the prediction in the absence of local adaptation. (b) F_{ST} for pairs of demes both belonging to the rare habitat ($F_{ST}^{(rr)}$), or both to the common habitat ($F_{ST}^{(cc)}$), or one to the rare and the other to the common habitat ($F_{ST}^{(rc)}$), vs. Nm, for the same parameters as in (a). The plots show simulation results (symbols) as well as theoretical predictions (lines: obtained from eq. (9c)-(9f) together with eq. (6)). (c)-(d) $\beta=\frac{\left(1/F_{ST}^{(r)}\right)-1}{\left(1/F_{ST}^{(c)}\right)-1}$ vs. Nm for (c) ρ=0.1 and (d) ρ=0.3, for different values of L and Ns (depicted by the different colors) and s=0.02. The quantity β gives the ratio of the effective number of immigrants per generation into islands within the rare habitat to the corresponding number for islands within the common habitat. Theoretical predictions (solid lines) match simulation results (symbols) across all parameter combinations. The black dashed line in each plot represents the threshold $\beta=\rho/(1-\rho)$, which is the expected value of β under complete RI (wherein immigrants from the dissimilar habitat have zero RV). The short horizontal coloured lines along the vertical axis represent the analytical $Nm\to 0$ prediction for β under the infinite-island model (eq. (10)). F_{ST} values in simulations are computed by averaging over 40 unlinked neutral markers.

'infinite-island' population. Thus, any two islands within the rare habitat exhibit high neutral divergence with respect to each other (since they will each be close to fixation for a random set of alleles segregating in the habitat as a whole), even as the diversity within each is strongly reduced because of very low net effect immigration. By contrast, islands that support the common habitat have fairly high within-island diversity as well as between-island divergence, causing F_{ST} values to be lower for any pairwise comparison that includes such islands.

Thus, differences in F_{ST}, for islands within the rare vs. common habitat, reflect the under-

lying asymmetry in the effective levels of immigration that the two kinds of islands experience, when both are at least partially locally adapted. This asymmetry, in turn, is due to the fact that the majority of immigrants into the rare habitat have low RV, while this is true of only a minority of immigrants into the common habitat. A direct measure of this asymmetry is provided by the quantity $\beta=\frac{\left(1/F_{ST}^{(r)}\right)-1}{\left(1/F_{ST}^{(c)}\right)-1}$, which gives the ratio of the effective number of immigrants per generation into islands within the rare habitat to the corresponding number for islands within the common habitat.

In the absence of local adaptation (i.e., if alleles favoured in the common habitat are fixed across the entire population at all divergently selected loci), the relative fitness of all immigrants between and within habitats is equal to that of residents for any island. This means that the effective number of immigrants into any island is equal to the actual number of migrants, irrespective of the habitat of the island (at least under soft selection), so that $\beta=1$. In the opposite limit, where both habitats are perfectly adapted and $sL\gg1$, individuals who migrate between habitats have RV close to zero, while individuals who migrate within a habitat have RV close to 1. Thus, in this case (which corresponds to nearly complete RI), the effective number of immigrants into any island is essentially equal to the number of immigrants that originate from other islands within the same habitat. Thus, in this limit, we have $\beta=\rho/(1-\rho)$.

Figs. 3(c) and 3(d) show β vs. Nm for $\rho=0.1$ (fig. 3(c)) and $\rho=0.3$ (fig. 3(d)), for various combinations of Ns and L (which are chosen to be the same as in fig. 2). As before, symbols represent results of individual-based simulations, while lines represent theoretical predictions; the horizontal dashed lines represent $\beta=\rho/(1-\rho)$, which is the expectation under complete RI. As expected, β increases as Nm increases and adaptive divergence between habitats decreases. Note also that β lies above the dashed line even for the smallest value of Nm that we simulate, indicating that RI is incomplete, even at this level of migration.

We can obtain an explicit (approximate) expression for β in the limit of very weak migration for the infinite-island model, by noting that $\overline{p}_c-\overline{p}_r$ approaches a migration-independent value as $Nm\to0$ (Szép et al. 2021). There is a critical threshold $Ns_c=\frac{1}{2}\log\left(\frac{1-\rho}{\rho}\right)$, such that no adaptive divergence is possible (i.e., $\overline{p}_r=\overline{p}_c$) even for very weak migration, for $Ns<Ns_c$. However, for $Ns>Ns_c$, the allele frequency difference $\overline{p}_c-\overline{p}_r$ approaches a non-zero value Δ as $Nm\to0$ (see below; also eq. 5 of Szép et al.,2021). Interestingly, in an *infinite*-island setting, divergence is incomplete even with very weak migration in , i.e., we have $\Delta<1$, due to a net bias towards alleles favoured in the common habitat.

Approximating the reduction in the effective migration rate between habitats by $e^{-2Ls(\overline{p}_c-\overline{p}_r)}$, we have:

$$\begin{aligned}\lim_{Nm\to0}\beta &\approx \frac{\rho+(1-\rho)\,e^{-2sL\Delta}}{\rho\,e^{-2sL\Delta}+1-\rho}\\ \Delta=\lim_{Nm\to0}\overline{p}_c-\overline{p}_r &= \frac{1}{1+\frac{1-\rho}{e^{2Ns}\rho-(1-\rho)}-\frac{\rho}{\rho-e^{2Ns}(1-\rho)}}\end{aligned} \tag{10}$$

These $Nm\to0$ predictions for β are depicted by short horizontal lines along the vertical axis in figs. 3(c) and 3(d). Simulations results for the lowest levels of migration ($Nm\sim0.2-0.4$) are consistent with this prediction. Note that the above prediction relies on the infinite-island assumption and is sensitive to the order in which we take the two limits $n_D\to\infty$ and $Nm\to0$. Thus, in simulations with large but finite n_D, there will be a (low) value of Nm, below which the infinite-island prediction breaks down.

Note that the typical values of F_{ST} found here— $F_{ST}^{(rc)}$ in the range $0.2-0.6$ for $Nm<2$ (fig. 3(b)), correspond to the so-called grey zone of speciation (Roux et al. 2016), where it may be

difficult to infer effective migration rates and the extent of RI from sequence data. The analysis here shows that such values of F_{ST} are consistent with incomplete RI (and thus low levels of ongoing gene flow between habitats), at least in the context of patchy populations.

Discussion

This paper introduces a simple heuristic for approximating the combined effects of LD and genetic drift on allele frequencies, when multiple genes are under divergent selection across habitats. The heuristic relies on the *separation of timescales* that emerges naturally when local adaptation is polygenic: local adaptation in a rare habitat requires that selection per locus be at least as strong as drift and migration, i.e., $1/N \sim s \sim m$, when loci are unlinked (Szép et al. 2021). However, drift must then be much weaker than net selection against introgressing genotypes that carry multiple deleterious alleles, i.e., $1/N \sim s \ll Ls$. This allows us to use fundamentally different mathematical descriptions for multi-locus and single-locus evolutionary dynamics, which we can then splice together by incorporating effective migration rates (obtained from a consideration of deterministic multi-locus dynamics) into the diffusion approximation (which accounts for the effects of drift at individual loci).

