Population-genomic inference of the strength and timing of selection against gene flow

Simon Aeschbacher[1,a], Jessica P. Selby[2], John H. Willis[2], and Graham Coop[1]

14 September 2016

[1]Department of Evolution and Ecology, University of California, Davis, CA 95616
[2]Department of Biology, Duke University, Durham, NC 27708
[a]saeschbacher@mac.com

Abstract

How strong is the natural selection that maintains species and locally adapted populations in the face of gene flow? To what extent is genomic divergence limited by gene flow? Here, we use DNA polymorphism data and the genome-wide variation in recombination rate to infer the strength and timing of selection, and the baseline level of gene flow under various demographic scenarios. To achieve this, we develop theory that merges the coalescent process with the concept of effective gene flow. The latter describes the reduction in gene flow at neutral loci due to divergent selection against maladapted immigrant alleles. This effect decreases with recombinational distance from the loci under selection, such that in regions of low recombination genetic divergence among populations is on average increased compared to regions of high recombination. Our inference procedure exploits this relationship in a genome-wide aggregate manner. We validate our approach using individual-based simulations and apply it to two datasets from the yellow monkeyflower (*Mimulus guttatus*). First, we infer a strong signal of adaptive divergence in the face of gene flow between populations growing on and off phytotoxic serpentine soils. We show that the genome-wide intensity of this selection is not exceptional compared to what *M. guttatus* may usually experience when adapting to local conditions. Second, we quantify and date selection against introgression from the selfing sister species *M. nasutus*. Our study provides a theoretical framework that explicitly links genome-wide patterns of divergence and recombination with the underlying evolutionary mechanisms.

When and how strongly divergent selection acts against gene flow to maintain species and locally adapted populations is a long-standing question in biology [1–3]. Answering it is fundamental to

understanding the evolution and persistence of organismal diversity, and to assessing the prevalence of different routes to speciation [4–6]. This requires quantitative estimates of the strength and timing of selection and gene flow. Population-genomic approaches, combined with a growing amount of DNA sequencing data, offer a powerful way to obtain such estimates and identify candidate speciation genes. The genome-wide perspective of these approaches is essential: genes underlying speciation and local adaptation act as barriers to gene flow, which translates into higher genetic divergence. But this is only detectable in contrast to genomic regions void of such barriers [7, 8]. In practice, a framework that explicitly links observable patterns of DNA polymorphism with the underlying evolutionary mechanisms and allows for robust parameter inference has so far been missing [8–10].

Existing population-genomic approaches to study adaptive genomic divergence in the face of gene flow come in three broad flavours. The first set of approaches apply methods for demographic inference to scenarios of speciation [11–17]. They inform our view of speciation by dating population splits and inferring the presence or absence of gene flow, yet generally do not explicitly account for natural selection [but see 18, 19].

A second set of approaches are genome scans for peaks of genetic divergence among populations or species. Such outliers are used to identify candidate loci underlying speciation or local adaptation [20–23]. This includes the search for so-called genomic islands of divergence [24–26], i.e. extended genomic regions of elevated divergence. Methods of this type can be confounded by other modes of selection and demography, and will always propose a biased subset of candidate loci [3, 27–30].

The third set are tests for a negative correlation between absolute genetic divergence and recombination rate across the genome [31–35]. This approach is based on the prediction that divergence is increased in regions of the genome where genetic linkage to loci under divergent selection is higher on average [31, 36, 37]. Such tests are in principle very powerful, as they aggregate information across the entire genome, and because the pattern of a negative correlation is highly specific to divergent selection with gene flow [38]. However, this approach is so far purely descriptive and much underexplored.

Here, we develop novel theory describing the pattern used by this latter type of approach, and an inference procedure based on this. Unlike the first two sets of approaches, ours explicitly accounts for selection and its effect on neutral variation, allows estimation of the strength and timing of selection and gene flow, and is robust to various confounding factors. We apply it to sequencing and recombination data from *Mimulus guttatus*, showing evidence for adaptive divergence and maintenance of a species barrier despite gene flow.

Idea of Approach and Population-Genomic Model

The key idea behind our approach is to exploit the genome-wide variation in recombination rate and its effect on genetic divergence between populations under selection against maladaptive gene flow. Recombination and genetic divergence are linked because divergent selection reduces effective gene flow at neutral sites, and this reduction depends on the recombinational distance from the loci under selection. In the following, we conceptualise this in terms of the effective migration rate and the expected pairwise between-population coalescence time (Fig. 1A). The latter directly relates to the absolute genetic diversity between populations, a quantity that is readily estimated from DNA sequence data. In the second part of the paper, we therefore devise an inference procedure and show its application. We start by describing our model.

We consider two populations (1, 2) of effective size N_1 and N_2, with diploid individuals and non-overlapping generations. A balance between one-way gene flow at rate m per generation from population 2 and local directional selection is maintained in population 1 for τ generations before the present. We call this the migration–selection (MS) phase (Fig. 1A). Selection against maladaptive immigrant alleles occurs at an arbitrary number of biallelic loci that we refer to as migration–selection polymorphisms (MSPs). At each MSP, allele A_1 is favoured in population 1 over allele A_2 by an average selection coefficient s, while A_2 is introduced by gene flow. We assume additive fitness interactions and no dominance.

Prior to the MS phase, we model a panmictic (P) phase in an ancestral population of effective size N_0 that starts τ generations ago and extends into the past (Fig. 1A). We call this the (MS)P demographic scenario. The P phase can be exchanged for an ancestral migration (M) phase with gene flow at rate m_0, resulting in what we call the (MS)M scenario. Here, we use the (MS)P and (MS)M scenarios to describe our approach. In the Supporting Information (SI) we provide extensions to more general scenarios that include an intermediate isolation (I) phase (SI Appendix A, Fig. S1.1, Table S1.2).

We assume that the MSPs occur at a constant rate ν per base pair, such that the distance between consecutive MSPs is approximately exponentially distributed with mean $1/\nu$ base pairs. Finally, we denote the per-base pair recombination rate by r_{bp}. In summary, the key parameters are the mean selection coefficient s, the genomic density ν of MSPs, the baseline migration rate m, and the duration τ of the MS phase.

