Effect of assortative mating and sexual selection on polygenic barriers to gene flow

Parvathy Surendranadh[1] and Himani Sachdeva[2]

[1]Insititute of Science and Technology Austria (ISTA)
[2]Department of Mathematics, University of Vienna, Austria

Abstract

Assortative mating and sexual selection are widespread in nature and can play an important role in speciation, through the buildup and maintenance of reproductive isolation (RI). However, their contribution to genome-wide suppression of gene flow during RI is rarely quantified. Here, we consider a polygenic 'magic' trait that is divergently selected across two populations connected by migration, while also serving as the basis of assortative mating, thus generating sexual selection on one or both sexes. We obtain theoretical predictions for divergence at individual trait loci by assuming that the effect of all other loci on any locus can be encapsulated via an effective migration rate, which bears a simple relationship to measurable fitness components of migrants and various early generation hybrids. Our analysis clarifies how 'tipping points' (characterised by an abrupt collapse of adaptive divergence) arise, and when assortative mating can shift the critical level of migration beyond which divergence collapses. We quantify the relative contributions of viability and sexual selection to genome-wide barriers to gene flow and discuss how these depend on existing divergence levels. Our results suggest that effective migration rates provide a useful way of understanding genomic divergence, even in scenarios involving multiple, interacting mechanisms of RI.

Keywords: assortative mating, sexual selection, effective migration rates, polygenic selection, reproductive isolation

1 Introduction

Reproductive isolation (RI) typically results from the interaction of multiple processes that cause loss of hybrid fitness, thus reducing gene flow between populations. Ecological specialisation arising from adaptation to different environmental niches, immigrant inviability, or intrinsic genetic incompatibilities may all generate postzygotic barriers to gene flow and maintain genetic differences between populations (Rundle and Nosil (2005); Coughlan and Matute (2020); Rice and Hostert (1993)). Additionally, processes such as assortative mating and sexual selection can act as both prezygotic and postzygotic barriers— by reducing heterospecific matings (that produce hybrids) as well as the mating success of hybrids. Indeed, it has been argued that the buildup of RI typically involves coupling between both postzygotic and prezygotic barriers (Butlin and Smadja (2018)).

Assortative mating (AM)– which involves positive phenotypic (and genetic) correlation between mating pairs (Lewontin et al. (1968)), is especially common in animals Jiang et al. (2013). Numerous studies have documented its role during RI, e.g., in cichlids (Elmer et al. (2009); Stelkens and Seehausen (2009)), swordtail fish (Schumer et al. (2017)), sticklebacks (Vines and Schluter (2006)), butterflies (Jiggins et al. (2001)), and marine snails (Johannesson et al. (2008)). AM decreases heterozygosity and genic variance within a population, but simultaneously increases total genetic variance (by generating higher levels of linkage disequilibria (LD)) (Lynch et al. (1998)), thus increasing the potential for divergence and speciation.

AM may occur either due to temporal or spatial separation of population groups, via preference-trait mechanisms, or via phenotype matching between potential mates (Kopp et al. (2018); Jiang et al. (2013); Kirkpatrick and Ravigné (2002)). Each mechanism differs in the extent of sexual selection it generates— ranging from strong sexual selection under phenotype matching to almost none when assortment is due to temporal or spatial separation. Sexual selection typically reduces mating success of rare phenotypes, and thus may counter the effects of AM (Servedio and Boughman (2017); Safran et al. (2013)). For example, when AM involves sexual selection on both sexes (e.g. if both males and females suffer reduced mating opportunities by being choosy), then its net effect is to generate stabilizing selection and inhibit divergence (Kirkpatrick and Nuismer (2004)).

Numerous studies have examined how different isolating mechanisms, namely assortative mating (Doebeli (1996); Kondrashov and Shpak (1998); Kopp et al. (2018)), ecological selection (Smith (1966); Maan and Seehausen (2011)), and sexual selection (Servedio (2016); Servedio and Kopp (2012); Schumer et al. (2017)) influence divergence and RI in the presence of ongoing gene flow. These suggest that RI is facilitated by coupling between traits that act as reproductive barriers (Coyne and Allen Orr (2004); Smadja and Butlin (2011); Kirkpatrick and Ravigné (2002)). Such coupling can arise if loci underlying different traits have pleiotropic effects or are tightly linked and not easily broken apart by recombination (Felsenstein (1981); Butlin and Smadja (2018); Smadja and Butlin (2011); Servedio (2009)). In this regard, speciation is thought to be most effective with magic traits, i.e., if the same trait is under divergent selection and also mediates AM. Magic traits are now known to be more common than previously thought (Servedio et al. (2011a)): well-known examples include wing color pattern in Heliconius (Merrill et al. (2012)) and body size in Littorina (Perini et al. (2020)).

Magic traits may differ in their mechanism of AM and the extent to which they are under sexual selection, and have been studied extensively to understand the evolution of mate choice during speciation (Servedio et al. (2011b); Servedio and Kopp (2012); Servedio and Boughman (2017)). While previous work has mostly focused on few-locus models of mate choice, here we consider a polygenic trait influenced by many small-effect alleles. Not only is this more realistic in view of the polygenic architecture of most traits but also much less understood, with previous studies relying largely on simulation of two populations (Sachdeva and Barton (2017); Muralidhar et al. (2022)) or of hybrid zones (Irwin (2020); Perini et al. (2020)). In general, the evolution of polygenic divergence and RI depends on the ease with which *multiple* locally advantageous alleles establish and/or are maintained across the genome despite maladaptive gene flow between populations. This in turn depends on whether selected loci evolve more-or-less independently or if LD between sets of maladapted alleles causes them to be eliminated *together* before they recombine onto fitter backgrounds. This was first studied in the context of hybrid zones, where it was found that the strength of LD between multiple

selected loci depends on the ratio of total selection and total recombination (Barton (1983)).

As we demonstrate here, a powerful way of analysing the effects of LD on divergence is by tracking how introgressing alleles at any one trait locus are transferred between genetic backgrounds, typically via multiple recombination events over multiple generations. The different genetic backgrounds on which the focal allele finds itself may differ in their hybrid index (i.e., the proportion of genetic material they derive from the parental populations) and thus also have different fitness values, which influences the transmission probability of the allele. This is captured by the notion of the *effective* migration rate (m_e), which is the rate at which neutral introgressing alleles are transferred between hybridising populations (Bengtsson (1985)). In essence, strong LD between large numbers of loci reduces m_e across the genome (Barton and Bengtsson (1986)). This increases divergence between hybridising populations, which reduces hybrid fitness, further decreasing effective migration rates. Indeed, it has been argued that this genomewide reduction in m_e between hybridising populations provides a natural way of quantifying RI (Westram et al. (2022)). Subsequent work has generalised theory based on m_e to weakly selected loci with partial and/or transient divergence between populations (Sachdeva (2022)), heterogeneous effect sizes and dominance (Zwaenepoel et al. (2024)) and sex-linkage (Fraïsse and Sachdeva (2021)). However, all of this work is restricted to randomly mating populations.

In this paper, we extend theory based on effective migration rates to a scenario with assortative mating due to female preference for phenotypically similar males, which leads to sexual selection on one or both sexes. The phenotype underlying assortment is polygenic and under divergent selection across a mainland and island: thus, we have a simple 'magic trait' scenario of speciation. We derive expressions for m_e at an unlinked neutral locus and then use this to predict allele frequency divergence at individual trait loci under migration-selection-assortment balance by assuming that the effect of all other loci on the focal locus is encapsulated by this effective migration rate. This allows us to predict mean trait divergence between mainland and island, and explore how assortment, viability and sexual selection jointly influence the critical threshold at which migration swamps adaptive divergence. We then use this to disentangle the relative contributions of viability and sexual selection to the genomewide barrier to gene flow in parameter regimes where partial RI persists despite limited hybridisation.

2 Model and Methods

2.1 Model

Consider an island subject to one-way migration from a large mainland, such that a fraction m of individuals on the island are replaced by migrants in every generation. Individuals are haploid and hermaphroditic (i.e. produce both male and female gametes), and can thus play the male or female role during reproduction. Each individual expresses an additive polygenic trait $z = \sum_{i=1}^{L} \alpha_i X_i$ influenced by a large number L of unlinked, biallelic loci, where $X_i = 0, 1$ denote alternative alleles and $\alpha_i > 0$ is the effect size at locus i. The trait is under directional selection on the island– the viability component of fitness of an individual with trait value z is proportional to $e^{-\beta z}$, where β denotes the strength of viability selection. We will not explicitly model the mainland population but assume that trait values follow a normal distribution, or alternatively, that migrant genotypes are maximally deleterious on the island (see sec. 2.2). Unless stated otherwise, we will also assume that the island is initially

perfectly adapted (with $z = 0$ for each individual).

Individuals on the island mate assortatively, with trait z also serving as the basis for AM via female choice. We assume a Gaussian choice function, such that the probability of mating between a male and female with trait values z_M and z_F is proportional to $e^{-\gamma(z_M - z_F)^2}$, where γ is the strength of assortment. In other words, females preferentially mate with males with trait values within a few $1/\sqrt{\gamma}$ of their own. Thus, the assortment trait is a 'magic trait', which simultaneously influences mate choice and is under both natural and sexual selection, thereby generating both postzygotic and prezygotic barriers to gene flow.

We consider two different models of AM– allowing for sexual selection on both sexes (Model I) or only on males (Model II)– sometimes referred to as the 'plant model' and 'animal model' of assortment (Kirkpatrick and Nuismer (2004)). The probability $M(z_M, z_F)$ of mating between a male and female with trait values z_M and z_F under the two models is:

$$M(z_M, z_F) = \frac{e^{-\gamma(z_M - z_F)^2}}{\int \int P(z)P(y)e^{-\gamma(z-y)^2} dy\, dz} \qquad \text{Model I}$$

$$M(z_M, z_F) = \frac{e^{-\gamma(z_M - z_F)^2}}{\int P(z)\, e^{-\gamma(z-z_F)^2}\, dz} \qquad \text{Model II}$$

Here, $P(z)$ is the trait value distribution just before mating. Note that under Model II, mating probabilities are normalized such that any female, regardless of trait value, mates with probability 1, while under Model I, females with rare phenotypes have lower mating probability. Thus, sexual selection acts only on males in the former case but on both sexes in the latter.

2.2 Effective migration rate at a neutral locus

Our analysis is based on effective migration rates (m_e)– defined as the rate at which neutral alleles entering a population via migration are incorporated into the resident gene pool (Bengtsson (1985); Barton and Bengtsson (1986)), where 'residents' are defined more precisely later. In our model, migrants and their descendants experience both viability and sexual selection and are less fit than residents. Consequently, in order to establish, neutral alleles must recombine away from migrant genetic backgrounds before these are eliminated by selection. Thus, m_e will vary along the genome, being lower at sites that are tightly linked to a selected locus or in genomic regions with a high density of selected loci. However, we will focus on m_e at a neutral locus that is *unlinked* to any selected locus, since we are interested in quantifying the average or genomewide reduction in gene flow.

Under rare migration ($m \ll 1$), the ratio m_e/m for an unlinked neutral allele, also called the gene flow factor g, is equal to the reproductive value (RV) of migrants relative to residents (Kobayashi et al. (2008)). Here, RV refers to the long-term genetic contribution of the migrant in the recipient population (Fisher (1999); Barton and Etheridge (2011)), and is given by:

$$g = \frac{m_e}{m} = \prod_{i=0}^{\infty} \overline{W}_i \tag{1}$$

where $\overline{W}_0$ denotes the average fitness of migrants relative to residents, and $\overline{W}_i$ the average relative fitness of ith generation descendants of migrants (see, e.g., Westram et al. (2022)). For $m \ll 1$,

Parameter/ Variable	**Description**
z	Additive magic trait; individual trait value $z = \sum_{i=1}^{L} \alpha_i X_i$, where $X_i = 0$ or 1 denote alternate alleles at locus i
α_i	Effect size at a locus i
L	Number of loci influencing magic trait
N	Population size
m	Rate of migration from mainland to island
β	Strength of viability selection on island; individual with trait value z has viability $\propto e^{-\beta z}$
γ	Strength of assortative mating
$M(z_M, z_F)$	Probability of mating between a male and female with trait values z_M and z_F; differs under the two models of assortative mating
s_i	Selection coefficient at locus i given by $s_i = \beta\alpha_i$; we consider equal-effect loci with $s_i = s$
m_e	Effective migration rate at any (equal-effect) trait locus
g	Gene flow factor $g = m_e/m$
Δ	Mean trait divergence between mainland and island, $\Delta = \sum_{i=1}^{L} \alpha_i(1-p_i)$; p_i is frequency of locally deleterious allele at locus i on island; mainland is fixed for the alternative allele
Δ_0	Maximum possible divergence between mainland and island, given by αL
Y	Allele frequency divergence per locus, $Y = \Delta/\Delta_0$, equals $1-p$ for equal-effect loci
$\beta\Delta_0$, $\gamma\Delta_0^2$	Scaled strength of viability and sexual selection; amount by which (log) fitness of migrants is reduced due to viability and sexual selection respectively, when $\Delta = \Delta_0$
m_c/s	Critical migration rate (relative to selection per locus) at which divergence Y becomes 0
Y_c	Allele frequency divergence associated with the tipping point
Quantities that appear in derivation of gene flow factor g	
$P_r(z), P_i(z)$	Trait value distribution among residents and i^{th} generation descendants of migrants; $i = 0$ corresponds to migrants.
$\overline{z}_r, \overline{z}_i$	Mean trait value among residents and i^{th} generation descendants of migrants
V_r, V_i	Variance of trait values among residents, i^{th} generation descendant of migrants
$V_{r,r}, V_{r,i}$	Within-family variance when one parent is a resident and the other a resident or an ith generation descendant of a migrant
$\overline{W}_i$	Mean fitness of ith generation descendants of migrants relative to residents

Table 1: Key parameters and variables

individuals with recent immigrant ancestry are rare, so that migrants and their descendants mate primarily with residents. As a result, first-generation descendants of the migrants are F_1 hybrids, second-generation descendants are typically first-generation backcrosses (with residents), i^{th} generation descendants are $(i-1)^{th}$ generation backcrosses (denoted by BC_{i-1}), and so on. Thus, computing m_e boils down to calculating the relative fitness of successive back-crosses.

In the following, we classify individuals on the island by the number of generations leading back to their most recent immigrant ancestor, i.e., into F_1, BC_1, BC_2, up to BC_n (where n is arbitrary). All other individuals, i.e., those with no migrant ancestor in the previous $n+1$ generations, are designated as 'residents'. We further assume trait values to be normally distributed *within* any group of individuals with the same level of recent migrant ancestry. In other words, the trait distributions $P_r(z)$, $P_1(z)$, $\ldots P_i(z)$, $\ldots$ amongst residents, F_1 hybrids, i^{th} generation descendants (who are BC_{i-1}) and so on, are assumed to be normal with means $\overline{z}_r, \overline{z}_1, \ldots \overline{z}_i, \ldots$ and variances $V_r, V_1, \ldots V_i, \ldots$ respectively (see figure 1A; also Table 1). For generality, we also take the migrant trait value distribution to be normal with mean $\overline{z}_0$ and variance V_0. However, in the Results, we will only consider a scenario where the mainland is fixed for alleles locally deleterious on the island (so that $\overline{z}_0 = \sum_{i=1}^{L} \alpha_i$ and $V_0 = 0$).

Assuming normally distributed trait values is justified for traits influenced by a large number of loci of small effect. The inheritance of such traits is described by the infinitesimal model (Fisher (1918)), which states that trait values of the offspring of any two individuals are normally distributed about the mean of the parents, with a 'within-family' variance V_*, which depends only on the genetic relatedness between parents, regardless of selection, non-random mating etc. (Barton et al. (2017)). We further assume that V_* depends only on the extent of migrant ancestry of the parental individuals, and does not vary significantly across (for instance) different resident $\times$ F_1 parental pairs, all of which thus have approximately the same V_*. We denote the within-family variance by $V_{r,r}$ when both parents are residents, $V_{r,0}$ when one parent is a resident and the other migrant, and $V_{r,i}$ when one parent is a resident and the other an ith generation descendant of a migrant.

Then, we can express the distribution $P_i(z)$ in terms of the parental trait value distributions $P_r(z_1)$ and $P_{i-1}(z_2)$ (eq. 2a). Further, we can express the mean fitness $\overline{W_i}$ of ith generation descendants (relative to residents) as an integral over the parental distributions (eq. 2b).

$$P_i(z) \propto \int\int P_r(z_1) P_{i-1}(z_2)\, e^{-\beta(z_1+z_2)} \left[\frac{M(z_1,z_2)+M(z_2,z_1)}{2}\right] \frac{e^{-\frac{(z-(z_1+z_2)/2)^2}{2V_{r,i-1}}}}{\sqrt{2\pi V_{r,i-1}}} dz_2\, dz_1 \tag{2a}$$

$$\overline{W}_i = \int\int \frac{e^{-\beta z_1}}{\int dz\, e^{-\beta z} P_r(z)} P_r(z_1) \frac{e^{-\beta z_2}}{\int dz\, e^{-\beta z} P_r(z)} P_i(z_2) \left[\frac{M(z_1,z_2)+M(z_2,z_1)}{2}\right] dz_2\, dz_1 \tag{2b}$$

In eq. 2a, z_1 and z_2 denote respectively the trait values of the resident parent and the parent who is an $(i-1)^{th}$ generation descendant of a migrant, with corresponding trait distributions $P_r(z_1)$ and $P_{i-1}(z_2)$. The term $e^{-\beta(z_1+z_2)}$ captures the effect of viability selection on the parents, while $M(z_1,z_2)$ (which is the probability of mating between a male and female with trait values z_1 and z_2 respectively) captures the effect of AM and sexual selection. Note that the non-resident parent may have a different probability of being chosen as a mate depending on whether it takes the female or male role during reproduction: this is captured by the two terms $M(z_1,z_2)$ and $M(z_2,z_1)$ (that are unequal under Model II). Finally, the last term gives the distribution of trait values of the offspring under the

infinitesimal model; it captures the effect of recombination between parental genotypes and random segregation. Supplementary Information (SI)A provides a detailed derivation of eq. 2.

