OXFORD GENETICS

GENETICS, 2024, 228(3), iyae140
https://doi.org/10.1093/genetics/iyae140
Advance Access Publication Date: 22 August 2024
Investigation

The genetic architecture of polygenic local adaptation and its role in shaping barriers to gene flow

Arthur Zwaenepoel [1,*], Himani Sachdeva [2,‡], Christelle Fraïsse [1,‡]

[1]UMR 8198 – Evo-Eco-Paleo, CNRS, Univ. Lille, Lille F-59000, France
[2]Department of Mathematics, University of Vienna, Vienna 1090, Austria

*Corresponding author: UMR 8198 – Evo-Eco-Paleo, CNRS, Univ. Lille, Lille F-59000, France. Email: arthur.zwaenepoel@univ-lille.fr
‡Contributed equally.

We consider how the genetic architecture underlying locally adaptive traits determines the strength of a barrier to gene flow in a mainland-island model. Assuming a general life cycle, we derive an expression for the effective migration rate when local adaptation is due to genetic variation at many loci under directional selection on the island, allowing for arbitrary fitness and dominance effects across loci. We show how the effective migration rate can be combined with classical single-locus diffusion theory to accurately predict multilocus differentiation between the mainland and island at migration–selection–drift equilibrium and determine the migration rate beyond which local adaptation collapses, while accounting for genetic drift and weak linkage. Using our efficient numerical tools, we then present a detailed study of the effects of dominance on barriers to gene flow, showing that when total selection is sufficiently strong, more recessive local adaptation generates stronger barriers to gene flow. We then study how heterogeneous genetic architectures of local adaptation affect barriers to gene flow, characterizing adaptive differentiation at migration–selection balance for different distributions of fitness effects. We find that a more heterogeneous genetic architecture generally yields a stronger genome-wide barrier to gene flow and that the detailed genetic architecture underlying locally adaptive traits can have an important effect on observable differentiation when divergence is not too large. Lastly, we study the limits of our approach as loci become more tightly linked, showing that our predictions remain accurate over a large biologically relevant domain.

Keywords: reproductive isolation; local adaptation; barriers to gene flow; genetic architecture; dominance

Introduction

When a population is subdivided across multiple habitats with different environmental conditions, the extent to which distinct subpopulations can maintain locally beneficial genetic variation depends on the rate of migration between them. Migration between populations that maintain divergently selected alleles can generate migration load (a reduction in mean fitness due to the influx of locally maladaptive alleles) or may lead to loss of local adaptation altogether (so-called *swamping* by gene flow) (e.g. Lenormand 2002). While local adaptation may be driven by a few conspicuous loci (e.g. adaptive melanism in peppermoths (van't Hof *et al.* 2016) or pocket mice (Nachman *et al.* 2003)), it is believed to typically be polygenic, involving alleles of different effect at many loci across the genome (Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012; Schumer *et al.* 2018; Westram *et al.* 2018; Martin *et al.* 2019; Barghi *et al.* 2020; Bomblies and Peichel 2022; Stankowski *et al.* 2023).

When local adaptation is polygenic, migration between populations adapted to different environmental conditions will generate linkage disequilibria (LD), i.e. statistical associations, between selected loci, as locally deleterious alleles will tend to reside in the genomes of individuals with recent migrant ancestry. The rate at which individual locally deleterious alleles are eliminated will be affected by these associations, a phenomenon often referred to as a "coupling" effect (Barton 1983; Kruuk *et al.* 1999; Feder *et al.* 2012; Yeaman 2015; Sachdeva 2022). Indeed, sets of loosely linked locally deleterious alleles introduced by migrants will be eliminated jointly in the first few generations after migration at a rate which depends essentially on the relative fitness of migrant individuals and their immediate descendants. These coupling effects will in turn affect the equilibrium migration load and swamping thresholds (i.e. the migration rate beyond which local adaptation is lost). Neutral variation may also come to be associated with locally selected alleles, so that the latter constitute a "barrier" to neutral gene flow, increasing neutral genetic differentiation (as quantified by F_{ST} for instance) beyond the single-locus neutral expectation (Petry 1983; Bengtsson 1985).

Barrier effects due to divergent selection at many loci may play an important role in the evolution of reproductive isolation (RI), and hence speciation (Nosil 2012; Barton 2020). The colonization of a new habitat will often involve selection on polygenic traits and give rise to a subpopulation that exhibits some divergence from its ancestors (Barton and Etheridge 2018). Conditional on the initial successful establishment of such a divergent

Received on 15 February 2024; accepted on 12 August 2024

subpopulation, whether or not speciation ensues depends on whether local adaptation can be maintained in the face of maladaptive gene flow (if any), and on the extent to which the partial RI deriving from local adaptation may promote further divergence and strengthen RI (e.g. through reinforcement, coupling of locally adaptive alleles with intrinsic incompatibilities, or the establishment of additional locally beneficial mutations; (Barton and De Cara 2009; Bierne et al. 2011; Butlin and Smadja 2018; Kulmuni et al. 2020)).

Despite mounting evidence that local adaptation is indeed often polygenic, little is known about the underlying genetic details and how these influence the maintenance of adaptive differentiation in the face of maladaptive gene flow (Yeaman and Whitlock 2011; Yeaman 2015; Bomblies and Peichel 2022). How many loci are involved? What are the typical effect sizes? Are divergently selected alleles typically closely linked or spread all over the genome? How nonadditive is local adaptation? Relatedly, it is currently unclear how the detailed genetic architecture of local adaptation affects the strength of the resulting barrier to gene flow across the genome. These issues do not only arise in the study of local adaptation per se, but are also central to speciation research, as we still have little insight in the genomic architecture underlying isolating mechanisms in nature (Jiggins and Martin 2017; Ravinet et al. 2017).

In a recent paper, Sachdeva (2022) showed that, when the loci under selection are unlinked, the effects of LD on equilibrium differentiation at any individual locus in a polygenic system can be well described by classical single-locus population genetic theory, provided that the migration rate m is substituted by an *effective* migration rate m_e (Petry 1983; Bengtsson 1985; Barton and Bengtsson 1986; Kobayashi et al. 2008), which captures the multilocus barrier effect, i.e. how gene flow at a focal locus is affected by selection against the associated genetic background. The effective migration rate for a neutral locus can furthermore serve as a quantitative measure of RI, i.e. $\mathrm{RI} = 1 - m_e/m$ (Westram et al. 2022). Crucially, m_e depends itself on the frequencies of divergently selected alleles, giving rise to feedback effects where a small increase in migration rate may cause a sudden collapse of local adaptation (i.e. swamping). In Sachdeva (2022), a detailed study was conducted of the joint effects of drift and LD on swamping thresholds and neutral differentiation in the mainland-island and infinite-island models of population subdivision, assuming a haploid sexual life cycle and divergently selected loci of equal effect.

In this paper, we focus on how the genetic architecture of a locally adaptive additive trait determines the strength of a barrier to gene flow and the conditions for maintaining local adaptation. Extending the theoretical framework of Sachdeva (2022), we derive an expression for the effective migration rate at a neutral locus under a polygenic architecture of local adaptation with arbitrary fitness and dominance effects across loci (referred to as a *heterogeneous* barrier), assuming a population with a general life cycle (which includes haplontic and diplontic life cycles as special cases) and weak linkage. We use this m_e to build an approximation for the marginal allele frequency distributions at migration–selection balance in a mainland-island model, allowing us to quantify the extent of adaptive differentiation at equilibrium and swamping thresholds at individual loci influencing the polygenic trait.

After showing the accuracy of our approximations, we use these to address a number of questions concerning the relationship between the genetic architecture of divergently selected traits and the strength of the resulting barrier effect across the genome. In particular, we ask: Are barriers to gene flow stronger when locally adaptive alleles are recessive or dominant? Does a more heterogeneous architecture of local adaptation lead to stronger barriers to gene flow or rather the converse? How does the distribution of fitness effects (DFE) of locally adaptive alleles affect equilibrium differentiation and swamping thresholds at individual loci? How does linkage affect the strength of a barrier to gene flow and swamping thresholds? Finally, in the Discussion section, we summarize our main findings and the limits of our approach, and consider the relevance of our theoretical work for the inference of the genetic architecture of local adaptation and barriers to gene flow from genomic data.

Model and methods

Mainland-island model

Here we outline a mainland-island model for a sexual population subject to selection in both the haploid and diploid phases. We assume a regular and synchronous alternation of generations, where an island population of N haploids (gametophytes) produces an effectively infinite pool of gametes from which $2Nk$ gametes are sampled that unite randomly to form Nk diploid individuals (sporophytes), k being the number of diploids per haploid individual (Supplementary section 2.1). The diploid generation produces in turn an effectively infinite pool of haploid spores through meiosis, of which N are drawn to form the next haploid generation. Throughout, we assume that sexes need not be distinguished. Our notation is summarized in Table 1.

We shall focus on the case of haploid-phase migration, where in each generation, M haploid individuals on the island are replaced by haploid mainland individuals, where M is Poisson distributed with mean Nm. Our models and approximations are however straightforwardly extended to allow for gametic or diploid migration, or indeed any combination of these (see Appendix A). Fitness on the island is determined by L biallelic loci which are under divergent selection relative to the mainland. The L loci may be linked or unlinked, and we denote the recombination rate between locus i and j by r_{ij}. The mainland population is

Table 1. Glossary of the main symbols (parameters, variables) and some of their interrelations.

Sym.	Equals	Description
0	—	Beneficial allele on the island
1	—	Deleterious allele on the island
s_1	—	Haploid-phase selection coefficient against the 1 allele
s_{01}	—	Diploid-phase selection coefficient against heterozygotes
s_{11}	—	Diploid-phase selection coefficient against homozygotes
s	$2s_1 + s_{11}$	Effective selection coefficient
h	$\frac{s_1+s_{01}}{2s_1+s_{11}}$	Effective dominance coefficient
N	—	Number of haploid individuals
k	—	Number of diploid individuals per haploid individual
N_e	$(\frac{1}{N}+\frac{1}{2Nk})^{-1}$	Effective population size
L	—	Number of selected loci
u	—	Mutation rate
m	—	Migration rate
r_{ij}	—	Recombination rate between locus i and j
p_i	$1-q_i$	Frequency of the 0 allele at locus i on the island
p_i^*	$1-q_i^*$	Frequency of the 1 allele at locus i on the mainland
p_{-i}	—	Vector of allele frequencies at all loci except locus i
pq_{-i}	—	Vector of heteroyzgosities at all loci except locus i
g_i	—	Gene flow factor at locus i

Additional subscripts i and j for selection coefficients refer to locus-specific parameters.

assumed to have a constant genetic composition. In the results section we shall assume the mainland to be fixed at each locus for the allele which is deleterious on the island. Throughout, we assume a scenario of secondary contact where the locally deleterious allele is initially rare on the island.

Fitness effects are allowed to vary arbitrarily across loci. The wild-type and deleterious alleles (denoted by `0` and `1`) at locus i have relative fitnesses on the island of 1 and $e^{-s_{i1}}$ in the haploid phase. The relative fitnesses for the three genotypes in the diploid phase `00`, `01`, and `11` at locus i are 1, $e^{-s_{i01}}$, and $e^{-s_{i11}}$. Throughout, we denote the frequency of the wild-type allele (on the island) at locus i by p_i, and the frequency of the locally deleterious allele by $q_i = 1 - p_i$. Fitness is multiplicative across loci, so that, for instance, the log relative fitness of a haploid individual fixed for all the `1` alleles is given by $\log w = -\sum_{i=1}^{L} s_{i1}$. We assume that each haploid (diploid) individual contributes gametes (spores) to the gamete (spore) pool in proportion to its fitness. We assume symmetric mutation at a small constant rate u per locus, occurring at meiosis. Individual-based simulations of this model are implemented in a Julia package (Bezanson *et al.* 2017) available at https://github.com/arzwa/Sewall.