Average Effective Gene Flow and Selection Denisty

Selection against maladapted immigrant alleles acts as a barrier to gene flow in the MS phase. At a focal neutral site, the baseline migration rate m is reduced to an effective migration rate m_e by a so-called gene-flow factor, $\mathrm{gff} = m_e/m$ [39–42]. The reduction increases with the strength of selection at the MSPs, and decreases with their recombinational distance from the neutral site (Fig. 1A). Considering the nearest I up- and J downstream MSPs, Aeschbacher and Bürger [43] showed that the effective migration rate at the neutral site can be approximated as $m_e^{(I,J)} \approx m\, g^{(I)}\, h^{(J)}$, where

$$g^{(I)} \approx \left(1 + \frac{a_1}{k_1 r_{\mathrm{bp}}}\right)^{-1} \prod_{i=2}^{I} \left(1 + \frac{a_i}{k_i r_{\mathrm{bp}} + \sum_{n=1}^{i-1} a_n}\right)^{-1}. \tag{1}$$

Here, k_i is the physical distance to, and a_i the selection coefficient at the ith upstream MSP ($h^{(J)}$ is analogous, Eq. S1.1).

To understand how this effect translates from a given neutral site to the entire genome, we average over the possible genomic locations and selection coefficients of the MSPs. In doing so, we make the simplifying assumption of an infinite chromosome with a linear relationship between physical and genetic map distance. Integrating over the distances to all MSPs and assuming an exponential distribution of selection coefficients, we find that the expected effective migration rate $\mathbb{E}[m_e^{(I,J)}]$ depends on s, ν, and r_{bp} exclusively through

$$\eta = \frac{\sigma}{r_{\mathrm{bp}}}, \tag{2}$$

where $\sigma = s\nu$ is the product of the mean selection coefficient times the density of MSPs. We call σ the 'selection density' (per base pair). For instance, with $I = J = 2$ we find

$$\mathbb{E}[m_e^{(2,2)}] \approx m(1 + 2\eta \ln \eta), \tag{3}$$

which is a good approximation if $\eta \lesssim 0.1$, i.e. if the recombination rate is at least ten times larger than the selection density, at which point effective gene flow is reduced by about 50 % (Fig. S1.2). Equation (3) shows that the average level of effective gene flow decreases with selection density and increases with recombination rate. We found that adding increasing numbers of MSPs has a diminishing effect on $\mathbb{E}[m_e^{(I,J)}]$ (Fig. S1.2A), so that (3) captures the essential pattern if $\eta \lesssim 0.1$.

As exclusive dependence of $\mathbb{E}[m_e^{(I,J)}]$ on selection and recombination through the compound parameter η holds for any I and J, it also applies to the genome-wide average of m_e (SI Appendix A, Eq. S1.8). Note that η is the selection density per genetic map unit. Its emergence here implies that doubling the number of MSPs has the same effect on average effective gene flow as doubling the mean selection coefficient. We therefore anticipate that, in practice, s and ν can be inferred only jointly as σ from population-genomic data in our framework.

Expected Pairwise Coalescence Time With Selection

To enable parameter inference from population-genomic data, we phrase our theory in terms of the expected coalescence time of two alleles, one from each population. This accounts for genetic drift at the neutral site and makes a link to the genetic divergence π_B between the two populations, which is readily estimated from DNA sequence data. The expected between-population coalescence time under neutrality, T_B, depends on the baseline migration rate m (Table S1.2). To incorporate the effect of selection, we substitute the effective migration rate m_e for m. Averaging over all numbers and locations of the MSPs, we obtain $\mathbb{E}[T_B]$, and can predict π_B as $2\,u\,\mathbb{E}[T_B]$, where u is the mutation rate per base pair and generation.

In the following, we obtain $\mathbb{E}[T_B]$ for two versions of our model, accounting either for only the nearest-neighbouring MSP (one-MSP), or all possible numbers of MSPs (multi-MSP). To better reflect real genomes, we now assume a finite genome size and define r_f as the recombination rate that corresponds to free recombination, such that MSPs located more than $k_f = r_f/r_{\mathrm{bp}}$ base pairs from a neutral site are unlinked to it.

Selection at one locus

Under the one-MSP model, we implicitly assume that ν is small, i.e. $\nu \ll r_{\mathrm{bp}}/s, m, \tau$, where s is now the selection coefficient at the single MSP. In the simplest case of the (MS)M scenario with $m_0 = m$ (Fig. S1.1C), we find that, for small ν, the expected pairwise between-population coalescence time at an average neutral site is

$$\mathbb{E}[T_B] \approx 2N_2 + \frac{1}{m} + \frac{1}{m}\frac{2s\nu}{r_{\mathrm{bp}}}\left(e^{-m\tau}D + F\right) + \frac{1}{m}\frac{s}{r_f}e^{-2\nu k_f}, \tag{4}$$

where D and F depend on m, τ, and ν, (Materials and Methods). The first two terms in (4) are the expectation without selection [45] (Table S1.2). The third and fourth term reflect the increase in coalescence time if the MSP is linked ($k_1 < k_f$) and unlinked ($k_1 \geq k_f$) to the neutral site,

respectively. Importantly, the term accounting for a linked MSP suggests that $\eta = \sigma/r_{\mathrm{bp}}$ strongly determines $\mathbb{E}[T_B]$, although s, ν, and r_{bp} also enter (4) independently. Indeed, given r_{bp} and in the parameter range where (4) is a good approximation, the effect of selection on $\mathbb{E}[T_B]$ is entirely captured by the selection density σ (Fig. S1.4). For details and other demographic scenarios, see SI Appendix A.

Selection at multiple loci

Under the multi-MSP model, we explicitly account for all MSPs possibly present in the genome and for the average physical chromosome length. Finding $\mathbb{E}[T_B]$ is in principle similar to the one-MSP model above, but amounts to averaging over all possible numbers and genomic locations of the MSPs. We wrote a Monte-Carlo integration program to do this (SI Appendix A). The result agrees very well with individual-based forward simulations (Figs. 1B and S1.7). As for the one-MSP model, if r_{bp} is given, $\mathbb{E}[T_B]$ depends on s and ν effectively only through the selection density σ (Fig. 1C). In fact, falling back on the idealising assumption of a global linear relationship between physical and genetic map distance (i.e. $r_f \to \infty$), we can prove that this holds exactly (SI Appendix A, Eq. S1.59). This corroborates σ as a key parameter and a natural metric to quantify genome-wide divergent selection in the face of gene flow.

Application to *Mimulus guttatus*

Based on our theory above, we developed an inference procedure and applied it to two datasets from the predominantly outcrossing yellow monkeyflower (*Mimulus guttatus*), an important model system for speciation [33, 46–55] and local adaptation [56–61] native to western North America. In the first dataset, we detected a strong signal of local adaptation in the face of gene flow. In the second, we inferred the strength and timing of selection against introgression from the selfing sister species *M. nasutus* (SI Text 1).

