SUPPORTING INFORMATION FOR

Divergence and Evolution of Assortative Mating in a Polygenic Trait Model of Speciation with Gene Flow

Himani Sachdeva and Nicholas H. Barton

Appendix S1: Hypergeometric model

The population consists of haploid organisms that express a phenotype X which is determined by L loci, each having two possible allelic states '$-$' and '$+$'. The diallelic loci additively determine X according to $X=\sum_{i=1}^{L}\Gamma_i$, where Γ_i can take on two values $-\gamma$ and γ, corresponding to the two allelic states. We choose $\gamma=\sqrt{2/L}$, so that trait values lie in the range $[-\sqrt{2L},\sqrt{2L}]$. Under the symmetry assumption (defined below) and in the fixed assortment scenario, the demes are fully described at any time t by the phenotypic probability distributions: $P_1(X)$ and $P_2(X)$ for demes 1 and 2 respectively. In a single generation, the probability distributions $P_1(X)$ and $P_2(X)$ evolve through the following steps:

Selection:

$$P'_k(X)=\frac{W_k(X)P_k(X)}{\sum_Y W_k(Y)P_k(Y)} \quad k=1,2 \tag{1}$$

Here, each niche is characterised by a distinct fitness function: $W_k(X)=e^{-(x-\mu_k)^2/2V_{s_k}}$.

Migration:

$$P''_k(X)=\sum_l M_{kl}P'_l(X) \quad \text{where} \quad M=\begin{pmatrix}1-m_{12} & m_{12}\\ m_{21} & 1-m_{21}\end{pmatrix} \tag{2}$$

Here, m_{12} is the fraction of the resident population of niche 1 that is replaced by immigrants from niche 2 and vice versa for m_{21}.

Assortative mating and free recombination:

$$P'''_k(X)=\sum_{Y,Z}\frac{P''_k(Y)P''_k(Z)e^{-\frac{(Y-Z)^2}{2\alpha^2}}R_{Y,Z\to X}}{\sum_{Z'}P''_k(Z')e^{-\frac{(Y-Z')^2}{2\alpha^2}}} \tag{3}$$

Females with phenotype Y mate with males carrying phenotype Z with a probability that is proportional to the Gaussian preference function $e^{-\frac{(Y-Z)^2}{2\alpha^2}}$, where α is the female preference range. Further, all females irrespective of phenotype, are assured equal mating success; this is encapsulated in the normalisation term in the denominator in eq. 3.

In general, it is not possible to predict the probability distribution of offspring phenotypes without knowing the exact genotypes of the parents, but under the assumption that all genotypes

corresponding to a particular phenotype are present with equal frequencies in the population, the distribution can be written purely in terms of parental phenotypes. This offspring distribution appears in eq. 3 as $R_{Y,Z\to X}$, which is the probability that parents with trait values Y and Z produce an offspring with trait value X upon free recombination.

For convenience, we recast the recombination term into the form $R_{n_1,n_2\to n_3}$ where n_1, n_2 and n_3 are the numbers of + alleles in sequences with trait values Y, Z and X respectively. An expression for $R_{n_1,n_2\to n_3}$ was derived in Kondrashov (1984) and Barton (1992). Here, we use an alternative (and equivalent) expression which is more amenable to taking the large L limit:

$$\begin{aligned} R_{n_1,n_2\to n_3} &= \sum_{m=0}^{L} \sum_{\substack{n=\max[0,\\ n_1+m-L,\\ n_3-n_2,\\ n_3+m-L]}}^{\min[m,n_3,\\ n_1,m+n_3-n_2]} \binom{L}{m}\frac{1}{2^L} \times \frac{\binom{n_1}{n}\binom{L-n_1}{m-n}}{\binom{L}{m}} \times \frac{\binom{n_2}{n_3-n}\binom{L-n_2}{L-m-n_3+n}}{\binom{L}{m}} \\ &= \sum_{m=0}^{L} \sum_{\substack{n=\max[0,\\ n_1+m-L,\\ n_3-n_2,\\ n_3+m-L]}}^{\min[m,n_3,\\ n_1,m+n_3-n_2]} \frac{\binom{n_1}{n}\binom{L-n_1}{m-n}\binom{n_2}{n_3-n}\binom{L-n_2}{L-m-n_3+n}}{\binom{L}{m}2^L} \end{aligned} \tag{4}$$

The expression under the summation in eq. 4 is a product of three terms: the probability of inheriting m loci from the first parent and $L-m$ loci from the second parent under free recombination, multiplied by the probability that there are precisely n '+' alleles among the m loci that the offspring inherits from the first parent (who has n_1 + alleles), which is finally multiplied by the probability that there are n_3-n + alleles among the $L-m$ loci inherited from the second parent (who has n_2 + loci). In writing the second and third probability terms, we assume that all genetic sequences with the same number (here n_1 or n_2) of + alleles have the same frequency in the population—this is the basic 'symmetry' assumption of the HM. Summing over allowed values of m and n then gives eq. 4.