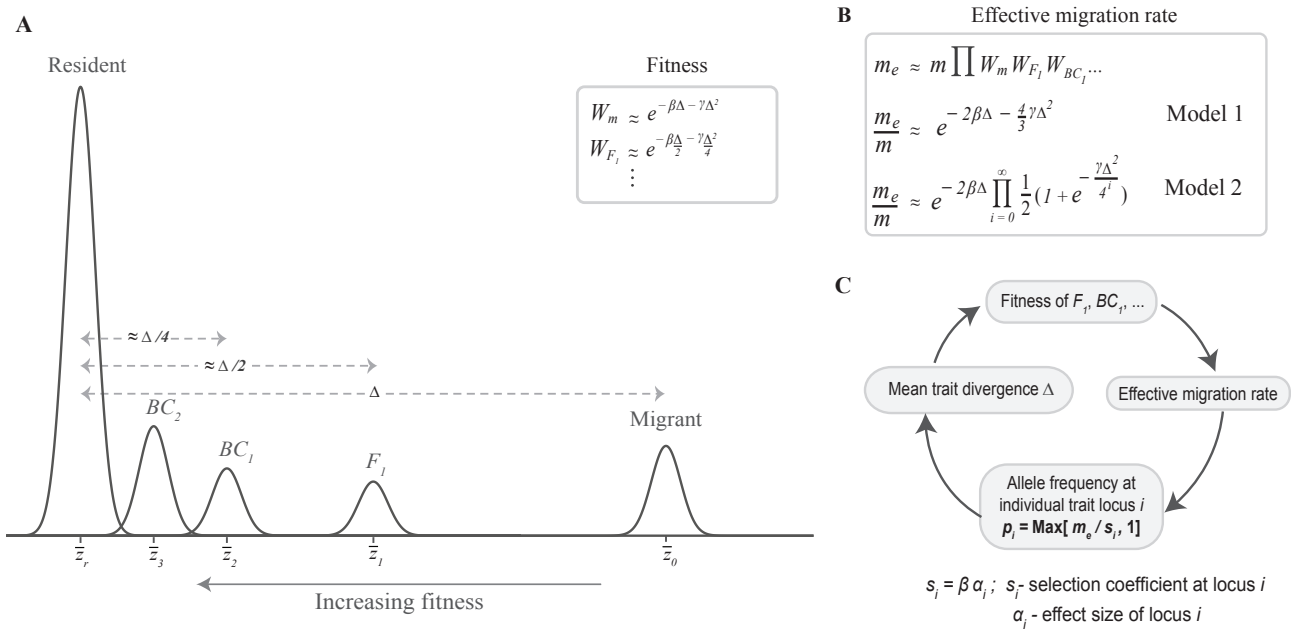

Figure 1: Schematic summarising our theoretical approach for predicting mean trait divergence Δ (or allele frequency divergence per locus Y) using effective migration rates (m_e). **A**: Distribution of trait values on the island is the sum of trait value distributions associated with residents, F_1s, BC_1s, and so on, where each such distribution is assumed to be approximately normal. Denoting the difference between trait means of migrants and residents by $\Delta = \overline{z}_0 - \overline{z}_r$, the difference between trait means of F_1s and residents is $\approx \Delta/2$, and more generally, between that of ith generation descendants (of migrants) and residents is $\Delta/2^i$. This allows us to calculate *average relative fitness* of F_1s, BC_1 etc. (inset). **B**: The *gene flow factor* g (which describes the reduction in neutral gene flow at a site unlinked to any selected site) can be calculated from the average relative fitness of migrants and of individuals with recent migrant ancestry (F_1s, BC_1s), as in eq. 1. This gives approximate expressions for g under Model I (sexual selection on both sexes) and Model II (sexual selection only on males) **C**: The *mean trait divergence* Δ can be calculated by first using m_e to predict the allele frequency p_i at trait locus $i = 1, \ldots, L$, then summing over loci to obtain the mean trait divergence Δ, then using this to predict relative fitnesses of F_1s, BC_1s and so on (as in A.), and finally using these to predict the effective migration rate m_e (as in B.). This kind of circular dependence, where m_e depends on Δ which depends on m_e (via allele frequencies), allows for a self-consistent solution for the mean trait divergence Δ.

Integrating over eq. 2a gives an expression for the mean and variance of trait values among the ith generation descendants in terms of the means and variances among $(i-1)^{th}$ generation descendants and among residents. This specifies a set of recursions for $\overline{z}_i$, V_i and $V_{r,i}$ which can be solved numerically (see SI, **A**). In turn, these predict the relative fitnesses $\overline{W}_i$ (using eq. 2b), from which one can calculate the gene flow factor g (using eq. 1) and m_e.

However, in the main paper, we will derive a more approximate and intuitive expression for g by simply tracking the difference between trait means of residents and other groups (i.e., F_1s, BC_1s and so on), while neglecting phenotypic variance *within* any group. More concretely, we assume that the mean trait value of hybrid offspring is the average of the trait means of the parental subgroups. Thus, for example, F_1 hybrids are assumed to be midway between residents and migrants, giving $\overline{z}_1 \approx (\overline{z}_r + \overline{z}_0)/2$, which gives: $\overline{z}_1 - \overline{z}_r \approx (\overline{z}_0 - \overline{z}_r)/2 = \Delta/2$. Here $\Delta = \overline{z}_0 - \overline{z}_r$ denotes the difference in the trait means of migrants and residents, which we refer to as mean trait divergence. It follows similarly that $\overline{z}_i - \overline{z}_r \approx (\overline{z}_{i-1} - \overline{z}_r)/2 = \Delta/2^i$ (see figure 1A). Thus, in effect, every successive

generation of backcrossing halves the average phenotypic distance of the (next-generation) backcross from the resident phenotype (Sachdeva (2022); Zwaenepoel et al. (2024)).

As we argue below and in SI, **A**, this simple argument holds as long as $\beta V_i \ll \Delta$ and $\gamma V_i \ll 1$ (for all i). Moreover, under these conditions, the relative fitness of any group of descendants is (from eq. 2b): $\overline{W_i} \approx e^{-\beta(\overline{z}_i-\overline{z}_r)-\gamma(\overline{z}_i-\overline{z}_r)^2}$ for Model I (sexual selection on both sexes) and $\approx e^{-\beta(\overline{z}_i-\overline{z}_r)}\frac{1}{2}(1+e^{-\gamma(\overline{z}_i-\overline{z}_r)^2})$ for Model II (sexual selection only on males). Now, using $\overline{z}_i - \overline{z}_r = \Delta/2^i$, where Δ is the mean trait divergence between mainland and island, we obtain the following approximate expressions (see figure 1B) for the gene flow factor $g = \prod_{i=0}^{\infty} \overline{W_i}$:

$$g(\Delta) \approx e^{-2\beta\Delta - \frac{4}{3}\gamma\Delta^2} \qquad \text{Model I} \tag{3a}$$

$$g(\Delta) \approx e^{-2\beta\Delta} \prod_{i=0}^{\infty} \left(\frac{1 + e^{-\gamma\Delta^2/4^i}}{2} \right) \qquad \text{Model II} \tag{3b}$$

A more detailed derivation of eq. 3 is provided in SI, **A**. A striking feature of eq. 3 is that under both models of sexual selection, the gene flow factor can be expressed as a product of two terms– the first describing the reduction in reproductive value of migrants due to viability selection, and the second due to sexual selection. The first term above, $e^{-2\beta\Delta}$, is the square of the ecological fitness of migrants relative to residents. The second term is related to the relative mating success of migrants and differs between the two models. For instance, under Model I, the sexual selection term is simply the relative mating success of migrants raised to the power 4/3. These powers (of 2 or 4/3) emerge from the specific fitness and mating functions assumed in our model. However, more generally, the fact that the viability and sexual selection components of g bear a simple relationship to the corresponding fitness components (of say migrants or F_1s) is due to there being a simple relationship between the immediate fitness and long-term reproductive value of individuals when fitness is polygenic.

What assumptions underlie our simple approximation for trait means, which is the basis of eq. 3? First, within any group of descendants, say F_1s, individuals closer to the resident phenotype will have higher viability, causing the trait mean of BC_1s (the hybrid offspring of F_1s and residents) to be slightly shifted towards residents rather than exactly midway between the two parental groups. The magnitude of this shift is proportional to βV_1; thus, in ignoring the effects of viability selection *within* subgroups, we assume that $\beta V_i \ll \Delta$, where Δ is the mean trait divergence between mainland and island. As we argue in SI, **A**, under migration-selection balance, the various βV_i must be comparable to the migration rate m, and can thus be neglected if $m \ll 1$.

Second, we ignore the effects of sexual selection within subgroups, which may lead to small differences in mating success among (say) F_1s, again slightly shifting the BC_1 trait mean towards one of the parental subgroups. Moreover, AM can also increase the phenotypic variance within each group (Fisher (1918); Wright (1921); Crow and Felsenstein (1968)), at least if sexual selection is weak, as under Model II (Kirkpatrick and Nuismer (2004)), further influencing the magnitude of these shifts. As we argue in SI, **A**, both these effects can be ignored if $\gamma V_i \ll 1$, i.e., if the preference function is much wider than the typical phenotypic spread within any subgroup. Thus, in summary, our simple approximation for trait means– $\overline{z}_i - \overline{z}_r \approx \Delta/2^i$ holds as long as $\beta V_i \ll \Delta$ and if $\gamma V_i \ll 1$. In effect, these two conditions amount to assuming that viability and sexual selection are strong enough

to affect the survival and mating success of migrants and their F_1 and BC descendants relative to residents, but not strong enough to cause differential mating success *within* any group of descendants.

In summary, eq. 3 relates the gene flow factor $g = m_e/m$ at an unlinked locus to the mean trait divergence Δ between mainland and island at equilibrium, where Δ now depends on the allele frequency differences across all trait loci. In the following, we relate the allele frequency difference at any locus back to the effective migration rate m_e, thus allowing us to obtain the mean trait divergence in a self-consistent way (see figure 1C).

2.3 Predicting allele frequency divergence at trait loci using m_e

We now consider allele frequency dynamics at individual trait loci, assuming the mainland to be fixed for alleles that are deleterious on the island. First, consider the equilibrium allele frequency at a locus with effect size α_i under *linkage equilibrium*, i.e., neglecting the effects of LD between the focal locus and other trait loci. In SI, **B**, we show that sexual selection does not contribute to *direct* selection at a locus (at least under our assumed quadratic preference function), provided α_i is sufficiently small. More concretely, at migration-selection balance, the deleterious allele frequency p_i is simply m/s_i, where $s_i = \beta\alpha_i$, provided $\beta\alpha_i \ll 1$ and $\alpha_i \ll \gamma/\beta$ (SI, **B**). The latter condition can be re-written as $\alpha_i/\Delta \ll (\gamma\Delta^2)/(\beta\Delta)$— this simply means that the relative contribution of any locus to trait divergence should be much smaller than the ratio of the sexual selection component of fitness to the viability selection component.

Following Sachdeva (2022), we now assume that the main effect of LD between immigrant alleles is to reduce m_e at any locus. This can be justified as long as trait loci are unlinked or weakly linked so that LD between these breaks down much faster than allele frequencies change. Further, if individual loci are weakly selected, then the effective migration rate at a trait locus is approximately equal to that at a neutral locus, so that in a large population, we have: $p_i \approx m_e/s_i \approx \frac{m}{s_i}\, g(\Delta)$. Here, the gene flow factor g is given by eq. 3, and captures the effect of both viability and sexual selection on migrants and their descendants.

Note that g depends on allele frequencies $\{p_i,\ i = 1, .. L\}$ at all L loci via the mean trait divergence $\Delta = \sum_{i=1}^{L} \alpha_i(1 - p_i)$. We can thus determine allele frequencies by numerically solving the L equations $p_i = (m/s_i)g(\Delta)$, all coupled via Δ (Zwaenepoel et al. (2024)). However, here we will focus on the simpler case of equal-effect loci, such that $\alpha_i = \alpha$ (and $s_i = s$) for all i. Then, in the absence of drift, equilibrium allele frequencies are equal across all loci (i.e., $p_i = p$), at least in parameter regimes where there is only one stable equilibrium. Thus, the mean trait divergence can be written as $\Delta = \alpha L(1-p)$. In the following, we will express our results in terms of $Y = \Delta/\Delta_0$, the trait divergence between the mainland and the island relative to the maximum possible divergence $\Delta_0 = \alpha L$. Note that for the case of equal-effect loci, we have: $Y = 1 - p$, so that Y is also the mean allele frequency divergence per locus. It follows from eq. 3 that Y satisfies:

$$Y \approx 1 - \frac{m}{s} e^{-2\beta\Delta_0 Y - \frac{4}{3}\gamma\Delta_0^2 Y^2} \qquad \text{Model I} \tag{4a}$$

$$Y \approx 1 - \frac{m}{s} e^{-2\beta\Delta_0 Y} \prod_{i=0}^{\infty} \left(\frac{1 + e^{-\gamma\Delta_0^2 \frac{Y^2}{4^i}}}{2} \right) \qquad \text{Model II} \tag{4b}$$

as long as a solution with $0 \leq Y \leq 1$ exists, and is zero otherwise. Eq. 4 can be solved numerically to obtain allele frequency divergence per locus Y or the trait divergence Δ. In the case of Model II, the product in eq. 4b converges rapidly and can be computed by taking ~ 20 terms.

Equation 4 is deterministic, implicitly assuming that $Ns \gg 1$, where N is the size of the island population. However, our basic approach can be generalised to account for genetic drift, by treating each frequency as a (continuous) *random* variable described by a probability density $\psi[p]$ (Sachdeva, 2022). Since a finite population subject to one-way migration requires non-zero mutation rates at trait loci in order to maintain long-term adaptive divergence, we allow for mutation at rate μ per individual per generation between alternative alleles at each locus. Then, under our approximation, the (unnormalised) equilibrium probability density $\psi[p]$ is given by:

$$\psi[p|\Delta] = p^{2N\mu+2Nmg[\Delta]-1}(1-p)^{2N\mu-1}e^{-2Nsp} \tag{5a}$$

$$\Delta/\Delta_0 = Y = \frac{\int_{p=0}^{1} p\,\psi[p|\Delta]\,dp}{\int_{p=0}^{1} \psi[p|\Delta]\,dp} \tag{5b}$$

Equation 5a is simply the allele frequency distribution at a single locus at migration-selection-mutation-drift equilibrium, as predicted by the diffusion approximation (Wright (1937)), but with m replaced by $m_e = m\,g[\Delta]$. As before, the effective migration rate m_e (or equivalently the gene flow factor g) is a function of the trait divergence Δ and is given by eq. 3. We can obtain Δ, in turn, by summing over allele frequencies at all trait loci. With equal-effect loci, we have: $\Delta = \Delta_0 Y$, where Y, the expected allele frequency divergence per locus, is obtained by integrating over $\psi[p]$ (eq. 5b), thus allowing us to self-consistently solve for Δ (as in the deterministic case).

2.4 Deterministic simulations

In the main paper, we analyse the deterministic limit and test our approximations (eqs. 3 and 4) against deterministic simulations based on the hypergeometric model, which describes the inheritance of a quantitative trait influenced by equal-effect loci (Barton (1992); Doebeli (1996)). In each generation, migration is followed by viability selection, which is followed by assortative mating (involving sexual selection on one or both sexes), and production of offspring via free recombination between parental genotypes. The mainland is fixed for alleles that are locally deleterious on the island. Additionally, unless stated otherwise, simulations are initialised with the island fixed for the locally adaptive allele at each trait locus.

Note that with equal-effect loci, the trait value z of any individual is just αj, where j is the number of locally deleterious alleles it carries. SI, **C** specifies the recursions that describe how the distribution $P(j)$ (or alternatively, $P(z)$) changes in any generation under the combined effects of migration, (viability and sexual) selection, AM and random segregation. Recursions for $P(j)$ are iterated until equilibrium, and allele frequency divergence per locus Y obtained using: $Y = \frac{1}{L}\sum_{j=0}^{L}(L-j)P(j)$. This can then be compared with theoretical predictions (eq. 4).

Individual-based simulations of finite populations (used to explore the effects of genetic drift on divergence) are described in SI, **D**.

3 Results

We now use predictions based on effective migration rates to investigate the combined effects of migration, viability and sexual selection, and assortative mating on the equilibrium allele frequency divergence between mainland and island as well as on the genome-wide reduction in neutral gene flow (as measured by the gene flow factor). In the main paper, we focus on the deterministic limit of the model; the effects of genetic drift are examined in SI, **D**. We first quantify how mean trait divergence declines with increasing migration (which is parameterized by m/s, migration rate relative to selective effect per locus), for different strengths of viability and sexual selection against migrants. This allows us to identify a critical migration threshold m_c beyond which local adaptation is swamped on the island, i.e., Y becomes zero. Next, we ask: how strong do viability and sexual selection have to be to shift this swamping threshold to *above* that for a single divergently selected locus? Finally, we examine the relative contributions of viability and sexual selection to RI under different parameter regimes, by using the increase in F_{ST} at an arbitrary site (unlinked to any trait locus) as a proxy for the genome-wide barrier effect.

Throughout, we quantify the strength of viability and sexual selection in terms of $\beta\Delta_0$ and $\gamma\Delta_0^2$, which denote respectively the reduction in log fitness of migrants (relative to residents) due to viability and sexual selection under conditions of maximum divergence between the mainland and island, i.e., when $\Delta = \Delta_0$. If traits are highly polygenic (large L and small effect sizes α), then in the deterministic limit, divergence (as measured by $Y = \Delta/\Delta_0$) depends essentially on the composite parameters m/s, $\beta\Delta_0$ and $\gamma\Delta_0^2$, where $\Delta_0 = \alpha L$ (see eq. 4). However, for smaller L (say tens of loci), divergence levels may be sensitive to the exact genetic basis of the trait, i.e., depend on α and L separately (SI, **E**). In the main text, we only focus on the large L limit (where eq. 4 applies). Additionally, we assume a scenario of secondary contact where mainland and island are initially maximally diverged; other initial conditions (with partial divergence) are explored in SI, **F**.

3.1 Effect of viability and sexual selection on adaptive divergence.

Figures 2A-2F show Y, the allele frequency divergence per locus (or equivalently, trait divergence between mainland and island relative to the maximum possible divergence) as a function of m/s, under the two models of AM in the deterministic limit. For each model, the plots depict divergence levels for $\beta\Delta_0 = 0.2, 0.5, 1$ (i.e., stronger viability selection from left to right). Further, for each $\beta\Delta_0$, we consider random mating (grey) as well as populations with increasing assortment ($\gamma\Delta_0^2 = 0.2$, 0.5 and 1), which generate increasing levels of sexual selection on both sexes (Model I; top) or only on males (Model II; bottom). The analytical predictions obtained by solving eq. 4 (lines) match well with the results of deterministic simulations of a trait influenced by $L = 80$ loci (points) unless net selection (viability and/or sexual selection) is very strong; see, e.g., red curves corresponding to $\gamma\Delta_0^2 = 1.5$ in figures 2E and 2F (with $\beta\Delta_0 = 1$). This discrepancy between simulations and theory becomes weaker with increasing number of trait loci (SI, **E**).

Thus, there is a significant range of parameter space where simple approximations based on m_e (eq. 4) predict equilibrium divergence accurately. Note that these approximations only account for the lower reproductive value of migrants due to sexual (and viability) selection, but ignore other effects of AM such as increased variance among residents, F_1s, etc, and the increased probability of mating

between individuals with recent migrant ancestry. That there is, nonetheless, close agreement between theory and simulations (which make no assumptions about AM), suggests that AM strengthens divergence in our model primarily by generating additional (sexual) selection against migrants and their descendants. This reduces m_e, thus increasing the genomewide barrier effect and facilitating divergence (see also *Discussion*).