Interestingly, even though the deterministic calculations of effective migration rates assume weak migration, in effect assuming that deleterious genotypes are too rare to encounter each other during mating, the approximations developed in this paper, which incorporate drift, apply over a much wider range of conditions. In particular, we can accurately predict critical migration thresholds, at which local adaptation is lost, across nearly all simulated parameter combinations. The key idea that allows us to extend the approximations to scenarios where the average frequencies of locally deleterious alleles are high is as follows: in small populations, the loss of local adaptation occurs via successive (quasi-)fixations of deleterious alleles, rather than small shifts in allele frequencies across multiple polymorphic loci. Thus, even at high migration levels, i.e., even for low to moderate divergence between the rare and common habitats, locally deleterious alleles will still be quite rare within any subpopulation in the rare habitat at the *subset* of divergently selected loci that actually contribute to divergence between habitats. This allows us to use effective migration rates (calculated by assuming low deleterious allele frequencies) by assuming that the effect of increasing migration is to primarily reduce the effective number of barrier loci contributing to divergence rather than the level of divergence per locus. Note that this condition of small population sizes is not as restrictive as it seems: the approximations remain reasonably accurate at fairly large Nm, for which intermediate-frequency polymorphisms are common (figs. 1(b), 1(c))

The fact that the effects of LD between sets of divergently selected alleles are captured by a single quantity— the deterministic effective migration rate, which depends only on the difference in the average (Malthusian) fitness of two hybridizing populations (for small s and large L), has several implications. First, the interpretation of effective migration rates in terms of RV (see above) suggests that at least in the infinitesimal or highly polygenic limit, where selection acts on large numbers of weak-effect loci across populations, the gross effects of multi-locus LD on allele frequencies may depend on very few quantities, e.g., the mean fitness of F_1 individuals and their within-family variance, even with an arbitrary effect size distribution (Barton and Etheridge 2011). Generalizing approximations based on effective migration rates to unequal effect sizes and arbitrary dominance (for diploids) is a promising direction as it will allow us to understand how hybridisation outcomes are influenced by LD between loci under different kinds of selective constraints, e.g., loci under divergent selection (which may generate hybridisation load) and those under background selection (which may contribute to heterosis).

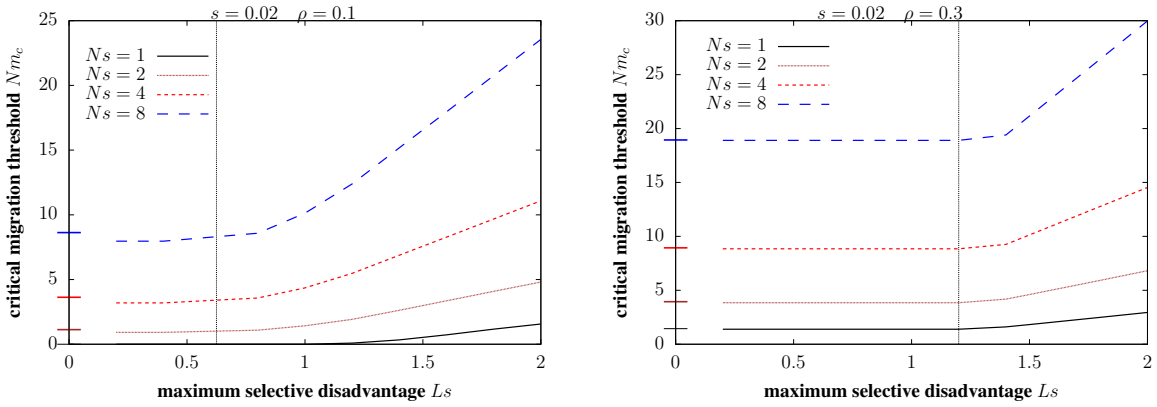
Second, the fact that the effects of LD are essentially summarised by a deterministic effective

migration rate suggests that migration is the main source of (positive) LD, at least between *unlinked* divergently selected loci, with negative LD due to drift (i.e., Hill-Robertson type effects) playing only a minor role. Note that this does not imply that drift has no effect on allele frequencies, but that these effects operate essentially at the level of individual loci. In particular, direct selection on deleterious alleles is more effective in larger populations (i.e., Ns greater), resulting in higher levels of adaptation and higher population fitness, which causes introgressing deleterious genotypes (whose relative fitness is consequently lower) to be eliminated more efficiently, thus further increasing adaptation. Thus, in larger populations there is a stronger positive feedback between the efficacy of selection at *individual* loci (which is impaired by drift) and the efficacy of LD-driven elimination of *groups* of introgressing deleterious alleles (which is largely unaffected by drift).

These observations raise the following question: when does the separation of timescales between the effects of drift and those of multi-locus selection no longer hold, so that approximations based on splicing deterministic effective migration rates with the diffusion approximation break down? In general, we expect this to occur only if the elimination of multi-locus deleterious genotypes (by selection or recombination) is sufficiently slow that genetic drift becomes relevant on the same time scales; however, it is far from obvious whether local adaptation would then be possible at all. For instance, linkage between sets of locally deleterious alleles causes them to remain associated with each other over longer periods, but this also allows selection to eliminate such sets more efficiently. Thus, for sets of locally deleterious alleles to persist in the population, not only must linkage between them be tight, but also multi-locus selection correspondingly weak and of the same order as drift: however, the combined effects of drift and migration would then easily overwhelm selection at individual loci, causing alleles favoured in the common habitat to fix throughout the population (Szép et al. 2021). This suggests that approximations based on incorporating deterministic effective migration rates (for linked loci; see Barton and Bengtsson, 1986) into the diffusion framework may also apply for linked loci at migration-selection-drift-recombination equilibrium, at least under conditions that allow for local adaptation. Extending the present analysis to include linkage will also allow us to assess the relative importance of divergence hitchhiking vs. genomic hitchhiking (see e.g., Feder et al. (2012)) in small populations in a systematic way, thus facilitating a better understanding of how drift influences genomic signatures of local adaptation.

A second factor that could potentially alter the separation of timescales associated with single-locus and multi-locus dynamics is epistasis, i.e., the non-multiplicative action of deleterious alleles. For a given F_1 fitness, diminishing-returns epistasis between deleterious alleles (wherein selection per deleterious allele is weaker when many such alleles act in combination) tends to further strengthen the barrier to gene flow, while synergistic epistasis (alleles more deleterious when in combination) weakens the barrier relative to multiplicative selection (Barton and Bengtsson 1986). This can be understood by noting that the fitness of genotypes with an intermediate number of deleterious alleles is lower in the former case (see fig. 5 of Barton and Bengtsson, 1986). While the prevalence of diminishing-returns epistasis in nature is not entirely clear (Halligan and Keightley 2009), it represents an important scenario (associated with potentially strong barrier effects) where our approximations may break down due to the relatively similar levels of selection on genotypes with a few vs. many deleterious alleles. Thus, generalizing the present analysis to arbitrary patterns of epistasis, including various forms of cryptic epistasis (which can generate strong RI even in the absence of strong ecological differentiation; see Blanckaert et al., 2020), remains an interesting direction for future work.

A key limitation of this approximation is that while it accurately captures the effects of LD on allele frequencies, it provides little insight into the nature of multi-locus LD itself, and more generally, fails to predict the distribution of genotype frequencies. At low levels of migration,



(a) (b)

Figure 4: Theoretical predictions for the critical migration threshold Nm_c, at which local adaptation collapses in the rare habitat, vs. Ls for (a) ρ=0.1 (b) ρ=0.3, for s=0.02 and various values of Ns. Theoretical predictions are obtained by numerically solving eq. (8) and using eq. (5). The short horizontal coloured lines along the vertical axis represent the LE prediction $Nm_c \approx \frac{Ns - \frac{1}{2}\log\left[\frac{1-\rho}{\rho}\right]}{1-2\rho}$ (see Szép et al.,2021). The vertical dotted lines represent the deterministic prediction for $(Ls)_*$ (see text), beyond which m_c/s starts increasing with Ls.

genotype frequencies of strongly deleterious multi-locus genotypes are close to deterministic predictions obtained by assuming segregating deleterious genotypes to be rare (fig. 1(d)). However, at higher migration, i.e., as deleterious genotypes become more common, this is no longer true. The fact that effective migration rates (calculated under the rare genotypes assumption) nevertheless predict the effects of LD on allele frequencies, even where they underestimate frequencies of multi-locus genotypes is thus somewhat surprising, and requires further investigation.

In summary, our approximations provide a quantitative description of 'genome-wide congealing' (Feder et al. 2014) in small populations, accurately capturing the sharp migration thresholds for loss of local adaptation that can emerge when selection is strongly divergent across populations ($Ls \sim 1$), even when individual loci are only weakly selected and significantly perturbed by drift ($Ns \gtrsim 1$). They show that thresholds are sharper when Ls (which governs net selection against maladapted genotypes) and Ns (which governs the strength of selection per locus relative to drift) are higher (figs. 1(a), 1(b) and 3), and when ρ (which governs the relative abundances of the two habitats) is lower (fig. 3).