In the large L limit, eq. 4 can be simplified by using Stirling's approximation for the factorials, and replacing the summations with integrals. This yields:

$$\begin{aligned} R_{n_1,n_2\to n_3} &\sim \frac{1}{\sqrt{2\pi V_f[n_1,n_2]}} \exp\left[-\frac{(n_3-(n_1+n_2)/2)^2}{2V_f[n_1,n_2]}\right] \\ \text{where}\quad V_f[n_1,n_2] &= \frac{n_1+n_2}{4} - \frac{n_1 n_2}{2L} \end{aligned} \tag{5}$$

If L is very large and $n_1/L \ll 1$ and/or $n_2/L \ll 1$, so that the probability of both parents having a + allele at the same locus is negligible, then eq. 5 can also be derived using the central limit theorem by assuming that the number of + alleles inherited from either parent, and hence the number of + alleles in the offspring genotype is normally distributed. Equation 5 turns out to be a highly accurate approximation of eq. 4 as long as the parents are not very close to the edges of the phenotypic range. Equation 5 can be recast in terms of phenotypic values as:

$$\begin{aligned} R_{Y,Z\to X} &\sim \frac{1}{\sqrt{2\pi V_f[Y,Z]}} \exp\left[-\frac{(X-(Y+Z)/2)^2}{2V_f[Y,Z]}\right] \\ \text{where}\quad V_f[Y,Z] &\sim \frac{\gamma^2}{2}L - \frac{YZ}{2L} \end{aligned} \tag{6}$$

For unimodal populations clustered about the generalist phenotype ($Y{\sim}Z{\sim}0$), the variance $V_f[Y,Z]$ is well approximated by $(\gamma^2/2)L$, which can be set to 1 by choosing $\gamma{=}\sqrt{2/L}$, thereby recovering the infinitesimal model in which the segregation variance is independent of parental phenotypes. More generally, for unimodal populations with phenotypic mean $\overline{Z}$, we can choose the segregation variance as $V_f{\sim}1-(\overline{Z}^2/2L)$ (obtained by setting $Y{\sim}Z{\sim}\overline{Z}$ in eq. (6)), which is again the same for all mating pairs but depends on the population mean, leading to a 'modified infinitesimal' model. In either case, $\sqrt{V_f}$ provides a natural unit in which to measure all other phenotypic 'distances' such as μ, $\sqrt{V_s}$, α and $\sqrt{V_0}$.

Evolving preference model. In the situation where a modifier allele is segregating at the preference range locus, we need to specify the probability distribution $P(X,\alpha)$ of individuals carrying phenotype X and having preference range α. As before, the change in $P(X,\alpha)$ over one generation can be broken down into changes due to selection, migration and assortative mating. The change due to selection and migration can be written in a similar way as in eqs. 1 and 2. The effect of assortative mating on the distribution in any deme k, in the case where r alleles corresponding to r different preference ranges α_i ($i{=}1\ldots r$) are segregating in the population, is given by:

$$P_k'''(X,\alpha){=}\sum_{Y,Z,\alpha_i,\alpha_j}\left[\left(\frac{1}{2}\delta_{\alpha,\alpha_i}+\frac{1}{2}\delta_{\alpha,\alpha_j}\right)\frac{P_k''(Y,\alpha_i)P_k''(Z,\alpha_j)e^{-\frac{(Y-Z)^2}{2\alpha_i^2}}R_{Y,Z\to X}}{\sum\limits_{Z',\alpha_l}P_k''(Z',\alpha_l)e^{-\frac{(Y-Z')^2}{2\alpha_i^2}}}\right] \tag{7}$$

Equation 7 is a generalization of eq. 3 to the case with evolving assortment. The scenario with a single modifier allele is obtained by setting $r{=}2$, and $\alpha_1{=}\alpha_{anc}$, $\alpha_2{=}\alpha_{mod}$, where α_{anc} and α_{mod} are the preference ranges associated with the ancestral and modifier alleles respectively.