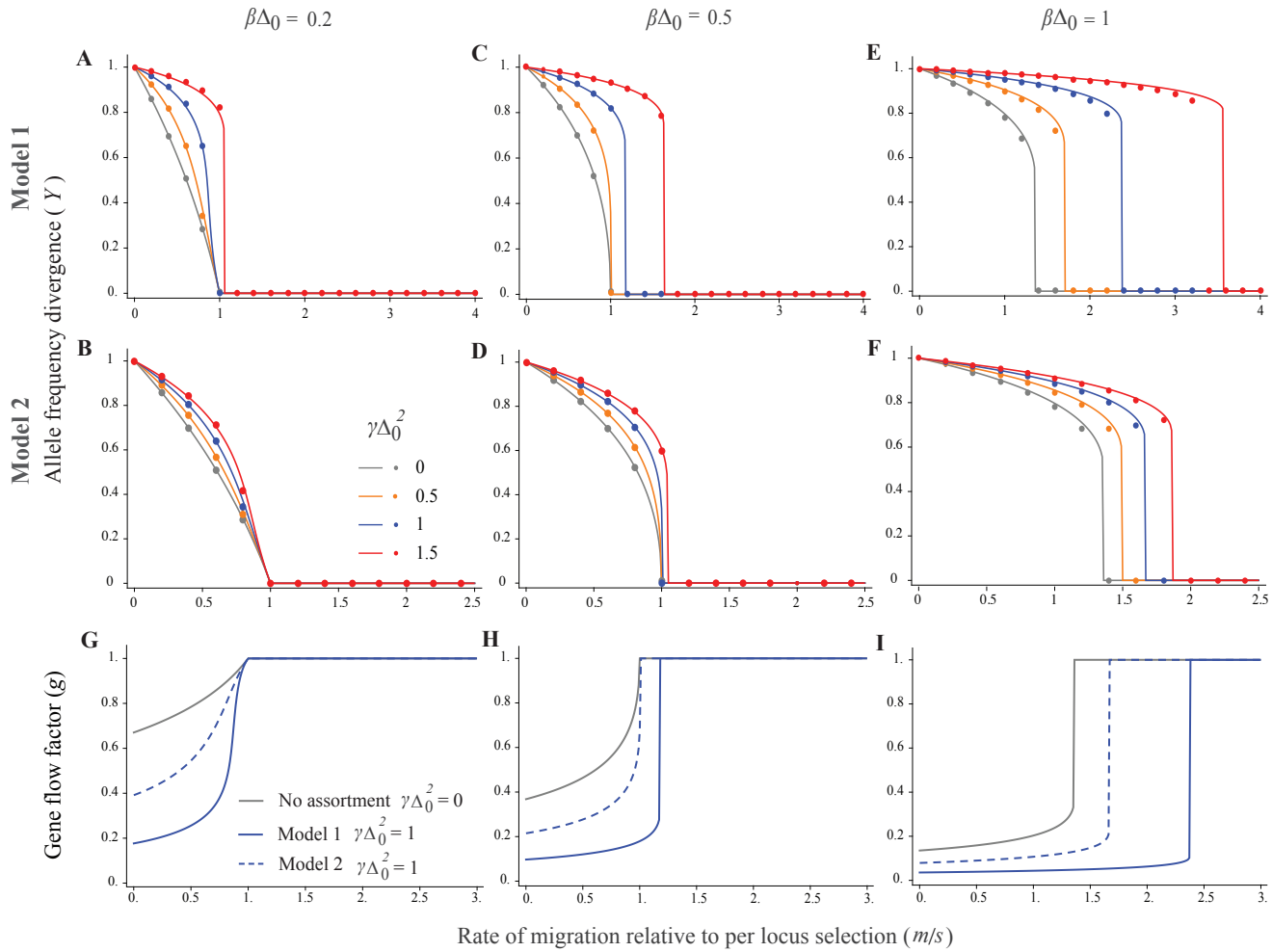

Figure 2: Allele frequency divergence per locus Y and gene flow factor g as a function of m/s, the migration rate relative to selection per locus. The top and middle rows depict respectively divergence levels for Model I (sexual selection on both sexes) and Model II (sexual selection only on males). Different subfigures within each row correspond to different strengths of viability selection ($\beta\Delta_0 = 0.2, 0.5, 1$ from left to right); different colors within each plot depict different strengths of AM and sexual selection ($\gamma\Delta_0^2 = 0, 0.2, 0.5, 1$ shown in grey, orange, blue and red respectively). Solid lines show theoretical predictions (obtained by numerically solving eq. 4) while dots represent the results of deterministic simulations of the hypergeometric model with $L = 80$ loci underlying the polygenic trait. Increasing viability and sexual selection cause sharper thresholds ('tipping points') for loss of divergence as well as an increase in the critical migration rate at which divergence is lost. The bottom row depicts theoretical predictions for the gene flow factor g for random mating (grey) vs. assortatively mating populations under Models I and Models II (solid vs. dashed lines), assuming $\gamma\Delta_0^2$. The gene flow factor and m_e are more strongly reduced under Model I than under Model II.

As evident from figures 2G-2I, this reduction in m_e (or equivalently g) is stronger under Model I (where sexual selection acts on both sexes) as compared to Model II (sexual selection only on males), and also depends on existing levels of divergence (which in turn depend on the strength of migration and viability selection).

Figure 2 also highlights how migration affects divergence differently in parameter regimes characterised by strong vs. weak selection against migrants. When sexual selection (due to assortment) and viability selection are weak, divergence declines *smoothly* with increasing m/s, going to zero at a critical threshold $m_c/s = 1$, regardless of the actual value of $\beta\Delta_0$ and $\gamma\Delta_0^2$ (e.g., in figure 2B). Note that $m_c/s = 1$ is also the critical migration threshold for loss of adaptation at a single divergently selected locus. Thus, this threshold remains unchanged even if divergence is polygenic, as long as viability and sexual selection on the trait are weak.

By contrast, with strong viability and/or sexual selection, divergence at first decreases mildly with increasing migration, but then exhibits an *abrupt* collapse once it crosses a threshold Y_c (at a migration level $m/s \gtrsim 1$). Such a 'tipping point', defined as a discontinuous change in divergence (though not necessarily to $Y = 0$), can be observed in figure 2C and 2D at higher assortment and in 2E and 2F at all levels of AM. We can use eq. 3 to derive expressions for Y_c, the allele frequency divergence per locus associated with tipping points under the two models of AM (details in SI, **G**):

$$Y_c = \frac{(4\gamma\Delta_0^2 - 3\beta\Delta_0) + \sqrt{(3\beta\Delta_0 + 4\gamma\Delta_0^2)^2 - 24\gamma\Delta_0^2}}{8\gamma\Delta_0^2} \quad \text{Model I} \tag{6a}$$

$$Y_c = 1 - \frac{1}{2\beta\Delta_0 + 2\gamma\Delta_0^2\, Y_c \sum_{i=0}^{\infty} \frac{4^{-i}}{1+e^{\gamma\Delta_0^2 (Y_c^2/4^i)}}} \quad \text{Model II} \tag{6b}$$

For Model II, eq. 6b specifies a transcendental equation for Y_c, which can be solved numerically (e.g., by approximating the sum by the first ~ 20 terms). Further, in SI, **G**, we show that the migration threshold associated with such a tipping point is: $m/s = (1 - Y_c)/g(\Delta_0 Y_c)$, where g, the gene flow factor, is given by eq. 3, and $\Delta_0 Y_c$ is the mean trait divergence at the tipping point.

The threshold Y_c can also be thought of as the minimum allele frequency difference per locus required to 'lock in' divergently selected alleles across the genome into distinct combinations that can be stably maintained by selection despite gene flow. Divergence levels lower than Y_c would result in an insufficient genomewide barrier to gene flow, causing a further decrease in divergence, further increasing m_e and so on, thus generating a tipping point. Interestingly, even though the tipping point occurs a specific level of migration (see, e.g., figure 2A-F), the divergence threshold Y_c associated with the tipping point is *independent* of this migration rate, and depends only on the strengths of viability and sexual selection (eq. 6).

In addition to generating tipping points, strong viability or sexual selection (or both) can also increase the critical migration threshold for loss of divergence *above* the single-locus threshold $m_c/s = 1$ (see e.g., figures 2E, 2F). In other words, if selection against migrants and their descendants is sufficiently strong, then polygenic divergence can be maintained at migration levels at which a single divergently selected locus (in the absence of LD) would have been swamped.

Both tipping points and increased swamping thresholds may be explained as follows: higher levels of migration reduce adaptive divergence between mainland and island, thereby increasing m_e (see eq. 3) and weakening multi-locus selection against locally deleterious alleles, which increases their frequency– further reducing adaptive divergence and increasing m_e, which finally results in the swamping of locally adaptive alleles above a certain migration threshold. Crucially, the positive feedback between loss of divergence at individual loci and increase in m_e across the genome is stronger when net selection

against migrant phenotypes (either viability or sexual selection) is higher. In the next section, we will explore how strong do viability and sexual selection have to be to produce these effects?

Note that while figure 2 and the analyses below pertain to the deterministic model, the basic results also hold for a finite population of size N, provided Ns is large. For instance, tipping points occur under both models of AM if $Ns \gtrsim 5$, for moderately strong (viability plus sexual) selection (SI, **D**). Moreover, theoretical approximations based on m_e (eq. 5) furnish reasonably accurate predictions for divergence levels in finite populations as well (Figure S1, SI).

3.2 When does assortative mating and sexual selection lead to tipping points and higher swamping thresholds?

We first illustrate the various thresholds described above using a concrete example assuming weak viability selection ($\beta\Delta_0 = 0.2$) and sexual selection on both sexes (Model I). Under *weak* AM and sexual selection (in this example, $\gamma\Delta_0^2 < 1.2$), divergence decreases smoothly with increasing migration and is completely swamped at $m_c/s = 1$: this behaviour is qualitatively similar to that of a single divergently selected locus. For *intermediate* assortment (here, $1.2 < \gamma\Delta_0^2 < 1.4$), increasing migration leads to a tipping point or discontinuous change in divergence: this occurs once allele frequency divergence per locus approaches a threshold Y_c (given by eq. 6) and involves a strong reduction in the genomewide barrier effect. However, in this intermediate assortment regime, the tipping point still occurs at a migration rate lower than the single-locus swamping threshold $m_c/s = 1$. For instance, for $\gamma\Delta_0^2 = 1.3$, the tipping point occurs at $m/s \approx 0.95$ and is associated with $Y_c \sim 0.6$: thus, a tiny increase in m/s (say from 0.949 to 0.95) causes allele frequency divergence per locus to fall from 0.6 to 0.1 (Figure S5). A further increase in migration further reduces divergence, and complete swamping occurs (as before) at $m_c/s = 1$. However, with even *stronger* assortment (in this example, $\gamma\Delta_0^2 > 1.4$), the genomewide barrier effect becomes strong enough to maintain divergence beyond $m/s = 1$, so that the tipping point now occurs at $m_c/s > 1$. Further, divergence is completely swamped (Y falls to zero) at the tipping point. Thus, in this strong assortment regime, tipping points and swamping thresholds coincide and occur at a critical migration threshold that exceeds the single-locus threshold (Figure S5). This is in contrast to the intermediate assortment regime– where the tipping point occurs (i.e., divergence collapses to a low but non-zero level) at a migration rate which is lower than the swamping threshold.

We now ask: in what parameter regimes do we observe these qualitatively different behaviours? Figures 3A and 3B show the divergence threshold Y_c associated with tipping points (i.e., sharp drops in divergence), while figures 3C and 3D show the critical migration threshold m_c/s beyond which divergence is completely swamped, as a function of the strength of viability selection (measured as $\beta\Delta_0$). As discussed above, the migration threshold associated with the tipping point is just the swamping threshold m_c/s if assortment is strong. Figures 3A-3D show Y_c and m_c/s for the two models of AM; colors in each plot correspond to different $\gamma\Delta_0^2$, which quantifies the strength of sexual selection (due to AM). These predictions all follow from eqs. 3-6 (see SI, **G** for details).

Under sufficiently strong viability selection ($\beta\Delta_0 > 0.5$), increased swamping thresholds ($m_c/s > 1$) emerge even in the absence of AM, as shown by the grey lines in figures 3C-3D (see also Sachdeva (2022)). Thus, in this regime, AM and sexual selection further increase the swamping threshold m_c/s

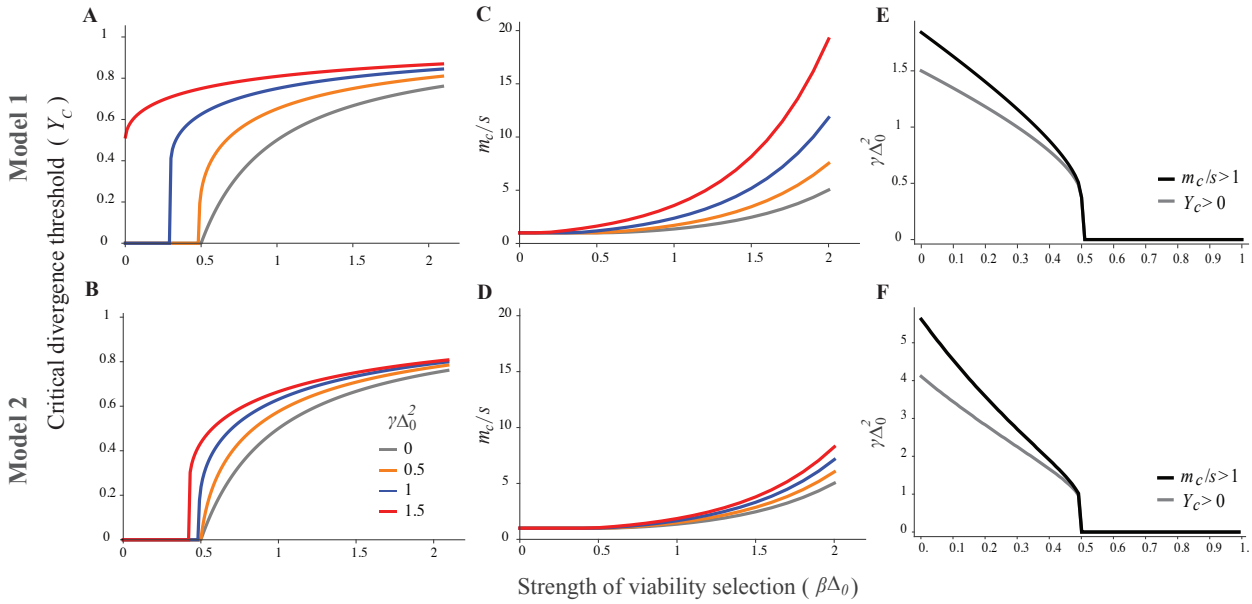

Figure 3: Deterministic predictions for swamping thresholds and tipping points. (A)-(B) Critical allele frequency divergence Y_c below which a tipping point occurs (i.e, divergence decreases sharply) vs. $\beta\Delta_0$, strength of viability selection. (C)-(D) Critical migration rate m_c/s at which divergence is completely lost vs. $\beta\Delta_0$. The different colors show results for different levels of AM and sexual selection ($\gamma\Delta_0^2$ =0, 0.5, 1 and 1.5). Upper panel (A, C, E) shows results for Model I (sexual selection on both sexes) and lower panel for Model II (sexual selection only on males). In panels A and B, parameter combinations with $Y_c = 0$ are those for which divergence decreases smoothly with increasing migration, i.e., there is no tipping point. In panels C and D, parameter combinations with $m_c/s = 1$ are those for which viability and sexual selection are too weak to shift the swamping threshold above the single-locus threshold ($m/s = 1$). (E)-(F) Grey curves show the minimum level of sexual selection (as measured by $\gamma\Delta_0^2$) required for $Y_c > 0$, i.e., to generate tipping points in divergence, as a function of $\beta\Delta_0$, the strength of viability selection. Black curves show the minimum level of sexual selection ($\gamma\Delta_0^2$) required for $m_c/s > 1$, i.e., to shift the swamping threshold above the single-locus threshold, as a function of $\beta\Delta_0$. Panels E and F correspond to Model I and II respectively. All results in this figure show (large L) theoretical predictions obtained from eqs. 3-6.

(beyond 1) as well as the critical level of divergence Y_c that can be stably maintained. This can be explained, as before, in terms of the effects of sexual selection on m_e and the overall genomewide barrier to gene flow. As expected, AM has a stronger effect in Model I (where sexual selection is stronger) than in Model II. For example, for $\beta\Delta_0 = 1$, AM with strength $\gamma\Delta_0^2 = 1$ increases the critical migration threshold (m_c/s) from 1.36 (with random mating) to 2.37 under Model I and 1.66 under Model II. Note also that AM makes it harder to maintain intermediate levels of divergence, with divergence becoming unstable and collapsing once mean allele frequency difference per locus approaches 0.75 under Model I and 0.66 under Model II. By contrast, under random mating (and with $\beta\Delta_0 = 1$), we have $Y_c = 0.5$, so that somewhat lower levels of divergence are maintained between the mainland and island, albeit at lower migration rates.

The situation is somewhat different if $\beta\Delta_0 < 0.5$; now, viability selection by itself (i.e., in the absence of AM) is too weak to generate tipping points or shift the swamping threshold m_c/s. However, sufficiently strong sexual selection due to AM will lead to one or both. More specifically, given viability selection $\beta\Delta_0 < 0.5$, we can determine an assortment threshold, i.e., a minimum value of $\gamma\Delta_0^2$ beyond which tipping points occur ($Y_c > 0$ in figures 3B and 3D), and a second threshold beyond which critical migration rates become higher than the single locus threshold (m_c/s exceeds 1 in figures

3A and 3C); see SI, **G** for details. These assortment thresholds are depicted by grey and black curves respectively in figures 3E and 3F.

As expected, much higher levels of AM are required to generate tipping points and increase swamping thresholds when viability selection is weak, and when sexual selection acts only on males (instead of both sexes). For instance, with $\beta\Delta_0 = 0.2$, we have $m_c/s > 1$ only if $\gamma\Delta_0^2$ exceeds 1.41 under Model I and 3.58 under Model II (see figures 3E, 3F). However, the sex-averaged mating success of migrants relative to residents is $e^{-\gamma\Delta_0^2} \sim 0.24$ (for $\gamma\Delta_0^2 = 1.41$) under Model I and and $(1 + e^{-\gamma\Delta_0^2})/2 \sim 0.512$ (for $\gamma\Delta_0^2 = 3.58$) under Model II, both calculated by assuming maximum divergence between mainland and island. Thus, critical assortment thresholds are associated with much stronger sexual selection on migrants under Model I (sexual selection on both sexes) than Model II, despite AM being weaker in the former case.

3.3 Effect of viability and sexual selection on neutral gene flow

In the previous section, we have shown that approximations based on m_e accurately predict long-term adaptive divergence between the mainland and island under AM and sexual selection. In essence, divergent selection at very many loci (whether due to viability or sexual selection) results in a genome-wide reduction in gene flow, which further strengthens divergence across the genome. As discussed above, this genome-wide reduction can be quantified via $g = m_e/m$, the gene flow factor at an unlinked locus, which may be viewed as a measure of RI between hybridising populations (see Westram et al. (2022)). Moreover, g is also related to commonly used measures of neutral differentiation such as F_{ST}, which depends on both the number of migrants, Nm, and their reproductive value (which is equal to g). More concretely, at equilibrium, neutral F_{ST} between the mainland and island is: $F_{ST} \sim 1/(1 + 2Nm_e) = 1/(1 + 2Nm\,g)$ on average.