We fitted our model to the empirical relationship between recombination rate (r_{bp}, estimated from a linkage map) and putatively neutral between-population diversity (π_B, estimated from 4-fold degenerate coding sites), after correcting the latter for genomic correlates and divergence to the outgroup *M. dentilobus* (SI Appendix B). Our inference procedure computes the sum of squared deviations (SSD) across genomic windows between these estimates of π_B and those predicted by our model given the estimate of r_{bp} for each window and for a given set of parameter values. Minimising the SSD over a large grid of parameter values, we obtained point estimates for the selection density (σ), baseline migration rate (m), and duration of the MS phase (τ). We estimated 95 % non-parametric confidence intervals (CIs) for the parameters by doing a block-bootstrap over genomic windows (SI Appendix B). For both datasets, we fitted the (MS)P demographic scenario (Figs. 1A and Fig. S1.1D). In the following, we report results obtained under the multi-MSP model with genomic windows of 500 kb. Results for windows of 100 and 1000 kb were very similar (SI Appendix B, SI Text 2). Genetic diversity *within* populations was not positively correlated with recombination rate, so that background selection is unlikely to bias our findings [62–64].

Adaptive divergence maintained in the face of gene flow

Like many species across different plant taxa [65–67], *M. guttatus* has locally adapted to serpentine outcrops occuring throughout its range [68, p. 4]. Serpentine soils are characterised by high concentrations of heavy metals and a low calcium-to-magnesium ratio [69, 70], conditions that are toxic to plants [71, 72]. While the mechanism and molecular basis of serpentine tolerance in *M. guttatus* are unresolved [73], strong differences in survival on serpentine soil exist between serpentine and non-serpentine *M. guttatus* ecotypes, both in the greenhouse and in nature [68]. To investigate a population-genomic signal of local adaptation to serpentine soil, we sequenced the whole genomes of 324 *M. guttatus* individuals collected from two pairs of geographically close populations growing on and off serpentine soil in California in a pooled approach (the serpentine dataset; Fig. 2A, SI Text 1). We started by estimating the strength of selection in serpentine populations (REM and SLP) against maladaptive immigrant alleles from the geographically closer off-serpentine population (SOD and TUL, respectively), assuming that the off-serpentine population in each pair is a proxy for the source of gene flow.

Fitting our model to the data, we found that the conditional surface of the $-$SSD (holding m and τ at their point estimates) showed a pronounced ridge for s and ν, with the 95 % confidence hull falling along this ridge (Fig. 2B). With parameters on a $\log_{10}$ scale, the slope of this ridge is -1, nicely confirming our theoretical result that s and ν can only be estimated jointly as their product, the selection density σ. We therefore adjusted our inference procedure to jointly infer m, τ, and σ instead of m, τ, s, and ν (SI Appendix B). This resulted in conditional $-$SSD surfaces for σ and m with a unique peak and tight confidence hulls (Fig. 2C).

For both serpentine $\times$ off-serpentine pairs, we found a strong genome-wide signal of divergent selection against gene flow, with point estimates for σ of about 10^{-3} and 10^{-4} per megabase (Mb) in REM $\times$ SOD and SLP $\times$ TUL, respectively, and tight 95 % CIs (Fig. 3A, C; SI File S4.2). Given an assembled genome size of about 320 Mb for *M. guttatus*, this would for instance be consistent with about 300 MSPs, each with a selection coefficient of about 10^{-3} to 10^{-4}. The strong impact of this selection on genome-wide levels of polymorphism is visible from the red curves in Figs. 3B, D that represent the model fit. The 95 % CI of the relative difference (δ_π) between the maximum and minimum of this fitted curve for π_B clearly excludes 0 (Fig. 3B, D; SI Appendix B). According to our estimates of m, selection maintains this divergence against an average baseline level of gene flow of about $10^{-5.4}$ in REM $\times$ SOD and $10^{-5.7}$ in SLP $\times$ TUL (Fig. 3B, C). Given the estimated effective population sizes of REM and SLP (SI Text 1), this implies high rates of about 2.5 and 1.3 diploid immigrants per generation, respectively.

We had no power to infer precise point estimates for τ, but lower bounds of the 95 % CIs were around 10 Mya. Repeating our analyses for the (MS)M demographic scenario (Fig. S4.15), as well as when considering all non-focal populations jointly as source of gene flow (Figs. S4.14, S4.16), we obtained very similar results to those above. Our inference about selection and gene flow therefore seems to be robust to the unknown specifics of the demography.

To assess if the selection against gene flow we found is specific to serpentine $\times$ off-serpentine comparisons (REM $\times$ SOD, SLP $\times$ TUL), we also fitted our model for the two long-distance off-serpentine $\times$ off-serpentine configurations (SOD $\times$ TUL, TUL $\times$ SOD) as well as the long-distance serpentine $\times$ off-serpentine pairs (REM $\times$ TUL, SLP $\times$ SOD). Interestingly, we inferred selection densities, durations of the migration-selection regime, and migration rates on the same order as those estimated for the short-distance serpentine $\times$ off-serpentine comparisons (Fig. S4.13, SI File S4.2). This suggests that the signal we detect may have little to do with local adaptation to

serpentine per se, and may not be specific to the history of particular pairs of populations. Rather, given the long time τ that this selection appears to have acted over, we view our estimates as reflecting adaptive divergence in response to a range of locally varying conditions that *M. guttatus* experiences across its entire range [56–61].

Persistence of species barrier to *M. nasutus*

Where *M. guttatus* has come into secondary contact with its selfing sister species *M. nasutus*, hybridisation occurs [33, 55] despite strong reproductive barriers [46–50, 52–55]. A genome-wide analysis found large genomic blocks of recent introgression from *M. nasutus* into *M. guttatus* [33, 55]. In the same study, absolute divergence ($\pi_B = \pi_{\text{Gut}\times\text{Nas}}$) was found to be negatively correlated with recombination rate (r_{bp}) in sympatric, but not allopatric Gut $\times$ Nas comparisons, consistent with selection against this introgression [33]. We reanalysed whole-genome sequences from single individuals and recombination data from this study (the GutNas dataset) with our new method. While we replicate the negative, albeit weak, partial correlation between π_B and r_{bp} in sympatric comparisons (Fig. S4.10, SI Text 2), our estimates of σ were generally very low, except for the allopatric SLP $\times$ Nas pair (Fig. S4.17). Our model fit showed an uptick of π_B at low values of r_{bp} in all comparisons, yet the 95 % CIs included the case of neutrality ($\delta_\pi = 0$), except for SLP $\times$ Nas (S4.17). After removal of genomic windows that fall inside the blocks of recent introgression [33], our estimates of σ were significantly different from neutrality in *both* allopatric pairs (AHQ $\times$ Nas: $10^{-3.8}$ per Mb; SLP $\times$ Nas: 10^{-3} per Mb), but remained non-significant in sympatric ones (CAC $\times$ Nas, DPR $\times$ Nas) (Figs. 4, S4.18).