We can also use the dynamical equations for $P_k(X,\alpha)$ to derive changes in the frequency of the modifier allele over time. In the symmetric version of the model, with $\mu_1{=}{-}\mu_2{=}\mu$, $V_{s_1}{=}V_{s_2}{=}V_s$ and $m_{12}{=}m_{21}{=}m$, and with the same initial frequency of the modifier allele in both demes, the frequency $p_{mod}(t)$ of the modifier allele in any deme, say deme 1, changes in one generation according to:

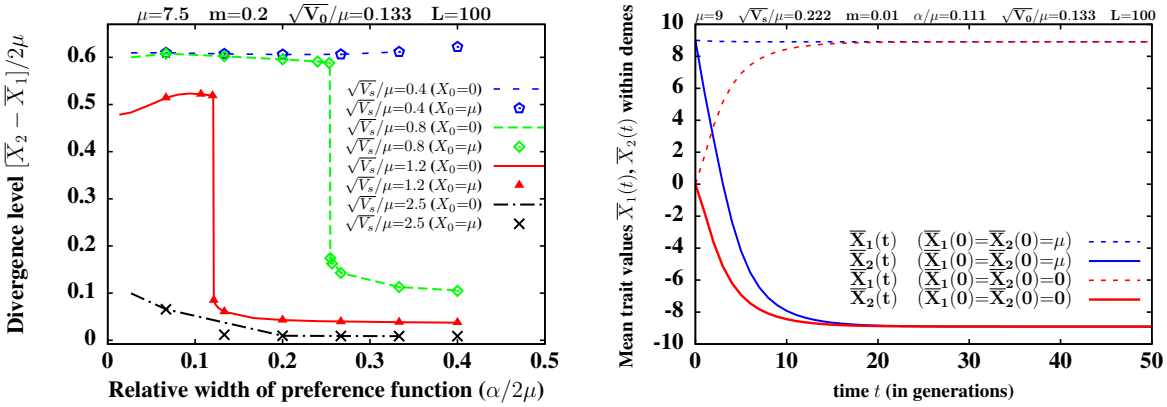
$$\frac{\tilde{p}_{mod}}{p_{mod}(t)}{=}\frac{\overline{W}_1(\alpha_{mod})}{\overline{W}_1}\qquad\text{where}$$
$$\overline{W}_1(\alpha_{mod}){=}\frac{\sum\limits_X W_1(X)P_1(X,\alpha_{mod})}{\sum\limits_X P_1(X,\alpha_{mod})}\quad\text{and}\quad\overline{W}_1{=}p_{mod}(t)\overline{W}_1(\alpha_{mod})+[1-p_{mod}(t)]\overline{W}_1(\alpha_{anc}) \tag{8a}$$

$$\frac{p_{mod}(t+1)}{\tilde{p}_{mod}}{=}\frac{1}{2}\left[1+\frac{1}{\tilde{p}_{mod}}\sum_{\substack{X,\\ \tilde{\alpha}=\alpha_{anc},\\ \alpha_{mod}}}P_1''(X,\tilde{\alpha})\frac{\sum\limits_Z e^{-\frac{(X-Z)^2}{2\tilde{\alpha}^2}}P_1''(Z,\alpha_{mod})}{\sum\limits_Z e^{-\frac{(X-Z)^2}{2\tilde{\alpha}^2}}[P_1''(Z,\alpha_{mod})+P_1''(Z,\alpha_{anc})]}\right] \tag{8b}$$

Here, $P_1(X,\alpha)$ is the probability distribution before selection, $\tilde{p}_{mod}$ is the modifier frequency after selection and migration (since migration produces no net change in modifier frequency in the symmetric scenario), and $P_1''(X,\alpha)$ the distribution after selection and migration.

Equations 8a and 8b encapsulate indirect selection acting on the modifier allele at two different stages of the lifecycle, and represent the effect of viability selection and sexual selection respectively. According to eq. 8a, if the marginal fitness of the modifier allele is greater than the population mean fitness (i.e., if the modifier allele gets preferentially associated with fitter phenotypes), then it increases in frequency. According to eq. 8b, if the phenotypes associated with the modifier allele are chosen for mating with a probability higher than their frequency (i.e., if their phenotypic distance to other individuals is shorter on average), then again the modifier undergoes positive selection, due to higher mating success of the associated phenotypes.

Appendix S2: Long-term HM evolution of populations having different initial states



(a) (b)

Figure S1: (a) Scaled divergence level $(\overline{X}_2 - \overline{X}_1)/2\mu$ vs. relative width $(\alpha/2\mu)$ of preference function, for various V_s and two different initial phenotypic distributions. Lines correspond to the case of two initially generalist population with $\overline{X}_2$=$\overline{X}_1$=0 at t=0, while points correspond to the case of two population initially adapted to one of the optima, having $\overline{X}_2$=$\overline{X}_1$=μ at t=0. Both lines and points represent HM predictions. (b) Dynamics of the trait means in the two demes (as predicted by the HM) for initially identical populations with initial trait mean at the generalist phenotype (red) vs. populations with initial trait mean at one of the optima (blue). Populations diverge and become bimodal (with $\overline{X}_1$ approaching $+\mu$ and $\overline{X}_2$ approaching $-\mu$) in both cases under the HM for the given set of parameters.