We now quantify the extent to which viability and sexual selection reduce gene flow across the genome in a scenario with high migration ($2Nm = 49$). Note that in the absence of divergent selection, such high levels of gene flow would erase most neutral differences between mainland and island, resulting in $F_{ST} = 0.02$. However, strong enough selection can maintain adaptive divergence despite high migration, thus suppressing gene flow and increasing genome-wide F_{ST}. Figure 4 illustrates this by plotting F_{ST} as a function of $\beta\Delta_0$ and $\gamma\Delta_0^2$ for $m/s = 1$, under the two models of AM. These figures are generated by solving for the allele frequency divergence per locus Y (using eq. 4) for each combination of $\beta\Delta_0$ and $\gamma\Delta_0^2$, then using this to compute the gene flow factor g (via eq. 3), and finally computing $F_{ST} = 1/(1 + 2Nm\,g)$.

In each plot, the region to the left of the solid curve shows parameter combinations for which adaptive divergence is lost ($Y = 0$) at $m/s = 1$, so that $m_e = m$ and $F_{ST} = \frac{1}{1+2Nm} = 0.02$. As before, increasing $\beta\Delta_0$ and $\gamma\Delta_0^2$ (which correspond to stronger viability and sexual selection respectively) allow non-zero adaptive divergence to be maintained, thereby reducing m_e and increasing F_{ST}. The dashed curve in each plot depicts parameter combinations for which mean allele frequency divergence is $Y = 0.95$, which corresponds to a gene flow factor of $g \approx \frac{1-Y}{m/s} = 0.05$ and an F_{ST} value of $\frac{1}{1+2Nm\,g} \approx 0.29$ under either model of AM. To the right of this curve, i.e, with even stronger viability and/or sexual selection, there is a further increase in F_{ST}, which is now *not* due to an increase in adaptive divergence (which is close to its maximum possible value). Instead higher levels of F_{ST} in this parameter regime reflect

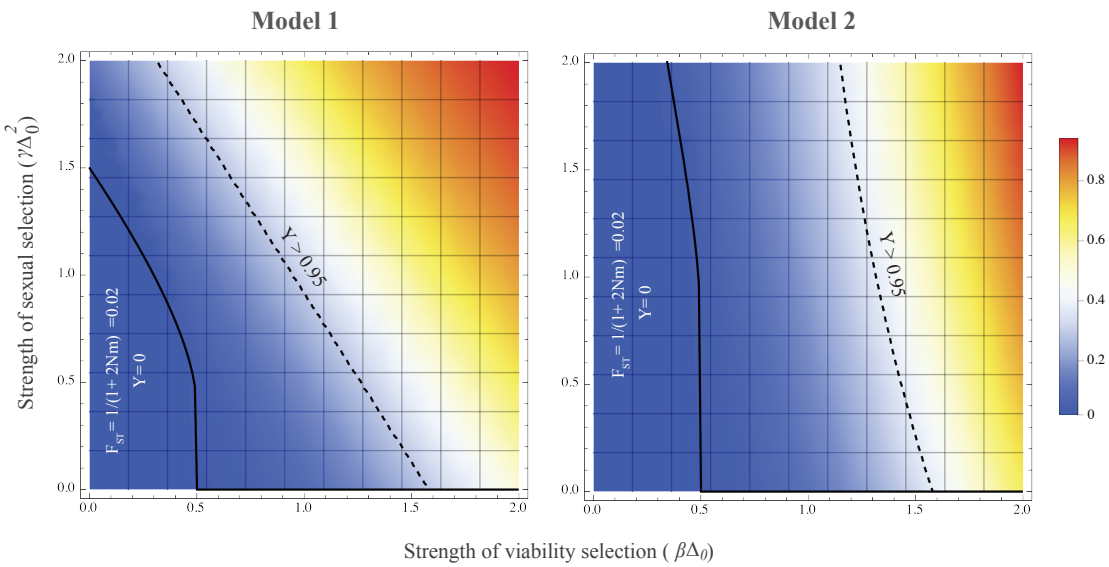

Figure 4: Heatmaps depicting F_{ST} at an unlinked neutral locus for different values of $\beta\Delta_0$ and $\gamma\Delta_0^2$ (which parameterize viability and sexual selection respectively) for Model I (sexual selection on both sexes) and Model II (sexual selection only on males) of assortative mating, assuming $m/s = 1$ and $2Nm = 49$. The color bar on the right shows F_{ST} values ranging from $F_{ST} = 0$ to 1. Parameter combinations to the left of the solid line are those for which there is no allele frequency divergence (i.e., $Y = 0$) so that $F_{ST} = 1/(1 + 2Nm) = 0.02$; parameters to the right of this line allow for divergence, so that $F_{ST} = 1/(1 + 2Nm\, g\,[\Delta_0 Y])$, where Y is obtained by solving eq. 4. Parameters to the right of the dashed line are those for which $Y > 0.95$: in this regime, stronger viability or sexual selection increases the genomewide barrier effect and F_{ST} without significantly affecting divergence at trait loci (which is close to its maximum possible value).

lower viability and mating success of (nearly maximally diverged) migrants, which further reduces neutral gene exchange between mainland and island.

We can also ask: do changes in the strength of viability and sexual selection multiplicatively (i.e., independently) affect gene flow (as measured by g), or are their combined effects stronger (or weaker) than expected from their individual effects? More concretely, consider a scenario with moderately strong viability and sexual selection ($\beta\Delta_0 = \gamma\Delta_0^2 = 0.5$), $m/s = 1$, so that mean allele frequency divergence is $Y = 0.36$ (from eq. 4) and $g = 0.64$ (from eq. 3) under Model I. A 20% increase in *either* viability or sexual selection, i.e., increasing $\beta\Delta_0$ to 0.6 while keeping $\gamma\Delta_0^2$ fixed at 0.5 (or vice versa) causes g to fall to 55% (or 74%) of its original value. Now, if viability and sexual selection were to independently reduce gene flow, then we would expect g to decrease by a factor of $0.55 \times 0.74 = 0.39$, i.e., to 39% of its original value in response to a 20% increase in *both* components of selection ($\beta\Delta_0 = \gamma\Delta_0^2 = 0.6$). Instead, g decreases to 43% of its original value. Thus, in this example, the two kinds of reproductive barriers (lower survival and lower mating success of immigrants and their descendants) cause a weaker reduction in gene flow than if they were to act independently.

Such non-independent effects arise because the gene flow factor is a product of two terms– one due to viability and the other due to sexual selection (see eq. 3), both of which depend on allele frequency divergence Y, which in turn is influenced by both components of selection. Thus, the strongest non-multiplicative effects emerge in parameter regimes where viability and sexual selection are strong enough to significantly influence divergence levels but not so strong as to drive divergence to its maximum possible value. In the latter scenario, i.e., as Y approaches 1 (to the right of the dashed

curves in figure 4), the effect of viability and sexual selection again becomes multiplicative.

4 Discussion

While it has long been recognised that sexual selection can maintain inter-specific differences, questions remain as to its importance relative to other processes during speciation (Safran et al. (2013)). Sexual selection varies significantly across species and even across environments within a species or over time (Schumer et al. (2017); Magurran and Ramnarine (2004); Rosenthal (2013)), making it difficult to identify broad patterns. Moreover, sexual selection typically acts together with other kinds of selection against heterospecifics and hybrids (e.g., due to ecological mismatch or expression of genetic incompatibilities), making it difficult to disentangle the contributions of different processes to reproductive isolation. This only underscores the broader challenges in quantifying RI– an issue that has attracted considerable attention recently (Westram et al. (2022)).

Here, we examine some of these issues in the context of a polygenic 'magic' trait that is under divergent ecological selection across populations, while also mediating AM and sexual selection. In such a scenario, long-term divergence depends crucially on LD between trait loci. Such LD is, in turn, a consequence of both (natural and sexual) selection and assortment, making it challenging to model. Previous analyses have relied on one of two approaches- either focusing on how trait variances are affected by selection and assortment while largely ignoring genetic details (e.g., Doebeli (1996); Kondrashov and Shpak (1998)), or explicitly modeling LD between sets of loci (Kirkpatrick (2000); Barton and De Cara (2009)), but assuming LD to be weak (to ensure tractability). Here, we employ a novel approach based on effective migration rates which bridges phenotypic descriptions (involving trait means and variances) with genic descriptions (based on divergence at individual selected sites).

An advantage of our approach is that it provides an economical description of tipping points in terms of just allele frequency differences between populations rather than a plethora of LD measures. The key idea behind our approach is that LD between introgressing alleles breaks down rapidly if these are unlinked or loosely linked (a standard assumption in quantitative genetics), allowing us to represent its effects via an effective migration rate m_e (or alternatively, the gene flow factor $g = m_e/m$) over the longer timescales over which allele frequencies change. The gene flow factor, being the average reproductive value of migrants, is a composite measure of various processes- prezygotic and postzygotic- that affect the fitness of migrants and their descendants. In particular, it captures both the prezygotic effects of AM– which causes fewer F_1 hybrids to be produced (due to sexual selection against migrants) as well as its postzygotic effects– as reflected in the lower mating success of F_1s and later generation hybrids (again, due to sexual selection).

As we show here, theory based on m_e accurately predicts divergence across a range of parameters, including when LD is strong enough to generate so-called tipping points (Nosil et al. (2017)), but also when genetic drift is strong enough to wash away these tipping points (see SI, **D**). Moreover, it captures the consequences of sexual selection on one vs. both sexes (Model I vs. II). These models exhibit qualitatively similar behaviours– tipping points and/or shifted swamping thresholds, which, however, require much higher levels of assortment under Model II (compared to Model I) due to net (sex-averaged) sexual selection being weaker in this case. This again highlights the fact that it is sexual selection (rather than AM by itself) that influences barriers to gene flow in our model.

We also show that to good approximation, g is a product of two terms- one reflecting viability selection and the other sexual selection on migrants and their descendants (see eq. 3). This is in line with previous work that quantifies sexual and natural selection via probabilities of conspecific vs. heterospecific mating, or fitness of offspring from different types of mating, and then assumes that the two components of selection multiplicatively determine the total barrier to gene flow (Ramsey et al. (2003); Sobel and Chen (2014)). Note however, that in our model, the two components (due to viability and sexual selection) both depend on the existing level of divergence Δ, and thus are not independent. For instance, stronger AM and sexual selection will typically increase divergence, which will also reduce the (relative) viability of migrants and their descendants, thus further strengthening the barrier effect of viability selection, even when there is no change in the strength of viability selection. This suggests that measurements of RI are highly context and state dependent, making it difficult to extrapolate from lab measurements or between replicate hybrid zones.

Nevertheless, in theory, eq. 3 provides an unambiguous way of quantifying the contributions of natural and sexual selection to RI. In practice, this requires estimates of fitness components of individuals with different levels of migrant ancestry, which are only available in a handful of well-studied populations with multi-generational pedigrees (Bérénos et al. (2014); Chen et al. (2019)). However, such populations are necessarily spatially limited and show very little divergence or RI. More typically, we might have indirect fitness estimates for hybrids (and less commonly, for backcrosses and recombinants) from hybrid zones between divergent ecotypes. It is an open question whether combining these with experimental approaches, e.g., reciprocal transplant or mate choice experiments (that individually measure viability or mating success) might provide reliable estimates of fitness components in natural populations, thus allowing us to disentangle the contributions of natural and sexual selection to RI.

Our work also bears upon the interplay between assortative mating and sexual selection during divergence with gene flow. The distinction between the two has been highlighted in earlier work (Kirkpatrick and Nuismer (2004); Polechová and Barton (2005); Servedio and Bürger (2014)), which points out that sexual selection can reduce the mating success of outlier individuals, thus constraining genetic variance and reducing the potential for divergence (Kirkpatrick and Nuismer (2004)). By contrast, pure assortment without sexual selection always increases variance, thus facilitating divergence.

The situation is somewhat different following secondary contact between diverged populations: now, sexual selection against migrants reduces introgression of deleterious alleles. This effect is only strengthened by assortative mating (based on ancestry) which increases LD between introgressing alleles, causing them to be eliminated more effectively. In recent work, Muralidhar et al. (2022) provide a rough estimate of the effect of AM following a pulse of admixture, where an individual's viability declines in proportion to the amount of introgressed material it carries. Their argument is as follows: the change in introgressed allele frequency per generation (due to viability and/or sexual selection) is proportional to the variance of introgressed ancestry across individuals. In an assortatively mating population with correlation coefficient ρ between mates (where ρ is a measure of the strength of AM), the variance in introgressed ancestry declines by a factor $\sim (1+\rho)/2$ per generation, *slower* than it would under random mating, allowing for more efficient selective purging of introgressing alleles. It then follows that AM *by itself* (as a process that is distinct from sexual selection) increases net purging by a factor of $1/(1-\rho)$ (Muralidhar et al. (2022)). This argument implicitly assumes that sexual selection is weak and does not cancel out the variance-increasing effect of AM; this assumption

would, for instance, break down under our Model I (see also Kirkpatrick and Nuismer (2004))

The conclusions of Muralidhar et al. (2022) differ from ours– in particular, we show that divergence is shaped essentially by sexual selection on various subgroups (i.e., F_1s, BC_1s etc.), with little to no effect of AM (e.g., on within-subgroup variances). What might explain these differing conclusions? Muralidhar et al. (2022) analyse the consequences of a single admixture event that replaces a sizable fraction of the native population by migrants from a diverged population. In this case, individuals with recent migrant ancestry are sufficiently common to mate assortatively with one another– this causes LD between introgressing alleles to persist longer, thus allowing selection to eliminate them efficiently. By contrast, we consider ongoing migration at rate m, and focus on the long-term equilibrium between selection and migration. In this case, long-term adaptive divergence requires m_e to be lower than the typical per locus selective effect s, so that m is at most a few times larger than s. Thus, migration is sufficiently low that individuals with recent migrant ancestry mate predominantly with residents rather than with each other. Then, introgression depends essentially on the mating success of (or alternatively, sexual selection on) various hybrids (F_1, BC_1 etc.) relative to residents. We emphasise that assuming low migration rates ($m \ll 1$) does not imply that neutral divergence between populations is high, since the latter depends on the number of migrants Nm per generation, which may be large. Nor does it imply high levels of adaptive divergence (see, e.g., figure 2); indeed, this can be arbitrarily low, depending on m/s, and the total strength of viability and sexual selection.

The preceding discussion highlights how the role of AM (as distinct from sexual selection) depends on the context of divergence. Broadly speaking, assortment is likely to be important if diverged subgroups are present together at comparable frequencies, as in the centre of a hybrid zone. Our theoretical approximations would not apply in this case, since F_1s would be sufficiently common as to mate amongst themselves and generate F_2 recombinants. Moreover, many hybrid zones are characterised by a bimodal distribution of parental ancestries or equivalently, of hybrid indices (Jiggins and Mallet (2000); Bridle and Butlin (2002); Schumer et al. (2017)). The effects of AM in such a scenario are not amenable to simple descriptions which assume trait variance to increase by a constant factor $(1+\rho)/2$ (Muralidhar et al. (2022)), since the correlation ρ between mating pairs in an assortatively mating population also depends on the distribution of phenotypic values and can be very different for unimodal vs. bimodal populations. The consequences of non-random mating in hybrid zones thus remain an important direction for future work (see also Irwin (2020)), which will require us to expand existing theoretical descriptions.

What are some limitations of our study? We have considered a model of AM via self-referencing, wherein individuals prefer mates with similar phenotypes. However, sexual selection may be less effective in suppressing gene flow if female preference and the male traits on which it acts are encoded by different genes, requiring these to be in LD. More generally, little is known about the prevalence of different mechanisms of AM and the extent to which these generate sexual selection in different ecological contexts (Safran et al. (2013); Kopp et al. (2018)), making it important to examine the robustness of theoretical findings to alternative assumptions about assortment mechanisms.

We also consider an extreme form of divergent ecological selection, where alternative alleles are favoured in the two populations at every locus. In this case, sexual selection is predominantly directional, reducing the mating success of individuals in proportion to their migrant ancestry. Al-

ternatively, the magic trait could be under stabilizing selection towards different optima in the two populations. While selection on migrants would still be directional if trait optima are far apart, adaptive divergence can now be maintained at higher migration rates (Sachdeva and Barton (2017)). This results in a more complicated interplay between assortment and sexual selection, since mating between individuals with recent migrant ancestry becomes more common. Importantly, in this case, we need to go beyond theory that treats introgressing alleles as cascading down successively fitter genetic backgrounds, and explicitly account for F2s etc.

Finally, we neglect linkage between trait loci, which generally strengthens barriers to gene flow, at least under ecological divergent selection (Barton and Bengtsson (1986)). However, if traits are under frequency-dependent sexual selection, then linkage between trait loci may have more subtle effects (Kirkpatrick and Nuismer, 2004), even weakening barriers to gene flow under some conditions (Aubier et al., 2024). Approximations based on effective migration rates should extend to loose linkage between trait loci (Zwaenepoel et al., 2024). However, the effects of tight linkage (e.g., between alleles clustered within supergenes or inversions) are more obscure.

We focus here on the interplay between natural and sexual selection at a *fixed* level of assortment and following secondary contact between diverged populations. Other questions remain as to how assortment evolves in the first place– e.g., when mating with conspecifics might be favoured by selection, thus reinforcing divergence (Servedio and Noor (2003)). While our study does not address these questions, it does suggest that analysing the evolution of assortment in terms of the reproductive value (or effective rate of exchange) of modifiers of mate choice could be a fruitful direction for future work. More generally, effective migration rates link a 'phenotypic' view of RI (based on fitness components of hybrids) with a 'genic' view (that conceptualises RI as a process that reduces gene flow and increases divergence across the genome), making them a powerful way of understanding RI.

Author contributions: P.S. and H.S. designed the study, did the mathematical analysis and simulations, and wrote the manuscript. The authors declare no conflict of interest.

Acknowledgements : We thank Nick Barton for useful comments on the manuscript. This research was supported by the Scientific Service Units (SSU) of IST Austria through resources provided by Scientific Computing (SciComp).

Data availability: Mathematica notebooks for numerical analysis and simulations are available at `https://doi.org/10.15479/AT:ISTA:17344`.

Conflict of Interest: The authors declare no conflict of interest.

References

Aubier, T. G., Kopp, M., Linn, I. J., Puebla, O., Rafajlović, M., and Servedio, M. R. (2024). Negative coupling: The coincidence of premating isolating barriers can reduce reproductive isolation. *Cold Spring Harbor Perspectives in Biology*, 16(10).

Barton, N. (1992). On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution*, pages 551–557.

Barton, N. and Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, 57(3):357–376.

Barton, N. H. (1983). Multilocus clines. *Evolution*, 37(3):454–471.

Barton, N. H. and De Cara, M. A. R. (2009). The evolution of strong reproductive isolation. *Evolution*, 63(5):1171–1190.