In the following sections, we build up a theoretical approximation to this model, and validate the approximations by comparing numerical results against individual-based simulations.

Single-locus theory

We start by considering the evolutionary dynamics at a single locus in the island population with and without drift. Next, we use this single-locus theory to approximate the multilocus dynamics by defining an appropriate *effective migration rate*, capturing the effects of all the other loci on any one focal locus. We then outline a numerical approach to study the equilibria of the multilocus model using this approximation.

We first consider a deterministic model for the allele frequency dynamics at a single locus, ignoring the influence of the other loci as well as genetic drift. As shown in detail in Supplementary section 2.1, for weak selection and migration, the dynamics of p can be described in continuous time by the nonlinear ordinary differential equation (ODE)

$$\frac{dp}{dt} = -m(p - p^*) + spq(h + (1 - 2h)q), \tag{1}$$

where p^* is the frequency on the mainland of the allele which is beneficial on the island, $s = 2s_1 + s_{11}$ the effective selection coefficient (combining haploid-phase and diploid-phase selection) and $h = \frac{s_1 + s_{01}}{2s_1 + s_{11}}$ the effective dominance coefficient. The equilibria of eq. (1) are analyzed in detail in Supplementary section 2.1. We assume throughout that s is positive, and p^* will be assumed to be small, so that selection increases p, whereas migration decreases p.

When evolutionary forces are sufficiently weak, diffusion theory can be applied to approximate the equilibrium allele frequency distribution at a single locus in a finite population. Note that our haplodiplontic model is akin to the standard Wright–Fisher (WF) model with a population size that regularly alternates between N and $2Nk$ gene copies. The corresponding effective population size is hence $N_e = (N^{-1} + (2Nk)^{-1})^{-1}$, twice the harmonic mean of the phase-specific number of gene copies (Hein *et al.* 2004) (twice because our unit of time is an alternation of generations, not a single generation). The equilibrium allele frequency distribution at a single locus is then given by

$$\phi(p) \propto p^{2N_e(u+mp^*)-1} q^{2N_e(u+mq^*)-1} e^{-N_e sq(2h+(1-2h)q)}, \tag{2}$$

where no closed-form expression is known for the normalizing constant. This is essentially Wright's distribution, generalized to a haplodiplontic life cycle (Wright 1937).

Effective migration rate

To make the bridge from single-locus to multilocus theory, we derive an expression for the *gene flow factor* (gff), i.e. the reduction in gene flow at a neutral locus (relative to the "raw" migration rate m) due to its being in LD with multiple selected loci (Bengtsson 1985; Barton and Bengtsson 1986). As shown formally in Kobayashi *et al.* (2008), for weak migration, the gff at an unlinked neutral locus equals the expected reproductive value (RV) of migrants in the resident background. This is the expected long-term genetic contribution of a migrant individual to the local population, relative to that of a resident individual. In order to calculate the expected RV of migrants, one has to track the average reproductive output of a migrant individual (relative to resident individuals), as well as that of its descendants over multiple generations. At any time, the proportion of individuals with recent migrant ancestry on the island is $O(m)$, so that the probability of individuals with migrant backgrounds mating with each other to produce, for instance, F2 crosses of the migrant and resident genotypes, is $O(m^2)$, and hence negligible for sufficiently weak migration. The descendants of a migrant individual will therefore most likely be F1s and subsequent backcrosses with the resident population, so that to a good approximation, the RV of a migrant depends only on the relative fitnesses of F1, BC1, BC2, etc. individuals.

Let $W_h^{(n)}$ and $W_d^{(n)}$ denote the relative fitness of an individual derived from an nth generation haploid, respectively diploid, backcross of a migrant with the resident population (i.e. $W_d^{(1)}$ is the relative fitness of an F1 diploid, $W_d^{(2)}$ of an offspring from a F1 × resident cross (BC1 generation), etc.). Assuming migration occurs in the haploid phase before selection, the gff for a neutral locus can be expressed as

$$g = \frac{m_e}{m} = \mathbb{E}\left[W_h^{(0)} \prod_{n=1}^{\infty} W_d^{(n)} W_h^{(n)} \right], \tag{3}$$

where $W_h^{(0)}$ is the relative fitness of the haploid migrant in the resident population (Barton and Etheridge 2018; Sachdeva 2022; Westram *et al.* 2022). Note that this involves an expectation over all possible lines of descent of an initial haploid migrant. In practice, g is determined only by the first 10 backcross generations or so, as subsequent backcrosses are essentially indistinguishable from residents. While this gff applies strictly only to neutral loci (Kobayashi *et al.* 2008), it gives a reasonable approximation for the reduction in gene flow at (weakly) selected loci as well.

In order to derive a useful approximate expression for g, we shall make two further important assumptions: (1) both the resident and migrant subpopulations, as well as each backcross generation, is in Hardy–Weinberg and linkage equilibrium (HWLE); (2) the expected allele frequency at any locus in any backcross generation is midway between that of the parents (e.g. the mean of the mainland and island allele frequencies for the F1 generation). In reality, due to Mendelian segregation, individuals will not inherit exactly half of the selected alleles of each parent, and this segregation variance will lead to variation within F1s, BC1s, etc. on which selection can act, resulting in deviations from the midparent value. However these deviations are $O(s^2)$, and can be ignored

for $s \ll 1$. As derived in Appendix A, in the unlinked case, the approximate gff at locus j under these assumptions is

$$g_j \approx \exp\left[-2\sum_{i\neq j} s_i h_i\left(q_i^* - \mathbb{E}[q_i]\right) - s_i(1-2h_i)\left(p_i^*\mathbb{E}[q_i] - \mathbb{E}[p_i q_i]\right)\right], \quad (4)$$

It is worth stressing that the gff is a function only of the expected allele frequencies and heterozygosities ($\mathbb{E}[p_i q_i]$) at the selected loci, and does not depend on any multilocus associations between selected loci. Also, although we assume migration is sufficiently rare ($m \ll 1$), alleles introduced by migrants may be common (if $m \sim s$), resulting in appreciable heterozygosities. We shall often highlight the dependence of the gff on the allele frequencies and heterozygosities by writing $g[\mathbb{E}[p], \mathbb{E}[pq]]$. Note further that eq. (4) corresponds to $(\mathbb{E}[W_h^{(0)}]\mathbb{E}[W_d^{(1)}])^2$, i.e. the product of the relative fitness of a haploid migrant in the haploid resident population and the relative fitness of the diploid F1 in the diploid resident population, squared. Hence, g can, in principle, be determined empirically. Equation (4) is straightforwardly adapted to allow for migration in the diploid phase (see Appendix A).

As shown in Appendix B, we can heuristically account for weak linkage by considering the allele frequency dynamics at a neutral locus linked to a single barrier locus. The resulting expression for the gff is

$$g_j \approx \exp\left[-\sum_{i\neq j} \frac{s_i h_i\left(q_i^* - \mathbb{E}[q_i]\right) - s_i(1-2h_i)\left(p_i^*\mathbb{E}[q_i] - \mathbb{E}[p_i q_i]\right)}{m + r_{ij} + s_i(h_i - \mathbb{E}[q_i] + 2(1-2h_i)\mathbb{E}[p_i q_i])}\right], \quad (5)$$

Note that when $r \gg s \sim m$ (i.e. recombination is much stronger than selection per locus) the denominator in the sum in the exponent becomes $\approx r_{ij}$, so that eq. (4) appears as a special case of eq. (5) when $r_{ij} = 0.5$ for all i, j.

Multilocus dynamics and equilibria

The gff captures the effect of LD among selected loci on the rate of gene flow from the mainland into the island at any individual locus. The key observation is that a certain separation of time scales applies: although selection against migrant *genotypes* can be very strong (of magnitude Ls, roughly), these genotypes are rapidly broken down by recombination (at a rate $r_{ij} \sim r$). Thus, as long as recombination among selected loci is stronger than selection per locus ($r \gg s$), LD decays much faster than allele frequencies change. The main effect of LD is then to reduce the effective rate of migration at any individual selected or neutral locus (Sachdeva 2022).

As a consequence, in the deterministic case, we expect that the effects of LD should be captured by substituting the effective migration rate $m_e = mg$ for m in eq. (1). Specifically, we get a system of L coupled differential equations, where for $1 \leq j \leq L$,

$$\frac{dp_j}{dt} = -mg_j[p_{-j}]p_j + s_j p_j q_j(h_j + (1-2h_j)q_j), \quad (6)$$

where we assumed the mainland to be fixed for the deleterious allele on the island at all loci. Here we write $g_j[p_{-j}]$ for the gff to highlight the dependence of the gff at locus j on the allele frequencies at the other $L-1$ loci. One can find the equilibria of this model by numerically solving for p at stationarity (as we do in Supplementary section 2.5), but here we will focus on the more general stochastic case with arbitrary $N_e s$.

In order to account for LD in the stochastic case, we can plug m_e into the single-locus diffusion approximation to determine the equilibrium allele frequency distribution for each locus on the island (Sachdeva 2022). Specifically, we can compute moments of the allele frequency distribution at each locus by solving self-consistently for $\mathbb{E}[p_j]$ and $\mathbb{E}[p_j q_j]$ in

$$\begin{aligned}\mathbb{E}[p_j] &= Z_j^{-1}\int p_j \phi(p_j;\, N_e,\, u,\, mg_j[\mathbb{E}[p_{-j}],\, \mathbb{E}[pq_{-j}]],\, s_j,\, h_j)dp_j \\ \mathbb{E}[p_j q_j] &= Z_{j'}^{-1}\int p_j q_j \phi(p_j;\, N_e,\, u,\, mg_j\left[\mathbb{E}[p_{-j}],\, \mathbb{E}[pq_{-j}]\right],\, s_j,\, h_j)dp_j,\end{aligned} \quad (7)$$

where the Z's are normalizing constants, and $\mathbb{E}[pq_{-j}]$ is the vector of expected heterozygosities at all loci excluding locus j. To solve this system of $2L$ nonlinear equations, we use the fixed point iteration outlined in Supplementary section 2.3. The numerical methods used in this paper are also implemented in the Julia package available at https://github.com/arzwa/Sewall.

Realized genetic architecture of divergent selection

For a given DFE of divergently selected loci, we can use the diffusion approximation to predict the distribution of s and h (understood as effective selection and dominance coefficients) for those loci that maintain divergent alleles at migration–selection balance. We refer to this as the *realized* architecture of local adaptation, and quantify it as the conditional probability density for s and h at locus i, given that a randomly sampled gene copy at locus i from the island population is the locally beneficial allele (the event $X_i = 1$, where X_i is an indicator random variable), i.e.

$$\begin{aligned} f(s_i,\, h_i \mid X_i = 1) &= \frac{\int_{\mathcal{B}} \Pr\{X_i = 1 \mid s_i,\, h_i,\, B\} f_{\mathrm{DFE}}(s_i,\, h_i,\, B)dB}{\Pr\{X_i = 1\}} \\ &\propto \int_{\mathcal{B}} \mathbb{E}[p_i \mid s_i,\, h_i,\, B] f_{\mathrm{DFE}}(s_i,\, h_i,\, B)dB. \end{aligned} \quad (8)$$

Here, f_{DFE} denotes the joint density of the selection and dominance coefficient across the L divergently selected loci, and B is a shorthand for the selection and dominance coefficients at the $L-1$ other loci ("B" for background), and we integrate over the set of all possible such backgrounds $\mathcal{B}$. For a given DFE model, we can characterize this conditional probability density using a Monte Carlo approach by sampling random L-locus genetic architectures from the DFE and calculating the expected beneficial allele frequency $\mathbb{E}[p_i \mid s_i,\, h_i,\, B]$ for each locus in the barrier as a weight for the associated $(s_i,\, h_i)$ pair. The weighted sample will be distributed according to f.