In the main paper, we followed the long-term evolution of populations starting out with an initial distribution $P(X)$=$\exp(-(X - X_0)^2/2V_0)/\sqrt{2\pi V_0}$ in both demes, with X_0 chosen to be zero (corresponding to a case where both populations are initially clustered around the generalist phenotype). Here, we consider alternative initial conditions, wherein both populations are locally adapted to one of the optima and have identical initial distributions with X_0=μ. Figure S1a shows the long-term divergence level as a function of α for various parameters, for populations with initial mean phenotype X_0=0 and X_0=μ under the HM. The divergence levels in both cases are nearly identical, pointing towards the lack of dependence of long-term evolution on the initial mean phenotype (though not necessarily the initial phenotypic spread). This stems from two causes– the high polygenic variation assumed in the HM and infinitesimal model, and the assumption of soft selection.

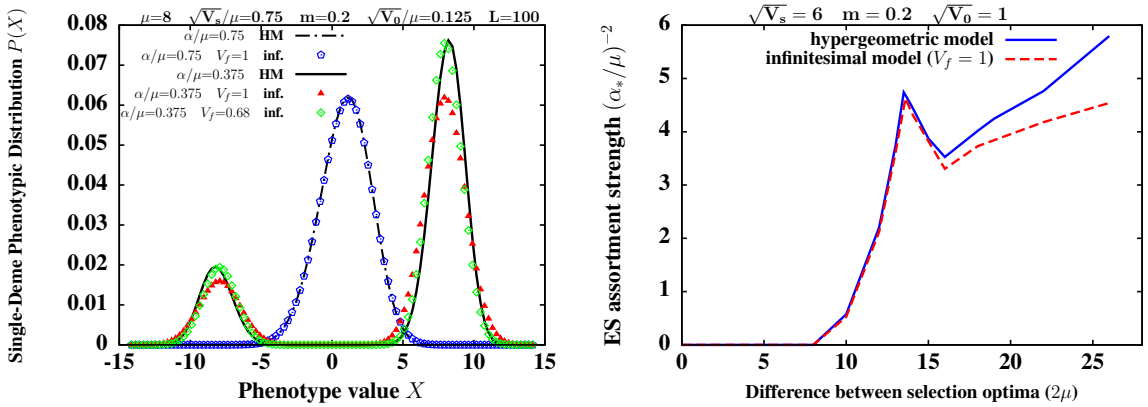
Due to soft selection within each deme, even fairly unfit phenotypes (far from the deme optimum) can survive and reproduce, as long as their relative fitness is high, i.e., they are

closer to the optimum than other phenotypes. Moreover, the large number of polymorphic loci within these surviving phenotypes, allows for the generation of new phenotypes via mating and segregation. These phenotypes are further selected for, if their relative fitness is high, allowing for a long-term selection response, irrespective of where the populations start out in phenotypic space. This sort of selection response would not occur if polymorphisms are depleted by stabilizing selection around the optima, but is expected to be robust if there is sufficient mutational variance.

Note that the evolutionary dynamics of the populations does depend on their initial phenotypic composition, as illustrated in fig. S1b, for the case of strong slection, weak migration and strong assortment. However, the long term state of the populations (and the mean phenotypic values) become identical in the long run for the two initial conditions depicted in the figure. This appears to be the case for a wide range of parameters (fig. S1a).

Appendix S3: Comparison of hypergeometric and infinitesimal models

In the large L limit, the recombination kernel $R(Y, Z \to X)$ approaches a Gaussian form $\exp[-(X-(Y+Z)/2)^2/2V_f]$ (eq. 6). Taking the segregation variance V_f to be independent of parental phenotypes yields the infinitesimal model. This can be analyzed in the two-deme setting in the same way as the HM, by using eq. 1-3 but now using eq. 6 (instead of eq. (4)) and treating V_f as a variable which describes the full population rather than the phenotypic variance of the offspring of specific parental pairs.



(a) (b)

Figure S2: (a) Phenotypic distribution $P(X)$ vs. X within a single deme for a hybridized population (α=0.75μ) and a diverged population (α=0.375μ), as predicted by the HM and the infinitesimal model. HM predictions closely match that of the infinitesimal model with V_f=1 for the unimodal population, and with V_f=1 − $\mu^2/2L$ for the bimodal population. (b) The scaled ES assortment strength $(\alpha_*/\mu)^{-2}$ vs. 2μ, as predicted by the HM and the infinitesimal model (with V_f=1). The two models show close agreement except for large μ where the infinitesimal model predicts a lower ES assortment strength in the diverged state for reasons discussed in text.