In addition to generating sharper thresholds for loss of adaptation, LD between sets of locally adaptive (or deleterious) alleles may, under certain conditions, *shift* the critical migration threshold itself (see e.g., figs. 3(a), 3(c)). To explore when LD results in shifted (in addition to sharper) thresholds, it is useful to plot the critical migration threshold Nm_c, at which local adaptation is lost, vs. Ls. Figure 4 shows these plots for various Ns, for ρ=0.1 (fig. 4(a)) and ρ=0.3 (fig. 4(b)). The migration threshold Nm_c is insensitive to Ls for small Ls, but starts increasing with increasing Ls above a threshold $(Ls)_*$ which depends on habitat fractions. This increase is sharper for larger Ns and smaller ρ, reflecting the stronger effects of migration-induced LD in larger demes, especially when these are subject to predominantly maladaptive gene flow from a more widespread, differently adapted habitat.

For low values of Ls, Nm_c is well-approximated by the LE prediction: $Nm_c \approx \frac{Ns - \frac{1}{2}\log\left[\frac{1-\rho}{\rho}\right]}{1-2\rho}$ (Szép et al. 2021). These predictions are indicated by short horizontal coloured lines along the vertical axis in fig. 4. We can also obtain an explicit expression for the threshold $(Ls)_*$, above which Nm_c starts increasing with Ls, by neglecting drift, i.e., in the deterministic limit (see Appendix D, SI). This is: $(Ls)_* \approx 1/[2(1-2\rho)]$ for $\rho \leq 1/4$, and $(Ls)_* \approx 4\rho$ for $1/4 \rho < 1/2$. These

thresholds are indicated by dotted vertical lines in fig. 4. Note that the actual thresholds in the presence of drift (i.e., for small to moderate Ns) are somewhat higher; however, the deterministic prediction for $(Ls)_*$ provides a reasonable lower bound.

Thus, strongly divergent selection (i.e., $Ls \gg 1/2$) can allow extremely marginal habitats to withstand comparable levels of migration as less marginal habitats, e.g., in scenarios involving secondary contact subsequent to initial divergence. This is especially pertinent in light of the rather high values of Nm estimated in natural populations across a range of taxa (Morjan and Rieseberg 2004); the present analysis (e.g., fig. 4) shows that adaptive divergence between habitats may be maintained despite such high levels of migration when multiple loci underlie divergence. It is less obvious if (or when) the stronger protective effects of LD in more marginal habitats might also allow new locally adaptive mutations to establish in these habitats despite ongoing gene flow at rates that are comparable to those in less marginal habitats. Thus, extending the present analysis to investigate how a reduction in effective migration rates influences the establishment probabilities of new mutations (at sites with different degrees of linkage to pre-existing barrier loci) in small populations is an interesting direction for future work.

A striking observation to emerge from the analysis of RI in the infinite-island scenario is that neutral, genomewide F_{ST} between pairs of subpopulations is actually highest when both subpopulations belong to the rare habitat (fig. 3(b)), despite the fact that there is *no* RI (or isolation-by-distance) between such subpopulations.

This counter-intuitive finding can be rationalised by noting that in pairwise comparisons of subpopulations embedded within a larger mosaic of interconnected subpopulations, F_{ST} between any pair of subpopulations does not measure the extent of *direct* genetic exchange between them, but instead depends on the level of exchange between any individual subpopulation and the larger population, averaged over the two subpopulations. Since demes belonging to the rare habitat are more effectively isolated from the larger population than demes belonging to the common habitat, it follows that the average level of genetic exchange between any two randomly chosen demes and the population as a whole is lowest if both demes belong to the rare habitat, and highest if both belong to the common habitat, so that $F_{ST}^{(rr)} > F_{ST}^{(rc)}, > F_{ST}^{(cc)}$ when both habitats are at least partially locally adapted.

The fact that neutral diversity in the rare habitat is reduced much more strongly than in the common habitat is, to some extent, a consequence of the infinite-island assumption, in which individuals migrate from any island to any other with equal probability, irrespective of the habitat of origin and destination. One could consider alternative models where islands are embedded in a 2D spatial matrix, and where habitats are spatially sorted, so that migration between islands belonging to the same habitat is more likely than between islands belonging to different habitats (models with habitat choice but no explicit space might result in qualitatively similar outcomes). However, as long as spatial sorting of habitats and/or habitat choice is not complete, we expect effective immigration into the rare habitat to be at least somewhat lower than into the common habitat and average F_{ST} between demes belonging to the rare habitat to be accordingly highest, despite the lack of RI between such demes.

In their re-analysis of so-called genomic islands of divergence across a variety of hybridizing populations, Cruickshank and Hahn (2014) pointed out that increased F_{ST} may reflect reduced diversity (e.g., due to purifying selection within populations) rather than increased divergence (e.g., due to reduced gene flow between populations). A consideration of local adaptation in metapopulations yields yet another caveat: where sampled subpopulations are embedded within a larger metapopulation, a reduction in gene flow between populations belonging to different ecological niches can manifest itself primarily via reduced diversity within populations rather than an increased divergence between subpopulations; moreover, this reduction is more severe for subpopulations which support the niche that is more marginal or less abundant in the

population as a whole. While this effect is likely to be less extreme in reality than it is under the infinite-island model analysed here, the analysis nevertheless points to the potential perils of neglecting the wider spatial context of two (or a few) sampled subpopulations, when interpreting relative divergence between them. Note that this caveat does not only apply to divergence at unlinked neutral loci but also to markers linked to barrier loci (since the underlying argument relates only to there being a rare and a common habitat), and is thus also relevant to the interpretation of localized elevations of F_{ST}.

Acknowledgments. I thank Joachim Hermisson and Nick Barton for very helpful comments on the manuscript.

References

Barton, N., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357–376.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.

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

Barton, N. H., and A. M. Etheridge. 2011. The relation between reproductive value and genetic contribution. Genetics 188:953–973.

Bengtsson, B. O. 1985. The flow of genes through a genetic barrier. Pages 31–42 Cambridge University Press Cambridge; New York.

Blanckaert, A., C. Bank, and J. Hermisson. 2020. The limits to parapatric speciation 3: evolution of strong reproductive isolation in presence of gene flow despite limited ecological differentiation. Philosophical Transactions of the Royal Society B: Biological Sciences 375:20190532.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, M. M. Goodman, C. Harjes, K. Guill, D. E. Kroon, S. Larsson, N. K. Lepak, H. Li, S. E. Mitchell, G. Pressoir, J. A. Peiffer, M. O. Rosas, T. R. Rocheford, M. C. Romay, S. Romero, S. Salvo, H. S. Villeda, H. S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M. D. McMullen. 2009. The genetic architecture of maize flowering time. Science 325:714–718.

Bürger, R., and A. Akerman. 2011. The effects of linkage and gene flow on local adaptation: A two-locus continentisland model. Theoretical Population Biology 80:272–288.

Butlin, R. K., and C. M. Smadja. 2018. Coupling, reinforcement, and speciation. The American Naturalist 191:155–172. PMID: 29351021.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical Research 70:155–174.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133–3157.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012. Establishment of new mutations under divergence and genome hitchhiking. Philosophical Transactions of the Royal Society B: Biological Sciences 367:461–474.

Feder, J. L., and P. Nosil. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729–1747.

Feder, J. L., P. Nosil, A. C. Wacholder, S. P. Egan, S. H. Berlocher, and S. M. Flaxman. 2014. Genome-Wide Congealing and Rapid Transitions across the Speciation Continuum during Speciation with Gene Flow. Journal of Heredity 105:810–820.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon Press.

Geroldinger, L., and R. Bürger. 2015. Clines in quantitative traits: The role of migration patterns and selection scenarios. Theoretical Population Biology 99:43–66.

Good, B. H., A. M. Walczak, R. A. Neher, and M. M. Desai. 2014. Genetic diversity in the interference selection limit. PLOS Genetics 10:1–1.

Halligan, D. L., and P. D. Keightley. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 40:151–172.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genetical Research 8:269–294.

Kobayashi, Y., P. Hammerstein, and A. Telschow. 2008. The neutral effective migration rate in a mainland-island context. Theoretical Population Biology 74:84–92.