Results

Our main aim is to elucidate how the genetic architecture of polygenic divergent selection determines the strength of a barrier to gene flow, and how, in turn, a polygenic barrier to gene flow affects swamping thresholds at individual loci under divergent selection. We first ignore linkage, and assess the accuracy of our theoretical predictions for the unlinked case by comparing numerical results based on the multilocus approximation against individual-based simulations. Having established the validity of our approximations, we next examine the effects of dominance and heterogeneous genetic architectures on the strength of a barrier to gene flow and swamping thresholds. Finally, we consider the effects of linkage between barrier loci for realistic genetic maps.

Throughout, we focus on parameter regimes where drift may be appreciable but does not overwhelm selection per locus

($N_e\bar{s} \geq 5$, where $\bar{s}$ is the average selection coefficient across the L-locus genetic architecture), where the rate of migration is roughly of the same order as the strength of selection per locus ($0 \leq m/\bar{s} \leq 2$, say), and where multilocus selection is appreciable but not so strong as to likely yield complete RI (roughly $0.5 \leq L\bar{s} \leq 2$). Mutation per locus is always considered weak ($u/\bar{s} \leq 1/100$).

Evaluation of the approximation for unlinked loci

We find that substituting $m_e = mg[\mathbb{E}[p], \mathbb{E}[pq]]$ for m in the single-locus diffusion theory and solving self-consistently for $\mathbb{E}[p]$ and $\mathbb{E}[pq]$ (see eqs. (4) and (7)) yields remarkably accurate predictions for allele frequencies and swamping thresholds as observed in individual-based simulations (Fig. 1 and Supplementary Fig. 1). Not only can we reliably obtain the expected allele frequencies, but we also obtain very good predictions for the entire allele frequency distribution (Fig. 1, right).

Effect of dominance on barrier strength and swamping thresholds

The effects of dominance on migration–selection balance at a single locus are well known (Haldane 1930; Nagylaki 1975, recapitulated in Supplementary section 2.2). However, it is less clear how dominance affects the ability to maintain adaptive differentiation and RI for polygenic architectures (but see Harris and Nielsen 2016 for an investigation in the context of selection against introgressed ancestry after an admixture pulse).

To explore the effects of dominance in a polygenic barrier, we assume all loci have the same fitness effects (a *homogeneous barrier*) and that the mainland is fixed for the locally deleterious allele on the island ($q_i^* = 1$). Under these assumptions, we obtain the simpler expression

$$g = e^{-2Lsh\mathbb{E}[p]} e^{-2Ls(1-2h)\mathbb{E}[pq]} \tag{9}$$

Note that the loci are indistinguishable under the stated assumptions, hence we can write $\mathbb{E}[p_i] = \mathbb{E}[p]$ for all $i \in [1..L]$. Here, the first factor is just the gff associated with a haploid L-locus system with selection coefficients sh. The second factor captures the effects of dominance and depends on the heterozygosity $\mathbb{E}[pq]$. Clearly, h has opposing effects on both factors. The immediate effect of dominance is therefore that the gff is decreased (barrier strength increased) relative to the additive case ($h = 1/2$) whenever invading alleles exhibit a dominant deleterious effect on the island ($h > 1/2$). Only when the heterozygosity becomes appreciable does the second factor contribute to the increase (when $h > 1/2$) or decrease (when $h < 1/2$) of the gff.

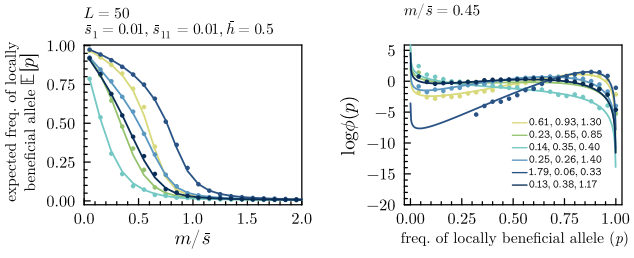

Fig. 1. The theoretical approximation for unlinked loci agrees well with individual-based simulations. Left: expected equilibrium frequencies for the locally adaptive allele on the island ($\mathbb{E}[p]$) for increasing migration rates for six loci in an unlinked multilocus barrier ($L = 50$). Right: frequency distributions ($\log_{10}\phi$, see eq. (2)) at $m/\bar{s} = 0.45$ for the same set of loci. Lines show predictions from the multilocus approximation, whereas dots show results from individual-based simulations (simulating for $200N_e$ generations, sampling every 10th generation after discarding the first $50N_e$ generations). We assume $N_e = 500$ and $L = 50$ loci, with selection coefficients in the haploid and diploid phase at each locus sampled from an exponential distribution with mean 0.01, and dominance coefficients for the diploid phase sampled from a Uniform(0, 1) distribution (the values of s_1, s_{01}, and s_{11} for the six highlighted loci are shown in the legend (×100)). We assume $u = \bar{s}/100$.

Effect of dominance on equilibrium frequencies

We now examine the effect of dominance in more detail (Fig. 2). We find, as expected, that stronger net selection against maladapted genotypes (larger Ls) increases the equilibrium frequency of the locally beneficial allele relative to the single-locus prediction, but that the magnitude of this effect depends quite strongly on dominance. When invading alleles are recessive (Fig. 2, $h = 0$), gene flow is not at all impeded while migration is weak and locally deleterious alleles are rare on the island. This is essentially because, F1s, BC1s, etc. will almost always be heterozygous for the deleterious allele (even when migrants are homozygous), so that deleterious alleles are not "seen" by selection and the gff is close to one. Only once deleterious alleles are segregating at appreciable frequencies on the island, are F1, BC1, etc. individuals likely to be homozygous at several loci, thus exposing locally deleterious alleles to selection and reducing the RV of migrants. Thus, recessive invading alleles contribute to the genomewide barrier to gene flow only above a certain level of migration.

The situation is clearly different when invading alleles are dominant (Fig. 2, $h = 1$), as these will immediately express their full load in the resident population even in the heterozygous state, causing them to be efficiently eliminated (the gff being at its minimum when migrant alleles are rare). Any increase in the frequency of the deleterious allele on the island will merely increase the expected relative fitness of individuals with migrant ancestry, and hence reduce the barrier effect (increase the gff). We observe a transition between these two qualitatively different types of behavior at intermediate values of h: when $h < 1/3$, the barrier strength *increases* with increasing introgression, decreasing the rate of gene flow, until differentiation falls below $(3h - 1)/(4h - 2)$. When $h > 1/3$ on the other hand, an increase in the frequency of introgressed alleles always reduces the barrier strength (Supplementary Fig. 2).

Effect of dominance on swamping thresholds

In the single-locus model, arbitrarily small frequencies of the locally beneficial allele can be maintained at migration–selection

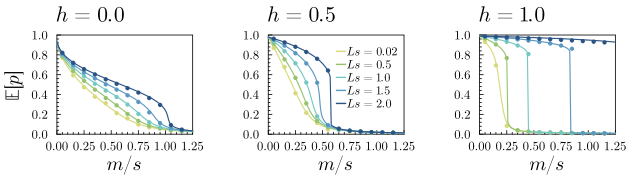

Fig. 2. Recessive local adaptation (invading alleles are dominant) yields stronger barriers to gene flow and sharper swamping thresholds. Equilibrium frequencies ($\mathbb{E}[p]$) of the locally beneficial alleles in a genetic architecture with L equal-effect loci are shown for increasing Ls for the case of recessive ($h = 0$), additive ($h = 0.5$) and dominant ($h = 1$) invading alleles. Note that the mainland is fixed for the alternative allele, so that $\mathbb{E}[p]$ corresponds to the expected allele frequency difference at equilibrium. The lines show the predictions based on the multilocus approximation, whereas the dots show results from individual-based simulations. We assume $s = 0.02$, $N_e s = 20$, $u = s/100$. Individual-based simulations were run for $50N_e$ generations, sampling every 10 generations after discarding the first $10N_e$.

balance when $h < 2/3$, whereas in the case of $h > 2/3$ (i.e. invading alleles are dominant), swamping occurs abruptly if the frequency of locally beneficial alleles falls below 0.5 (Supplementary section 2.2). Sachdeva (2022) showed that such sharp thresholds for swamping also appear in the absence of dominance due to LD between immigrant alleles. This is because a small increase in locally deleterious allele frequencies across many loci can substantially increase the relative fitness of migrant individuals and hence the effective migration rate, which further increases gene flow and deleterious allele frequencies, leading to a positive feedback effect (Sachdeva 2022). As the number of loci, and consequently Ls, increases, this effect will become more pronounced, increasing both the critical migration rate at which swamping occurs and the minimum level of differentiation that can be maintained before the swamping point is reached, leading to sharper swamping thresholds.

Our results indicate that dominance has a considerable influence on how LD sharpens and displaces swamping thresholds (Fig. 2). Sharp thresholds for swamping always appear when Ls becomes sufficiently large, irrespective of dominance, and the migration rate at which swamping occurs increases with increasing Ls. However, for $h < 2/3$ (i.e. when invading alleles are not strongly dominant), the critical migration threshold for swamping hardly shifts for $Ls < 2$. This is in sharp contrast with the case where local adaptation is due to strongly recessive alleles (invading alleles are dominant, $h > 2/3$), where the threshold for swamping increases rapidly with Ls (Fig. 2).

Importantly, the critical differentiation level below which local adaptation collapses is strongly affected by dominance. For instance, for $Ls = 2$ and $h = 0$, the differentiation between mainland and island decreases smoothly until a value of $\mathbb{E}[p] \approx 0.3$ is reached, after which local adaptation collapses (Fig. 2). For the $h = 0.5$ (additive case), this occurs at a value of $\mathbb{E}[p] \approx 0.6$, whereas for $h = 1$, differentiation is either almost complete or absent (Fig. 2). To see why this happens, it is helpful to contrast the expression for the gff (eq. (9)) for the $h = 0$ and $h = 1$ cases. When invading alleles are completely recessive (local adaptation is fully dominant, $h = 0$), we have $g = e^{-2Ls\mathbb{E}[pq]}$, i.e. the gff is only affected by the heterozygosity $\mathbb{E}[pq]$, which first increases as locally deleterious alleles become more common, thereby *decreasing* the gff. Hence, for weak migration there is a negative feedback: a small increase in deleterious allele frequency increases the heterozygosity, decreasing the rate of gene flow. Only once the frequency of deleterious alleles exceeds 0.5 does the heterozygosity start to decline again, and does the positive feedback described above for the additive case emerge. When invading alleles have a fully dominant effect ($h = 1$), $g = e^{-2Ls(\mathbb{E}[p]-\mathbb{E}[pq])}$, so that there is always a positive feedback between increasing deleterious allele frequency (falling p) and increasing gff. A more detailed analysis of swamping thresholds in the deterministic model defined by eq. (6) is provided in Supplementary section 2.5. Likewise, an analysis of the effect of the relative strength of selection in the haploid and diploid phase is included in Supplementary section 2.6.

We stress that our approach accounts for genetic drift. Drift reduces the efficiency of selection at each locus, reducing, on average, the allele frequency divergence between the mainland and island. This in turn increases the expected relative fitness of migrants and their descendants, thereby reducing the barrier effect and reducing the extent to which allele frequency divergence at any individual locus is increased by LD with the other selected loci (Supplementary Figs. 3 and 4). As a result, swamping thresholds are both decreased and made less sharp by drift, and the detailed behavior depends on the dominance coefficient (Supplementary Fig. 3B).

Heterogeneous genetic architectures

We now depart from the unrealistic assumption of equal selection and dominance coefficients across the polygenic barrier. In the methods section above, we developed the multilocus theory for general heterogenous genetic architectures (see eq. (4)), where the selection coefficients s_1, s_{01} and s_{11} can vary arbitrarily across loci, and we already verified that we do indeed obtain accurate predictions in this setting (Fig. 1). This allows us to address in more detail a number of questions pertaining to the genetic architecture of local adaptation at migration–selection balance, while accounting for both LD and genetic drift. All results in the following paragraphs apply to a general life cycle with arbitrary selection in both phases, provided that s and h are interpreted as *effective* selection and dominance coefficients (Table 1). Given the accuracy of our theoretical approximation (e.g. Fig. 1 and Supplementary Fig. 31), we restrict ourselves to numerical predictions in the following sections.