Figure S2a compares the HM prediction (which uses eq. 4) with the infinitesimal model prediction (which uses eq. 6) for the phenotypic distribution $P(X)$ for both unimodal and bimodal populations in a deme with selection optimum at $+\mu$. For the unimodal population, we set V_f=1 (assuming that the median parental phenotype is $X \sim 0$), while for the bimodal

phenotype we set V_f=$1-\mu^2/2L$ (using eq. 6 and assuming that most successful pairings occur within but not between the resident and migrant pools, where the median parental phenotype is $+\mu$ and $-\mu$ respectively). The resultant infinitesimal model predictions for $P(X)$ match the HM predictions very closely. Note that even with V_f=1, the infinitesimal model is able to qualitatively predict the (bimodal) nature of $P(X)$ for the diverged population (S2a); however, for quantitative agreement between the two, V_f must be chosen as above.

We also compare HM and infinitesimal model predictions (with V_f=1) for the evolving assortment scenario by plotting the ES assortment strength vs. 2μ (fig. S2b), as in fig. 5a of the main paper. These are also in close agreement, except for very large μ, when the ES assortment level is determined by selection on modifiers in bimodal populations for which the typical segregation variance is less than 1 (see above), leading to a discrepancy between the HM and infinitesimal prediction for the ES assortment strength.

The above comparison thus illustrates how the main predictions of the HM agree with those of the infinitesimal model at least at a qualitative level (and even quantitatively in a significant region of parameter space), when V_f is chosen appropriately. The quantitative agreement between the two models improves across all parameter regimes as the number of loci determining the trait in the HM increases (results not shown).

Appendix S4: A Gaussian approximation for the infinitesimal model

The analysis of our model can be simplified considerably by approximating the phenotypic distribution in each deme by a Gaussian function (Charlesworth 1990), and further assuming that the distribution remains Gaussian at each stage of the lifecycle. This assumption typically works very well for unimodal populations, even in the presence of disruptive selection (Turelli and Barton 1994). Thus, we expect that the Gaussian approximation will adequately describe hybridized populations and may also predict the threshold for divergence. For ease of calculation, we will apply the Gaussian approximation to the infinitesimal model.

Consider the symmetric scenario μ_1=$-\mu_2$=μ, V_{s1}=V_{s2}=V_s and m_{12}=m_{21}=m. Let the phenotypic distribution in each deme be a Gaussian function with mean x_0 (or $-x_0$ in the other deme) and variance V.

Replacing $P_k(X)$ in eq. 1 by these functions, and converting the discrete sum into an integral, we obtain $P'_k(X)$ after selection, which is also normally distributed. Using this expression for $P'_k(X)$ in eq. 2 now gives the distribution $P''_k(X)$ after migration. This distribution is the weighted sum of two Gaussian distributions, but for hybridized populations (high m, V_s), it can be approximated by a single Gaussian function with mean and variance equal to the weighted sums of the mean and variance of the two Gaussians. Thus, it follows that under the Gaussian approximation, the mean and variance after selection and migration are given by:

$$\begin{aligned} x'_0 &= (1-2m)\frac{V_s x_0 + V\mu}{V_s + V}; \\ V' &= \frac{V_s V}{V_s + V} + 4m(1-m)\left[\frac{V_s x_0 + V\mu}{V_s + V}\right]^2 \end{aligned} \tag{9}$$

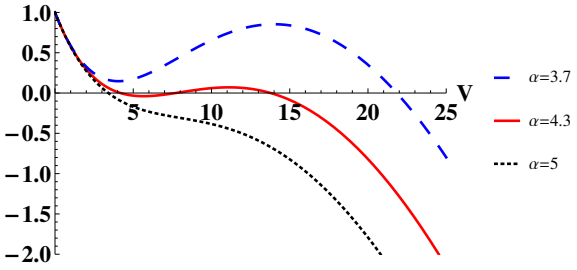


Figure S3: Numerical solution of eq. 12 for $\sqrt{V_s}$=6, μ=7.5, m=0.2, V_f=1. Value(s) of V at which any curve crosses zero (the horizontal axis) represent the solution(s) of the equation. At high α, the equation has a single small-V solution, corresponding to the hybridized state; at small α, it has a single large-V solution corresponding to the diverged state. For intermediate α, the equation has multiple solutions, indicating multi-stability, or divergence levels that depend on the initial variance of the population.