Barton, N. H. and Etheridge, A. M. (2011). The relation between reproductive value and genetic contribution. *Genetics*, 188(4):953–973.

Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73.

Bengtsson, B. (1985). The flow of genes through a genetic barrier. In *Evolution: Essays in honour of John Maynard Smith*, pages 31–42. Cambridge University Press, Cambridge.

Bridle, J. R. and Butlin, R. K. (2002). Mating signal variation and bimodality in a mosaic hybrid zone between *Chorthippus* grasshopper species. *Evolution*, 56(6):1184–1198.

Butlin, R. K. and Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *American Naturalist*, 191(2):155–172.

Bérénos, C., Ellis, P. A., Pilkington, J. G., and Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology*, 23(14):3434–3451.

Chen, N., Juric, I., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., Schoech, S. J., Clark, A. G., and Coop, G. (2019). Allele frequency dynamics in a pedigreed natural population. *Proceedings of the National Academy of Sciences*, 116(6):2158–2164.

Coughlan, J. M. and Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1806):20190533.

Coyne, J. A. and Allen Orr, H. (2004). *Speciation*. Sunderland, Mass.: Sinauer Associates.

Crow, J. F. and Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly*, 15(2):85–97.

Doebeli, M. (1996). A quantitative genetic competition model for sympatric speciation. *Journal of Evolutionary Biology*, 9(6):893–909.

Elmer, K. R., Lehtonen, T. K., and Meyer, A. (2009). Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution*, 63(10):2750–2757.

Felsenstein, J. (1981). Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution*, 35(1):124–138.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Proc. Roy. Soc. Edinburgh*, 52:399–433.

Fisher, R. A. (1999). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Fraïsse, C. and Sachdeva, H. (2021). The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes *vs* autosomes. *Genetics*, 217(2):iyaa025.

Irwin, D. E. (2020). Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. *The American Naturalist*, 195(6):E150–E167.

Jiang, Y., Bolnick, D. I., and Kirkpatrick, M. (2013). Assortative mating in animals. *The American Naturalist*, 181(6):E125–138.

Jiggins, C. D. and Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, 15(6):250–255.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., and Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, 411(6835):302–305.

Johannesson, K., Havenhand, J. N., Jonsson, P. R., Lindegarth, M., Sundin, A., and Hollander, J. (2008). Male discrimination of female mucous trails permits assortative mating in a marine snail species. *Evolution*, 62(12):3178–3184.

Kirkpatrick, M. (2000). Reinforcement and divergence under assortative mating. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1453):1649–1655.

Kirkpatrick, M. and Nuismer, S. L. (2004). Sexual selection can constrain sympatric speciation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1540):687–693.

Kirkpatrick, M. and Ravigné, V. (2002). Speciation by Natural and Sexual Selection: Models and Experiments. *The American Naturalist*, 159(S3):S22–S35.

Kobayashi, Y., Hammerstein, P., and Telschow, A. (2008). The neutral effective migration rate in a mainland-island context. *Theoretical Population Biology*, 74(1):84–92.

Kondrashov, A. S. and Shpak, M. (1998). On the origin of species by means of assortative mating. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1412):2273–2278.

Kopp, M., Servedio, M. R., Mendelson, T. C., Safran, R. J., Rodríguez, R. L., Hauber, M. E., Scordato, E. C., Symes, L. B., Balakrishnan, C. N., Zonana, D. M., and van Doorn, G. S. (2018). Mechanisms of assortative mating in speciation with gene flow: Connecting theory and empirical research. *The American Naturalist*, 191(1):1–20.

Lewontin, R., Kirk, D., and Crow, J. (1968). Selective mating, assortative mating, and inbreeding: definitions and implications. *Eugenics Quarterly*, 15(2):141–143.

Lynch, M., Walsh, B., et al. (1998). *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA.

Maan, M. E. and Seehausen, O. (2011). Ecology, sexual selection and speciation. *Ecology Letters*, 14(6):591–602.

Magurran, A. E. and Ramnarine, I. W. (2004). Learned mate recognition and reproductive isolation in guppies. *Animal Behaviour*, 67(6):1077–1082.

Merrill, R. M., Wallbank, R. W. R., Bull, V., Salazar, P. C. A., Mallet, J., Stevens, M., and Jiggins, C. D. (2012). Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):4907–4913.

Muralidhar, P., Coop, G., and Veller, C. (2022). Assortative mating enhances postzygotic barriers to gene flow via ancestry bundling. *Proceedings of the National Academy of Sciences*, 119(30):e2122179119.

Nosil, P., Feder, J. L., Flaxman, S. M., and Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*, 1(2):0001.

Perini, S., Rafajlović, M., Westram, A. M., Johannesson, K., and Butlin, R. K. (2020). Assortative mating, sexual selection, and their consequences for gene flow in *Littorina*. *Evolution*, 74(7):1482–1497.

Polechová, J. and Barton, N. H. (2005). Speciation through competition: A critical review. *Evolution*, 59(6):1194–1210.

Ramsey, J., Bradshaw Jr, H., and Schemske, D. W. (2003). Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (phrymaceae). *Evolution*, 57(7):1520–1534.

Rice, W. R. and Hostert, E. E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, 47(6):1637–1653.

Rosenthal, G. G. (2013). Individual mating decisions and hybridization. *Journal of Evolutionary Biology*, 26(2):252–255.

Rundle, H. D. and Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3):336–352.

Sachdeva, H. (2022). Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLOS Genetics*, 18(9):e1010297.

Sachdeva, H. and Barton, N. H. (2017). Divergence and evolution of assortative mating in a polygenic trait model of speciation with gene flow. *Evolution*, 71(6):1478–1493.

Safran, R. J., Scordato, E. S., Symes, L. B., Rodríguez, R. L., and Mendelson, T. C. (2013). Contributions of natural and sexual selection to the evolution of premating reproductive isolation: a research agenda. *Trends in Ecology & Evolution*, 28(11):643–650.

Schumer, M., Powell, D. L., Delclós, P. J., Squire, M., Cui, R., Andolfatto, P., and Rosenthal, G. G. (2017). Assortative mating and persistent reproductive isolation in hybrids. *Proceedings of the National Academy of Sciences*, 114(41):10936–10941.

Servedio, M. R. (2009). The role of linkage disequilibrium in the evolution of premating isolation. *Heredity*, 102(1):51–56.

Servedio, M. R. (2016). Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. *Evolutionary Applications*, 9(1):91–102.

Servedio, M. R. and Boughman, J. W. (2017). The role of sexual selection in local adaptation and speciation. *Annual Review of Ecology, Evolution, and Systematics*, 48(1):85–109.

Servedio, M. R. and Bürger, R. (2014). The counterintuitive role of sexual selection in species maintenance and speciation. *Proceedings of the National Academy of Sciences*, 111(22):8113–8118.

Servedio, M. R., Doorn, G. S. V., Kopp, M., Frame, A. M., and Nosil, P. (2011a). Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution*, 26(8):389–397.

Servedio, M. R., Doorn, G. S. V., Kopp, M., Frame, A. M., and Nosil, P. (2011b). Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution*, 26(8):389–397.

Servedio, M. R. and Kopp, M. (2012). Sexual selection and magic traits in speciation with gene flow. *Current Zoology*, 58(3):510–516.

Servedio, M. R. and Noor, M. A. (2003). The Role of Reinforcement in Speciation: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):339–364.

Smadja, C. M. and Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, 20(24):5123–5140.

Smith, J. M. (1966). Sympatric Speciation. *The American Naturalist*, 100(916):637–650.

Sobel, J. M. and Chen, G. F. (2014). Unification of Methods for Estimating the Strength of Reproductive Isolation. *Evolution*, 68(5):1511–1522.

Stelkens, R. B. and Seehausen, O. (2009). Phenotypic divergence but not genetic distance predicts assortative mating among species of a cichlid fish radiation. *Journal of Evolutionary Biology*, 22(8):1679–1694.

Vines, T. H. and Schluter, D. (2006). Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. *Proceedings of the Royal Society B: Biological Sciences*, 273(1589):911–916.

Westram, A. M., Stankowski, S., Surendranadh, P., and Barton, N. (2022). What is reproductive isolation? *Journal of Evolutionary Biology*, 35(9):1143–1164.

Wright, S. (1921). Systems of Mating. III. Assortative mating based on somatic resemblance. *Genetics*, 6(2):144–161.

Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences*, 23(6):307–320.

Zwaenepoel, A., Sachdeva, H., and Fraïsse, C. (2024). The genetic architecture of polygenic local adaptation and its role in shaping barriers to gene flow. *Genetics*, page iyae140.

SUPPLEMENTARY INFORMATION for
Effect of assortative mating and sexual selection on polygenic barriers to gene flow

A. Derivation of gene flow factor

Let h_i represent the fraction of individuals in the population with a migrant ancestor precisely i generations ago. Thus, h_1 represents the fraction of $F1$ individuals and h_i the fraction of $(i{-}1)^{th}$ generation backcrosses (denoted as BC_{i-1}). As specified in the main text, an individual with *no* migrant ancestor in the last n generations is designated as a 'resident', where n is an arbitrary threshold. Thus, by definition, the fraction of residents in the population is $h_r{=}1{-}\sum_{i=1}^{n} h_i$. We further assume that trait values are normally distributed *within* any population subgroup: thus the distribution $P_r(z)$ of trait values among residents is normal with mean $\overline{z}_r$ and variance V_r; similarly, the distribution $P_i(z)$ amongst i^{th} generation descendants of migrants is normal with mean z_i and variance V_i (for any i). The trait value distribution across the entire population is thus the sum of $n{+}1$ normal distributions and is given by $P(z){=}h_rP_r(z){+}\sum_{i=1}^{n} h_iP_i(z)$. For simplicity, we will assume that the distribution of trait values amongst migrants is also normal with mean $\overline{z}_0$ and variance V_0; however, this assumption is not crucial. Finally, we will assume that $m{\ll}1$ and perform various computations to first order in m. This is justified on the grounds that non-zero adaptive divergence between mainland and island requires m to be comparable to the typical *per locus* selection coefficient s, where $s{\ll}1$ for a polygenic trait influenced by many small-effect loci.

We now consider how the proportions h_i and the distributions $P_i(z)$ associated with different population subgroups change within a single generation due to migration, selection and assortative mating. *Migration* from the mainland results in a fraction m of each subgroup being replaced by migrants. Thus, following migration: $h_0^{'}{=}m$, $h_r^{'}{=}(1{-}m)h_r$, and $h_i^{'}{=}(1{-}m)h_i$ for $i{=}1,\ldots n$. Note that migration does not change the trait distributions of existing sub-groups, so that: $P_r^{'}(z){=}P_r(z)$ and $P_i^{'}(z){=}P_i(z)$.

Migration is followed by *viability selection* on the trait, wherein an individual with trait value z contributes to the pool of reproducing adults in proportion to $e^{-\beta z}/\overline{W}_{vs}$. Here, $\overline{W}_{vs}$ is the viability component of mean fitness and is given by: $\overline{W}_{vs}{=}\int e^{-\beta y}\left[h_r^{'}P_r^{'}(y){+}\sum_{i=0}^{n} h_i^{'}P_i^{'}(y)\right] dy$. Thus, following viability selection, the proportions and trait value distributions of the different subgroups are:

$$h_i^{''} = h_i^{'}\,\frac{\int e^{-\beta z}P_i^{'}(z)\,dz}{\overline{W}_{vs}} \qquad h_r^{''} = h_r^{'}\,\frac{\int e^{-\beta z}P_r^{'}(z)\,dz}{\overline{W}_{vs}} \tag{S1a}$$

$$P_i^{''}(z){=}\frac{e^{-\beta z}P_i^{'}(z)}{\int e^{-\beta y}P_i^{'}(y)\,dy} \qquad P_r^{''}(z){=}\frac{e^{-\beta z}P_r^{'}(z)}{\int e^{-\beta y}P_r^{'}(y)\,dy} \tag{S1b}$$

Equation S1b specifies the distribution of trait values among different population subgroups following viability selection. These distributions are still normal, but with a shifted mean $\overline{z}_i^{''}=\overline{z}_i-\beta V_i$ for sub-group i (and $\overline{z}_r^{''}=\overline{z}_r-\beta V_r$ for residents). The variances within sub-groups remain unchanged ($V_i^{''}=V_i$ and $V_r^{''}=V_r$)– this is because directional selection (fitness proportional to $e^{-\beta z}$) acting on a (sub-)population with *normally* distributed trait values only shifts the mean without affecting the trait variance. Thus, the trait value distribution across the full population after viability selection is the sum of $n+2$ normal distributions: $P^{''}(z)=h_r^{''}P_r^{''}(z)+\sum_{i=0}^{n} h_i^{''}P_i^{''}(z)$. Note that this sum now includes the group '0', i.e., individuals who have immigrated at the start of the generation.

In the following, we will show that βV is $\mathcal{O}(m)$; thus, the various βV_i must also be $\mathcal{O}(m)$ (assuming V_i to be comparable to V). Further, as we argue below, the various h_i, i.e., the proportion of $F1$s, BC_1 and so on, are also $\mathcal{O}(m)$. Thus, to first order in m, we can ignore the shifts in trait means (due to viability selection) within subgroups, i.e., within migrants, $F1$s and BC_i, since the contribution of these shifts to the shift in the overall trait mean of the full population is $\mathcal{O}(m^2)$. For the same reason, we cannot ignore the shift in trait mean among residents, since $h_r^{''}=1-\sum_{i=0}^{n} h_i^{''}$ is $\mathcal{O}(1)$.

Finally, consider the effects of *assortative mating* and *sexual selection* on population sub-groups. As described in the main text, the mating function $M(x,y)$ specifies the probability of mating between a male and a female with trait values x and y respectively. Since an individual with a given trait value is equally likely to be male or female, the probability of mating between two randomly chosen individuals with trait values x and y is $[M(x,y)+M(y,x)]/2$. Under the weak migration ($m\ll 1$) assumption, migrants mate predominantly with residents to produce $F1$s, which mate with residents to produce BC_1 and so on. Further, under our definition of resident, mating between BC_n and residents also produces residents. Thus, after mating, we have:

$$h_r^{'''}=\int\int \frac{M(x,y)+M(y,x)}{2}\left[\left(h_r^{''}\right)^2 P_r^{''}(x)P_r^{''}(y)+2h_r^{''}h_n^{''}P_r^{''}(x)P_n^{''}(y)\right]dx\,dy \tag{S2a}$$

$$h_i^{'''}=\int\int \frac{M(x,y)+M(y,x)}{2}\left[2h_r^{''}h_{i-1}^{''}P_r^{''}(x)P_{i-1}^{''}(y)\right]dx\,dy \qquad i=1,\ldots,n \tag{S2b}$$

Finally, the trait value distribution among the various population sub-groups are given by:

$$P_r^{'''}(z)=\int\int \frac{M(x,y)+M(y,x)}{2}\left[\frac{\left(h_r^{''}\right)^2}{h_r^{'''}}P_r^{''}(x)P_r^{''}(y)\frac{e^{-\frac{\left(z-\frac{x+y}{2}\right)^2}{2V_{r,r}}}}{\sqrt{2\pi V_{r,r}}}+\frac{2h_r^{''}h_n^{''}}{h_r^{'''}}P_r^{''}(x)P_n^{''}(y)\frac{e^{-\frac{\left(z-\frac{x+y}{2}\right)^2}{2V_{r,n}}}}{\sqrt{2\pi V_{r,n}}}\right]dx\,dy \tag{S3a}$$

$$P_i^{'''}(z)=\int\int \frac{M(x,y)+M(y,x)}{2}\left[\frac{2h_r^{''}h_{i-1}^{''}}{h_i^{'''}}P_r^{''}(x)P_{i-1}^{''}(y)\frac{e^{-\frac{\left(z-\frac{x+y}{2}\right)^2}{2V_{r,i-1}}}}{\sqrt{2\pi V_{r,i-1}}}\right]dx\,dy \qquad i=1,\ldots,n \tag{S3b}$$

Equation S3 assumes that the trait is sufficiently polygenic that the offspring of any two individuals are normally distributed about the midparent value with a within-family variance that depends on the relatedness between parents, as in the standard infinitesimal model (Barton et al. (2017)). As discussed in the main text, we further assume that the within-family variance only depends on the subgroups to which the two parents belong. Thus, for example, the within-family variance of the offspring of any resident and BC_i pair is assumed to be $V_{r,i+1}$.

At *equilibrium*, the proportions of various subgroups and their trait value distributions at the end of each generation must be the same as at the beginning, so that $h_i^{'''}=h_i$, $h_r^{'''}=h_r$, and $P_i^{'''}(z)=P_i(z)$, $P_r^{'''}(z)=P_r(z)$. These conditions, together with eqs. S1-S3, give us a set of equations for $P_i(z)$ and $h_i(z)$. These equations are somewhat complex, but simplify considerably if we assume that $m\ll 1$, and only calculate the various h_i to lowest order in m. In particular, it follows that h_i are $\mathcal{O}(m)$ for all i. One can see this from eq. (S2b), where $h_1=h_1^{'''}$ is proportional to h_0'' which is $\mathcal{O}(m)$ (being the proportion of migrants in the population post viability selection), so that h_1 is also $\mathcal{O}(m)$. Similarly, $h_2=h_2^{'''}$ (being proportional to h_1'') is also $\mathcal{O}(m)$, and so on. The equilibrium conditions $h_i^{'''}=h_i$, $h_r^{'''}=h_r$, $P_i^{'''}(z)=P_i(z)$, and $P_r^{'''}(z)=P_r(z)$ also lead to eq. 2a of the main text, which recursively relates $P_i(z)$ at equilibrium to $P_{i-1}(z)$. We can then use this to derive expressions for $\{\,\overline{z}_i,\ V_i\,\}$, the equilibrium means and variances of trait values in each sub-group, to lowest order in m. In the following, we present these derivations separately for Models I and II.