Effect of variation in fitness effects on overall differentiation

We first consider the case with variable selection coefficients across the L loci in the barrier, assuming no dominance. Fig. 3a shows the average per-locus expected differentiation across the barrier ($\bar{\Delta} = \sum_i^L \mathbb{E}[p_i - p_i^*]/L$) when selection coefficients are sampled from a Gamma distribution (note that $\bar{\Delta} = \sum_i^L \mathbb{E}[p_i]/L$, since we assume the mainland is fixed for the locally deleterious allele on the island, i.e. $p_i^* = 0$ for all i). When migration is weak relative to selection (roughly $m/\bar{s} < 1/4$), increasing the variance in fitness effects, while keeping $\bar{s}$ constant, yields on average lower equilibrium differentiation than a homogeneous barrier of strength $L\bar{s}$, although the barrier strength (as measured by the average gff across the L selected loci, $\bar{g}$) is hardly affected (Fig. 3b). By contrast, at higher migration rates, where loci with selection coefficients close to $\bar{s}$ become prone to swamping, heterogeneous architectures tend to yield higher equilibrium differentiation and a stronger barrier effect than a homogeneous one with the same total effect $L\bar{s}$ (Fig. 3a,b).

One should be careful, however, in the interpretation of $\bar{\Delta}$. As shown in Fig. 3(c,d), differentiation across loci in the barrier is often strongly bimodal, especially when Var[s] is large, where most loci are either strongly differentiated or not at all, and with rather few loci having $\mathbb{E}[p]$ near $\bar{\Delta}$. This implies that empirically, instead of detecting L selected loci with an average differentiation of $\bar{\Delta}$, we are more likely to observe about $L\bar{\Delta}$ strongly differentiated loci, at least when migration is relatively weak. For low $m/\bar{s}$, the weaker differentiation observed for more heterogeneous barriers is due to a smaller number of loci effectively contributing to local adaptation, with about half of the locally beneficial alleles swamped at $m/\bar{s} = 0.1$ while the other half is strongly differentiated (Fig. 3c, $\mathrm{Var}[s]/\bar{s}^2 > 1$) As expected, for stronger migration, the increased differentiation relative to the homogeneous case is driven by a subset of strongly selected loci that resist swamping (Fig. 3d).

These results are not significantly affected when there is variation in h, at least when h and s are uncorrelated (Supplementary Fig. 5). The effect of increasing Var[h], while keeping the s_i fixed across the barrier, is less dramatic than the effect of heterogeneity in selection coefficients, although we do see systematic increases or decreases in equilibrium differentiation depending on whether the migration rate exceeds the swamping

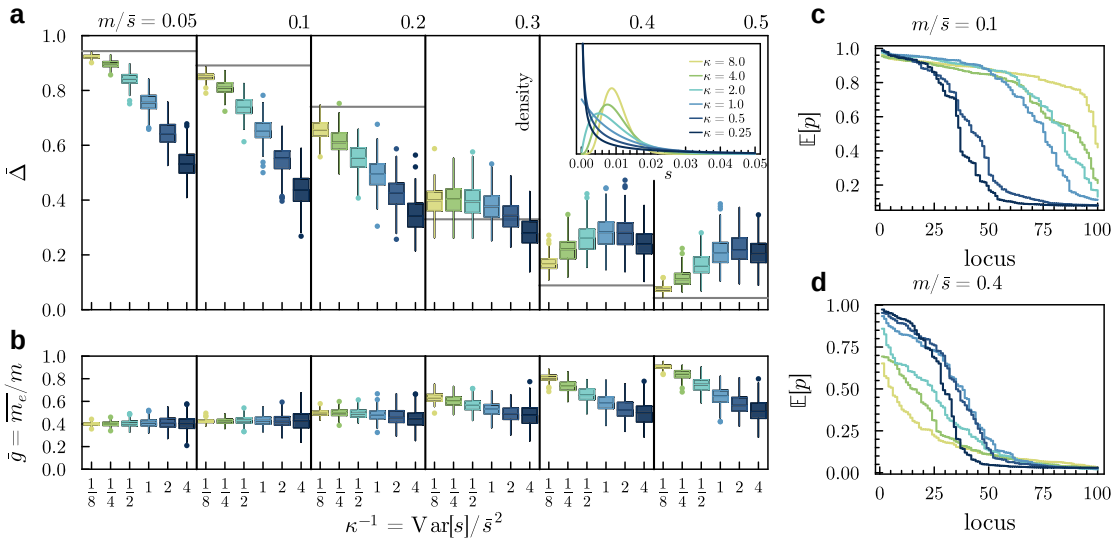

Fig. 3. More heterogeneous genetic architectures yield stronger barriers to gene flow when migration is high, but not when migration is low. a) The boxplots show the mean per-locus differentiation ($\bar{\Delta} = \sum_i^L \mathbb{E}[p_i](1-\mathbb{E}[p_i])/L$) across the $L = 100$ divergently selected loci, for 200 random L-locus architectures with no dominance and selection coefficients sampled from a Gamma(κ, $\kappa/\bar{s}$) distribution, with $\mathbb{E}[s] = \bar{s} = 0.01$ and six different values of κ (note that $\mathrm{Var}[s] = \bar{s}^2/\kappa$, i.e. $\kappa^{-1} \propto \mathrm{Var}[s]$). The solid horizontal line shows the predicted equilibrium differentiation per locus for a homogeneous barrier of strength $L\bar{s}$ (i.e. the prediction accounting for LD but not for heterogeneity in selection coefficients). (a, inset) Density functions for the six different Gamma distributions used in (a). b) Average locus-specific gff ($\bar{g}$) across the L divergently selected loci for the same random architectures as in (a). c,d) Expected beneficial allele frequencies across the barrier in a single simulation replicate for each of the six assumed distributions, sorted by allele frequency, assuming (c) $m/\bar{s} = 0.1$ and (d) $m/\bar{s} = 0.4$. Colors are as in (a). Other parameters are $N_e\bar{s} = 10$, $u/\bar{s} = 0.005$. All results are based on the multilocus approximation.

threshold for recessives (which are associated with higher equilibrium frequencies) or not (Supplementary Fig. 6).

Differentiation and swamping at individual loci in a heterogeneous barrier

Focusing on a single locus embedded within a heterogeneous barrier, we find that, for weak migration, variation in selection coefficients across the barrier has a negligible effect on differentiation at a focal locus with fixed selective effect, whereas (as already shown in Fig. 3) it does have a strong effect on average differentiation across the L loci (Supplementary Figs. 7 and 8). On the other hand, when migration is strong, a given locus shows on average higher equilibrium differentiation when embedded in a heterogeneous barrier than in a homogeneous one, even when the average differentiation across the barrier is lower in the former (Supplementary Figs. 7 and 8). These results are in line with the observed effect of Var[s] on $\bar{g}$ (Fig. 3b) and indicate that the presence of a small number of loci of large effect can have a strong impact on the total barrier strength, and hence on the expected differentiation at a focal selected locus when migration is strong (Supplementary Fig. 7). As the distribution of selection coefficients is generally believed to be at least somewhat leptokurtic (note that excess kurtosis $\propto\kappa^{-1}$ for our Gamma DFE model), we conclude that heterogeneity in selection coefficients can have important consequences for observable differentiation at migration selection-balance that would not be adequately captured by substituting an average selection coefficient in either single-locus or multilocus theory.

Figure 4 highlights how different loci in a heterogeneous barrier are affected differently by the genome-wide barrier effect. For each locus, the expected differentiation at equilibrium is compared against the corresponding single-locus prediction, showing the magnitude of the barrier effect. Selection and dominance coefficients across the L loci are assumed to be independently distributed according to a Gamma and Uniform distribution respectively (see Supplementary section 2.7). As expected, the barrier effect is strongly dependent on both $m/\bar{s}$ and the total strength of selection $L\bar{s}$ (Fig. 4), as well as the strength of genetic drift (Supplementary Fig. 9). Ignoring swamped loci, differentiation is most strongly affected for loci with recessive locally beneficial alleles, whereas the deviation from the single-locus prediction for dominant variants is considerably less (Fig. 4 and Supplementary Fig. 7). This is in line with our results for the effects of dominance in homogeneous architectures. Again we find that when migration is strong, increased heterogeneity of selection coefficients leads to stronger barrier effects, where an appreciable proportion of alleles are protected from swamping due to association with a few strongly selected barrier loci (Supplementary Figs. 10 and 11).

The realized architecture of local adaptation

Although the above results indicate that, when $L\bar{s}$ is appreciable, the barrier effect is strongest for recessive locally adaptive alleles (Fig. 4), this does not imply that the distribution of dominance coefficients among alleles that are divergently maintained at equilibrium will be shifted towards more recessive alleles (larger h). Although strongly selected recessives will tend to show strong differentiation, weakly selected recessive alleles will be more prone to swamping than partially dominant ones (e.g. Supplementary Fig. 7). One way to quantify how these two phenomena interact to yield the *realized* genetic architecture of divergent adaptation (related to the concept of *adaptive architecture*, as defined in Barghi *et al.* 2020) is by considering the conditional probability density for the selection and dominance coefficient at a locus, given that a locally adaptive allele is observed on the island at that locus (see Methods, eq. (8)). Fig. 4b shows Monte Carlo approximations to the marginal distributions $f(s_i \mid X_i = 1)$ and $f(h_i \mid X_i = 1)$ obtained in this way for the heterogeneous barrier model assumed in the preceding section.

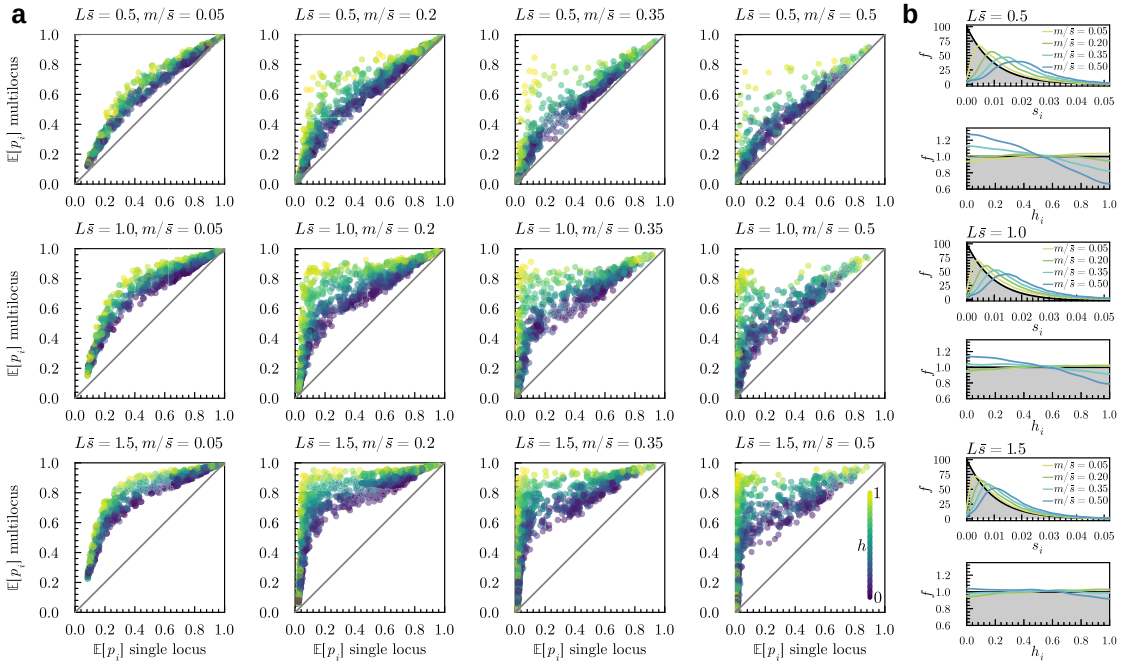

Fig. 4. Characterizing barrier effects for heterogeneous genetic architectures of local adaptation for different levels of migration and total barrier strength. a) Deviation of predicted allele frequencies for loci in heterogeneous polygenic barriers when accounting for LD (i.e. predictions using the multilocus theory, *y*-axis) from single-locus predictions that neglect LD (*x*-axis). The rows show results for different total strengths of selection ($L\bar{s}$), whereas the columns show results for increasing rates of migration relative to selection ($m/\bar{s}$). We assume the s_i to be exponentially distributed with mean $\bar{s}$ and dominance coefficients are sampled uniformly from the [0, 1] interval. Each dot is associated with a single locus in an L-locus barrier, and is colored according to its dominance coefficient (yellow for locally beneficial recessives ($h = 1$), purple for dominants ($h = 0$)). Each plot shows results for 1,000 such loci, subsampled from a total of 150,000/L randomly sampled L-locus architectures. b) Monte Carlo approximation to the marginal distribution of the selection and dominance coefficient conditional on observing a divergent allele on the island (i.e. $f(s_i \mid X_i = 1)$ and $f(h_i \mid X_i = 1)$, see eq. (8)). The distribution graphed in gray shows f_{DFE}, i.e. the marginal distribution of the selection and dominance coefficient for a random locus in the L-locus barrier in the absence of migration. We assumed $N_e\bar{s} = 20$ and $u/\bar{s} = 0.005$ for all results.