With blocks of recent introgression excluded, we estimated m to be on the order of 10^{-6} with fairly tight 95 % CIs (Figs. 4, S4.18). Although small in absolute value, when scaled by the effective population size, these estimates would imply an average net influx of about 0.1 to 1 diploid genomes per generation in the absence of selection. Lower confidence bounds for τ were consistently above 250 kya, with point estimates between about 500 kya (SLP $\times$ Nas) and 1.1 Mya (DPR $\times$ Nas), and no systematic difference between allopatric and sympatric comparisons (Figs. 4, S4.18). These estimates are somewhat above a previous estimate of about 196 kya for the onset of divergence between *M. guttatus* and *M. nasutus* [33]. Our older estimates of τ are compatible with divergent selection acting already in the ancestral, geographically structured, *M. guttatus* clade before speciation [33].

In summary, we found evidence for divergent selection maintaining a species barrier against gene flow over at least the last 250 to 500 ky. The current extent of range overlap with *M. nasutus* is not predictive of the strength of selection that we infer. Our results also suggest that a signal of selection against historical gene flow [74] might be partially masked by blocks of recent introgression that have not yet converged to the pattern predicted by our model (Fig. S4.12; SI Text 2).

Discussion

The genomes of incompletely isolated species and locally adapted populations have long been thought of as mosaics of regions with high and low divergence [55, 75–79]. This pattern is due in part to varying levels of effective gene flow along the genome, created by an interaction of divergent selection and recombination-rate variation [39–41, 80, 81]. This idea has shaped the modern view of local adaptation and speciation since its inception [7, 82]. The recent explosion of genome-

wide DNA sequencing data has created the opportunity to directly observe this mosaic. It has spurred theoretical and empirical studies aiming to better understand the mechanisms underlying local adaptation and speciation [4, 83–87], as well as identify the genetic architecture of the traits involved [88–96]. Yet, an explicit, model-based framework linking observed genome-wide patterns of divergence with the underlying mechanism has hitherto been missing [10, 97].

Here, we developed such a framework by merging the concept of effective migration rate with coalescence theory. One key insight that emerges from this work is that the empirical pattern of a genome-wide negative relationship of between-population diversity with recombination rate [e.g. 33, 38] can be described by the compound parameter 'selection density'. Its implication is that, at a genome-wide level, very different mosaic patterns lead to the same average outcome: a large number of weak genetic barriers to gene flow (MSPs) is equivalent to a much smaller number of strong barriers. Our approach therefore complements existing genome scans for empirical outliers of population divergence that can only hope to identify loci that are strong barriers to gene flow [29, 98–100]. It also could provide a better null model for such genome scans, as outliers could be judged against an appropriate background level of divergence for their local recombination rate.

The idea behind our approach is inspired by earlier work exploiting the *positive* genome-wide relationship between recombination rate and genetic diversity *within* a population for quantitative inference about genetic hitchhiking [101, 102] and background selection [36, 64]. While in *Mimulus* we did not observe a correlation between recombination rate and genetic diversity within populations, this earlier work would offer a natural way to correct for the potentially confounding effects of these types of selection at linked sites within our framework. Specifically, we could jointly fit a model of background selection or selective sweeps within pouplations [e.g. 103] and our existing model for divergent selection against gene flow.

We have assumed that the genetic barriers to gene flow occur at a constant rate (ν) along the genome. We could improve on this by making ν depend on the functional annotation of genomes, e.g. exon coordinates, which might allow ν and s to be estimated separately [see 104]. Our model also does not account for the clustering of locally adaptive mutations arising in tight linkage to previously established MSPs [43, 105], and the synergistic sheltering effect among MSPs that helps them against being swamped by gene flow [106]. If accounted for, this clustering would lead to an even more pronounced uptick of between-population diversity in regions of low recombination. Therefore, one might be able to use deviations from our current model in regions of low recombination as a way of detecting the presence of clustering in empricial data. Or, at the very least, our parameter estimates would offer a guide to whether we should see such an effect, and in which genomic regions clustering might be expected to have evolved.

An inherent limitation to our approach is that the populations or species must have had sufficient time for neutral divergence to accumulate. Otherwise, there is no power to detect variation in divergence among regions. This constrains the temporal resolution of our model, in particular if the duration of the migration–selection (MS) phase is short, which would be the case if the divergence time is low, or if strong reproductive isolation evolved so quickly that gene flow was completely and rapidly reduced across the entire genome. Another potential limitation is a relatively low resolution to infer the duration of the MS phase. A genome-wide negative correlation of recombination rate with between-population diversity will persist for a long time even after gene flow has come to a complete halt, as subsequent neutral divergence should then just add uniformly to the existing pattern. Our inference approach should therefore still provide good estimates of the strength of selection and gene flow even after speciation has completed, as long as these estimates are interpreted as averages over the inferred time τ. In this sense, our approach is likely robust to the

specifics of the most recent demographic history of the populations or species of interest. To better resolve the timing of events, we suggest using the additional information contained in the entire distribution of pairwise coalescence times, rather than relying on their mean, as we currently do.

The contrasting roles of gene flow and selection in speciation and local adaptation have a long and contentious history in evolutionary biology and population-genetics theory [4, 107]. In our view, a promising way forward is to provide quantitative, genome-wide statements about the strength and timing of selection and gene flow, so that empricial observations can be placed on a firm theoretical footing. We anticipate that the type of approach developed here, applied to the growing number of datasets with genome-wide polymorphism and recombination data, may help to resolve the role of gene flow in constraining divergent selection.

Materials and Methods

The terms D and F in (4) are $D = \mathrm{Ei}[(1-g_f)m\tau] - \mathrm{Ei}[(1-g_\circ)m\tau]$ and $F = \mathrm{Ei}[-g_\circ m\tau] - \mathrm{Ei}[-g_f m\tau] + \mathrm{Ei}[-\nu k_f] - \mathrm{Ei}[-\nu k_\circ]$, where $\mathrm{Ei}[z] = -\int_{-z}^{\infty} e^{-t}/t\,\mathrm{d}t$ is the exponential integral. Here, $g_f = [1+s/r_f]^{-1}$ and $g_\circ = [1+s/(k_\circ r_{\mathrm{bp}})]^{-1}$ are the contributions to the gff if the MSP is unlinked ($k_1 = r_f/r_{\mathrm{bp}}$) or if it is maximally linked ($k_1 = k_\circ$, $0 < k_\circ \lesssim 1/[r_{\mathrm{bp}}\tau]$), respectively, and $k_\circ$ is a small positive lower limit for the physical distance to the MSP.