Under **Model I** of assortative mating, we have:

$$M(x,y)=M(y,x)=\frac{e^{-\gamma(x-y)^2}}{\int\int P''(\tilde{x})P''(\tilde{y})e^{-\gamma(\tilde{x}-\tilde{y})^2}\,d\tilde{y}\,d\tilde{x}}$$

where $P''(x)$ is the distribution of trait values in the population just after viability selection, which is simply: $P''(x)=(1-\sum_{i=0}^{n} h_i'')P_r''(x)+\sum_{i=0}^{n} h_i''P_i''(x)$. Substituting into the mating function $M(x,y)$ and using eq. S2a, it follows that to lowest order in m, we have: $h_i=h_i^{'''}=2h_{i-1}''\frac{\int\int e^{-\gamma(\tilde{x}-\tilde{y})^2}P_r''(\tilde{x})P_{i-1}''(\tilde{y})\,d\tilde{y}\,d\tilde{x}}{\int\int e^{-\gamma(\tilde{x}-\tilde{y})^2}P_r''(\tilde{x})P_r''(\tilde{y})\,d\tilde{y}\,d\tilde{x}}$ and $h_r=1-\sum_{i=1}^{n} h_i$. Using eq. (S1) for $P_i''(z)$ and h_i'' finally gives the following expressions for the equilibrium proportions of the various subgroups:

$$h_0=m;\qquad h_i=2h_{i-1}\overline{W}_{i-1}\quad(\text{for}\quad i=1,\ldots,n);\qquad h_r=1-\sum_{i=1}^{n}h_i$$

$$\text{where}\quad \overline{W}_i=\frac{\int\int e^{-\gamma(\tilde{x}-\tilde{y})^2}\,e^{-\beta(\tilde{x}+\tilde{y})}P_r(\tilde{x})P_i(\tilde{y})\,d\tilde{y}\,d\tilde{x}}{\int\int e^{-\gamma(\tilde{x}-\tilde{y})^2}\,e^{-\beta(\tilde{x}+\tilde{y})}P_r(\tilde{x})P_r(\tilde{y})\,d\tilde{y}\,d\tilde{x}} \tag{S4}$$

$$=\sqrt{\frac{1+4\gamma V_r}{1+2\gamma(V_r+V_i)}}\exp\left[-\frac{\beta(1+4\gamma V_r)(\overline{z}_i-\overline{z}_r)+\gamma(\overline{z}_i-\overline{z}_r)^2-\frac{\beta^2}{2}(1+4\gamma V_r)(V_i-V_r)}{1+2\gamma(V_r+V_i)}\right]$$

It follows from eq. (S4) above that $h_1=2m\overline{W}_0$ (since $h_0=m$), $h_2=4m\overline{W}_0\overline{W}_1$, and (more generally) that: $h_i=2^i m\prod_{k=1}^{i-1}\overline{W}_k$. This establishes that the various h_i are all $\mathcal{O}(m)$.

Equation (S4) simply states that the equilibrium proportion of ith generation descendants of migrants (e.g., BC_1 for $i=2$) is equal to two times the equilibrium proportions of the two parental subgroups (e.g., residents and $F1$s for BC_1s) multiplied by the mean relative fitness of the two parental subgroups. The proportion of residents is $h_r=1-\mathcal{O}(m)$ and their relative fitness is $1+\mathcal{O}(m)$, whereas the various h_i are $\mathcal{O}(m)$. Thus, to first order in m, we have: $h_i=2h_{i-1}\overline{W}_{i-1}$, as above. Here, $\overline{W}_i$ denotes the mean fitness of ith generation descendants of migrants relative to residents, and is thus the same $\overline{W}_i$ that enters the expression for the gene flow factor g (equation 1 of the main text). From the above equation, we see that $\overline{W}_i$ is influenced by the strength of (viability and sexual) selection as well as the trait value distributions within subgroups.

However, eq. (S4) by itself does not allow us to calculate g, as the means ($\overline{z}_r$ and $\{\overline{z}_i,\ i=1,\ldots,n\}$) and variances ($\overline{V}_r$ and $\{\overline{V}_i,\ i=1,\ldots,n\}$) of the sub-group trait value distributions are unknown. These

can be determined using eq. (S3). To do this, we will further assume that the threshold n in eq. (S3a) is sufficiently large that $\overline{z}_n - \overline{z}_r \ll \sqrt{V_r}$, so that $P_r(z)$ (which is strictly a mixture of normal distributions; see eq. (S3a)) can be approximated by a single normal distribution with moments equal to the weighted sum of moments of the constituent distributions. Recall that n here is an arbitrary threshold such that an individual with *no* migrant ancestor in the last n generations is designated as a 'resident'.

Also note that we are interested in deriving expressions for $\overline{z}_r$, V_r and $\{\overline{z}_i, V_i\}$) that are correct to first order in migration rate m. Since the effective migration rate m_e is simply m multiplied by the gene flow factor g, this means that we need to derive expressions for g or alternatively for $\{\overline{z}_i, V_i\}$) (which determine $\{\overline{W}_i\}$, which determine g) that are correct to *zeroth* order in m. To do this, it is useful to first enumerate all the quantities that are $\mathcal{O}(m)$. As argued above, this includes all the equilibrium proportions h_i. Further, βV is also $\mathcal{O}(m)$: this will be shown for the special case $\gamma V \ll 1$ below, but follows more generally from the fact that under migration-selection balance, the change in trait mean due to selection (which is equal to βV) must be balanced by the change due to migration (which is proportional to m). Further, since $V_i \sim V$ (see below), it follows that all βV_i are also $\mathcal{O}(m)$. Thus, assuming that $\{h_i\}$, $\{\beta V_i\}$ and βV are all $\mathcal{O}(m)$, and using eq. (S3), we obtain the following equations for $\overline{z}_r$, V_r, and $\{\overline{z}_i, V_i\}$ (to lowest order in m):

$$\beta V_r = 2h_n \frac{(1+4\gamma V_r)}{1+2\gamma(V_r+V_n)} \frac{\overline{z}_n - \overline{z}_r}{2} \qquad V_r = 2V_{r,r} \tag{S5a}$$

$$\overline{z}_i - \overline{z}_r = \frac{1+4\gamma V_r}{1+2\gamma(V_r+V_{i-1})} \frac{\overline{z}_{i-1} - \overline{z}_r}{2} \qquad V_i = \frac{V_r + V_{i-1} + 8\gamma V_r V_{i-1}}{4(1+2\gamma(V_r+V_{i-1}))} + V_{r,i-1} \tag{S5b}$$

$$\overline{W}_i = \sqrt{\frac{1+4\gamma V_r}{1+2\gamma(V_r+V_i)}} \exp\left[-\frac{\beta(1+4\gamma V_r)(\overline{z}_i - \overline{z}_r) + \gamma(\overline{z}_i - \overline{z}_r)^2}{1+2\gamma(V_r+V_i)}\right] \tag{S5c}$$

Equation (S5a) can be derived by integrating eq. (S3a) to compute the various moments of $P_r'''(z)$, then using the equilibrium condition $\overline{z}_r''' = \overline{z}_r$ and $V_r''' = V$, and finally retaining terms that are lowest order in m. Equation (S5b) follows similarly from eq. (S3b). Equation (S5c) is simply obtained from the expression for $\overline{W}_i$ in eq. (S4) by retaining only $\mathcal{O}(1)$, i.e., lowest (zeroth) order terms in m.

Note that eq. (S5) depends on the various segregation (within-family) variances $V_{r,r}$ and $V_{r,i}$, which are (so far) unknown, but must depend on allele frequencies $\{p_{i,j},\, j=1,\ldots,L\}$ (and $\{p_{r,j},\, j=1,\ldots,L\}$) at the various trait loci in the various population sub-groups. Here, $p_{i,j}$ (or $p_{r,j}$) represents the allele frequency at the jth locus among the ith generation descendants of migrants (or among residents). More concretely, we have: $V_{r,r} = \sum_{j=1}^{L} \frac{\alpha_j^2}{2} p_{r,j}(1-p_{r,j})$ and $V_{r,i} = \sum_{j=1}^{L} \frac{\alpha_j^2}{2} [p_{r,j}(1-p_{i,j}) + p_{i,j}(1-p_{r,j})]$. In the absence of assortment, allele frequencies among i^{th} generation descendants will simply be the average of the parental allele frequencies (i.e., of residents and $(i-1)^{th}$ generation descendants). With assortment, $\{p_{i,j}\}$ are the weighted average of parental allele frequencies, which gives (under Model I): $p_{i,j} - p_{r,j} = \frac{1+4\gamma V_r}{1+2\gamma(V_{i-1}+V_r)} \frac{p_{i-1,j} - p_{r,j}}{2}$ (this is analogous to equation (S5b) for $\overline{z}_i - \overline{z}_r$). From this, we finally have:

$$V_{r,i} = V_{r,r} + \left(\prod_{k=0}^{i-1} \frac{1+4\gamma V_r}{1+2\gamma(V_k+V_r)}\right) \sum_{j=1}^{L} \alpha_j^2 \frac{p_{0,j} - p_{r,j}}{2^{i+1}} (1-2p_{r,j}), \quad \text{where} \quad V_{r,r} = \sum_{j=1}^{L} \frac{\alpha_j^2}{2} p_{r,j}(1-p_{r,j}) \tag{S6}$$

Thus, we can express all segregation variances in terms of the (as yet) unknown allele frequencies $\{p_{r,j}\}$ in the resident pool.

The main idea behind our theoretical approach is that if trait loci are unlinked (or loosely linked), then the allele frequency at any one locus is still (approximately) predicted by the theory for a *single* locus, but with the raw migration rate m replaced by an effective migration rate m_e, which depends on allele frequencies at all the other trait loci. More concretely, in section B, we argue that at migration-selection balance (and neglecting drift), allele frequencies in the resident pool satisfy: $s_j p_{r,j}(1-p_{r,j})+m_e(p_{0,j}-p_{r,j})$, where $s_j=\alpha\beta_j$ and $p_{0,j}$ is the allele frequency among migrants at the jth locus. If the mainland is fixed for alleles that are locally deleterious on the island (which is the scenario we consider in the main paper), we have $p_{0,j}=1$, which yields $p_{r,j}=m_e/s_j$.

We now have all the ingredients necessary for self-consistently and iteratively determining allele frequencies and mean trait divergence. The iterative procedure can be implemented as follows: one starts with an arbitrary 'guess' for the allele frequencies $\{p_{r,j}\}$; this allows one to determine the segregation variance $V_{r,r}$ (of offspring produced by mating between residents) using eq. (S6), as well as the trait mean ($\overline{z}=\sum_{j=1}^{L}\alpha_j p_{r,j}$) and trait variance ($V_r=2V_{r,r}$) among residents. These can then be used to determine the segregation variance $V_{r,1}$ (of $F1$ offspring produced by resident $\times$ migrant mating) using eq. (S6), as well as $\overline{z}_1$ and V_1 using eq. (S5b). These can now be used to determine $V_{r,2}$ (using eq. (S6)), which can be used to determine $\overline{z}_2$ and V_2 (using eq. (S5b)), allowing us to sequentially determine $\{\overline{z}_i, V_i, V_{r,i}\}$ for $i=1\ldots n$. The trait means and variances can now be substituted into eq. (S5c) to obtain $\{\overline{W}_i,\ i=1\ldots n\}$, the product of which yields the gene flow factor g. This can be used to calculate the new m_e, which yields the new allele frequencies via $p_{r,j}=m_e/s_j$. This procedure can be iterated until the allele frequencies $\{p_{r,j}\}$ converge to a fixed set of values.

However, we can obtain a simple expression for g if γV_r and γV_i are $\ll 1$ (see below for a discussion of this condition). Then, equation (S5) reduces to:

$$\beta V_r=h_n(\overline{z}_n-\overline{z}_r) \qquad V_r=2V_{r,r} \tag{S7a}$$

$$\overline{z}_{i+1}-\overline{z}_r=\frac{\overline{z}_i-\overline{z}_r}{2} \qquad V_{i+1}=\frac{V_r+V_i}{4}+V_{r,i} \tag{S7b}$$

$$\overline{W}_i=\exp\left[-\beta(\overline{z}_i-\overline{z}_r)-\gamma(\overline{z}_i-\overline{z}_r)^2\right] \tag{S7c}$$

In other words, the mean phenotypic difference between each successive backcross generation and residents is half that of the previous backcross generation (eq. (S7b)), so that: $\overline{z}_i-\overline{z}_r=(\overline{z}_0-\overline{z}_r)/2^i$. Substituting into eq. (S7c) gives: $\overline{W}_i=\exp\left[-\beta\frac{(\overline{z}_0-\overline{z}_r)}{2^i}-\gamma\frac{(\overline{z}_0-\overline{z}_r)^2}{4^i}\right]$, which can be substituted into $g=\lim_{n\to\infty}\prod_{i=1}^{n}\overline{W}_i=\prod_{i=1}^{\infty}\overline{W}_i$. This gives the approximate expression for g under Model I in equation 3a of the main paper.

A second observation is that substituting $h_n=2^n m\prod_{k=1}^{n-1}\overline{W}_k$ and $\overline{z}_n-\overline{z}_r=(\overline{z}_0-\overline{z}_r)/2^n$ into eq. (S7a) yields $\beta V_r=m\left(\prod_{k=1}^{n-1}\overline{W}_k\right)(\overline{z}_0-\overline{z}_r)$. If we now take the limit $n\to\infty$, we have: $\beta V_r=m_e(\overline{z}_0-\overline{z}_r)$. This is precisely what one would obtain by multiplying our heuristic equation $s_j p_r(1-p_{r,j})=m_e(p_{0,j}-p_{r,j})$ for allele frequency dynamics by the effect size α_j (where $s_j=\beta\alpha_j$), and summing over j. This demonstrates the self-consistent nature of our solution.

In summary, our simple approximation for the gene flow factor g (eq. 3a of the main text) follows from the more general equations (S2) and (S3) by making two additional assumptions– first, that migration rates are sufficiently low that we need only consider $\mathcal{O}(m)$ terms, and second, that γV_r , $\gamma V_i \ll 1$. As discussed above, the low migration assumption implies that: (i) the probability of mating between individuals with recent immigrant ancestry (e.g., $F1 \times F1$ or $F1 \times BC_1$ mating) is negligible, thus allowing us to express g in terms of the average fitness of successive backcrosses (see also Westram et al. (2022)), and (ii) the effect of viability selection within any sub-group of ith generation descendants can be neglected, since this only has an $\mathcal{O}(m^2)$ effect on the population trait mean.

We now turn to the second assumption, namely that $\gamma V \ll 1$. To understand what this implies, note that in a population with normally distributed values (with variance V), the phenotypic correlation between mating individuals is equal to $\frac{2\gamma V}{1+2\gamma V}$ under Model I and $\frac{2\gamma V}{\sqrt{1+2\gamma V(1+2\gamma V)}}$ under Model II. Thus, the correlation between mates (which can also be used as a measure of the strength of assortative mating) is $\approx 2\gamma V$, if $\gamma V \ll 1$, under both models. The condition that $\gamma V_r \ll 1$ and $\gamma V_i \ll 1$ can thus be interpreted as assortative mating being sufficiently weak that mate choice is ineffective and phenotypic correlation between mates is insignificant for mating *within* any subgroup. However, as long as $\gamma(\overline{z}_0 - \overline{z}_r)^2$ is not too small, mating *between* subgroups (e.g., between residents and $F1$s) is still much less likely than resident$\times$resident mating.

We now outline the derivation of the gene flow factor under **Model II** of assortative mating. This follows the same logic as above but involves somewhat more cumbersome expressions. Under Model II, the mating function is:

$$M(x,y) = \frac{e^{-\gamma(x-y)^2}}{\int P''(\tilde{x}) e^{-\gamma(\tilde{x}-y)^2}\, d\tilde{x}} = \frac{e^{-\gamma(x-y)^2}}{\int P''_r(\tilde{x}) e^{-\gamma(\tilde{x}-y)^2}\, d\tilde{x}} + \mathcal{O}(m)$$

Note that in this case, $M(x,y) \neq M(y,x)$, i.e., the probability of mating between a male and female with trait values x and y respectively is different from that of a female-male pair with values x and y. As before, from eq. (S2), it follows that: $h_i = h''_{i-1}\left[1 + \int P''_r(y) \frac{\int e^{-\gamma(x-y)^2} P''_{i-1}(x)\, dx}{\int e^{-\gamma(\tilde{x}-y)^2} P''_r(\tilde{x})\, d\tilde{x}}\, dy\right]$ and $h_r = 1 - \sum_{i=1}^{n} h_i$ to lowest order in m. Using equation (S1) for $P''_i(z)$ and h''_i then yields the following expressions for the equilibrium proportions of the various subgroups:

$$h_i = 2h_{i-1}\overline{W}_{i-1} \quad (\text{for} \quad i=1,\ldots,n) \quad \text{where} \quad h_0 = m; \qquad h_r = 1 - \sum_{i=1}^{n} h_i$$

$$\begin{aligned}
\overline{W}_i &= \frac{\int e^{-\beta x} P_i(x) \left[\frac{1}{2} + \frac{1}{2} b(x)\right] dx}{\int e^{-\beta z} P_r(z)\, dz} \quad \text{where} \quad b(x) = \int e^{-\beta y} P_r(y) \frac{e^{-\gamma(x-y)^2}}{\int e^{-\gamma(\tilde{x}-y)^2} e^{-\beta\tilde{x}} P_r(\tilde{x})}\, dy \\
&= \exp\left[-\beta\left(\overline{z}_i - \overline{z} - \frac{\beta}{2}(V_i - V)\right)\right]\left(\frac{1}{2} + \frac{1}{2}\frac{1+2\gamma V}{\sqrt{1+2\gamma(V+2\gamma V^2+V_i)}} \exp\left[-\frac{\gamma(\overline{z}_i - \overline{z} - \beta(V_i - V))^2}{1+2\gamma(V+2\gamma V^2+V_i)}\right]\right) \\
&= \exp\left[-\beta(\overline{z}_i - \overline{z})\right]\left(\frac{1}{2} + \frac{1}{2}\frac{1+2\gamma V}{\sqrt{1+2\gamma(V+2\gamma V^2+V_i)}} \exp\left[-\frac{\gamma(\overline{z}_i - \overline{z})^2}{1+2\gamma(V+2\gamma V^2+V_i)}\right]\right) + \mathcal{O}(m)
\end{aligned} \tag{S8}$$

As in the case of Model I, this yields: $h_i = 2^i m \prod_{k=0}^{i-1} \overline{W}_k$.

We can now use eq. (S3) together with the function $M(x,y)$ for Model II (as given on the previous

page) to derive expressions relating $\{\overline{z}_i, V_i\}$ and $(\overline{z}_r, V_r)$ to each other, restricting ourselves (as before) to computations that are correct to lowest order in m. However, there is now an important difference (as compared to Model I): under Model II, if trait values are normally distributed among residents and (say) among $F1$s, then the distribution of trait values among their (BC_1) offspring will be the *sum* of two different normal distributions– one corresponding to offspring of resident father and F1 mothers and the other to offspring of resident mothers and F1 fathers. These distributions may be slightly shifted away from the midparent value towards the F1 mean (for offspring of F1 mothers and resident fathers) or towards the resident mean (for offspring of resident mothers and F1 fathers, and provided $V_1 > V$) and thus are not identical. However, one can show that the squared difference between the mean of these two distributions relative to the variance is proportional to γV, and will thus be small for $\gamma V \ll 1$. Then, we can approximate the mixture of two normal distributions by a single normal distribution with moments equal to the weighted sum of the constituent distributions. This is the additional assumption that we need to make in order to derive expressions for $\{\overline{z}_i, V_i\}$ and $(\overline{z}_r, V_r)$ under Model II (in analogy to eq. (S5) for Model I).