Kulmuni, J., R. K. Butlin, K. Lucek, V. Savolainen, and A. M. Westram. 2020. Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. Philosophical Transactions of the Royal Society B: Biological Sciences 375:20190528.

Lamichhaney, S., A. M. Barrio, N. Rafati, G. Sundström, C.-J. Rubin, E. R. Gilbert, J. Berglund, A. Wetterbom, L. Laikre, M. T. Webster, M. Grabherr, N. Ryman, and L. Andersson. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in atlantic herring. Proceedings of the National Academy of Sciences 109:19345–19350.

Li, W.-H., and M. Nei. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. Theoretical Population Biology 6:173–183.

Morjan, C. L., and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Molecular Ecology 13:1341–1356.

Nagylaki, T. 1976. Clines with variable migration. Genetics 83:867–86.

Petry, D. 1983. The effect on neutral gene flow of selection at a linked locus. Theoretical Population Biology 23:300–313.

Robertson, A. 1961. Inbreeding in artificial selection programmes. Genetical Research 2:189194.

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 14:1–22.

Roze, D. 2021. A simple expression for the strength of selection on recombination generated by interference among mutations. Proceedings of the National Academy of Sciences 118.

Sachdeva, H., and N. H. Barton. 2018. Replicability of introgression under linked, polygenic selection. Genetics 210:1411–1427.

Slatkin, M. 1975. Gene flow and selection in a two-locus system. Genetics 81:787–802.

———. 1991. Inbreeding coefficients and coalescence times. Genetical Research 58:167–175.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Szép, E., H. Sachdeva, and N. H. Barton. 2021. Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. Evolution 75:1030–1045.

Weissman, D. B., and N. H. Barton. 2012. Limits to the rate of adaptive substitution in sexual populations. PLOS Genetics 8:1–18.

Wright, S. 1937. The distribution of gene frequencies in populations. Proceedings of the National Academy of Sciences 23:307–320.

Yeaman, S. 2015. Local adaptation by alleles of small effect. The American Naturalist 186:S74–S89. PMID: 26656219.

SUPPLEMENTARY INFORMATION
Reproductive isolation via polygenic local adaptation in sub-divided populations: effect of linkage disequilibria and drift

Himani Sachdeva

Department of Mathematics, University of Vienna, Vienna 1090, Austria.

Appendix A: Effective migration rates (neglecting drift).

Effective migration rate for deleterious alleles. We can solve eq. 3 of the main text iteratively to obtain P_y, the deterministic frequencies of genotypes that carry $y{=}1,2,\ldots L$ deleterious alleles, in a population at migration-selection equilibrium. These can be expressed as:

$$
\begin{aligned}
&P_L{=}m\frac{g_L}{1{-}g_L}\\
&P_{L-y}{=}\binom{L}{y}\frac{P_L}{1{-}g_{L-y}}\Bigg[1{+}\sum_{k=1}^{y-1}\binom{y}{k}\frac{g_{L-k}}{1{-}g_{L-k}}{+}\sum_{k=1}^{y-2}\sum_{j=1}^{y-k-1}\binom{y}{k}\binom{y-k}{j}\frac{g_{L-k}}{1{-}g_{L-k}}\frac{g_{L-k-j}}{1{-}g_{L-k-j}}{+}\\
&\sum_{k=1}^{y-3}\sum_{j=1}^{y-k-2}\sum_{v=1}^{y-k-j-1}\binom{y}{k}\binom{y-k}{j}\binom{y-k-j}{v}\frac{g_{L-k}}{1{-}g_{L-k}}\frac{g_{L-k-j}}{1{-}g_{L-k-j}}\frac{g_{L-k-j-v}}{1{-}g_{L-k-j-v}}{+}\ldots\Bigg],1{\le}y{\le}L{-}1\\
&\qquad\text{where}\quad g_k{=}\frac{e^{-s\,k}}{2^{k-1}}
\end{aligned}
\tag{S1}
$$

The above equation expresses the genotype frequency P_{L-y} for $y{=}1,2,\ldots L{-}1$ as a finite sum over y terms (i.e., the series within the square brackets contains $1,2,3\ldots$ terms in the expressions for P_{L-1}, P_{L-2}, P_{L-3}, $\ldots$ respectively). It is also useful to represent these as infinite sums by Taylor expanding the various $g_k/(1{-}g_k)$ in powers of g_k. Note that $g_k{=}\frac{2\,e^{-ks}}{2^k}$ has a simple interpretation: it is the average number of offspring of an individual with k deleterious alleles, who inherit none of the k alleles; alternatively, it is the average number of offspring who inherit all k alleles. Then, we have the alternative representation:

$$
P_y{=}m\left[\binom{L}{y}g_L{+}\sum_{k=0}^{L-y}\binom{L}{k}\binom{L-k}{y}g_Lg_{L-k}{+}\sum_{k=0}^{L-y}\sum_{j=0}^{L-k-y}\binom{L}{k}\binom{L-k}{j}\binom{L-k-j}{y}g_Lg_{L-k}g_{L-k-j}{+}\ldots\right]
\tag{S2}
$$

Note that the above representation involves an *infinite* series; the kth term in the series is a k-fold convolution. The average deleterious frequency $p_{det}{=}\sum_{k=1}^{L}(k/L)P_k$ can be obtained by using either eq. (S1) or (S2). Here, we use eq. (S2), thus expressing p_{det} as an infinite sum:

$$
\begin{aligned}
p_{det} = m \Bigg[& 2c \sum_{y=0}^{L-1} \binom{L-1}{y} c^{L-1} + 4c^2 \sum_{k=0}^{L-1} \binom{L-1}{k} c^{L-1} \sum_{j=0}^{L-1-k} \binom{L-k-1}{j} c^{L-k-1} + \\
& + 8c^3 \sum_{k=0}^{L-1} \binom{L-1}{k} c^{L-1} \sum_{j=0}^{L-k-1} \binom{L-k-1}{j} c^{L-k-1} \sum_{m=0}^{L-k-j-1} \binom{L-k-j-1}{m} c^{L-k-j-1} + \dots \Bigg] \\
= & \, m(2c)^L \left[1 + (2c) \left(\frac{1+2c}{2} \right)^{L-1} + (2c)^2 \left(\frac{1+c+2c^2}{2} \right)^{L-1} + (2c)^3 \left(\frac{1+c+c^2+2c^3}{2} \right)^{L-1} + \dots \right] \\
= & \, m \, (2c)^L \sum_{k=0}^{\infty} (2c)^k \left(\frac{1+c^k-2c^{k+1}}{2(1-c)} \right)^{L-1} \quad \text{where} \quad c = \frac{e^{-s}}{2} \\
= & \, m \, \frac{e^{-sL}}{(2-e^{-s})^{L-1}} \sum_{k=0}^{\infty} e^{-s\,k} \left(1 + \frac{e^{-sk}}{2^k} - \frac{e^{-s(k+1)}}{2^k} \right)^{L-1}
\end{aligned}
\tag{S3}
$$

The final expression is also provided in eq. 4 of the main text. For $L{=}1$, the above expression reduces to: $p_{det}(L{=}1){=}m/(e^s{-}1) \approx m/s$. Following the main text, we define the effective migration rate $m_e(s, L)$ as the migration rate which would cause the deleterious allele frequency at a single locus under migration-selection equilibrium (with selection coefficient s for the deleterious allele) to be equal to p_{det}, which is the average equilibrium frequency that emerges in the multi-locus model, where immigrant genotypes carrying L such deleterious alleles are introduced at rate m per generation. Then, we have: $m_e(s, L){=}p_{det}(e^s{-}1){\approx}sp_{det}$, where p_{det} is given by eq. (S3). This finally yields the expression for $m_e(s, L)$ in eq. 5 in the main text.