As expected, we find that as migration rates go up (colors in Fig. 4b), the distribution of selection coefficients in the barrier at migration–selection balance shifts towards higher values of s. This effect becomes weaker with increasing $L\bar{s}$, which increases the extent by which small-effect alleles are protected from swamping. Notably, recessives contribute *less* to the DFE at migration–selection balance compared to the DFE of the selected loci when migration is sufficiently strong (note the blue curves in Fig. 4b, corresponding to $m/\bar{s} = 0.5$). This is despite the fact that, conditional on no swamping, equilibrium frequencies of recessives are most affected by LD (Fig. 4a). This is most notable when $L\bar{s}$ is not large (top row in Fig. 4). When $L\bar{s} = 1.5$ for instance, the depression in the conditional density at $h = 1$ becomes very slight even for relatively large migration rates (bottom row in Fig. 4b). We observe a similar shift in the distribution of dominance coefficients when h is Beta distributed with mean 2/3 instead of uniformly on the unit interval (Supplementary Fig. 12).

Interestingly, we observe a nonzero correlation between s and h among selected loci that maintain divergent alleles, even when no such correlation exists *a priori* across the L divergently selected loci. At equilibrium, observed variants of relatively large effect are more likely to act recessively than variants of small effect (Fig. 5 and Supplementary Fig. 15). The correlation is negligible for small migration rates, but as the strength of migration increases so that swamping effects become relevant, the correlation coefficient can become as large as 0.25, depending on $L\bar{s}$.

The simple DFE model assumed above, where selection and dominance coefficients are independent, is almost certainly inadequate. Theoretical and empirical work has indicated that, on the one hand, a correlation between selective effect and degree of dominance can be expected in standing genetic variation, with segregating deleterious alleles more likely to be recessive if they have large s (Caballero and Keightley 1994; Zhang *et al.* 2004; Agrawal and Whitlock 2011). On the other hand, during adaptation, dominant beneficial alleles have higher establishment probabilities than (partially) recessive ones with the same homozygous effect (Haldane's sieve, Haldane 1927; Turner 1981). These two aspects interact when adaptation is from standing variation: while more recessive beneficial alleles are less likely to fix than dominant alleles, they are more likely to be segregating at appreciable frequencies when the population faces an adaptive challenge (Orr and Betancourt 2001, but see Muralidhar and Veller 2022).

To examine how the realized genetic architecture of local adaptation depends on the correlation between s and h among divergently selected loci, we consider two alternative, admittedly *ad hoc*, DFE models, outlined in Supplementary section 2.7. Both models assume Gamma distributed selection coefficients and incorporate a positive correlation between s and h, so that alleles of large effect tend to be more recessive (recall once more that h in our case is the dominance coefficient of the invading allele, so $h = 1$ corresponds to recessive local adaptation). We keep the average dominance coefficient fixed to 2/3 for each model. In contrast with the independent model, we find that for the model in which s and h are positively correlated, recessives ($h > 1/2$) are typically more likely to contribute to the realized differentiation at equilibrium (Fig. 5 and Supplementary Figs. 12–14 and 16–18). When we make the opposite assumption that locally beneficial alleles tend to be dominant ($h < 1/2$, corresponding, for instance, to the

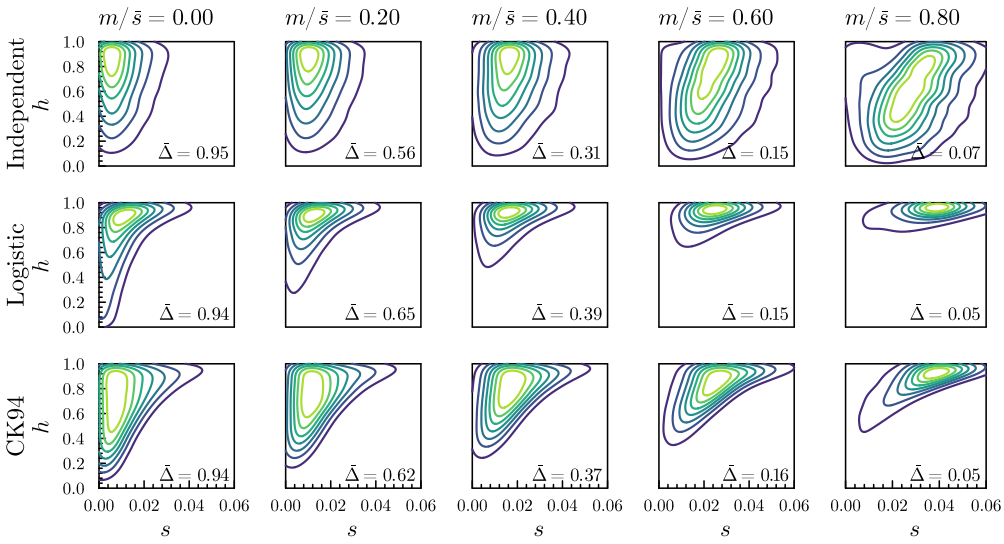

Fig. 5. The DFE affects the realized genetic architecture of local adaptation at migration–selection balance. Contour plots for the joint density of h and s conditional on observing a divergent allele on the island (see eq. (8)) are shown for the three DFE models (rows) for increasing rates of migration (columns). Values of $\bar{\Delta}$ in the lower right corner denote the mean expected differentiation per locus. We assume $L\bar{s} = 0.8$, $\bar{s} = 0.01$, $N_e\bar{s} = 20$, $u/\bar{s} = 0.005$ and exponentially distributed selection coefficients, and parameterize the DFE models so that $\mathbb{E}[h] = 2/3$, assuming $\alpha = 2$, $\beta = 1$ for the independent model, $a = 7.2$, $b = 1.2$, $\sigma = 1$ for the logistic model and $K = 50$ for the CK94 model (see Supplementary section 2.7 for details on the different DFE models considered here). The densities are approximated using a Monte Carlo approach, simulating 500 replicate L locus genetic architectures from the assumed DFE model, determining the equilibrium allele frequencies for each replicate, and fitting a kernel density estimate to the sample so obtained.

case with a strong Haldane's sieve effect during adaptation), we find a somewhat less dramatic shift in the joint density as migration rates go up (Supplementary Fig. 19). The distribution also shifts towards higher selection coefficients, but somewhat less so than in the model with the opposite correlation. Swamping of partially recessive alleles of small effect further shifts the distribution towards smaller values of h. These examples show how correlations between s and h among the loci under divergent selection can have a rather important influence on the realized genetic architecture at migration–selection balance (and hence on which loci actually contribute to adaptive differentiation).

The effects of linkage

So far, we have ignored physical linkage of the L loci under selection. The importance of linkage will depend on the total map length, the number of chromosomes, and the number and location of the selected loci. For organisms with a small number of chromosomes (e.g. *Drosophila*), linkage may have important consequences, whereas with larger chromosome numbers (for instance, in humans), most of the barrier effect may be due to unlinked loci.

The overall effect of linkage should be to increase the strength of a barrier to gene flow, as linkage increases the variance in introgressed ancestry among a migrant's descendants, yielding more efficient purging of sets of introgressing maladaptive alleles (Barton 1983; Veller *et al.* 2023). However, it is less clear what happens in finite populations and close to swamping thresholds, where linked combinations of adaptive and deleterious alleles may persist due to Hill–Robertson interference. We expect that when recombination is strong relative to selection, i.e. $r_{ij} \gg s_i$ for all j, the basic separation of timescales between the breakdown of LD and the establishment of migration–selection equilibrium at individual loci still applies, and eq. (5) should yield reasonably accurate predictions. However, associations between tightly linked loci, for which $r \lesssim s$, will be broken down by recombination at rates comparable to or slower than their elimination by selection, so that the strength of coupling is increased (Barton 1983; Kruuk *et al.* 1999), and the approach based on inserting an effective migration rate in single-locus theory becomes inappropriate.

These predictions are verified in our simulations (Supplementary Fig. 20). As the strength of recombination relative to selection (r/s) decreases, the barrier strength increases. We find that for $r/s \geq 4$, equilibrium frequencies and swamping thresholds appear to be accurately predicted using single-locus diffusion theory with the approximate gff given by eq. (5), although a slight but systematic overprediction of the equilibrium frequencies is apparent. As expected, when the strength of recombination decreases further ($r/s < 4$), the approximation breaks down.

We now consider what this means for realistic genetic maps. In Fig. 6 (see also Supplementary Fig. 21), we compare our numerical predictions against individual-based simulations for $L = 100$ equal-effect loci, uniformly distributed along the human and *Drosophila* genomes (i.e. each selected locus is at a randomly sampled position in the genome), assuming $Ls = 1$ and $N_e s = 10$. We obtain very good predictions for both equilibrium frequencies and swamping thresholds for the human genetic map, where tight linkage is rare (>99% of the r_{ij}/s values exceed 4). For the *Drosophila* genetic map (where 97% of the r_{ij}/s values exceed 4, and 92% exceed 10), we obtain fair predictions for equilibrium differentiation when the migration rate is not too high. However, we tend to predict swamping at lower m/s than observed in individual-based simulations, in line with our observations in Supplementary Fig. 20.