To calculate the change in the distribution due to assortative mating, we replace $P_k''(X)$ etc. in eq. 3 by Gaussian functions with the above mean and variance, then use eq. 6 for $R_{Y,Z\to X}$ with V_f set to a constant (infinitesimal model), and finally replace the sums in eq. 3 by integrals. This gives the phenotypic distribution after assortative mating, which is a Gaussian with mean and variance given by:

$$\begin{aligned} x_0''&=x_0' \\ V''&=\left[4(V')^2(V_f+V')+\alpha^2V'(8V_f+5V')+2\alpha^4(2V_f+V')\right]/4(\alpha^2+V')^2 \end{aligned} \tag{10}$$

In steady state, we can equate x_0'' with x_0, and V'' with V. The relation $x_0''=x_0$, in particular, gives:

$$x_0=\frac{(1-2m)\mu V}{V+2mV_s} \tag{11}$$

Substituting eq. 11 into eq. 9 and using the resultant expression for V' in eq. 10 finally yields the following transcendental equation for V:

$$\begin{aligned} &\left[4(f[V])^2(V_f+f[V])+\alpha^2 f[V](8V_f+5f[V])+2\alpha^4(2V_f+f[V])\right]/[4(\alpha^2+f[V])^2]-V=0 \\ &\text{where}\quad f[V]=\frac{V_sV}{V_s+V}+4m(1-m)\left[\frac{V\mu}{2mV_s+V}\right]^2 \end{aligned} \tag{12}$$

Equation 12 can be solved numerically to obtain V for any choice of parameters V_s, μ, m, α and V_f. This is illustrated in fig. S3 which shows a plot of the left hand side of eq. 12 as a function of V, with the zeroes of this function yielding the solutions of the equation. Note how the nature of the solution changes on varying α. For high values of α (dotted curve), the function has a single zero at a small value of V. For low values of α (dashed curve), the function again has a single zero, but this now occurs at a large value of V. For intermediate α (solid curve), the function has multiple zeros, corresponding to both small and large V solutions. Assuming that small-V solutions correspond to unimodal populations and large-V solutions to bimodal populations, this suggests the following: (i) There exists a preference threshold α_{c_1} (at which the small-V solution disappears), such that divergence always occurs if $\alpha<\alpha_{c_1}$, even when the initial variance of the population is very low. (ii) There exists another threshold α_{c_2} (at which the large-V solution disappears), such that there is no divergence for $\alpha>\alpha_{c_2}$, even

when the initial variance of the population is very large. (iii) Between these two thresholds, i.e., for $\alpha_{c_1}<\alpha<\alpha_{c_2}$, divergence depends on the initial variance of the populations, with populations diverging if initial variance is high, but remaining unimodally distributed if initial variance is low.

We test these predictions regarding the dependence of assortment thresholds on initial variance against the exact solution of the HM, and find that there is qualitative agreement between the two (fig. 2d of main paper). Moreover, the Gaussian approximation is able to accurately predict the preference threshold α_{c_1} for divergence of populations starting with low initial variance, at least if selection is not too strong or migration not too weak (see fig. 2c of main paper). The second threshold α_{c_2} for populations with high initial variance is, however, not very well predicted by the Gaussian approximation (results not shown). This reflects the fact that when initial phenotypic variance is very high and populations include optimal phenotypes at the outset, a transient divergence occurs, which is subsequently lost due to hybridization if assortative mating is weak. Thus, this is more similar to a secondary contact scenario, for which we do not expect a Gaussian approximation to work well.

Appendix S5: Long-term stability of extreme phenotypes in the large V_s, small α limit

As discussed in the main text, for weak selection (large V_s), the HM predicts that the populations remain unimodal even for very strong assortment, while undergoing slight shifts towards the phenotypic extremes. (This corresponds to the $\sqrt{V_s/\mu}$=2.5 curve in fig. 2a., main text). For still larger values of $\sqrt{V_s/\mu}$, i.e., weaker selection, the two extreme phenotypes are further amplified (in both demes) even though most of the population is clustered around X=0. For even larger $\sqrt{V_s/\mu}$ (fig. S4), the HM predicts that the two extreme phenotypes build up to such an extent that they form the majority constituents in each deme, resulting in high phenotypic variance in both demes (though the divergence between demes is typically low since both extreme phenotypes are present in both demes with only slight differences in frequencies). This buildup of extreme phenotypes is independent of ecological selection and occurs due to a mutually reinforcing increase in male and female phenotypic values, akin to runaway divergence in preference-trait scenarios of sexual selection (Lande 1981; Higashi et al. 1999); in fact with assortative mating, there is perfect correlation between the male trait and female preference, both being represented by the same trait X.

Individual-based simulations of finite populations show that assortment does indeed cause extreme phenotypes to proliferate in the population at short times (see the dashed line in fig. S4), also resulting in a transient increase in the phenotypic variance in the deme (inset, fig. S4). However, this state is not stable, in that deviations from the symmetry assumption of the HM model tend to destroy it over longer time scales. In particular, weak stabilizing selection within demes (large V_s) or weakly divergent selection across demes (low μ) effectively generates stabilizing selection towards X=0 on the whole population, which amplifies such deviations in finite populations over time, leading to loss of the extreme phenotypes and increase in the frequency of phenotypes closer to X=0 (see data points in fig. S4). This is also reflected in a corresponding decrease in phenotypic variance over longer time scales (inset, fig. S4).