In the infinitesimal limit, i.e., taking the limit $s{\to}0$, $L{\to}\infty$ with $sL{=}\theta$ held constant, we can approximate $m_e(s, L)/m$ as:

$$
\begin{aligned}
\frac{m_e(s,L)}{m} = & \frac{e^{-sL}\,(e^s-1)}{(2-e^{-s})^{L-1}} \sum_{k=0}^{\infty} e^{-s\,k} \left(1 + \frac{e^{-sk}}{2^k} - \frac{e^{-s(k+1)}}{2^k} \right)^{L-1} \\
\approx & \; s\,e^{-2\theta} \sum_{k=0}^{\infty} e^{-s\,k} \left(1 + s\frac{e^{-sk}}{2^k} \right)^{L-1} \\
= & \; s\,e^{-2\theta} \sum_{m=0}^{L-1} \binom{L-1}{m} s^m \sum_{k=0}^{\infty} \frac{e^{-sk\,(m+1)}}{2^{k\,m}} \\
= & \; s\,e^{-2\theta} \sum_{m=0}^{L-1} \binom{L-1}{m} \frac{s^m}{1 - \frac{e^{-s\,(m+1)}}{2^m}} \\
\approx & \; s\,e^{-2\theta} \left[\frac{1}{s} + \sum_{m=1}^{L-1} \frac{\theta^m}{m!(1-\frac{1}{2^m})} \right] \\
\approx & \; e^{-2\theta} \quad \text{as } s{\to}0,\ L{\to}\infty \text{ with } \theta{=}sL \text{ constant}
\end{aligned}
\tag{S4}
$$

Thus, the effective migration rate of selected alleles is reduced relative to the raw migration rate by a factor that converges (in the infinitesimal limit) to the average RV e^{-2sL} of migrants (see also the explanation following eq. 5 of the main text).

Effective migration rate for neutral alleles. Following Bengtsson (1985), we define the effective migration rate $m_e^*(s,L)$ associated with a neutral locus to be the rate at which neutral alleles at that locus, that immigrate as part of a genome with L unlinked deleterious alleles of effect s, escapes via one or more recombination events onto a genetic background with no deleterious alleles in a population at migration-selection equilibrium.

To calculate this, we must first calculate the equilibrium frequencies $\{f_y\}$ of genotypes that carry the neutral allele in conjunction with y deleterious alleles (each with selective effect s), where all alleles (neutral and deleterious) are unlinked. As before, we assume that the frequency of any genotype with deleterious alleles is sufficiently low that quadratic terms of the kind $f_y f_{y'}$ can be neglected. Then f_y satisfy the following equations in steady state:

$$f_y = \frac{1}{2}\left[m w_L \binom{L}{y}\left(\frac{1}{2}\right)^L + \sum_{k=y}^{L} f_k w_k \binom{k}{y}\left(\frac{1}{2}\right)^k\right] \quad \text{where} \quad w_i = 2e^{-is} \tag{S5}$$

Note that these are very similar to eq. 3 in the main text, except that the right hand side now has an additional factor of 1/2, which is the probability that an individual carrying the neutral allele transmits it to an offspring. Equation (S5) can be solved as before; it is again useful to express the various f_y as an infinite series in powers of g_k, so that we have:

$$\begin{aligned} f_y = m\Bigg[&\frac{1}{2}\binom{L}{y} g_L + \frac{1}{2^2}\sum_{k=0}^{L-y}\binom{L}{k}\binom{L-k}{y} g_L g_{L-k} + \\ &\frac{1}{2^3}\sum_{k=0}^{L-y}\sum_{j=0}^{L-k-y}\binom{L}{k}\binom{L-k}{j}\binom{L-k-j}{y} g_L g_{L-k} g_{L-k-j} + \dots\Bigg] \qquad \text{where} \quad g_i = \frac{w_i}{2^i} \end{aligned} \tag{S6}$$

The above expression allows us to read off the distribution of t_{esc}, the time taken for the focal (neutral) allele to 'escape' the deleterious background, by setting y=0. Thus, we have:

$$f_0 = m\sum_{k=1}^{\infty}\mathcal{P}(t_{esc}=k) \quad \text{where } \mathcal{P}(t_{esc}=1) = \frac{1}{2}\binom{L}{y} g_L, \quad \mathcal{P}(t_{esc}=2) = \frac{1}{2^2}\sum_{k=0}^{y}\binom{L}{k}\binom{L-k}{y-k} g_L g_{L-k}, \quad \dots \tag{S7}$$

These can be used to further calculate various moments $\mathbb{E}[t_{esc}^n]$ of the time to escape.

We can also use $\{f_y\}$ to obtain the effective migration rate $m_e^*(s,L)$ for introgressing neutral alleles, by recalling that this is the net rate at which alleles are transferred onto a wildtype background by recombination from deleterious backgrounds. Thus, we have:

$$
\begin{aligned}
\frac{m_e^*(s,L)}{m} &= \frac{1}{2m}\left[g_L m + \sum_{k=1}^{L} g_k f_k\right] \\
&= c^L\Bigg[1+\sum_{k=0}^{L-1}\binom{L}{k}c^{L-k}+\sum_{k=0}^{L-1}\sum_{j=0}^{L-1-k}\binom{L}{k}\binom{L-k}{j}c^{L-k}c^{L-j-k}+ \\
&\qquad \sum_{k=0}^{L-1}\sum_{j=0}^{L-1-k}\sum_{m=0}^{L-1-k-j}\binom{L}{k}\binom{L-k}{j}\binom{L-k-j}{m}c^{L-k}c^{L-j-k}c^{L-j-k-m}+\ldots\Bigg] \\
&= c^L\Big(1+[(1+c)^L-1]+[(1+c(1+c))^L-(1+c)^L]+[(1+c(1+c(1+c)))^L-(1+c(1+c))^L]+\ldots\Big) \\
&= \left(\frac{c}{1-c}\right)^L\sum_{k=0}^{\infty}\left[(1-c^{k+1})^L-(1-c^k)^L\right] \quad \text{where} \quad c=\frac{e^{-s}}{2} \\
&= \left(\frac{e^{-s}}{2-e^{-s}}\right)^L
\end{aligned}
\tag{S8}
$$

As before, in the limit $s\to 0$, $L\to\infty$ with $sL=\theta$ held constant (infinitesimal limit), this simplifies to $m_e^*(s,L)/m \approx e^{-2\theta}$.

Appendix B: Calculating average coalescence times and expected F_{ST} in the infinite-island model using the structured coalescent.

To calculate coalescence times, it is useful to consider a population with a finite number n_D of demes, and then take the $n_D\to\infty$ limit. There are 5 distinct possibilities for sampling 2 lineages at random in a population with n_D demes, of which a fraction $\rho<1/2$ belong to the rare habitat and a fraction $1-\rho$ to the common habitat. With probability ρ/n_D, both lineages will be sampled from the *same* deme (arbitrarily labeled i) in the rare habitat; with probability $(1-\rho)/n_D$, both will be from the *same* deme i in the common habitat; with probability $\rho(1-1/n_D)$, lineages will be sampled from two *different* demes (arbitrarily labeled i and j) both in the rare habitat; with probability $(1-\rho)(1-1/n_D)$, lineages will be from two different demes i and j both belonging to the common habitat; with probability $2\rho(1-\rho)$, lineages will be sampled from two different demes i and j, one belonging to the rare and the other to the common habitat. We denote the 5 coalescence times corresponding to these 5 different configurations for 2 lineages by: $T_{i,i;\,r}$, $T_{i,i;\,c}$, $T_{i,j;\,r,r}$, $T_{i,j;\,c,c}$, and $T_{i,j;\,r,c}$.