These results highlight that, for similar DFEs and population genetic parameters, divergent local adaptation is expected to lead to a much stronger barrier to gene flow (and hence stronger RI) in organisms with a *Drosophila*-like recombination landscape than in organisms that resemble humans in this regard (Fig. 6). Prediction accuracy is not noticeably different for heterogeneous barriers (Supplementary Fig. 22); nor does it depend in any obvious way on effective selection and dominance coefficients (Supplementary Fig. 23). This suggests that the effects of linkage

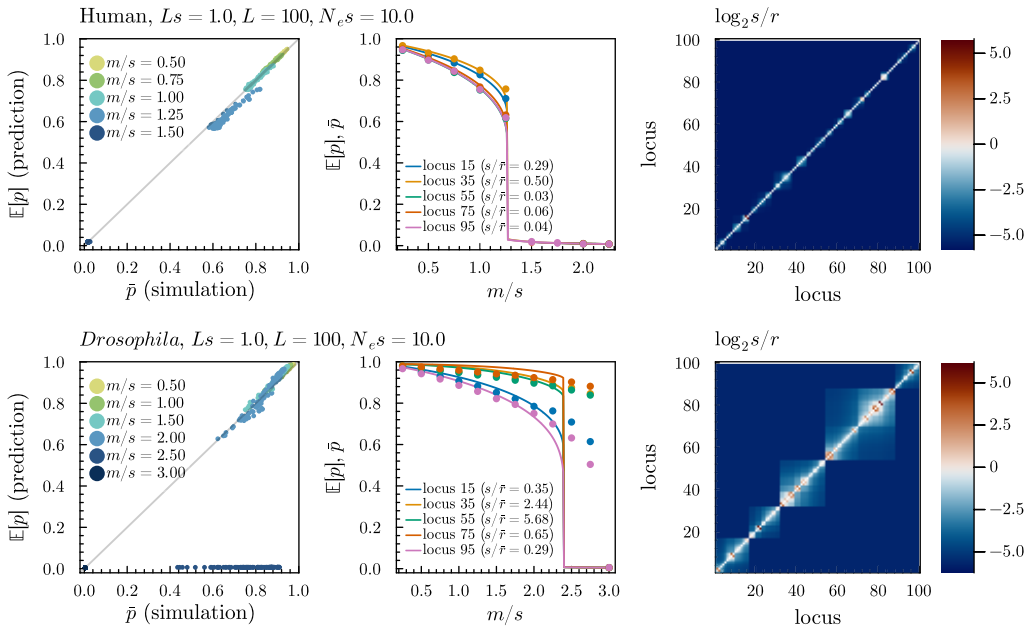

Fig. 6. Approximations based on effective migration rates accurately predict divergence along the genome when linkage among barrier loci is weak relative to the strength of selection per locus. Effects of linkage on the equilibrium frequencies of locally adaptive alleles for $L = 100$ loci randomly scattered on the Human (loose linkage on average) and *Drosophila* (more tight linkage on average) genomes are investigated. From left to right: scatter plot of the predicted equilibrium frequencies of the locally beneficial allele ($\mathbb{E}[p]$) at the 100 loci versus those observed in individual-based simulations ($\bar{p}$); observed (dots) and predicted (lines) allele frequencies for five loci for increasing migration rate; strength of selection relative to recombination across the genome ($\log_2 s/r$), i.e. for each pair of loci i and j, $\log_2 s/r_{ij}$ is shown. We assume haploid selection with equal selection coefficients across all loci, assuming $Ls = 1$, $N_e s = 10$, $u = s/100$.

are largely orthogonal to the effects of dominance and barrier heterogeneity studied above, at least in the absence of epistasis.

Discussion

LD and polygenic migration–selection balance

We study genetic barriers to gene flow comprising an arbitrary number of unlinked or weakly linked divergently selected loci with heterogeneous selective and dominance effects, assuming a mainland-island model and a general haplodiplontic life cycle. We derive expressions for the effective migration rate at any given locus in terms of the allele frequency divergence at all other loci, and then use this together with classical single-locus theory to self-consistently predict divergence at all selected loci. Importantly, this allows us to study the effects of coupling among barrier loci without assuming that locally deleterious alleles are somehow rare, enabling us to study swamping by gene flow in the polygenic setting.

Our results show how the maintenance of adaptive differentiation in the face of gene flow depends jointly on the extent of LD, drift, dominance and variation in selective effects across loci. The general success of the approach indicates two important features of polygenic migration–selection balance. Firstly, it suggests that the "separation of time scales" argument that is at the root of the approach indeed works, and generalizes well beyond the case of haploid equal-effect loci to more realistic architectures with dominance, heterogeneous fitness effects and weak linkage. Unless loci are tightly linked (i.e. $r/s < 4$), strong selection against multilocus genotypes occurs only in the first few generations after a migrant arrives, and the long-term fate of a migrant allele is unaffected by LD conditional on having survived these initial generations. As a consequence, the effects of LD are well described by the usual single-locus dynamics, but with a reduced migration rate. Secondly, it indicates that our rather crude approximation to the expected reproductive value of a migrant individual on the island (which assumes that migrants only cross with residents, that in each such cross the proportion of migrant alleles is exactly halved, and that each backcross generation is in HWLE), is an adequate estimator of the gff.

The effect of dominance on a polygenic barrier to gene flow

Our analyses for homogeneous genetic architectures indicate that, when there is selection in the diploid phase, dominance can have a considerable impact on the expected level of adaptive differentiation and swamping thresholds. In the multilocus setting, depending on the total strength of divergent selection (Ls), partially recessive variants may lead to strongly increased swamping thresholds and produce a much stronger barrier to gene flow than dominant variants with the same homozygous effect (Fig. 2), in sharp contrast with the single-locus case (Haldane 1930). The reason for this is that for recessive local adaptation, the dominant invading alleles are immediately exposed to selection, whereas for dominant local adaptation, the recessive invading alleles can introgress easily as long as the frequency of locally deleterious alleles is low. We find that dominance has a strong effect on the feedback between the level of differentiation and the strength of selection against migrants (which leads to sharp swamping thresholds).

It should be emphasized, however, that all our results assume a mainland-island model of migration and a scenario of secondary contact. The effects of dominance may be more subtle in models with multiple demes and more symmetric patterns of migration, in which case assumptions on environmental dependence of dominance may become important (e.g. Bürger 2013).

Heterogeneous architectures of polygenic barriers

When migration is weak relative to the average strength of selection per locus, increased variation of s in the DFE underlying

locally adaptive traits gives rise to lower overall adaptive differentiation at equilibrium, while generating a barrier to gene flow of similar strength (Fig. 3). On the other hand, for high rates of migration, more heterogeneous architectures generate a stronger barrier to gene flow due to the presence of a larger number of strongly selected loci that resist swamping. As a consequence, we can expect a more heterogeneous architecture of divergent selection to be more favorable for the evolution of RI between lineages coming into secondary contact.

The presence of a few loci of relatively large effect can cause a substantial reduction in gene flow even at unlinked loci, raising neutral differentiation across the genome. Hence, from an empirical perspective, it may become harder to map loci underlying adaptation and RI when the genetic architecture becomes more heterogeneous. At the same time, our results also emphasize that, despite appreciable genome-wide coupling effects, considerable variation in equilibrium differentiation across nonswamped loci remains, depending on the locus-specific fitness effects (Fig. 4).

The genetic architecture of local adaptation at equilibrium

When there is appreciable gene flow, only a subset of the divergently selected loci that underlie a locally adaptive trait will actually exhibit substantial differentiation at migration–selection balance, and the DFE of these loci need not be representative for the DFE associated with all loci underlying the trait (that is, the genetic architecture of local adaptation may differ to greater or lesser extent from the genetic architecture of locally adaptive traits Yeaman and Whitlock 2011).

We find that when selection is fairly weak ($L\bar{s}$ is small), the subset of divergently selected loci that exhibit significant differentiation at migration–selection balance constitutes a more biased subset than when $L\bar{s}$ is large (Fig. 4). This has implications for our ability to map locally adaptive loci: when $L\bar{s}$ is large, RI is effectively complete, and genome-wide differentiation will be high, so that mapping locally adaptive loci becomes essentially impossible. When $L\bar{s}$ is small, adaptive loci will be more easily detected, but they will constitute a more biased subset of the loci that contribute to variation in locally adaptive traits. Similarly, the DFE at equilibrium shifts more and more to larger selection coefficients when migration increases (Fig. 4), whereas the effect on the distribution of dominance coefficients depends on the correlation between s and h in the DFE underlying locally adaptative traits (Fig. 4 and Supplementary Figs. 12–14).

Importantly, the maintenance of polygenic migration–selection equilibrium itself *generates* correlations between selection and dominance coefficients. Such correlations are often discussed in the context of adaptation, with different factors influencing the relationship between the homozygous effect and dominance deviation in the DFE of new mutations (Agrawal and Whitlock 2011; Manna *et al.* 2011), the DFE of the standing genetic variation (Orr and Betancourt 2001; Zhang *et al.* 2004), and the DFE of variants fixed during adaptation (Orr 2010). We show that migration–selection balance may be another source of s-h correlation: especially when RI is low and gene flow rather strong, recessive alleles that are divergently maintained at equilibrium tend to have higher selection coefficients than dominant alleles, even when no such correlation exists *a priori* (Fig. 5).

Linkage and heterogeneous gene flow across the genome

We developed a heuristic approximation for the gff when loci are linked across the genome using two-locus theory (Appendix B), and we showed that it accurately predicts multilocus equilibrium differentiation and swamping thresholds when linkage is not too tight (roughly $r/s \geq 4$). Comparison with individual-based simulations for the human and *Drosophila* genetic maps, shows that our approximations are reasonably accurate for realistic genetic maps, at least when divergently selected loci are not tightly clustered in the genome. For the *Drosophila* genetic map however, where tight linkage is common, we predict swamping at lower migration rates than observed in individual-based simulations, whereas for the human genetic map we can accurately predict swamping thresholds (Fig. 6).

Having an approximation which captures both the effects of physical linkage and large-scale genome-wide LD on patterns of genetic differentiation, we can use our theoretical work to illuminate the conditions under which one would expect to see so-called "genomic islands" of divergence rather than distinct isolated divergent sites or genome-wide elevated differentiation (Feder *et al.* 2012; Via 2012; Yeaman 2013; Shi *et al.* 2023). Indeed, for a given migration rate, the genetic architecture of divergent selection (the number of selected loci, the DFE, and their distribution across the genome) will jointly determine how m_e varies across the genome. The latter will in turn determine the expected pattern of neutral differentiation across the genome, as reflected for instance in an F_{ST} "genome scan." Our approach allows the detailed prediction of such patterns of heterogeneous differentiation for complicated polygenic architectures, accounting for LD among selected loci.

As an illustration, we show in Fig. 7 a predicted F_{ST} landscape for a stretch of genome under divergent selection, highlighting the importance of taking into account LD among selected loci when predicting neutral differentiation. When LD is accounted for, allele frequency divergence at selected sites is higher than when LE is assumed, which feeds back into the m_e approximation as it decreases the reproductive value of migrants. In the example in Fig. 7, we clearly observe three effects of this. First of all, a clear genome-wide increase in F_{ST} is observed. Secondly, some isolated F_{ST} peaks in the yellow curve are absent from the green curve, corresponding to loci that would be subject to swamping in the

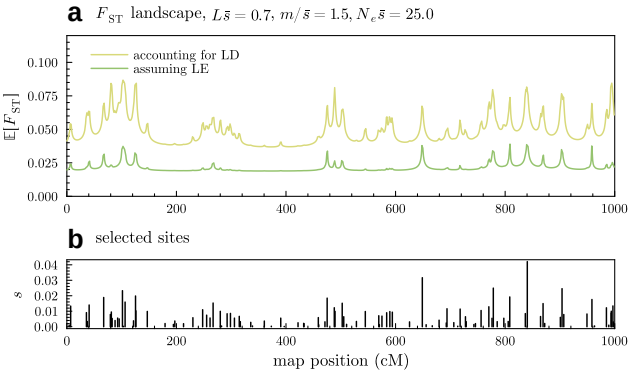

Fig. 7. The effects of LD among barrier loci on heterogeneous gene flow across the genome. a) Predicted neutral $F_{ST} = (1 + 2N_e m_e)^{-1}$ across a genomic segment with $L = 100$ selected sites uniformly scattered across the segment. For the yellow (top) line, we use eq. (5) to calculate local m_e across the genome with expected allele frequencies for the selected sites calculated accounting for LD among selected loci (i.e. using eq. (7)). For the green (bottom) line, we ignore LD, and use eq. (5) with expected allele frequencies for the selected sites calculated using single-locus theory, effectively assuming the selected loci are in LE and attain migration–selection balance independently of each other. b) Positions and fitness effects of selected sites in (a). We assume no dominance and $u = \bar{s}/100$.

absence of the protective effects of LD. Thirdly, in regions where multiple rather weakly selected alleles occur close to each other (e.g. near 280 cM, or 600 cM), LD leads to the emergence of a region with elevated differentiation relative to the genome-wide F_{ST} level.