We have considered here a case where $\sqrt{V_s/\mu}\gg 1$ for which effective stabilizing selection towards X=0 is very weak. The loss of extreme phenotypes even under such restrictive conditions

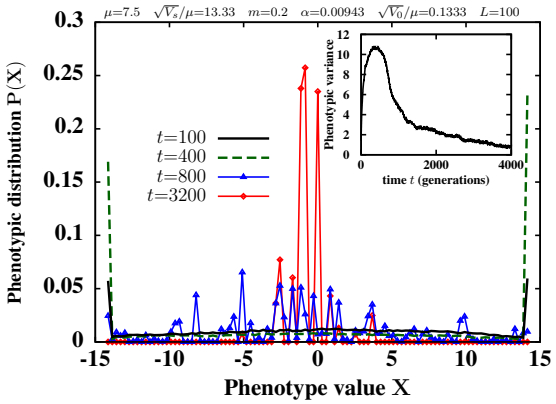


Figure S4: Phenotypic distributions $P(X)$ vs. X in a single deme in the weak selection, strong assortment limit (from individual-based simulations) at various time points (measured in generations) after the onset of divergent selection. There is an initial buildup of extreme phenotypes (dashed and solid lines) as also predicted by the HM, which is reversed over longer time scales (points). Inset: The phenotypic variance in the deme shows a transient increase due to this buildup, but then declines as extreme phenotypes are lost.

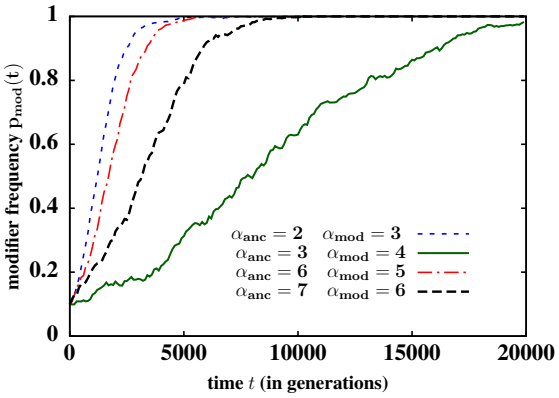


Figure S5: Modifier frequency dynamics obtained from individual-based simulations of a large population (N=10^6) with recurrent mutation at rate 10^{-3} per locus per generation: Modifiers that increase (or decrease) preference range towards α_*∼4.5 are positively selected and undergo fixation in both strongly diverged populations (with α_{anc}<α_*) as well as hybridized populations (with α_{anc}>α_*).

makes it all the more plausible that a similar effect operates when V_s is smaller, though still large enough to lie within the weak selection regime of the model.

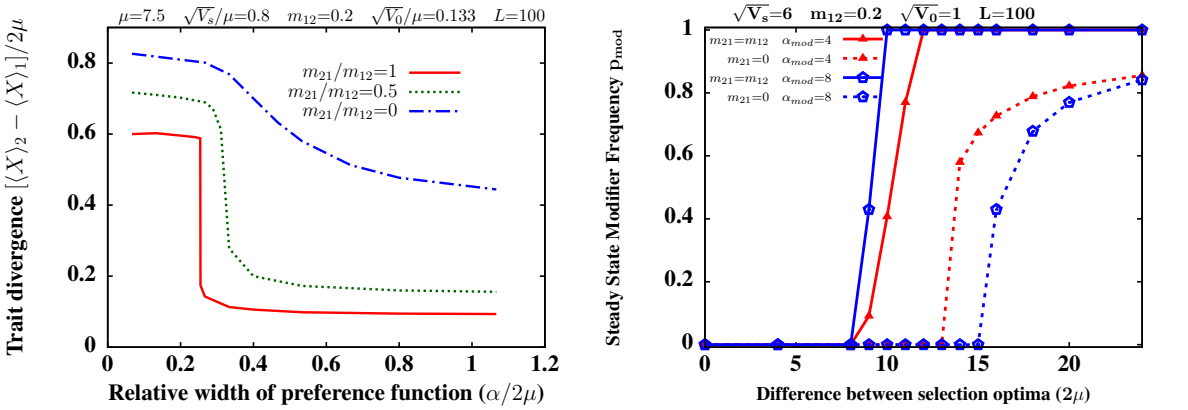
Appendix S6: Modifier evolution in the presence of mutation

We also consider a version of the model with mutation, which allows finite populations to maintain some phenotypic variation around the optima in spite of stabilizing selection and low gene flow. We study modifier dynamics for this model using individual-based simulations, focusing on the question: In the presence of mutational variation, is there a well-defined ES level of assortment α_*, such that populations with α_{anc}>α_* and α_{anc}<α_* are both susceptible to modifiers that shift the preference range towards α_*? In particular, do modifiers that increase α in strongly assortative populations undergo positive selection, as opposed to the neutral evolution (fig. 6(b) of main paper) that occurs in finite populations in the absence of mutation?