However, these expressions are somewhat cumbersome and we do not provide them here. Instead we focus on a parameter regime in which γV_r, γV_i are $\ll 1$. Then, as in Model I, the expressions for the trait means and variances and average relative fitness of the subgroups simplify considerably, and we have:

$$\beta V_r = h_n(\overline{z}_n - \overline{z}_r) \qquad V_r = 2V_{r,r} \tag{S9a}$$

$$\overline{z}_{i+1} - \overline{z}_r = \frac{\overline{z}_i - \overline{z}_r}{2} \qquad V_{i+1} = \frac{V_r + V_i}{4} + V_{r,i} \tag{S9b}$$

$$\overline{W}_i = e^{-\beta(\overline{z}_i - \overline{z}_r)} \frac{1 + e^{-\gamma(\overline{z}_i - \overline{z}_r)^2}}{2} \tag{S9c}$$

Note that equations (S9a) and (S9b) are identical to eqs. (S7a) and (S7b). However, the mean fitness of the ith subgroup is different under the two models (eq. (S7c) vs. eq. (S9c)), reflecting the fact that mating success is reduced for both male and female non-resident individuals under Model I but only for non-resident males under Model II. It follows from eq. (S9b) that $\overline{z}_i - \overline{z}_r = (\overline{z}_0 - \overline{z}_r)/2^i$. Substituting into eq. (S9c) gives: $\overline{W}_i = \exp[-\beta(\overline{z}_0 - \overline{z}_r)/2^i] \frac{(1+\exp[-\gamma(\overline{z}_0 - \overline{z}_r)^2/4^i])]}{2}$, which can be substituted into $g = \lim_{n\to\infty} \prod_{i=1}^{n} \overline{W}_i = \prod_{i=1}^{\infty} \overline{W}_i$. This gives the approximate expression for g under Model II (equation 3b of the main paper).

B. Deterministic allele frequency dynamics at a trait locus under Linkage Equilibrium (LE)

In this section, we derive an expression for *direct* selection at a trait locus by assuming Linkage Equilibrium (LE), i.e., by neglecting statistical associations or LD between loci. The key result here is that sexual selection contributes negligibly to direct selection at any trait locus if the trait is sufficiently polygenic (see below for a more precise statement). However, as we show in section A, sexual selection does reduce the gene flow factor, g (or $m_e = m\,g$), which accounts for the effects of LD between deleterious alleles (which we ignore in this section).

Consider the change in allele frequency $(\Delta p)_{sel}$ due to a single generation of natural and sexual selection at a focal locus with additive effect α and allele frequency p. This can be expressed as:

$$(\Delta p)_{sel} = p(1-p)\frac{\int W_1(z')P_*(z')dz' - \int W_0(z')P_*(z')dz'}{p\int W_1(z')P_*(z')dz' + (1-p)\int W_0(z')P_*(z')dz'} \tag{S10}$$

where $W_1(z')$ (or $W_0(z')$) is the fitness of an individual that carries the '1' (or '0') allele at the focal locus and has trait value $z{=}z'{+}\alpha$ (or $z{=}z'$). Thus, z' is the contribution to trait value due to all loci minus the focal locus, and $P_*(z')$ is the distribution of z' in the population. Note that under the LE assumption, this distribution must be the same for individuals carrying the '0' vs. '1' allele at the focal locus. Now, if L is large, then $P_*(z')$ is approximately normal with mean $\overline{z}_*{=}\overline{z}{-}\alpha\, p$ and variance $V_*{=}V{-}\alpha^2 p(1{-}p)$, where $\overline{z}$ and V are the mean and variance of trait values (as determined by all L loci) in the population.

Under **Model I**, an individual with trait value z has fitness $W(z)\propto e^{-\beta z}\int e^{-\beta y-\gamma(z-y)^2}P(y)dy$, where $P(z){=}\frac{1}{\sqrt{2\pi V}}e^{-\frac{(z-\overline{z})^2}{2V}}$ is the distribution of trait values in the population. This gives: $W(z)\propto \exp[-\frac{1+4\gamma V}{1+2\gamma V}\beta z-\frac{\gamma}{1+2\gamma V}(z-\overline{z})^2]$. Now substituting $W_1(z'){=}W(z'{+}\alpha)$ and $W_0(z'){=}W(z')$ into eq. (S10), performing the integrals over the (normal) distribution $P_*(z')$, and Taylor expanding in powers of α, one obtains:

$$(\Delta p)_{sel} = -\beta\alpha\, p(1-p)+\left(\frac{\beta^2\alpha^2}{2}-\frac{\gamma\alpha^2}{1+4\gamma V_r}\right)\, p(1-p)(1-2p)+\mathcal{O}(\alpha^3) \tag{S11}$$

Under **Model II**, the fitness of an individual with trait value z is:
$W(z)\propto e^{-\beta z}\int e^{-\beta y}\left[1+\frac{e^{-\gamma(z-y)^2}}{\int e^{-\beta\tilde{z}}e^{-\gamma(\tilde{z}-y)^2}P(\tilde{z})d\tilde{z}}\right]P(y)dy = e^{-\beta z}\left(1+\frac{1+2\gamma V}{\sqrt{1+2\gamma V+4\gamma^2V^2}}\, e^{-\frac{\gamma(z-\overline{z}+\beta V)^2}{2\gamma V(1+2\gamma V)}}\right)$. As before, we can substitute $W_1(z'){=}W(z'{+}\alpha)$ and $W_0(z'){=}W(z')$ into eq. (S10), integrate over $P_*(z')$, and Taylor expand in powers of α. This gives:

$$(\Delta p)_{sel} = -\beta\alpha\, p(1-p)+\left(\frac{\beta^2\alpha^2}{2}-\frac{\gamma\alpha^2}{2(1+2\gamma V_r)^2}\right)\, p(1-p)(1-2p)+\mathcal{O}(\alpha^3) \tag{S12}$$

Thus, if α is sufficiently small (see below for a more precise condition), then to leading order in α, we have: $(\Delta p)_{sel}\approx -\beta\alpha\, p(1-p)$ under both models of assortative mating. If we now assume the mainland to be fixed for alleles that are locally deleterious on the island, then the per generation change in allele frequency due to migration is simply $(\Delta p)_{mig}{=}m(1{-}p)$. Then, at migration-selection balance (and assuming LE): $\Delta p{=}(\Delta p)_{sel}{+}(\Delta p)_{mig}{=}{-}sp(1{-}p){+}m(1{-}p){=}0$, where $s{=}\beta\alpha$. Thus, the equilibrium allele frequency is $p{=}m/s$ if $m{<}s$, and $p{=}1$ otherwise.

How small does α need to be for us to be able to neglect $\mathcal{O}(\alpha^2)$ terms in equations (S11) and (S12)? This requires that both $\beta^2\alpha^2$ and $\gamma\alpha^2$ be much smaller than $\beta\alpha$, which in turn requires: $\beta\alpha{\ll}1$ and $\alpha{\ll}\beta/\gamma$. The first condition, namely, $s{=}\beta\alpha{\ll}1$ simply means that the change in allele frequency in any generation due to selection (or migration) should be sufficiently small that allele frequency dynamics can be approximated by continuous time equations (as is standard in population genetics). The second condition $\alpha{\ll}\beta/\gamma$ can be rewritten as $\frac{\alpha}{\overline{z}_0-\overline{z}}{\ll}\frac{\beta(\overline{z}_0-\overline{z})}{\gamma(\overline{z}_0-\overline{z})^2}$: this translates into the requirement that the relative contribution of any locus to mean trait divergence is much smaller than the ratio of the sexual selection component of migration load relative to the viability selection component. This condition is satisfied as long as the two components are comparable in magnitude.

C. Deterministic simulations under the hypergeometric model

We perform deterministic simulations of the mainland-island model based on the hypergeometric model. The order of events on the island is migration, viability selection, assortative mating (which generates sexual selection), and free recombination between parental genotypes. The mainland is assumed to be fixed for the alleles that are locally deleterious on the island. Additionally, unless specified otherwise, the island is initially fixed for the favorable allele at all trait loci, as in a secondary contact scenario.

Let $P(z)$ be the probability distribution of trait values $z{=}\sum_{i=0^L} \alpha_i X_i$ on the island, where α_i is the effect size and $X_i{=}0,1$ denote alternate alleles at locus i. Under the hypergeometric model, all L loci have equal effect (denoted by α) on the trait and all multi-locus genotypes corresponding to a given trait value are equally abundant in the population. Thus, we can express the trait value of an individual as $z{=}\alpha k$, where k is the number of '1' alleles carried by the individual. Further, the distribution $P(z)$ is equivalent to the distribution $P(k)$ on the island. Starting with a distribution $P_t(k)$ at the end of generation t, the distribution is evolved under migration, viability selection and assortative mating as follows: :

Migration: $P'(k){=}(1{-}m)P_t(k){+}m\delta_{k,L}$, where $\delta_{k,L}{=}1$ if $k{=}L$ and is 0 otherwise.

Viability selection: $P''(k){=}\dfrac{W(k)P'(k)}{\sum\limits_{j=0}^{L} W(j)P'(j)}$, where $W(k){=}e^{-\beta\alpha k}$

Assortative mating and free recombination: $P'''(k){=}\sum\limits_{i=0}^{L}\sum\limits_{j=0}^{L} P''(i)P''(j)M(x{=}i\alpha, y{=}j\alpha)R_{i,j\to k}$.

Here, $R_{i,j\to k}$ is the probability that parents carrying i and j alleles (say, of type '1') produce an offspring with k '1' alleles. This is given by: $R_{i,j\to k} = \sum\limits_{h=0}^{min(j,k,i+j-k)} \frac{\binom{i}{h}\binom{L-i}{j-h}}{\binom{L}{j}}\binom{i+j-2h}{k-h}(\frac{1}{2})^{i+j-2h}$, where $j{\leq}i$, and $max(0, i{+}j{-}L) < k < min(i{+}j, L)$, and L is the number of loci (see equation A2 in Barton (1992)). Also, $M(x, y)$ is the probability of mating between a male and female with trait values x and y respectively, where $x{=}i\alpha$ and $y{=}j\alpha$ (see section A for concrete formulae under Models I and II). Thus, at the end of the $(t{+}1)^{th}$ generation, we have: $P_{t+1}(k){=}P'''(k)$.

The steps above are iterated in each generation until equilibrium is attained, i.e., the average allele frequency $\overline{p}{=}\frac{1}{L}\sum_{k=0}^{L} kP(k)$ (averaged over all trait loci) no longer changes with time. In the following, we will depict our results in terms of the average allele frequency divergence per trait locus, $Y{=}1{-}\overline{p}$.

D. Accounting for the effects of genetic drift on divergence

In the main paper, we focus on the deterministic limit of the model. However, our analytical approach generalises to finite populations, where we can obtain the expected trait divergence by integrating over the *distribution* of allele frequencies on the island. The distribution of allele frequencies at any locus is assumed to be governed by the balance between mutation, selection, genetic drift and *effective* migration (eq. 5 of the main paper), where the last depends on the expected trait divergence between mainland and island (eq. 3, main paper), allowing for a self-consistent solution. We now test the

validity of our finite population approximations by comparing against individual-based simulations.

These simulations track N haploid individuals on the island, each carrying L unlinked, biallelic trait loci. At the start of the simulation, the island is perfectly adapted, with all individuals carrying the locally adaptive allele at all L loci. In each generation, we choose a Poisson-distributed number of individuals (with mean Nm), and replace them with mainland individuals that carry the locally deleterious allele at each trait locus. Migration is followed by mutation, where the allelic state at each locus in each individual is flipped independently with probability μ. Mutation is followed by selection of parents, which is performed differently under the two models. In Model I, we sample (with replacement) N *pairs* of parents with sampling weights proportional to the relative fitness of the pair, which is: $e^{-\beta(Z+Z')+\gamma(Z-Z')^2}$ for a pair of individuals with trait values Z and Z'. In Model II, we first sample (with replacement) N individuals in the *female role*, i.e., with sampling weights proportional to female fitness, which is: $e^{-\beta Z}$ for an individual with trait value Z. For each such individual (having, say, trait value Z'), we then choose a mate by sampling individuals (with replacement) with weights proportional to male fitness, which is: $e^{-\beta Z+\gamma(Z-Z')^2}$ for an individual with trait value Z. Thus, in both Models I and II, we have N parental pairs at the end of the sampling step. Haploid offspring are then created by free recombination between pairs, i.e., by sampling alleles from one or other parent with probability 1/2 independently at each locus. These steps are iterated over successive generations, until the mean and variance of the trait value distribution on the island stop changing in time.

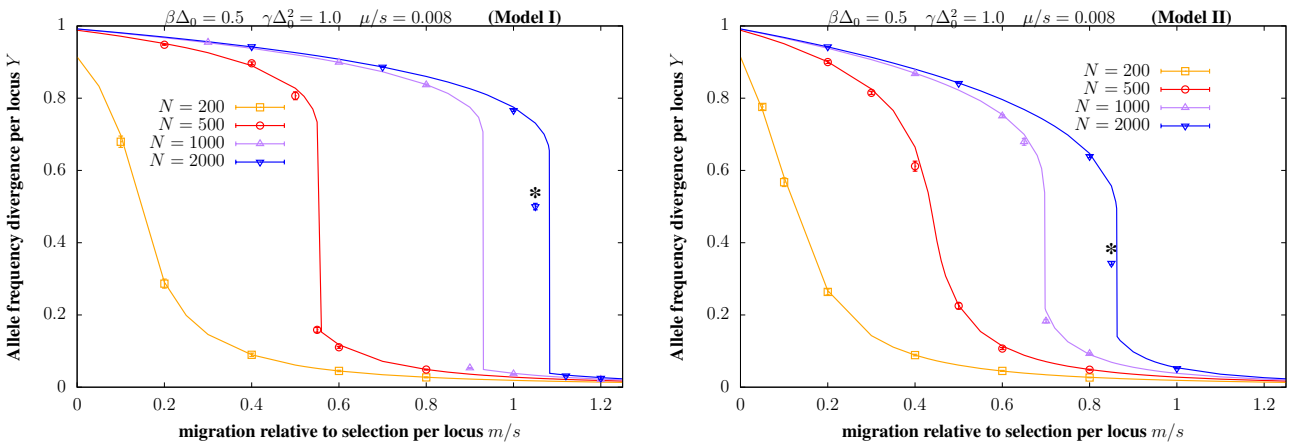

Figure S1: Comparison of individual-based simulations with theoretical predictions. Left and right panels show Y, the mean allele frequency divergence per locus as a function of m/s, the migration rate scaled by per locus selection coefficient, under models I and II of assortative mating. The magic trait is encoded by L=80 unlinked loci, with $\beta\alpha$=s=6.25×10^{-3} and $\gamma\alpha^2$=1.5625×10^{-4}, so that the scaled strengths of viability and sexual selection are respectively: $\beta\Delta_0$=$\beta\alpha L$=0.5 and $\gamma\Delta_0^2$=$\gamma\alpha^2 L^2$=1. Each locus undergoes bidirectional mutation at rate μ=5×10^{-5} per generation, so that μ/s=0.008. The different colors depict results for different population sizes N. Lines show theoretical predictions (obtained by numerically solving equation 5 of the main text); symbols show results of individual-based simulations. Each simulation point is obtained by averaging over allele frequencies at every 40th generation over a period of 5000 generations (after equilibration) in each simulation, and further averaging this over 20 simulation replicates; error bars depict standard error across replicates. The two points marked by $*$ have not equilibrated in the run time of the simulation, with the equilibrium values likely to be lower than what is depicted in the figures.

Figure S1 shows allele frequency divergence per locus as a function of m/s, for various values of N, for moderately strong viability and sexual selection ($\beta\Delta_0$=0.5, $\gamma\Delta_0^2$=1) under the two models of AM. Symbols depict results of individual-based simulations, while lines show theoretical predictions obtained by self-consistently solving equation 5 of the main text, while using equation 3 of the main

text to approximate the gene flow factor.

Three comments about figure S1 are in order. First, we find that theoretical approximations that account for drift are in fairly good agreement with simulation results across an order of magnitude of N (and Ns) values. Modest deviations between theory and simulations can be observed close to tipping points (as are also seen in the deterministic case; see figure 2 of the main text)– these are likely due to the relatively crude approximation that we use for the gene flow factor. Second, as expected, smaller populations are swamped at lower levels of gene flow. This is true even in the absence of multi-locus barriers, with local adaptation at a single divergently selected locus being appreciably reduced if $Ns \lesssim 20$, for (say) m/s=0.5. Third, tipping points (characterised by an abrupt collapse in divergence) can only be observed for sufficiently large Ns ($N \geq 500$ or $Ns \geq 3.125$ in Model I and $N \geq 1000$ or $Ns \geq 6.25$ in Model II). In smaller populations (i.e., with smaller Ns), divergence decays smoothly with increasing gene flow under both models. The absence of tipping points in small populations is explained by the fact that the positive feedback between increasing adaptation at individual loci and declining m_e across the genome (which is ultimately responsible for tipping points) is weaker when evolutionary dynamics at any locus are dominated by genetic drift.

E. Dependence of divergence per locus on the number of loci (deterministic predictions)

The theoretical predictions in equations 3 and 4 of the main paper (see also section A) are independent of L, the number of trait loci, and depend on only three composite parameters— the rate of migration relative to per locus selection m/s (where $s=\beta\alpha$), the strength of viability selection on migrants (as quantified by $\beta\Delta_0$) and the strength of sexual selection on migrants (as quantified by $\gamma\Delta_0^2$). These predictions are correct to first order in small parameters m, s etc., and are thus expected to become increasingly inaccurate as selection per locus becomes stronger. In particular, for a given strength of viability selection $\beta\Delta_0$ and sexual selection $\gamma\Delta_0^2$ (where $\Delta_0=\alpha L$ is the maximum possible trait divergence between mainland and island), we expect our predictions to become more accurate as we consider traits determined by larger and larger number of loci of weaker effect, i.e., as we approach the limit $\alpha \to 0$, $L \to \infty$ with $\Delta_0=\alpha L$ held constant.

We now verify these expectations by comparing theoretical predictions with the results of deterministic simulations of traits influenced by $L=10, 40, 100$ loci, for a specific strength of viability and sexual selection ($\beta\Delta_0=\gamma\Delta_0^2=1$) for the two models of assortative mating. In order to keep the maximum possible divergence $\Delta_0=\alpha L$ (as well as $\beta\Delta_0$ and $\gamma\Delta_0^2$) constant, we simultaneously decrease α while increasing L. For small L, say L=10, the effect of LD at any locus is weaker than what is predicted by our expressions, so that divergence is lost at lower values of m/s. However, as expected, there is better agreement between simulations and theory with increasing L, so that m_c/s for a trait with L=100 loci only slightly lower than the analytical prediction (compare purple lines with black line in figure S2). More generally, we observe larger deviations from theory when $\beta\Delta_0$ and/or $\gamma\Delta_0^2$, L are large. Thus, if net selection is strong, then L must be correspondingly large to observe good agreement between theory and simulations.