Inference of barriers to gene flow and the genetic architecture of RI

Our results bear relevance to model-based inference of barriers to gene flow from genomic data. Recent approaches to quantify gene flow in pairs of diverging populations based on observed genetic variation have accounted for heterogeneity in barrier effects by assuming a demographic model in which the migration rate m (interpreted as m_e) varies along the genome (Fraïsse *et al.* 2021; Laetsch *et al.* 2023). However, none of these approaches explicitly models the underlying genetic architecture of divergent selection that is assumed to cause variation in m_e across the genome.

So far, the only approach which has attempted this is the one by Aeschbacher *et al.* (2017), which combines information about local recombination rates with deterministic population genetic theory to obtain predictions of m_e across the genome. They assume a homogeneous genetic architecture with no dominance, where selected loci occur with a constant density across the genome, and infer the selection density per unit of map length (together with the migration rate) through its effect on observable neutral differentiation. Importantly, they assumed $m \ll s$ and no genetic drift at selected loci, so that there is essentially complete divergence at each selected locus. We wonder, naturally, whether our approximations could be fruitfully employed in a similar inferential approach, allowing for drift and heterogeneous barrier architectures, while avoiding the assumption that introgressing alleles are rare and accounting for the effect of LD on allele frequency divergence at selected loci. Such an approach that explicitly connects variation in m_e across the genome to the numbers and fitness effects of selected loci could allow for more detailed inferences about the genetic architecture of local adaptation and RI, and provide insights about the limits to such inference.

Limitations of the model

Several important limitations of the model, besides the issues arising from tight linkage, should be noted. Firstly, we have assumed that local fitness is determined by an additive trait under directional selection, considering a history where a rapid polygenic selection response has driven up allele frequencies to near fixation. Alternatively, populations may be under stabilizing selection towards different optima. With stabilizing selection and abundant standing variation, a polygenic selection response may initially only involve subtle changes in allele frequencies (Sella and Barton 2019; Hayward and Sella 2022), and there may be considerable genetic redundancy (Yeaman 2015; Barghi *et al.* 2020), leading to a scenario that is quite different from the one assumed in this paper. The extent of RI that can be maintained when these aspects of polygenic adaptation become important remains unclear and may require different approaches.

Secondly, our focus on the maintenance of polygenic local adaptation and the RI it causes provides only part of the picture, as we have ignored both the initial polygenic response, and the later build-up of divergence in the face of gene flow. We have considered how a *given* genetic architecture underlying divergent selection results in observable patterns of adaptive differentiation at equilibrium, but remain largely ignorant about what a plausible genetic architecture of divergent selection looks like, or how it evolves during divergence with gene flow (Yeaman and Whitlock 2011; Yeaman 2013). However, if we consider a history of rapid polygenic adaptation from standing genetic variation, followed by secondary contact between divergently adapted populations, our theory should readily apply. Such a divergence history will result in large-scale genome-wide patterns of LD among effectively unlinked regions, and this is often observed empirically (e.g. Schumer *et al.* 2014). We have also completely ignored epistasis, which may be an important additional source of multilocus selection against hybrids and their descendants (e.g. Dobzhansky–Muller incompatibilities, see for instance Schumer *et al.* 2018). All these are important topics deserving further study if we are to understand how populations can remain locally adapted when subjected to maladaptive gene flow, and, ultimately, the adaptive processes that could drive the origin of new species.

Data availability

The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and supplementary material (available at 10.6084/m9.figshare.26771119). Software implementing the numerical methods and individual-based simulations is available at https://github.com/arzwa/Sewall. Code for generating the figures in the main text is available in the same repository.

Supplemental material available at GENETICS online.

Acknowledgments

We thank Nick Barton, Nicolas Bierne, and six anonymous reviewers for helpful comments on an earlier version of the manuscript that substantially improved the presentation of our work.

Funding

This work was funded by the European Union (ERC BryoFit 101041201 granted to CF). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Conflicts of interest

The author(s) declare no conflicts of interest.

Author contributions

AZ, HS, and CF conceived the project; AZ conducted mathematical analyses and simulations; AZ wrote the manuscript with input and feedback from HS and CF.

Literature cited

Aeschbacher S, Selby JP, Willis JH, Coop G. 2017. Population-genomic inference of the strength and timing of selection against gene flow. Proc Natl Acad Sci USA. 114(27):7061–7066. doi:10.1073/pnas.1616755114.

Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics. 187(2):553–566. doi:10.1534/genetics.110.124560.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: A unifying framework to understand positive selection. Nat Rev Genet. 21(12):769–781. doi:10.1038/s41576-020-0250-z.

Barton NH. 1983. Multilocus clines. Evolution. 37(3):454–471. doi:10.2307/2408260.

Barton NH. 2020. On the completion of speciation. Philos Trans R Soc B. 375(1806):20190530. doi:10.1098/rstb.2019.0530.

Barton NH, Bengtsson BO. 1986. The barrier to genetic exchange between hybridising populations. Heredity. 57(3):357–376. doi:10.1038/hdy.1986.135.

Barton NH, De Cara MAR. 2009. The evolution of strong reproductive isolation. Evolution. 63(5):1171–1190. doi:10.1111/evo.2009.63.issue-5.

Barton NH, Etheridge A. 2018. Establishment in a new habitat by polygenic adaptation. Theor Popul Biol. 122:110–127. doi:10.1016/j.tpb.2017.11.007.

Bengtsson B. 1985. The flow of genes through a genetic barrier. In: Greenwood P J, Harvey P H , Slatkin M, editors. Evolution. Essays in honour of John Maynard Smith. Vol. 1. Cambridge, UK: Cambridge University Press. p. 357–376.

Bezanson J, Edelman A, Karpinski S, Shah VB. 2017. Julia: A fresh approach to numerical computing. SIAM Rev. 59(1):65–98. doi:10.1137/141000671.

Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: Why genome scans may fail to map local adaptation genes. Mol Ecol. 20(10):2044–2072. doi:10.1111/mec.2011.20.issue-10.

Bomblies K, Peichel CL. 2022. Genetics of adaptation. Proc Natl Acad Sci USA. 119(30):e2122152119. doi:10.1073/pnas.2122152119.

Bürger R. 2013. A survey on migration-selection models in population genetics. arXiv:1309.2576, https://doi.org/10.48550/arXiv.1309.2576, preprint: not peer reviewed.

Butlin RK, Smadja CM. 2018. Coupling, reinforcement, and speciation. Am Nat. 191(2):155–172. doi:10.1086/695136.

Caballero A, Keightley PD. 1994. A pleiotropic nonadditive model of variation in quantitative traits. Genetics. 138(3):883–900. doi:10.1093/genetics/138.3.883.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28(7):342–350. doi:10.1016/j.tig.2012.03.009.

Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, Romiguier J, Loire E, Simon A, Galtier N, Duret L, *et al.* 2021. Dils: Demographic inferences with linked selection by using abc. Mol Ecol Resour. 21(8):2629–2644. doi:10.1111/1755-0998.13323.

Haldane JBS. 1927. A mathematical theory of natural and artificial selection, part v: Selection and mutation. Math Proc Camb Philos Soc. 23(7):838–844. doi:10.1017/S0305004100015644.

Haldane JBS. 1930. A mathematical theory of natural and artificial selection, part vi, isolation. Math Proc Camb Philos Soc. 26(2):220–230. doi:10.1017/S0305004100015450.

Harris K, Nielsen R. 2016. The genetic cost of neanderthal introgression. Genetics. 203(2):881–891. doi:10.1534/genetics.116.186890.

Hayward LK, Sella G. 2022. Polygenic adaptation after a sudden change in environment. Elife. 11:e66697. doi:10.7554/eLife.66697.

Hein J, Schierup M, Wiuf C. 2004. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. USA: Oxford University Press.

Jiggins C, Martin S. 2017. Glittering gold and the quest for Isla de Muerta. J Evol Biol. 30(8):1509–1511. doi:10.1111/jeb.2017.30.issue-8.

Kobayashi Y, Hammerstein P, Telschow A. 2008. The neutral effective migration rate in a mainland-island context. Theor Popul Biol. 74(1):84–92. doi:10.1016/j.tpb.2008.05.001.

Kruuk L, Baird S, Gale K, Barton NH. 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics. 153(4):1959–1971. doi:10.1093/genetics/153.4.1959.

Kulmuni J, Butlin RK, Lucek K, Savolainen V, Westram AM. 2020. Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers.

Laetsch DR, Bisschop G, Martin SH, Aeschbacher S, Setter D, Lohse K. 2023. Demographically explicit scans for barriers to gene flow using gimble. PLoS Genet. 19(10):e1010999. doi:10.1371/journal.pgen.1010999.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 21(7):1548–1566. doi:10.1111/mec.2012.21.issue-7.

Lenormand T. 2002. Gene flow and the limits to natural selection. Trends Ecol Evol. 17(4):183–189. doi:10.1016/S0169-5347(02)02497-7.

Manna F, Martin G, Lenormand T. 2011. Fitness landscapes: An alternative theory for the dominance of mutation. Genetics. 189(3):923–937. doi:10.1534/genetics.111.132944.

Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. 17(2):e2006288. doi:10.1371/journal.pbio.2006288.

Muralidhar P, Veller C. 2022. Dominance shifts increase the likelihood of soft selective sweeps. Evolution. 76(5):966–984. doi:10.1111/evo.v76.5.

Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci USA. 100(9):5268–5273. doi:10.1073/pnas.0431157100.

Nagylaki T. 1975. Conditions for the existence of clines. Genetics. 80(3):595–615. doi:10.1093/genetics/80.3.595.

Nosil P. 2012. Ecological Speciation. Oxford, UK: Oxford University Press.

Orr HA. 2010. The population genetics of beneficial mutations. Philos Trans R Soc B Biol Sci. 365(1544):1195–1201. doi:10.1098/rstb.2009.0282.

Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics. 157(2):875–884. doi:10.1093/genetics/157.2.875.

Petry D. 1983. The effect on neutral gene flow of selection at a linked locus. Theor Popul Biol. 23(3):300–313. doi:10.1016/0040-5809(83)90020-5.

Pritchard JK, Di Rienzo A. 2010. Adaptation–not by sweeps alone. Nat Rev Genet. 11(10):665–667. doi:10.1038/nrg2880.

Ravinet M, Faria R, Butlin R, Galindo J, Bierne N, Rafajlović M, Noor M, Mehlig B, Westram A. 2017. Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. J Evol Biol. 30(8):1450–1477. doi:10.1111/jeb.2017.30.issue-8.

Sachdeva H. 2022. Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. PLoS Genet. 18(9):e1010297. doi:10.1371/journal.pgen.1010297.

Schumer M, Cui R, Powell DL, Dresner R, Rosenthal GG, Andolfatto P. 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. Elife. 3:e02535. doi:10.7554/eLife.02535.

Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, *et al.* 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 360(6389):656–660. doi:10.1126/science.aar3684.

Sella G, Barton NH. 2019. Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu Rev Genomics Hum Genet. 20:461–493. doi:10.1146/genom.2019.20.issue-1.

Shi Y, Bouska KL, McKinney GJ, Dokai W, Bartels A, McPhee MV, Larson WA. 2023. Gene flow influences the genomic architecture of local adaptation in six riverine fish species. Mol Ecol. 32(7): 1549–1566. doi:10.1111/mec.v32.7.

Stankowski S, Chase MA, McIntosh H, Streisfeld MA. 2023. Integrating top-down and bottom-up approaches to understand the genetic architecture of speciation across a monkeyflower hybrid zone. Mol Ecol. 32(8):2041–2054. doi:10.1111/mec.v32.8.

Turner JR. 1981. Adaptation and evolution in heliconius: A defense of neodarwinism. Annu Rev Ecol Syst. 12(1):99–121. doi:10.1146/ecolsys.1981.12.issue-1.

van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 534(7605):102–105. doi:10.1038/nature17951.