Figure S5 shows frequency dynamics for modifiers that shift the assortment strength towards intermediate levels in both weakly assortative (large α) populations as well as strongly assortative (small α) populations undergoing mutation at a rate of 10^{-3} per locus per generation (trait-wide mutation rate 0.1). Unlike in the zero mutation model (fig. 6(b) of main paper), in this case, modifiers that increase the preference range of small α populations undergo positive selection and invade. We also consider lower mutation rates and find that such modifiers are positively selected with trait-wide mutation rate $\sim$0.025, but undergo nearly-neutral evolution when the corresponding rate is $\sim$0.01. This suggests that in the presence of mutation, at least when mutation rate is not too low, predictions of the HM for modifier dynamics are likely to hold over a wide range of conditions.

Appendix S7: Asymmetric models

We first consider the model with asymmetric migration $m_{12} \neq m_{21}$. Figure S6a shows how divergence level varies with the preference range α in the established assortment case for three patterns of gene flow: symmetric migration, continent-island migration and an intermediate situation with bidirectional but unequal migration. A striking feature of the plot is that the assortment strengths required for divergence are actually lower (or α_c higher) when migration is asymmetric. Moreover, even the basal level of divergence (in the random mating limit) is higher for continent-island migration than for symmetric migration.



(a) (b)

Figure S6: (a) Divergence level or difference between mean trait values in the demes vs. relative preference range $\alpha/2\mu$ in the established assortment scenario for three different levels of asymmetry in gene flow. (b) Evolving assortment scenario for symmetric migration model vs. continent-island model: Long term modifier frequency p_{mod} after introduction of a modifier with preference range α in randomly-mating populations (in both demes vs. only on island), as a function of 2μ, the difference between selection optima of the two demes. Both plots show HM predictions.

This can be understood as follows. In the symmetric migration case, influx of individuals from deme 2 into deme 1 and subsequent gene flow, pulls the trait mean of the resident population in deme 1 closer to X=0. Thus individuals from deme 1, when they migrate to deme 2, are now phenotypically closer to the residents there and hybridize to a greater extent with them, which further reduces the trait mean in deme 2, setting up a mutually-reinforcing loss of local adaptation in the two demes. By contrast, with one-way migration, although islanders experience hybridization due to influx from the continent, this cannot alter the continental phenotype and trigger the sort of positive feedback between the two demes that occurs in the

symmetric case. As a result, all other parameters being the same, divergence in the continent-island setting may be more selection-driven and less assortment-dependent (as evident from the larger α_c and less abrupt change in divergence levels at α_c) than in the two-way migration case. However, with weaker ecological selection on the island, divergence again becomes strongly assortment-dependent, increasing abruptly at a threshold α_c (results not shown).

We also study the evolving assortment scenario for the two extreme limits, symmetric (m_{21}= m_{12}) and one-way migration (m_{21}=0), and compare the fate of a modifier that appears in both demes in the former case, with a modifier that only appears on the island in the latter (assuming random mating on the continent). Figure S6b shows the long-term frequency of the modifier in an ancestral population with random mating ($\alpha_{anc}\to\infty$), vs. 2μ. Both small-effect (α_{mod}=8) and large-effect (α_{mod}=4) modifiers establish over a more restricted range of μ in the continent-island case than in the symmetric migration case. This is especially true for the small-effect modifier for which indirect selection is very weak, and easily swamped by migration (as also observed in few-locus models, see Servedio (2000)). Even modifiers that invade the island population do not go to fixation, instead approaching a polymorphic frequency which reflects the underlying migration-selection balance with the random-mating allele introgressing from the continent.

We also explore other asymmetric versions involving unequal V_s in the two demes, and find that these do not differ qualitatively from the symmetric model. In particular, asymmetries in selection appear to have much less of an impact on assortment evolution as compared to asymmetric migration, with modifier dynamics being very similar in two demes subject to differing strengths of selection.

References

Kondrashov, A. S. (1984). On the intensity of selection for reproductive isolation at the beginnings of sympatric speciation. Genetika, 20:408-415.

Barton, N.H. 1992. On the spread of new gene combinations in the third phase of Wrights shifting balance. Evolution 46:551-557.

Charlesworth, B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. 55, 199-221.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. (USA) 78:3721-3725.

Higashi, M., G. Takimoto, and N. Yamamura. 1999. Sympatric speciation by sexual selection. Nature 402:523-526.

Servedio, M. R. 2000. Reinforcement and the genetics of nonrandom mating. Evolution 54:21-29.

Turelli, M., and N. H. Barton. 1994. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? Genetics 138:913-941.