For a detailed description of our model and theory and of the individual-based simulations, see SI Appendix A. Statistical data analyses, bias corrections, and the inference procedure are described in detail in SI Appendix B. The *Mimulus* datasets (sampling design, DNA sequencing, quality filtering), and the linkage map are discussed in SI Text 1. For complementary results, including tests of partial correlation between diversity and recombination rate, see SI Text 2.

Acknowledgments

We thank Yaniv Brandvain, Lex Flagel, Amanda Kenney, Andrea Sweigart, Kevin Wright, Chenling Xu, as well as members of the Coop, Ross-Ibarra, and Schmitt labs at UC Davis for helpful discussions and comments that improved our analyses. We thanks Ben Blackman for help with collections. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award numbers NIH RO1GM83098 and RO1GM107374 awarded to GC, grant no. 1353380 from the U.S. National Science Foundation to GC and JW, Ddig grant no. 1110753 from the National Science Foundation to JS, and an Advanced Postdoc.Mobility fellowship from the Swiss National Science Foundation P300P3_154613 to SA. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).

References

[1] Endler JA (1973) Gene flow and population differentiation. *Science* 179(4070):243–250.

[2] Mayr E (1996) What is a species, and what is not? *Philosophy of Science* 63(2):262–277.

[3] Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23:3133–3157.

[4] Coyne JA, Orr HA (2004) *Speciation*, vol 37 (Sinauer Associates Inc, Sunderland, MA), 1st edn.

[5] Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution* 20(5):229–237.

[6] Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics* 15(3):176–192.

[7] Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* 14(6):851–865.

[8] Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1587):332–342.

[9] Gavrilets S (2014) Models of speciation: Where are we now? *Journal of Heredity* 105(S1):743–755.

[10] Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology* 25(11):2337–2360.

[11] Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167(2):747–760.

[12] Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution* 63(10):2547–2562.

[13] Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution* 27(6):1425–1435.

[14] Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* 23(11):1817–1828.

[15] Frantz LAF, Madsen O, Megens HJ, Groenen MAM, Lohse K (2014) Testing models of speciation from genome sequences: divergence and asymmetric admixture in island South-East Asian *Sus* species during the Plio-Pleistocene climatic fluctuations. *Molecular Ecology* 23(22):5566–5574.

[16] Hearn J, Stone GN, Bunnefeld L, Nicholls JA, Barton NH, et al. (2014) Likelihood-based inference of population history from low-coverage de novo genome assemblies. *Molecular Ecology* 23(1):198–211.

[17] Filatov DA, Osborne OG, Papadopoulos AS (2016) Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna. *Molecular Ecology* 25(11):2467–2481.

[18] Bazin E, Dawson KJ, Beaumont MA (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics* 185(2):587–602.

[19] Kousathanas A, Leuenberger C, Helfer J, Quinodoz M, Foll M, et al. (2016) Likelihood-free inference in high-dimensional models. *Genetics* 203(2):893–904.

[20] Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74(1):175–195.

[21] Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B: Biological Sciences* 263(1377):1619–1626.

[22] Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180(2):977–993.

[23] Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *American Naturalist* 186(S1):S24–S36. PMID: 26656214.

[24] Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1587):343–353.

[25] Marques DA, Lucek K, Haesler MP, Feller AF, Meier JI, et al. (2016) Genomic landscape of early ecological speciation initiated by selection on nuptial color. *Molecular Ecology* pp n/a–n/a.

[26] Rougemont Q, Gagnaire PA, Perrier C, Genthon C, Besnard AL, et al. (2016) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and non-parasitic lamprey ecotypes. *Molecular Ecology* pp n/a–n/a.

[27] Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity* 103(4):283–284.

[28] Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* 23(9):2178–2192.

[29] Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* 24(5):1031–1046.

[30] Jensen JD, Foll M, Bernatchez L (2016) The past, present and future of genomic scans for selection. *Molecular Ecology* 25(1):1–4.

[31] Nachman MW (1997) Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* 147(3):1303–1316.

[32] Keinan A, Reich D (2010) Human population differentiation is strongly correlated with local recombination rate. *PLoS Genetics* 6(3):e1000886.

[33] Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics* 10(6):e1004410 EP.

[34] Carneiro M, Ferrand N, Nachman MW (2009) Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 181(2):593–606.

[35] Geraldes A, Basset P, Smith KL, Nachman MW (2011) Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Molecular Ecology* 20(22):4722–4736.

[36] Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* 70(2):155–174.

[37] Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* 15(5):538–543.

[38] Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1587):409–421.

[39] Petry D (1983) The effect on neutral gene flow of selection at a linked locus. *Theoretical Population Biology* 23(3):300–313.

[40] Bengtsson BO (1985) The flow of genes through a genetic barrier. *Evolution – Essays in honour of John Maynard Smith*, eds Greenwood PJ, Harvey P, Slatkin M (Cambridge University Press, New York, NY), vol 1, chap 3, pp 31–42.

[41] Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. *Heredity* 57(3):357–376.

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

[43] Aeschbacher S, Bürger R (2014) The effect of linkage on establishment and survival of locally beneficial mutations. *Genetics* 197(1):317–336.

[44] Harrison S, Safford H, Wakabayashi J (2004) Does the age of exposure of serpentine explain variation in endemic plant diversity in California? *International Geology Review* 46(3):235–242.

[45] Notohara M (1990) The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology* 29(1):59–75.

[46] Vickery RK (1964) Barriers to gene exchange between members of the mimulus guttatus complex (scrophulariaceae). *Evolution* 18(1):52–69.

[47] Kiang YT, Hamrick JL (1978) Reproductive isolation in the mimulus guttatus m. nasutus complex. *American Midland Naturalist* 100(2):269–276.

[48] Sweigart AL, Fishman L, Willis JH (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics* 172(4):2465–2479.

[49] Martin NH, Willis JH (2007) Ecological divergence associated with mathing system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61(1):68–82.

[50] Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* 62(5):1026–1039.

[51] Wu CA, Lowry DB, Cooley AM, Wright KM, Lee YW, et al. (2008) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220–230.

[52] Martin NH, Willis JH (2010) Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365(1552):2469–2478.