Veller C, Edelman NB, Muralidhar P, Nowak MA. 2023. Recombination and selection against introgressed DNA. Evolution. 77(4):1131–1144. doi:10.1093/evolut/qpad021.

Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos Trans R Soc B Biol Sci. 367(1587):451–460. doi:10.1098/rstb.2011.0260.

Westram AM, Rafajlović M, Chaube P, Faria R, Larsson T, Panova M, Ravinet M, Blomberg A, Mehlig B, Johannesson K, *et al.* 2018. Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. Evol Lett. 2(4):297–309. doi:10.1002/evl3.74.

Westram AM, Stankowski S, Surendranadh P, Barton N. 2022. What is reproductive isolation? J Evol Biol. 35(9):1143–1164. doi:10.1111/jeb.14005.

Wright S. 1937. The distribution of gene frequencies in populations. Proc Natl Acad Sci USA. 23(6):307–320. doi:10.1073/pnas.23.6.307.

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci USA. 110(19): E1743–E1751. doi:10.1073/pnas.1219381110.

Yeaman S. 2015. Local adaptation by alleles of small effect. Am Nat. 186(S1):S74–S89. doi:10.1086/682405.

Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration–selection balance. Evolution. 65(7): 1897–1911. doi:10.1111/evo.2011.65.issue-7.

Zhang X-S, Wang J, Hill WG. 2004. Influence of dominance, leptokurtosis and pleiotropy of deleterious mutations on quantitative genetic variation at mutation-selection balance. Genetics. 166(1):597–610. doi:10.1534/genetics.166.1.597.

Appendix A: Detailed derivation of eq. (4)

Under the assumptions stated above eq. (4), each of the $W^{(n)}$ is determined by the allele frequencies and heterozygosities at the selected loci in the mainland and the island populations at the assumed equilibrium. This allows us to determine $\mathbb{E}[q_i^{(n)}]$, the expected frequency of the locally deleterious allele (in the island) at locus i among nth generation descendants from a migrant, in terms of the allele frequencies in the mainland and island populations. Indeed, we have the recursive relation $q_i^{(n)} = \frac{1}{2}(q_i^{(n-1)} + q_i)$, i.e. the expected frequency of the deleterious allele at locus i in an nth generation descendant is the mean of the corresponding frequencies in the resident population and the $(n-1)$ generation backcrosses. Hence, we have $\mathbb{E}[q_i^{(n)}] = \frac{1}{2^n}(q_i^* + (2^n - 1)\mathbb{E}[q_i])$. Denoting the selection coefficient at locus i for the haploid phase by s_{i1}, the expected relative fitness of an nth generation haploid descendant is

$$\mathbb{E}\left[W_h^{(n)}\right] \approx \frac{\exp\left[-\sum_{i=1}^{L} s_{i1}\mathbb{E}[q_i^{(n)}]\right]}{\exp\left[-\sum_{i=1}^{L} s_{i1}\mathbb{E}[q_i]\right]} = \exp\left[-2^{-n}\sum_{i=1}^{L} s_{i1}(q_i^* - \mathbb{E}[q_i])\right],$$

where we have assumed that per-locus selection is sufficiently weak that $O(s^2)$ terms can be ignored. For the diploid phase, a similar argument shows that for the $(n+1)$th generation,

$$\mathbb{E}\left[W_d^{(n+1)}\right] = e^{-2^{-n}\sum_{i=1}^{L} s_{i01}(q_i^* - \mathbb{E}[q_i]) - (s_{i11} - 2s_{i01})\left(p_i^*\mathbb{E}[q_i] - \mathbb{E}[p_i q_i]\right)},$$

Putting everything together, the approximate gff becomes

$$\begin{aligned} g &\approx \mathbb{E}\left[W_h^{(0)}\right]\prod_{n=1}^{\infty}\left(\mathbb{E}\left[W_d^{(n)}\right]\mathbb{E}\left[W_h^{(n)}\right]\right) \\ &= \prod_{k=0}^{\infty} e^{-2^{-k}\sum_{i=1}^{L} s_i h_i (q_i^* - \mathbb{E}[q_i]) - s_i(1-2h_i)(p_i^*\mathbb{E}[q_i] - \mathbb{E}[p_i q_i])} \\ &= e^{-2\sum_{i=1}^{L} s_i h_i (q_i^* - \mathbb{E}[q_i]) - s_i(1-2h_i)\left(p_i^*\mathbb{E}[q_i] - \mathbb{E}[p_i q_i]\right)}, \end{aligned} \tag{A1}$$

where $s_i h_i = s_{i1} + s_{i01}$ and $s_i(1-2h_i) = s_{i11} - 2s_{i01}$ (see Table 1).

Two remarks are due. Firstly, the gff as derived above yields the effective migration rate at an unlinked neutral locus. We can calculate the gff at a selected locus by making the assumption that it is the same as that of a hypothetical neutral locus at the same location—an assumption which is only expected to work well if selection at the focal locus is sufficiently weak. Hence, if we wish to calculate the effective migration rate for a selected locus in the barrier, say locus j, the relevant gff is obtained by excluding index j from the sum in eq. (A1).

Secondly, we have assumed that migration occurs at the start of the haploid phase, e.g. due to spore dispersal. While the details of when migration occurs in the life cycle do not matter for the single-locus model as long as selection and migration are sufficiently weak (so that the continuous-time limit is appropriate), these details *do* matter for the effective migration rate. This is because, although selection *per locus* is weak (s being small), selection against migrant genotypes can be strong (Ls being appreciable). Thus, when migration is due to dispersal of gametes (e.g. pollen dispersal), the first generation experiencing selection on the island will be the diploid F1 generation, so that the appropriate gff under the same approximation is $g/\mathbb{E}[W_h^{(0)}]$. Secondly, when migration occurs at the beginning of the diploid phase (e.g. seed dispersal), the first generation experiencing selection will consist of diploid migrant individuals, so that $g\mathbb{E}[W_d^{(0)}]$ is the appropriate gff, where

$$\begin{aligned} \mathbb{E}[W_d^{(0)}] &\approx \frac{e^{-\sum_i^L 2p_i^* q_i^* s_{i,01} + q_i^{*2} s_{i,11}}}{e^{-\sum_i^L 2\mathbb{E}[p_i q_i] s_{i,01} + \mathbb{E}[q_i^2] s_{i,11}}} \\ &= \exp\left[-\sum_{i=1}^{L} s_{i11}(q_i^* - \mathbb{E}[q_i]) - s_i(1-2h_i)(p_i^* q_i^* - \mathbb{E}[p_i q_i])\right]. \end{aligned}$$

If the haploid, diploid, and gametic migration rates are m_1, m_2, and m_3, respectively, the effective migration rate will be $(m_1 + \mathbb{E}[W_d^{(0)}]m_2 + \mathbb{E}[W_h^{(0)}]^{-1}m_3)g$.

Appendix B: Accounting for (weak) linkage

To heuristically account for weak linkage between L loci, we shall extrapolate from a two-locus model at quasi-linkage equilibrium (QLE). We shall hence first consider the deterministic dynamics of

a two-locus haplodiplontic model with mainland-island migration in continuous-time.

We assume a locus A with alleles A_0 and A_1 and a linked locus B with alleles B_0 and B_1, with recombination between the two loci occurring at rate r. We assume arbitrary dominance and no epistasis, with relative Malthusian fitnesses of all possible two-locus genotypes in the two phases given by the following tables

Haploid phase

	B_0	B_1
A_0	0	β_1
A_1	α_1	$\alpha_1+\beta_1$

(B1)

Diploid phase

	B_0B_0	B_0B_1	B_1B_1
A_0A_0	0	β_{01}	β_{11}
A_0A_1	α_{01}	$\alpha_{01}+\beta_{01}$	$\alpha_{01}+\beta_{11}$
A_0A_1	α_{11}	$\alpha_{11}+\beta_{01}$	$\alpha_{11}+\beta_{11}$

(B2)

Let $\alpha_a=\alpha_1+\alpha_{01}$ and $\alpha_b=\alpha_{11}-2\alpha_{01}$ (and similarly for β_a and β_b). Let x_{ij} and y_{ij} denote the frequency of the A_iB_j haplotype on the island and mainland respectively. The two-locus dynamics in continuous-time are given by

$$\dot{x}_{ij}=m(y_{ij}-x_{ij})+(\omega_{ij}-\bar{\omega})x_{ij}-\eta_{ij}rD \tag{B3}$$

where $D=x_{00}x_{11}-x_{01}x_{10}$ is the usual measure of two-locus LD, ω_{ij} the marginal fitness of the A_iB_j haplotype, $\bar{\omega}$ the mean Malthusian fitness and $\eta_{ij}=1$ when $i=j$ and -1 otherwise. We write $p_A=1-q_A=x_{00}+x_{01}$ for the frequency of the A_0 allele, and $p_B=1-q_B=x_{00}+x_{10}$ for the frequency of the B_0 allele. Defining

$$\begin{aligned} Q_A&=\alpha_a+\alpha_b q_A \quad Q_B=\beta_a+\beta_b q_B\\ P_A&=\alpha_a+\alpha_b p_A \quad P_B=\beta_a+\beta_b p_B, \end{aligned}$$

one can find that $\bar{\omega}=q_AQ_A+q_BQ_B$ and write the marginal fitnesses as

$$\begin{aligned} \omega_{00}-\bar{\omega}&=-q_AQ_A-q_BQ_B\\ \omega_{01}-\bar{\omega}&=\beta_a-q_AQ_A-q_BP_B\\ \omega_{10}-\bar{\omega}&=\alpha_a-q_AP_A-q_BQ_B\\ \omega_{11}-\bar{\omega}&=\alpha_a+\beta_a-q_AP_A-q_BP_B. \end{aligned} \tag{B4}$$

Using eq. (B4) with eq. (B3), and assuming the mainland is at HWLE with allele frequencies p_A^* and p_B^* at the two loci, one can derive the dynamics for p_A, p_B, and D:

$$\begin{aligned} \dot{p}_A&=m(p_A^*-p_A)-p_Aq_AQ_A-Q_BD\\ \dot{p}_B&=m(p_B^*-p_B)-p_Bq_BQ_B-Q_AD\\ \dot{D}&=m\big((p_A^*-p_A)(p_B^*-p_B)-D\big)\\ &\quad+(Q_A(p_A-q_A)+Q_B(p_B-q_B))D-rD. \end{aligned} \tag{B5}$$

We can use eq. (B5) to derive the effective migration rate at a neutral locus linked to a barrier locus maintained at migration-selection balance. Let A be the selected locus, and B the linked neutral locus. The dynamics of the system are given by eq. (B5), where $Q_B=0$. Assuming r is sufficiently large relative to the strength of selection (linkage is sufficiently weak) so that D equilibrates much faster than the allele frequencies (QLE assumption), we can solve the system for D at equilibrium to find

$$\tilde{D}=\frac{m(p_A^*-p_A)(p_B^*-p_B)}{m+r-Q_A(p_A-q_A)} \tag{B6}$$

We can plug this into the ODE for the neutral locus to find

$$\dot{p}_B=m\left(1-\frac{(p_A^*-p_A)Q_A}{m+r-Q_A(p_A-q_A)}\right)(p_B^*-p_B) \tag{B7}$$

Suggesting that the effective migration rate under the stated assumptions should be

$$m_e=m\left(1-\frac{(p_A^*-p_A)Q_A}{m+r-Q_A(p_A-q_A)}\right) \tag{B8}$$

$$\approx me^{-\frac{(p_A^*-p_A)Q_A}{m+r-Q_A(p_A-q_A)}}\approx me^{-\frac{(p_A^*-p_A)Q_A}{r}} \tag{B9}$$

where the second approximation holds well for weak linkage ($r\gg s\sim m$), which we assumed when we derived eq. (B7). If we now consider a multilocus system with L loci, and assume that the effects of the $L-1$ other loci on gene flow at a focal locus act multiplicatively, we can approximate the gff at locus i as in eq. (5).

Editor: D. Roze