The expected values of the various F_{ST} measures (introduced in the main text) can be expressed in terms of the expected coalescence times as follows:

$$
\begin{aligned}
F_{ST}^{(r)} &= 1-\frac{\mathbb{E}[T_{i,i;\,r}]}{\mathbb{E}[T_{tot}]} \qquad F_{ST}^{(c)} = 1-\frac{\mathbb{E}[T_{i,i;\,c}]}{\mathbb{E}[T_{tot}]} \\
\text{where} \quad \mathbb{E}[T_{tot}] &= \frac{\rho}{n_D}\mathbb{E}[T_{i,i;\,r}]+\frac{1-\rho}{n_D}\mathbb{E}[T_{i,i;\,c}]+\rho\left(1-\frac{1}{n_D}\right)\mathbb{E}[T_{i,j;\,r,r}] \\
&\quad +(1-\rho)\left(1-\frac{1}{n_D}\right)\mathbb{E}[T_{i,j;\,c,c}]+2\rho(1-\rho)\mathbb{E}[T_{i,j;\,r,c}]
\end{aligned}
\tag{S9a}
$$

$$
\begin{aligned}
F_{ST}^{(r,r)} &= 1-\frac{\mathbb{E}[T_{i,i;\,r}]}{\frac{1}{2}\mathbb{E}[T_{i,i;r}]+\frac{1}{2}\mathbb{E}[T_{i,j;r,r}]} \qquad F_{ST}^{(c,c)} = 1-\frac{\mathbb{E}[T_{i,i;\,c}]}{\frac{1}{2}\mathbb{E}[T_{i,i;c}]+\frac{1}{2}\mathbb{E}[T_{i,j;c,c}]} \\
F_{ST}^{(r,c)} &= 1-\frac{\frac{1}{2}\mathbb{E}[T_{i,j;\,r,r}]+\frac{1}{2}\mathbb{E}[T_{i,j;\,c,c}]}{\frac{1}{4}\mathbb{E}[T_{i,i;r}]+\frac{1}{2}\mathbb{E}[T_{i,j;r,c}]+\frac{1}{4}\mathbb{E}[T_{i,i;c}]}
\end{aligned}
\tag{S9b}
$$

In order to calculate the expected coalescence times, we consider how lineages trace back in time through different demes belonging to one or other habitat. Let m_{rc} and m_{rr} denote the probability per unit time that a lineage in a deme within the rare habitat traces back to the rare and common habitats respectively. We can define analogous backward migration rates m_{cc} and m_{cr} for lineages in demes belonging to the common habitat. We use uppercase letters to denote the corresponding population-size-scaled migration rates $M_{rr}{=}Nm_{rr}$, $M_{rc}{=}Nm_{rc}$, $M_{cc}{=}Nm_{cc}$, $M_{cr}{=}Nm_{cr}$. Note that the rates m_{rr} and m_{cc} (or M_{rr} and M_{cc}), as defined here, include backward migration events where a lineage emigrates from a deme but then immigrates back into the same deme (though this occurs with a small probability $1/n_D$, which vanishes in the limit $n_D{\to}\infty$).

Going back one generation into the past, there are three possibilities for a pair of lineages that are, at present, in the same deme within the rare habitat: the two lineages can coalesce with probability $1/N$; one or other of the two lineages can trace back to a different deme within the rare habitat with probability $2m_{rr}\left(1{-}\frac{1}{n_D\rho}\right)$ (here, the factor $\left(1{-}\frac{1}{n_D\rho}\right)$ accounts for the fact that the migration rate m_{rr} also includes 'migration' events which lead a lineage back to its deme of origin); finally, one or other lineage can trace back to a deme belonging to the common habitat with probability $2m_{rc}$; note that we neglect events involving simultaneous migration of both lineages in a single timestep as these occur with vanishingly small rates under the usual coalescent scaling. Similarly, by enumerating all possible single-generation events for the other 4 two-lineage configurations, we arrive at the following set of recursions for the expected coalescence times:

$$\mathbb{E}[T_{i,i;\,r}]{=}\frac{1{+}M_{rc}\mathbb{E}[T_{i,j;\,r,c}]{+}M_{rr}\left(1{-}\frac{1}{\rho\,n_D}\right)\mathbb{E}[T_{i,j;\,r,r}]}{1{+}M_{rc}{+}M_{rr}\left(1{-}\frac{1}{\rho\,n_D}\right)} \tag{S10a}$$

$$\mathbb{E}[T_{i,i;\,c}]{=}\frac{1{+}M_{cr}\mathbb{E}[T_{i,j;\,c,r}]{+}M_{cc}\left(1{-}\frac{1}{(1{-}\rho)n_D}\right)\mathbb{E}[T_{i,j;\,c,c}]}{1{+}M_{cr}{+}M_{cc}\left(1{-}\frac{1}{(1{-}\rho)n_D}\right)} \tag{S10b}$$

$$\mathbb{E}[T_{i,j;\,r,r}]{=}\frac{1{+}M_{rr}\frac{1}{\rho n_D}\mathbb{E}[T_{i,i;\,r}]{+}M_{rc}\mathbb{E}[T_{i,j;\,r,c}]}{M_{rr}\frac{1}{\rho\,n_D}{+}M_{rc}} \tag{S10c}$$

$$\mathbb{E}[T_{i,j;\,c,c}]{=}\frac{1{+}M_{cc}\frac{1}{(1{-}\rho)n_D}\mathbb{E}[T_{i,i;\,c}]{+}M_{cr}\mathbb{E}[T_{i,j;\,r,c}]}{M_{cc}\frac{1}{(1{-}\rho)n_D}{+}M_{cr}} \tag{S10d}$$

$$\mathbb{E}[T_{i,j;\,r,c}]{=}\frac{2{+}M_{cr}\left(1{-}\frac{1}{\rho\,n_D}\right)\mathbb{E}[T_{i,j;\,r,r}]{+}M_{rc}\left(1{-}\frac{1}{(1{-}\rho)n_D}\right)\mathbb{E}[T_{i,j;\,c,c}]{+}M_{cr}\frac{1}{\rho\,n_D}\mathbb{E}[T_{i,i;\,r}]{+}M_{rc}\frac{1}{(1{-}\rho)n_D}\mathbb{E}[T_{i,i;\,c}]}{M_{cr}{+}M_{rc}} \tag{S10e}$$

Solving equation S8 for the expectations of the 5 coalescence times, substituting these expressions into eq. (S7) for the expected divergence measures, and then taking the limit $n_D{\to}\infty$ (which allows us to neglect all terms that are $\mathcal{O}(1/n_D)$) finally gives eq. 9 of the main text.

Appendix C: Convergence to small s, large L limit.

In the limit $s{\to}0$, $L{\to}\infty$ with $\theta{=}Ls$ fixed, the effective migration rates $m_e(s,L)$ and $m_e^*(s,L)$ converge to $m\,e^{-2sL}$ (see eq. (S4) and text following eq. (S8) in Appendix A). Figure 1(a) illustrates this convergence by plotting $m_e(s,L)/(m\,e^{-2sL}){-}1$ (dashed lines) and $m_e^{(*)}(s,L)/(m\,e^{-2sL}){-}$

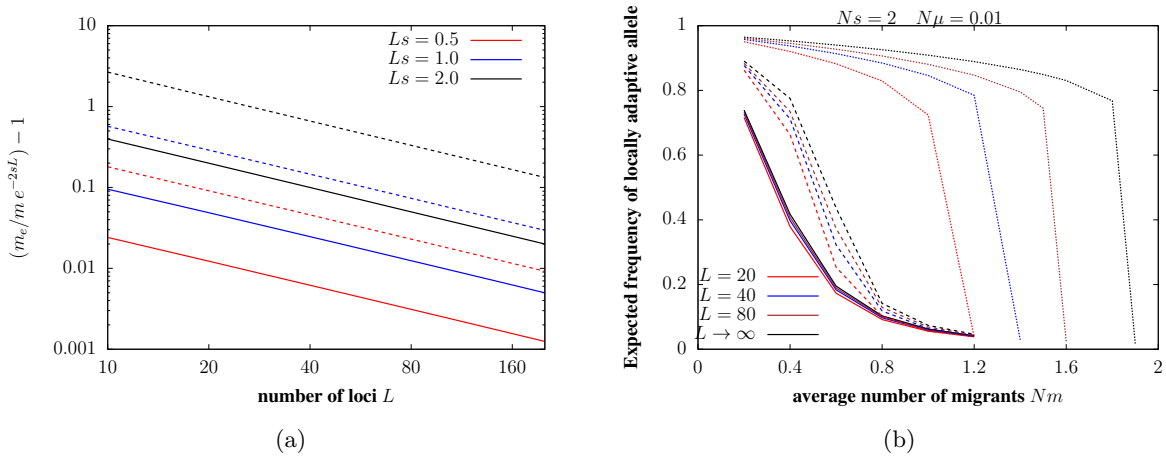


Figure S1: (a) Relative deviations $m_e(s,L)/(me^{-2sL})-1$ (dashed lines) and $m_e^*(s,L)/(me^{-2sL})-1$ (solid lines) of the effective migration rates associated with deleterious and neutral alleles from the asymptotic ($s\to 0$ $L\to\infty$) value, vs. L, for different values of Ls (different colors). The effective migration rates are calculated using eqs. 5 and 6 of the main text. (b) The expected frequency of the locally adaptive allele on the island vs. Nm, under mainland-island migration. Solid, dashed and dotted lines correspond to Ls=0.5, Ls=1 and Ls=2 respectively; the different colors correspond to different values of L; other parameters are Ns=2 and $N\mu$=0.01. All finite L predictions in (b) are obtained by numerically solving eq. 7 in conjunction with eq. 5 of the main text; the $L\to\infty$ predictions are obtained by numerically solving eq. 7 and using the asymptotic expression for $m_e(s,L)$.