[53] Fishman L, Sweigart AL, Kenney AM, Campbell S (2014) Major quantitative trait loci control divergence in critical photoperiod for flowering between selfing and outcrossing species of monkeyflower (*Mimulus*). *New Phytologist* 201(4):1498–1507.

[54] Sweigart AL, Flagel LE (2015) Evidence of natural selection acting on a polymorphic hybrid incompatibility locus in *Mimulus*. *Genetics* 199(2):543–554.

[55] Kenney AM, Sweigart AL (2016) Reproductive isolation and introgression between sympatric *Mimulus* species. *Molecular Ecology* 25(11):2499–2517.

[56] Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* 62(9):2196–2214.

[57] Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist* 183(3):776–788.

[58] Wu CA, Lowry DB, Nutter LI, Willis JH (2010) Natural variation for drought-response traits in the mimulus guttatus species complex. *Oecologia* 162(1):23–33.

[59] Oneal E, Lowry DB, Wright KM, Zhu Z, Willis JH (2014) Divergent population structure and climate associations of a chromosomal inversion polymorphism across the *Mimulus guttatus* species complex. *Molecular Ecology* 23(11):2844–2860.

[60] Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK (2015) Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual *Mimulus guttatus*. *New Phytologist* 206(1):152–165.

[61] Wright KM, Hellsten U, Xu C, Jeong AL, Sreedasyam A, et al. (2015) Adaptation to heavy-metal contaminated environments proceeds via selection on pre-existing genetic variation. *bioRxiv* .

[62] Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.

[63] Hudson RR, Kaplan NL (1995) Deleterious background selection with recombination. *Genetics* 141(4):1605–17.

[64] Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on background selection. *Genetics Research* 67(2):159–174.

[65] Jurjavcic N, Harrison S, Wolf A (2002) Abiotic stress, competition, and the distribution of the native annual grass *Vulpia microstachys* in a mosaic environment. *Oecologia* 130(4):555–562.

[66] Wright JW, L SM, Rosa S (2006) Local adaptation to serpentine and non-serpentine soils in *Collinsia sparsiflora*. *Evolutionary Ecology Research* 8(1):1–21.

[67] Sambatti JBM, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* 60(4):696–710.

[68] Selby J (2014) *The genetic basis of local adaptation to serpentine soils in* Mimulus guttatus. Doctoral dissertation, Duke University.

[69] Kruckeberg AR (1951) Intraspecific variability in the response of certain native plant species to serpentine soil. *American Journal of Botany* 38(6):408–419.

[70] O'Dell RE, James JJ, Richards James H (2006) Congeneric serpentine and nonserpentine shrubs differ more in leaf Ca:Mg than in tolerance of low N, low P, or heavy metals. *Plant and Soil* 280(1-2):49–64.

[71] Safford HD, Viers JH, Harrison SP (2005) Serpentine endemism in the California flora: a database of serpentine affinity. *Madroño* 52(4):222–257.

[72] Kay KM, Ward KL, Watt LR, Schemske DW (2011) Plant speciation. *Serpentine: The Evolution and Ecology of a Model System*, eds Harrison S, Rajakaruna N (University of California Press, Berkeley, CA), pp 71–96.

[73] Palm E, Brady K, Van Volkenburgh E (2012) Serpentine tolerance in *Mimulus guttatus* does not rely on exclusion of magnesium. *Functional Plant Biology* 39:679–688.

[74] Sweigart AL, Willis JH (2003) Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution* 57(11):2490–2506.

[75] Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics* 24:119–152.

[76] Carneiro M, Blanco-Aguiar JA, Villafuerte R, Ferrand N, Nachman MW (2010) Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution* 64(12):3443–3460.

[77] Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1587):451–460.

[78] Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, et al. (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* 22(3):842–855.

[79] Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology* 25(11):2454–2466.

[80] Barton N (1979) Gene flow past a cline. *Heredity* 43(3):333–339.

[81] Harrison RG (1986) Pattern and process in a narrow hybrid zone. *Heredity* 56(3):337–349.

[82] Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* 105(S1):795–809.

[83] Kronforst M, Salazar C, Linares M, Gilbert L (2007) No genomic mosaicism in a putative hybrid butterfly species. *Proceedings of the Royal Society of London Series B: Biological Sciences* 274:1255–1264.

[84] Bürger R, Akerman A (2011) The effects of linkage and gene flow on local adaptation: A two-locus continent–island model. *Theoretical Population Biology* 80(4):272–288.

[85] Via S, Conte G, Mason-Foley C, Mills K (2012) Localizing F_{ST} outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular Ecology* 21(22):5546–5560.

[86] Hemmer-Hansen J, Nielsen EE, Therkildsen NO, Taylor MI, Ogden R, et al. (2013) A genomic island linked to ecotype divergence in atlantic cod. *Molecular Ecology* 22(10):2653–2667.

[87] Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America* 110(19):1743–1751.

[88] Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152(2):713–727.

[89] Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363(1506):3009–3021.

[90] Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11(3):175–180.

[91] Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, et al. (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64(2):472–485.

[92] Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD-sequencing. *Evolution* 67(9):2483–2497.

[93] Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, et al. (2014) The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genetics* 10(8):e1003519 EP.

[94] Burgarella C, Chantret N, Gay L, Prosperi JM, Bonhomme M, et al. (2016) Adaptation to climate through flowering phenology: a case study in *Medicago truncatula*. *Molecular Ecology* 25(14):3397–3415.

[95] Ferris KG, Barnett LL, Blackman BK, Willis JH (2016) The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. *Molecular Ecology* pp n/a–n/a.

[96] Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology* 25(10):2144–2164.

[97] Byers KJRP, Xu S, Schlüter PM (2016) Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology* pp n/a–n/a.

[98] Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13(4):969–980.

[99] Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, et al. (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1587):364–373.

[100] Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology* 25(1):5–23.

[101] Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356(6369):519–520.

[102] Wiehe TH, Stephan W (1993) Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Molecular Biology and Evolution* 10(4):842–854.

[103] Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, et al. (2016) A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics* 12(8):e1006130.

[104] Jurić I, Aeschbacher S, Coop G. The strength of selection against Neanderthal introgression. Preprint available at http://biorxiv.org/content/early/2015/10/30/030148.

[105] Yeaman S, Aeschbacher S, Bürger R (2016) The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology* 25(11):2542–2558.

[106] Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* 65(7):1897–1911.

[107] Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology and Evolution* 16(7):330–343.