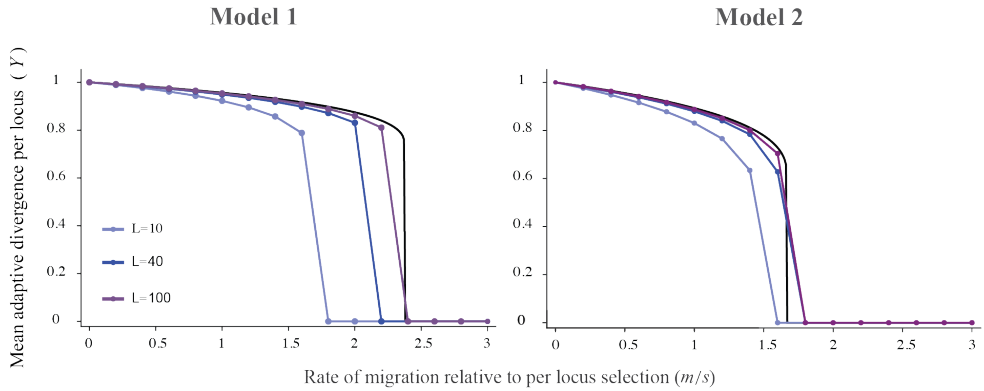

Figure S2: Mean allele frequency divergence per locus (Y) between mainland and island vs. migration rate relative to per locus selection (m/s), for Models I and II of assortative mating, with $\beta\Delta_0{=}\gamma\Delta_0^2{=}1$ in both panels. Different colours show the results from the deterministic simulation with L=10, 40, and 100 loci, where α is decreased as L is increased to keep $\Delta_0{=}\alpha L$ constant. The black curve shows the analytical (large L or, equivalently, small s) predictions for the two models, obtained by numerically solving eq. 4 in the main text.

F. Dependence of equilibrium divergence on initial divergence levels (deterministic predictions)

Throughout the main text, we have considered a secondary contact scenario with high initial divergence between mainland and island. However, under intermediate or strong assortment, where two stable equilibria are possible (marked by crosses in figure S4B and C), the island population can evolve towards either equilibrium under gene flow, depending on initial conditions. More specifically, starting with high initial divergence ($Y{\sim}1$), the population approaches the stable equilibrium associated with the higher level of divergence, provided $m{<}m_c$. By contrast, if Y is initially low, the population approaches the low (or zero) divergence equilibrium. In other words, in this regime, gene flow is strong enough to prevent genetically similar populations from diverging (despite divergent selection), but not strong enough to erode divergence in populations that are already strongly diverged.

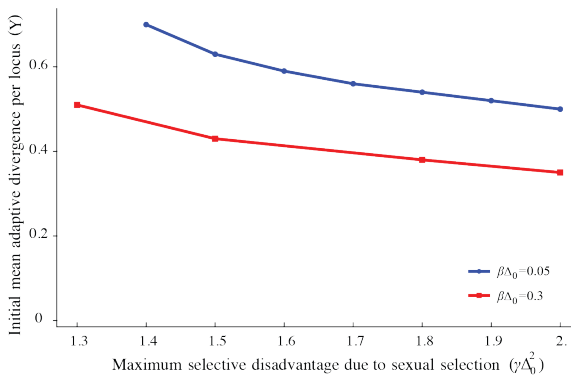

Figure S3: Minimum initial allele frequency divergence per locus required for populations to evolve towards the high-divergence equilibrium, plotted against the strength of sexual selection ($\gamma\Delta_0^2$) for Model I of assortative mating. Different colours show results for two different strengths of viability selection $\beta\Delta_0$=0.05 and 0.3. All results are obtained from deterministic simulations with L=40 loci.

Figure S3 explores how equilibrium divergence depends on initial divergence using deterministic simulations initialised with different values of Y, the mean per locus allele frequency divergence. Note that under the hypergeometric model (where all trait loci are exchangeable), this amounts to assuming an

allele frequency divergence Y at each locus. The population is then allowed to equilibrate under gene flow. We consider assortment levels at which alternative stable equilibria exist, so that populations can evolve towards either equilibrium, depending on the initial divergence Y. Figure S3 shows the minimum initial divergence per locus required for populations to attain the 'high-divergence' equilibrium, as a function of $\gamma\Delta_0^2$, the strength of sexual selection, for two different values of $\beta\Delta_0$. As expected, long-term divergence can be maintained more easily, i.e., even with moderate levels of initial divergence, if net (viability or sexual) selection is strong.

G. Critical divergence thresholds and migration rates (deterministic predictions)

As described in the main text (see also section B), the dynamics of the average allele frequency divergence per locus, $Y{=}1{-}\overline{p}$ can be approximated as:

$$dY/dt \approx sY(1{-}Y){-}m_e\,Y \tag{S13}$$

This is the same as the equation for a single selected locus, but with m replaced by the effective migration m_e, which captures the effect of LD between the focal locus and all other trait loci. Furthermore, $m_e{=}m\,g(Y)$, where g is the gene flow factor (given by equation 3 in the main text), which itself depends on the average divergence level (and on $\beta\Delta_0$ and $\gamma\Delta_0^2$, which quantify respectively the strength of viability and sexual selection). Thus, at migration-selection-assortment equilibrium, we have $dY/dt{=}0$, so that the equilibrium divergence level is given by the solution(s) of:

$$f(Y){=}Y(1{-}Y){-}\frac{m}{s}g(Y)Y{=}0 \tag{S14}$$

This yields one or more equilibria Y^*, which are stable provided $f'(Y^*){<}0$. This can be visualised by plotting $f(Y)$ vs Y (see figure S4). For any m/s, the point at which the curve intersects the horizontal axis corresponds to an equilibrium; it is stable if the curve is downward-sloping at that point and unstable otherwise. $Y^*{=}0$ is always an equilibrium and is stable if $m/s{>}1$. Additionally, there may be other equilibria with $Y^*{>}0$ which satisfy $1{-}Y^*{=}\frac{m}{s}g(Y^*)$.

For weak selection, say when $\beta\Delta_0{=}0.2$, $\gamma\Delta_0^2{=}0.5$ (see figure S4A), there is an unstable equilibrium at 0 and a non-zero stable equilibrium (as marked by a cross) when $m/s{<}1$. Increasing m/s causes this equilibrium to shift to smaller values; in other words, there is a smooth decline in adaptive allele frequencies until divergence is lost at the critical migration threshold, $m_c/s{=}1$. For $m/s{\geq}1$, the only stable equilibrium is at $Y^*{=}0$.

The behaviour of equilibria is qualitatively different if the strength of (viability and/or sexual) selection is intermediate or high. For example, for $\beta\Delta_0{=}0.2$ and $\gamma\Delta_0^2{=}1.3$ (figure S4B), there is a single non-zero stable equilibrium at low m/s, as with weak selection. Increasing m/s, say to 0.93 gives rise to two stable non-zero equilibria (marked as crosses), separated by an unstable equilibrium (in addition to an unstable equilibrium at 0). As m/s increases further (to 0.949), the stable 'high-divergence' equilibrium collides with the unstable equilibrium at a critical divergence threshold Y_c (marked by a circle). A further increase in m/s causes this high-divergence equilibrium to vanish, so that only

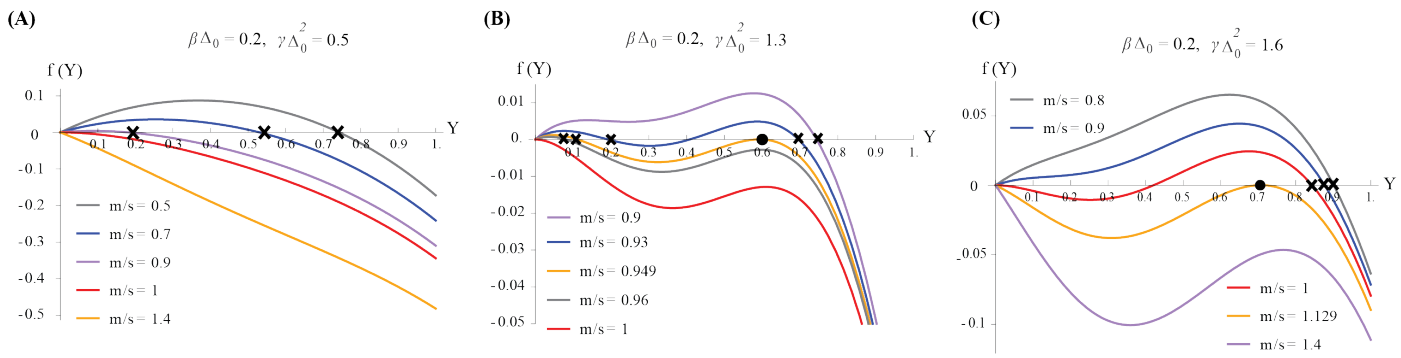

Figure S4: Different behaviours of the equilibria are visualized by plotting $f(Y)$ vs. Y for Model I of assortative mating, where Y is the mean allele frequency divergence per locus and $f(Y)=\frac{1}{s}\,dY/dt$ is given by equation S14. For a given m/s, equilibria Y^* are the points of the curve that intersect the horizontal axis, i.e., for which $f(Y^*)=0$, and are stable if the curve is downward sloping, i.e., if $f'(Y^*)<0$. (A) Weak selection ($\beta\Delta_0=0.2$, $\gamma\Delta_0^2=0.5$): there is a non-zero stable equilibrium (marked by a cross) if $m/s<1$, which disappears at a critical migration threshold, $m_c/s=1$, beyond which 0 is the only stable equilibrium. (B) Intermediate selection ($\beta\Delta_0=0.2$, $\gamma\Delta_0^2=1.3$): there are two non-zero stable equilibria for certain migration thresholds, e.g., for $m/s=0.93$ (see crosses), separated by an unstable equilibrium. With increasing m/s, the unstable and stable equilibria come closer and collide at a critical divergence threshold Y_c (marked by a circle for $m/s=0.949$). This equilibrium vanishes with a slight increase in m/s, say when $m/s=0.95$, resulting in a sudden drop in divergence to a low but non-zero value (marked by a cross). A further increase in m causes divergence to drop gradually until 0 becomes the only stable equilibrium beyond the critical migration threshold $m_c/s=1$. (C) Strong selection ($\beta\Delta_0=0.2$, $\gamma\Delta_0^2=1.6$): As with weak and intermediate selection, there is one non-zero stable equilibrium when $m/s<1$ and 0 is an unstable equilibrium. When $m/s\geq1$, there are two stable equilibria– one at 0, and the other at a high divergence level, separated by an unstable equilibrium. At a critical migration threshold $m_c/s=1.129$, the stable and unstable equilibria collide– this occurs at a divergence level Y_c (marked by a circle). Beyond this critical migration threshold, 0 is the only stable equilibrium. Note that the curve $f(Y)$ always has a maximum at $Y=Y_c$ (marked by circles in 1B and 1C).

the alternative 'low-divergence' stable equilibrium remains. We refer to this abrupt collapse of the high-divergence equilibrium as a 'tipping point'. A further increase in m/s causes the low-divergence equilibrium to approach 0, until $Y^*=0$ becomes the only stable equilibrium at $m_c/s\geq1$. Thus, with intermediate selection, we observe a tipping point (at which divergence collapses abruptly from high to low but non-zero levels) at a migration threshold m^*/s that is slightly below $m_c/s=1$, which is the critical migration threshold for complete loss of divergence (blue curve in figure S5).

With higher selection, say for $\beta\Delta_0=0.2$ and $\gamma\Delta_0^2=1.6$ (see figure S4C), there is a single stable high-divergence equilibrium for $m/s<1$. As m/s crosses 1, the equilibrium $Y^*=0$ also becomes stable, so that there are now two stable equilibria– the zero divergence and the high-divergence equilibrium separated by an unstable equilibrium. A further increase in m/s causes the unstable and stable (high-divergence) equilibria to approach each other, until at a critical migration threshold (here, $m_c/s=1.129$), the two equilibria collide– this occurs at a critical divergence value Y_c (marked by a circle). Beyond this migration threshold, $Y^*=0$ becomes the only stable equilibrium. Thus, in the strong selection regime, there is a critical migration threshold ($m_c/s>1$) at which a tipping point occurs such that divergence collapses abruptly from a high level to zero (figure S5).

Below, we use equation S14 to derive the divergence threshold (Y_c) beyond which there occurs a tipping point or sudden drop in divergence, and the critical migration threshold m_c/s, beyond which divergence goes to zero. From figures S4B and S4C, we can see that the critical divergence threshold Y_c (at which the high-divergence equilibrium suddenly vanishes) has the property that $f(Y_c)=0$ and

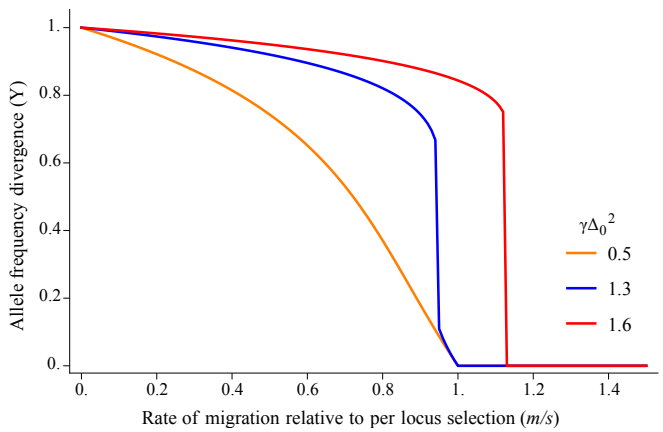

Figure S5: Allele frequency divergence Y vs. rate of migration relative to per locus selection (m/s) for a given strength of viability selection ($\beta\Delta_0$=0.2) and varying strengths of sexual selection ($\gamma\Delta_0^2$=1, 1.3 and 1.6). The plot illustrates different possible behaviours associated with loss of local adaptation: with moderate sexual selection ($\gamma\Delta_0^2$=1.3; blue curve), the tipping point occurs at a migration rate lower than m_c/s=1 (at which divergence is completely lost). With stronger sexual selection ($\gamma\Delta_0^2$=1.6; red curve), the tipping point and loss of divergence occur at the same threshold m_c/s>1.

$f'(Y_c)$=0. This gives:

$$1-Y_c=\frac{m}{s}g(Y_c) \tag{S15a}$$

$$1-2Y_c-\frac{m}{s}g(Y_c)-\frac{m}{s}Y_c g'(Y_c)=0 \tag{S15b}$$

Note that m/s in equation S15 is the migration threshold associated with Y_c. This migration threshold is less than m_c/s (which is equal to 1) when assortment is weak or moderate (see figure S4B) and the same as m_c/s under strong assortment (see figure S4C).

Random mating: With no assortment (γ=0), we have: $g(Y)=e^{-2\beta\Delta_0 Y}$, so that $g'(Y_c)=-2\beta\Delta_0\, g(Y_c)$, which when substituted into eq. S15 gives: $Y_c=\frac{2\beta\Delta_0-1}{2\beta\Delta_0}$.

Model I of assortative mating: In this case, we have: $g(Y)=e^{-2\beta\Delta_0 Y-\frac{4}{3}\gamma\Delta_0^2 Y^2}$, so that $g'(Y_c)=(-2\beta\Delta_0-\frac{8}{3}\gamma\Delta_0^2 Y_c)\, g(Y_c)$. Substituting this into eq. S15b, and using eq. S15a gives: $\frac{8}{3}\gamma\Delta_0^2 Y_c^2+(2\beta\Delta_0-\frac{8}{3}\gamma\Delta_0^2)Y_c+(1-2\beta\Delta_0)=0$. This can be solved to get two roots for Y_c of which

$$Y_c=\frac{4\gamma\Delta_0^2-3\beta\Delta_0+\sqrt{9(\beta\Delta_0)^2+16(\gamma\Delta_0^2)^2+24\beta\Delta_0\gamma\Delta_0^2-24\gamma\Delta_0^2}}{8\gamma\Delta_0^2} \tag{S16}$$

is the biologically meaningful solution.

Model II of assortative mating: In this case, we have: $g(Y)=e^{-2\beta\Delta_0 Y}\prod_{i=0}^{\infty}\left(\frac{1+e^{-\gamma\Delta_0^2\frac{Y^2}{4^i}}}{2}\right)$, so that: $g'(Y_c)/g(Y_c)=\frac{d\log g(Y)}{dY}\big|_{Y=Y_c}=-2\beta\Delta_0-2\gamma\Delta_0^2 Y_c\left[\sum_{i=0}^{\infty}4^{-i}\left(1+e^{\gamma\Delta_0^2\frac{Y_c^2}{4^i}}\right)^{-1}\right]$. Substituting this into eq. S15b and using eq. S15a gives the following equation for Y_c:

$$Y_c=1-\frac{1}{2\beta\Delta_0+2\gamma\Delta_0^2 Y_c\left[\sum_{i=0}^{\infty}4^{-i}\left(1+e^{\gamma\Delta_0^2\frac{Y_c^2}{4^i}}\right)^{-1}\right]} \tag{S17}$$

which is solved numerically by approximating the infinite sum by the first 20 terms (which is sufficient for convergence).

For both models of assortative mating, Y_c can be substituted in equation S15a to get m^*/s, which is the migration threshold associated with the tipping point. This is the same as m_c/s (at which divergence is completely lost) under strong selection (e.g., in figure S4C). However, under more moderate selection, tipping points occur at $m_*/s < m_c/s$, where m_c/s=1 (e.g., in figure S4B). Thus, we have: m_c/s= Max$[1, m^*/s]$. These predictions for Y_c and m_c/s are depicted in figure 3A-D of the main text.

We can also use the above predictions for Y_c and m_c/s to compute the minimum strength of sexual selection required for: (a) tipping points in divergence (associated with a non-zero Y_c), and (b) shifted critical migration thresholds, i.e., m_c/s>1, given a certain strength of viability selection. These predictions are shown in figure 3E-F of the main text. With random mating, Y_c>0 requires $\beta\Delta_0$>1/2 and is also associated with m_c/s>1 (also see Sachdeva (2022)). Put another way, if the strength of viability selection exceeds $\beta\Delta_0$=1/2, then we have Y_c>0 and m_c/s>1 even with zero sexual selection. Thus, we will focus on the parameter regime with $\beta\Delta_0$<1/2.

Under Model I of assortative mating, Y_c>0 requires that the term under the square root in equation S16 be positive. This gives the minimum strength of sexual selection required to observe a tipping point (given $\beta\Delta_0$) as:

$$\frac{3}{4}(1+\sqrt{1-2\beta\Delta_0}-\beta\Delta_0) \tag{S18}$$

To find the minimum strength of sexual selection required to shift the critical migration threshold beyond 1, we substitute m/s=1 into eq. S15a, which gives $g(Y_c)$=1−Y_c. If we now substitute the expression for Y_c under Model I (as given by eq. S16), then we obtain an equation that only involves $\beta\Delta_0$ and $\gamma\Delta_0^2$. This can be solved numerically to obtain the value of $\gamma\Delta_0^2$ at which m_c/s just exceeds 1, for any given value of $\beta\Delta_0$. Obtaining similar predictions under Model II is non-trivial. Thus, in this case, we numerically solve equation S17 to find the minimum value of $\gamma\Delta_0^2$ at which Y_c>0 and similarly for m_c/s>1. These predictions are depicted in figure 3E-F of the main text.

References

Barton, N. (1992). On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution*, pages 551–557.

Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73.

Sachdeva, H. (2022). Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLOS Genetics*, 18(9):e1010297.

Westram, A. M., Stankowski, S., Surendranadh, P., and Barton, N. (2022). What is reproductive isolation? *Journal of Evolutionary Biology*, 35(9):1143–1164.