1 (solid lines) vs. L for different values of Ls (different colors). As expected, both kinds of effective migration rates converge towards the asymptotic limit as s decreases (or equivalently, L increases). Effective migration rates for finite L are higher than the asymptotic $L\to\infty$ value, with the deviation being more significant for larger values of Ls, and also for the rate $m_e(s,L)$, which is associated with deleterious alleles.

One can also ask: to what extent do critical migration thresholds for loss of local adaptation depend on L and s separately, or only on the composite parameter Ls (for a given Ns), as one would expect in the infinitesimal limit. Figure 1(b) shows analytical predictions for the frequency of the locally adaptive allele on the island vs. Nm for Ls=0.5 (solid lines), Ls=1.0 (dashed lines) and Ls=2 (dotted lines), with the different colors corresponding to different numbers L of divergently selected loci (for each Ls). The finite L predictions are obtained by solving eq. 7 of the main text, using the expression for m_e in eq. 5, while the $L\to\infty$ predictions employ the asymptotic expression $m_e(s,L)=m\,e^{-2sL}$.

Figure 1(b) shows that critical migration thresholds are more sensitive to L at larger values of Ls, with the threshold being lower for lower L, i.e., if load is due to a smaller number of strongly selected loci. This is consistent with the observation that the deviation of $m_e(s,L)$ from the asymptotic $L\to\infty$ value is more significant for larger Ls (fig. 1(a)); further, small changes in m_e per locus can have a substantial effect on maladaptation when Ls is larger.

Appendix D: Deterministic predictions for the threshold $(Ls)_*$ above which the critical migration threshold increases with Ls

In the deterministic limit (i.e., neglecting drift), we can write down coupled equations for the time evolution of p_r and p_c, the allele frequencies in the rare and common habitats. As in the main text, p denotes the frequency of the allele that is advantageous in the common and disadvantageous in the rare habitat. We will also use $\Delta=p_c-p_r$ to denote the allele frequency

difference between the two habitats. Then, we have:

$$\frac{dp_r}{dt} = -sp_rq_r + m_e(1-\rho)(p_c - p_r) \tag{S11a}$$

$$\frac{dp_c}{dt} = sp_cq_c + m_e\rho(p_r - p_c) \tag{S11b}$$

For simplicity, we will approximate the effective migration rate by its asymptotic (infinitesimal) form: $m_e \approx m\, e^{-2sL(p_c-p_r)}$. The above equations can be simplified by assuming that the common habitat is always reasonably well-adapted, such that $q_c{=}1{-}p_r$ is small and $\mathcal{O}(q_c^2)$ terms can be neglected. This allows us to approximate: $p_cq_c{\approx}q_c$ and $p_rq_r{\approx}\Delta(1{-}\Delta){+}q_c(1{-}2\Delta)$, which gives the following coupled equations for the time evolution of p_c and Δ.

$$\frac{dp_c}{dt} = sq_c - m\, e^{-2sL\Delta}\rho\Delta \tag{S12a}$$

$$\frac{d\Delta}{dt} = 2sq_c(1-\Delta) + s\Delta(1-\Delta) - m\, e^{-2sL\Delta}\Delta \tag{S12b}$$

At migration-selection equilibrium, i.e., setting the time derivatives to zero, we can use eq. (S12a) to express q_c in terms of Δ, and then substitute into (S12b) to obtain the following transcendental equation for Δ:

$$\frac{d\Delta}{dt} = 0 = m\, e^{-2sL\Delta}\Delta[2\rho(1-\Delta)-1] + s\Delta(1-\Delta) \tag{S13}$$

There is always an equilibrium at $\Delta{=}0$ (corresponding to zero adaptive differentiation between habitats): this is unstable as long as $s{>}m(1{-}2\rho)$ and is stable otherwise. Note that $s{>}m(1{-}2\rho)$ is simply the criterion for polymorphism at a single locus at deterministic migration-selection equilibrium. In addition, there may be other equilibria at $\Delta{=}\Delta_*$, where Δ_* satisfies $(m/s)e^{-2sL\Delta_*}[2\rho(1{-}\Delta_*){-}1]{+}1{-}\Delta_*{=}0$. While this does not allow us to explicitly solve for Δ_*, we can obtain an explicit condition (in terms of m/s, Ls and ρ) when no biologically meaningful equilibrium (with $\Delta_*{>}0$) exists.

As a prelude to deriving this condition, it is useful to visualize the behaviour of $d\Delta/dt$ as a function of Δ (using eq. (S13)). Figures 2(a) and 2(b) show $d\Delta/dt$ (scaled by s) vs. Δ for $Ls{=}0.5$ and $Ls{=}1$ respectively, for $\rho{=}0.1$, with the different colors in each plot corresponding to different values of m/s. Those points at which the curves intersect the horizontal axis (i.e., at which $d\Delta/dt{=}0$) correspond to equilibria. An equilibrium is *stable* if $d\Delta/dt$ is a decreasing function of Δ, i.e., if the curve is downward sloping at the equilibrium, and is unstable otherwise.

For small Ls (fig. 2(a)) and low values of m/s (red and brown curves), we note that there are two equilibria— an unstable equilibrium at $\Delta{=}0$ and a stable equilibrium at $\Delta{=}\Delta_*{>}0$ (marked by the $\times$ symbol on the plots). Increasing m/s causes Δ_* to fall; above a critical migration threshold $m_c/s{=}1/(1{-}2\rho)$, the equilibrium at $\Delta{=}0$ becomes stable while the alternative equilibrium Δ_* becomes unstable and is also negative (and hence is no longer biologically meaningful). The transition threshold $m_c/s{=}1/(1{-}2\rho)$ (blue curve in fig. 2(a)) at which local adaptation in the rare habitat is lost, or alternatively, divergence between habitats collapses, corresponds to a so-called transcritical bifurcation.

For large Ls (fig. 2(b)), we observe qualitatively different transitions of the equilibria with changing m/s. At low levels of migration, i.e., for $m/s{<}1/(1{-}2\rho)$ (red curve), there is (as before) an unstable equilibrium at $\Delta{=}0$ and a stable equilibrium at $\Delta{=}\Delta_*{>}0$. At intermediate migration levels (brown curve), i.e., for $1/(1{-}2\rho){<}m/s{<}m_c/s$, the $\Delta{=}0$ equilibrium becomes stable; however, the stable $\Delta_*{>}0$ equilibrium still persists, and there now appears a third