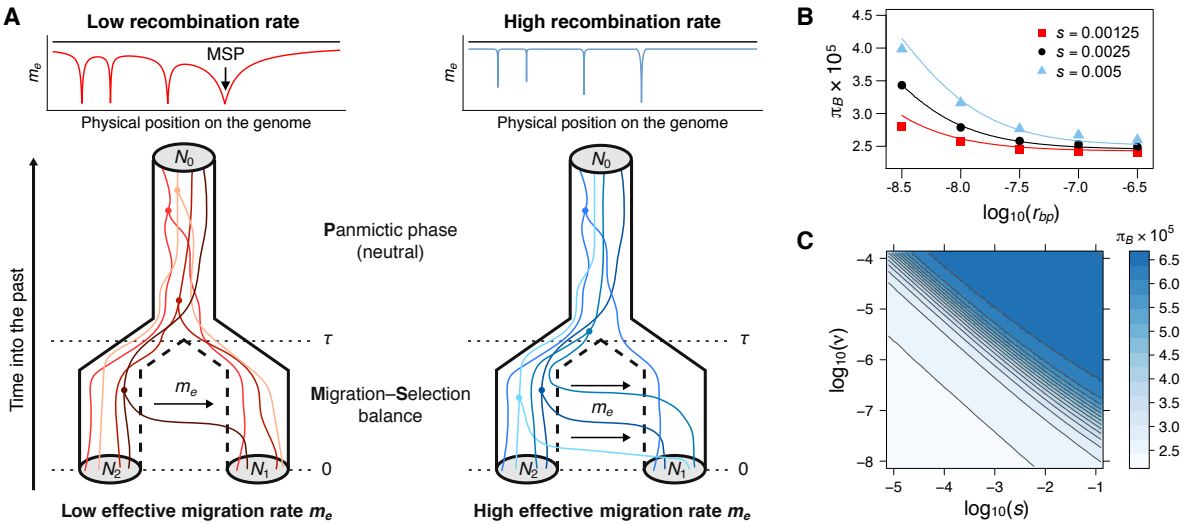

Figure 1. Divergent selection reduces gene flow and increases genetic differentiation between populations. (A) Selection against locally maladapted alleles at migration–selection polymorphisms (MSPs) reduces the effective migration rate m_e into population 1. The reduction is stronger in regions of low recombination (red, top left) and decreases the probability that lineages sampled in different populations migrate and coalesce. Multiple realisations of the coalescent process are shown in the bottom left for the (MS)P scenario (Fig. S.1.1). In regions of high recombination, m_e is reduced much less (blue, top right), such that migration events and earlier between-population coalescence times are more likely (bottom right). (B) Comparison of the predicted between-population diversity $\pi_B = 2u\mathbb{E}[T_B]$ (curves) with individual-based simulations (dots) as a function of the recombination rate r_{bp} for different selection coefficients s. Error bars ($\pm$SE) are too short to be visible. The (MS)M multi-MSP scenario was used with $N_2 = 5000$, $u = 10^{-9}$, $\nu = 2.5 \times 10^{-7}$, $m = m_0 = 5 \times 10^{-4}$, and $\tau = 2 \times 2N_2$. (C) Approximately linear contour lines with slope -1 in the surface of π_B as a function of $\log_{10}(s)$ and $\log_{10}(\nu)$ support the compound parameter selection density, $\sigma = s\,\nu$. Here, $r_{\text{bp}} = 10^{-8}$ (1 cM/Mb), $r_f = 0.5$, and the other parameters are as in (B).

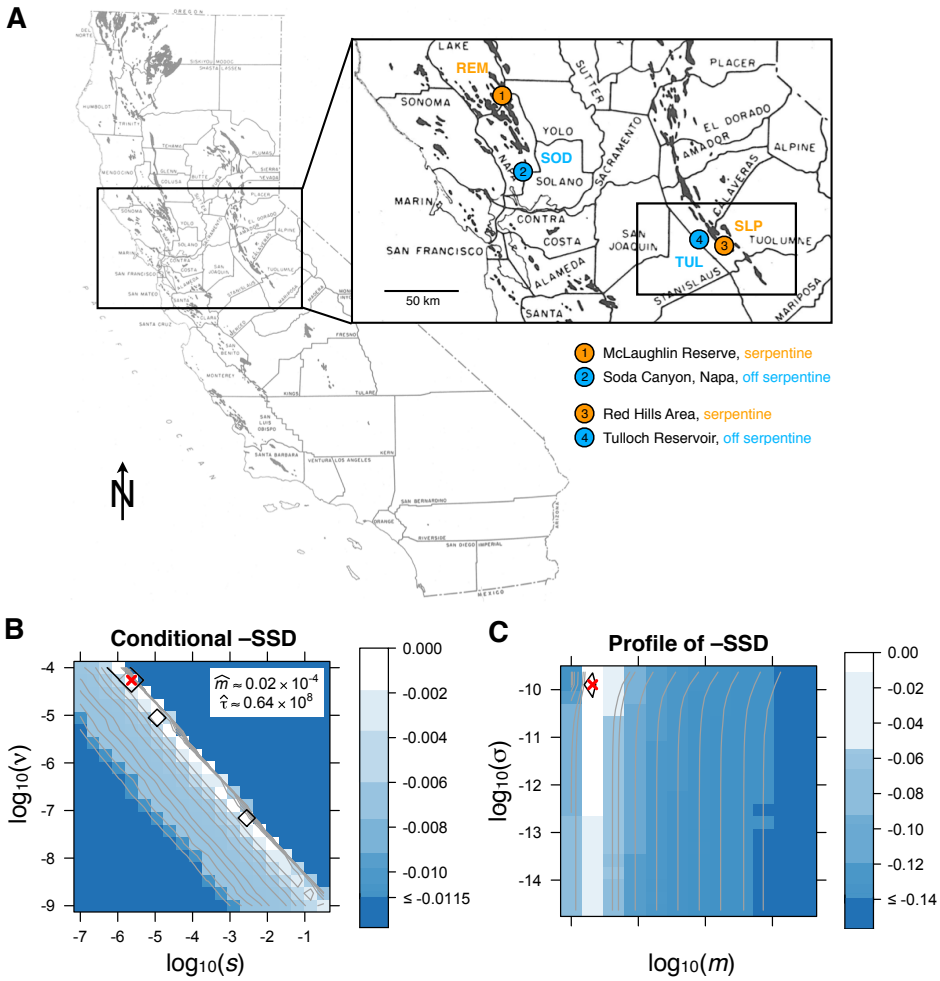

Figure 2. Geographic context of serpentine dataset and quasi-likelihood surfaces for population pair SLP × TUL. (A) Map of sampling sites in California, USA (modified with permission [44]). (B) Surface of the negative sum of squared deviations (−SSD) for the selection coefficient s and the genomic density ν of MSPs, conditional on point estimates of m and τ. The ridge with slope −1 confirms the compound parameter selection density, $\sigma = s\nu$. The cross indicates the point estimate and black hulls comprise the area of 95% bootstrap confidence. (C) Joint profile surface of the −SSD for the baseline migration rate m and the selection density σ, maximised over τ. Results are shown for the SLP × TUL pair under the multi-MSP model and the (MS)P scenario, and with genomic windows of size 500 kb.