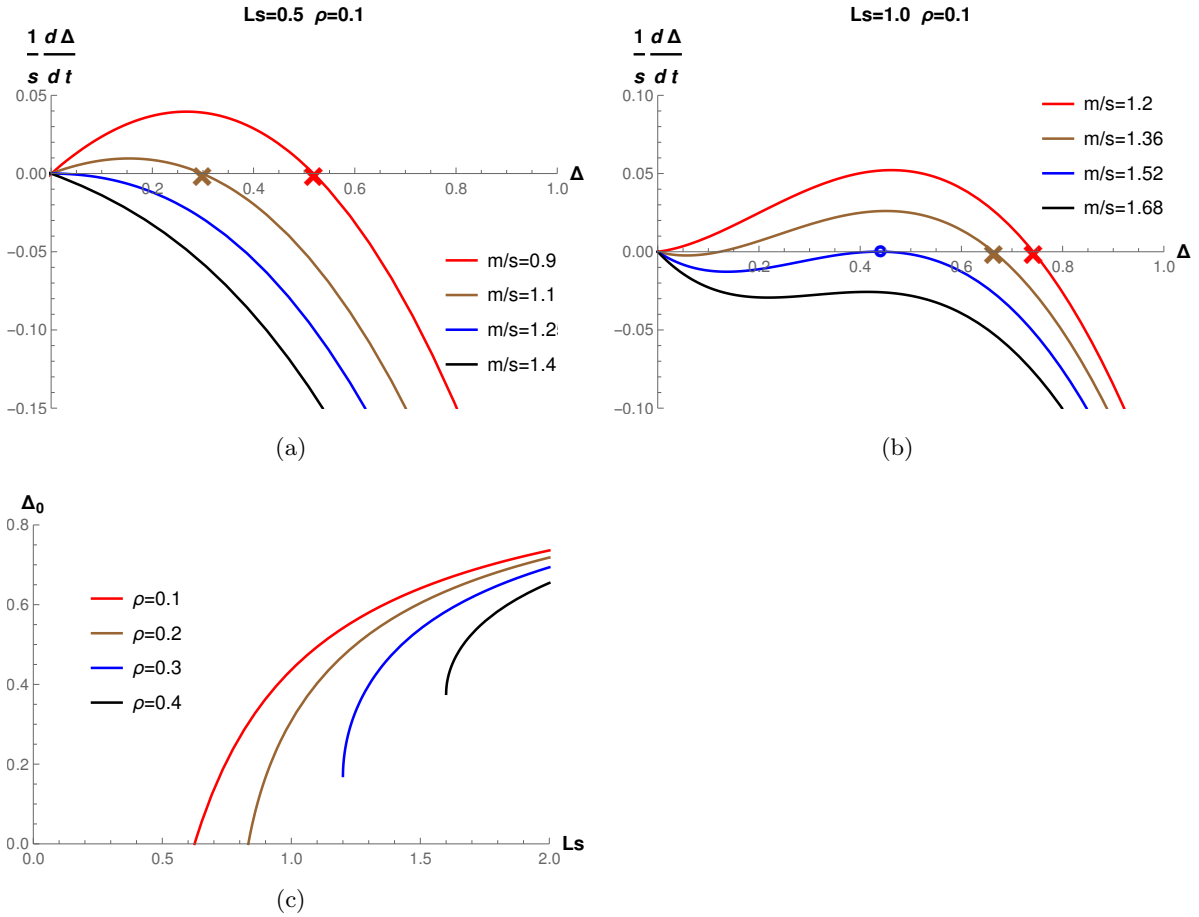


Figure S2: Deterministic analysis of local adaptation in the infinite-island model. (A)-(B) $d\Delta/dt$ (scaled by s) vs. Δ for various values of m/s for ρ=0.1 and (A) Ls=0.5 (B) Ls=1. Here, $\Delta=p_c-p_r$ is the allele frequency difference between the two habitats, and $d\Delta/dt$ the corresponding rate of change (eq. (S12b)) obtained by neglecting drift and assuming deleterious allele frequencies in the common habitat to be low. Values of Δ at which $d\Delta/dt$ equals 0 correspond to equilibria; an equilibrium is stable only if the curve is downward sloping, i.e., if the first derivative of $d\Delta/dt$ w.r.t. Δ is negative, at the equilibrium. For low Ls (fig. A), there is a critical migration level $m_c/s=1/(1-2\rho)$ (blue curve), above which the Δ=0 equilibrium becomes stable, while the $\Delta>0$ equilibrium (marked by $\times$ symbols) vanishes. For high Ls (fig. B), there are two thresholds: the Δ=0 equilibrium becomes stable for $m/s>1/(1-2\rho)$ but the $\Delta>0$ equilibrium also persists, along with a third unstable equilibrium. At a second migration threshold (blue curve), the stable and unstable equilibrium collide: the corresponding value of Δ at which this occurs is Δ_0 (marked by a circle in fig. B), which is the minimum level of stable divergence between habitats that is possible. Above this migration threshold, only a single stable equilibrium at Δ=0 exists. (C) The minimum possible level of adaptive divergence, Δ_0, vs. Ls for various ρ. Here, Δ_0 is calculated using the expression in eq. (S14), and is the divergence level corresponding to the critical migration threshold m_c/s. There is a threshold $(Ls)_*$ (which increases with increasing ρ) such that a non-zero Δ_0 exists only for $Ls>(Ls)_*$. Thus, the critical migration rate depends on Ls only for $Ls>(Ls)_*$ (see also text).

(unstable) positive equilibrium separating the two stable equilibria. Increasing m/s now causes the unstable and stable equilibria to converge towards a common value Δ_0 (marked by a triangle in fig 2(b)). At a critical migration threshold m_c/s (blue curve), the two equilibria collide (this corresponds to a so-called saddle-node bifurcation), and we have $\Delta_*=\Delta_0$ (marked by a circle in fig. 2(b)). Above this threshold, the two non-zero (stable and unstable) equilibria vanish, and there exists only a single stable equilibrium at $\Delta=0$ (black curve). Note that m_c/s is greater than the threshold $1/(1-2\rho)$ at which the $\Delta=0$ equilibrium becomes stable.

We can determine Δ_0 by noting that it is that value of Δ at which both $d\Delta/dt$ and its first derivative with respect to Δ become zero: note the shape of the blue curve at this equilibrium. The first condition, namely $\frac{d\Delta}{dt}\big|_{\Delta=\Delta_0}=0$, implies: $(m/s)e^{-2sL\Delta_0}[2\rho(1-\Delta_0)-1]+1-\Delta_0=0$ (as above). The second condition, namely $\frac{\partial}{\partial\Delta}\left(\frac{d\Delta}{dt}\right)\big|_{\Delta=\Delta_0}=0$, implies $(m/s)e^{-2Ls\Delta_0}[\rho(4\Delta_0(Ls(\Delta_0-1)-1)+2)+2Ls\Delta_0-1]-2\Delta_0+1=0$: this simply follows from eq. (S13), by differentiating $d\Delta/dt$ with respect to Δ and setting the derivative to zero. Finally, we can obtain an explicit expression for Δ_0 by substituting the first condition into the second condition and solving the resultant quadratic equation. This yields:

$$\Delta_0=\frac{Ls(4\rho-1)+\sqrt{Ls(Ls-4\rho)}}{4Ls\rho} \tag{S14}$$

which depends only on Ls and ρ. Recall that Δ_0 is the level of per locus divergence between habitats at the critical migration threshold. Thus, the existence of a non-zero Δ_0 implies that loss of divergence occurs at a critical migration threshold m_c/s, which is *greater* than the corresponding single locus threshold $1/(1-2\rho)$, and involves a saddle-node bifurcation (as in fig. 2(b)). Conversely, if no non-zero Δ_0 exists, then this implies that loss of divergence occurs at a critical migration threshold m_c/s, which is equal to the corresponding single locus threshold $1/(1-2\rho)$, and involves a transcritical bifurcation (as in fig. 2(a)).

Figure 2(c) shows Δ_0 as a function of Ls for various values of ρ. We note that there exists a non-zero Δ_0 only above a threshold value of Ls, which we denote by $(Ls)_*$. For $\rho<1/4$, the critical divergence level Δ_0 decreases as Ls decreases, approaching 0 as $Ls\to(Ls)_*$. This allows us to solve for $(Ls)_*$ by setting $\Delta_0=0$ in eq. (S14), which gives $(Ls)_*=1/[2(1-2\rho)]$ for $\rho<1/4$. For $\rho>1/4$ we note that the critical divergence level Δ_0 approaches a non-zero value as $Ls\to(Ls)_*$: from eq. (S14), we note that this bifurcation must occur at $Ls=4\rho$. Thus, for $\rho>1/4$, we have $(Ls)_*=4\rho$, with $\Delta_0\to1-1/(4\rho)$ as $Ls\to(Ls)_*$. This thus provides the deterministic predictions for the threshold $(Ls)_*$ above which the critical migration threshold starts increasing with increasing Ls (dashed vertical lines in fig. 4 of the main text).

References

Bengtsson, B. O. 1985. The flow of genes through a genetic barrier. Pages 31–42 Cambridge University Press Cambridge; New York.