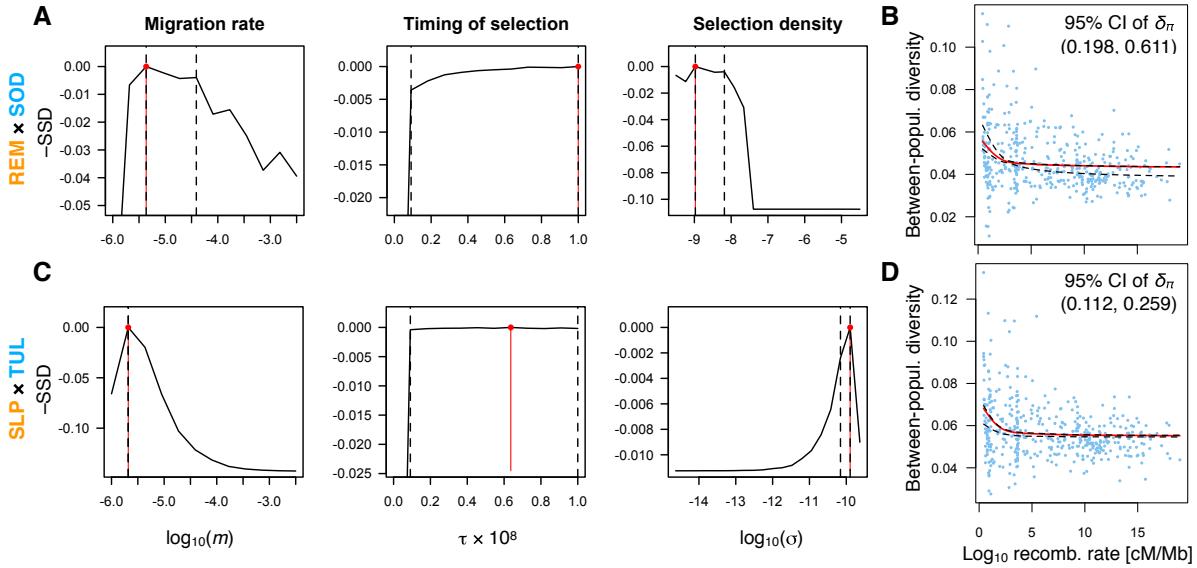

Figure 3. Parameter estimates and model fit for the serpentine dataset. (A, C) Profile curves of the quasi-likelihood (−SSD) for each parameter, maximising over the two remaining parameters, for the serpentine × off-serpentine comparisons REM × SOD (A) and SLP × TUL (C) (Fig. 2). Vertical red and black dashed lines indicate the point estimate and 95% bootstrap confidence intervals, respectively. (B, D) Raw data (blue dots) and model fit (red curve) with 95% confidence range (black dashed curves). The 95% confidence interval of the distribution of the relative difference between the maximum and minimum value of the model fit across all bootstrap samples, δ_π, is also given. Other details as in Fig. 2B–C. For other population pairs and the (MS)M scenario, see Figs. S4.13–S4.16 in SI Text 2.

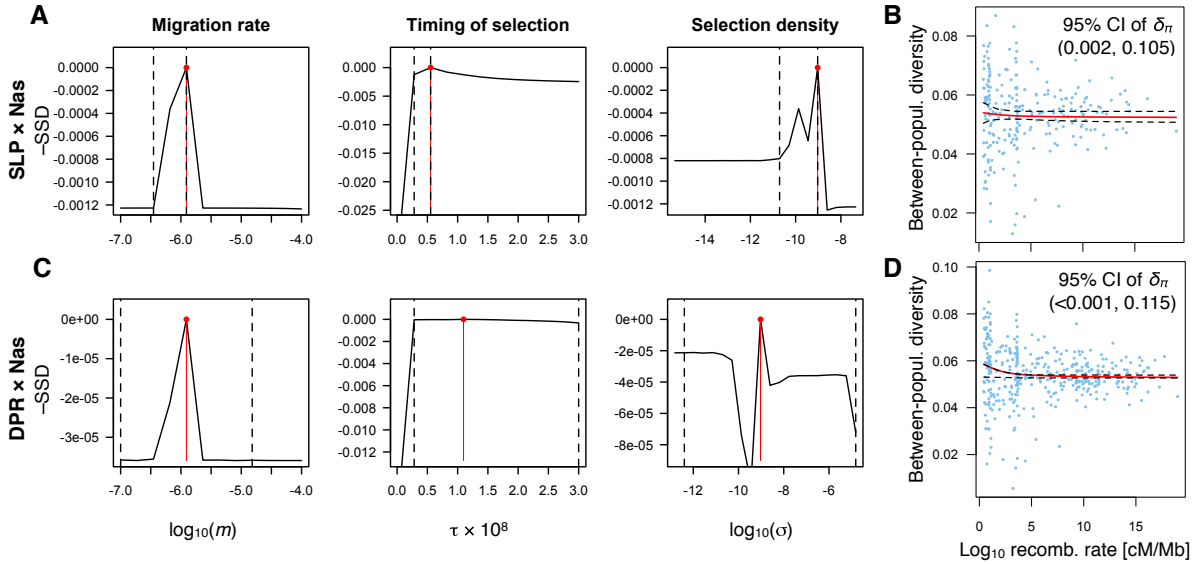

Figure 4. Parameter estimates and model fit for the Southern population pairs in the GutNas dataset. (A, C) Profile curves of the quasi-likelihood (−SSD) for each parameter, maximising over the two remaining parameters, for SLP × Nas (micro-allopatric) and DPR × Nas (sympatric), respectively. (B, D) Raw data, model fit with 95% confidence range, and the 95% confidence interval of the distribution of the relative difference δ_π between the maximum and minimum value of the model fit across bootstrap samples. Results are shown for the multi-MSP model under the (MS)P scenario after removal of blocks of recent introgression. Other details as in Fig. 3. For the Northern clade, see Fig. S4.18, and for results with blocks of recent introgression included, see Fig. S4.17.