# The signal of admixture can decay rapidly when using clustering-based methods

Josia Shemuel[1], Sonal Singhal[2], Craig Moritz[1], Stephen M. Zozaya[1*]

[1]Division of Ecology & Evolution, The Australian National University, Canberra, ACT, Australia.

[2]Department of Biology, California State University, Dominguez Hills, Carson, CA, United States.

*Corresponding author: stephen.zozaya@anu.edu.au

**Data accessibility:** All data and code are available on GitHub at https://www.github.com/joshemuel/admixture-decay

## Abstract

Gene flow shapes evolutionary trajectories by introducing novel alleles, facilitating or retarding adaptation, or eroding divergence among populations. Studies commonly infer gene flow through estimates of genetic admixture from methods that cluster individuals by allele-frequency similarity. However, the ability of these methods to reliably detect historical admixture remains poorly understood, particularly under spatially restricted dispersal and across a range of migration rates. Here we evaluate how signals of admixture arise and decay during and after gene flow across a range of within- and between-population dispersal rates, using three common inference methods: ADMIXTURE, sNMF, and PopCluster. Using forward-time simulations in a two-dimensional landscape, we modelled two metapopulations that diverged in isolation, experienced a pulse of secondary contact, and subsequently returned to isolation. All three methods systematically underestimated admixture, with the bias increasing once gene flow ceased: inferred admixture averaged 56% of the true value at the end of gene flow, declined to 23% after 1,000 generations, and fell below 5% by 2,000 generations, even where true admixture exceeded 20%. This underestimation was accompanied by corresponding overestimation of barrier strength, assessed using a metric analogous to cline width. Both patterns reflect the rapid re-establishment of Hardy–Weinberg and linkage equilibrium following secondary contact, which erodes the allele-frequency differences upon which these methods depend. Consequently, the absence of inferred admixture in present-day samples should not be taken as evidence against historical gene flow.

# 1 Introduction

Gene flow—the exchange of genetic material between populations or species—can have diverse evolutionary consequences. Limited gene flow can introduce beneficial alleles (adaptive introgression: Hedrick 2013; Arnold and Kunte 2017; Suarez-Gonzalez et al. 2018) and even trigger speciation and adaptive radiation (Anderson and Stebbins 1954; Seehausen 2004; Wogan et al. 2023). In contrast, high levels of gene flow may disrupt locally adapted gene complexes (Lenormand, 2002) or lead to genetic swamping (Rhymer and Simberloff 1996; Todesco et al. 2016), resulting in the merging of divergent populations or species. Estimating gene flow therefore has broad relevance across evolutionary biology and conservation, with applications spanning population structure inference, species delimitation, detecting hybridisation, and wildlife management.

A widely used approach for assessing gene flow involves clustering-based methods that estimate the posterior probability of an individual belonging to one population versus another—effectively interpreted as the proportion of the genome inherited from each parental population. The first implementation of this method (STRUCTURE; Pritchard et al. 2000) employs a Bayesian framework to assign individuals to a specified number of genetic clusters while estimating the contribution of each cluster to an individual's genome. This approach has influenced fields ranging from human genetics to molecular ecology by revealing complex patterns of admixture within and between populations or species (Rosenberg et al., 2002). By addressing some limitations of the original Bayesian approach, subsequent methods have improved computational efficiency and scalability. For instance, ADMIXTURE (Alexander et al., 2009) uses a likelihood-based optimisation strategy that produces ancestry estimates comparable to those from STRUCTURE, yet with substantially reduced computational demands. More recent advances include methods such as sNMF, which applies matrix factorisation to rapidly and accurately estimate individual ancestry coefficients (Frichot et al., 2014), and PopCluster, which is designed for the analysis of large genomic datasets (Wang

2022; Wang 2024).

Clustering methods infer ancestry by partitioning allele-frequency variation among individuals into discrete source populations. Despite their wide adoption, recent simulation studies have revealed several important sources of bias and estimation error. Spatial allele-frequency gradients arising from isolation by distance can be misidentified as admixture, while genuinely introgressed alleles go undetected once they have diffused to equilibrium frequencies across the landscape (Wiens and Colella, 2025). Even during active gene flow, admixture can be underestimated when the sample lacks individuals from unadmixed source populations, because clustering methods cannot correctly reconstruct ancestral allele frequencies from data dominated by admixed individuals (Ravagni et al., 2021). And once gene flow ceases, the admixture signal decays rapidly as post-admixture drift erodes the allele-frequency differences on which inference depends, rendering admixture undetectable within relatively few generations (Pang and Zhang, 2025). Together, these findings underscore that clustering methods are most reliable for detecting recent, ongoing admixture between well-differentiated populations. Yet how quickly the admixture signal fades after the cessation of gene flow, and how that rate varies with dispersal within and between metapopulations, remains poorly understood. This has direct consequences for how we interpret estimates from clustering methods. For example, if no admixture is detected at a parapatric boundary, does that necessarily imply long-term genetic isolation? Here we address this by simulating two metapopulations that diverge in isolation, exchange genes during a period of secondary contact, and then return to isolation. We then assess how accurately ADMIXTURE, sNMF, and PopCluster recover the spatial gradient of admixture proportions during and after gene flow across a range of within- and between-metapopulation migration rates.

# 2 Methods

## *2.1 Simulation models*

We simulated two metapopulations that first diverged in isolation, then exchanged genes in a period of secondary contact, and then resumed isolation. We implemented this model using forward-time genetic simulations with SLiM 4 (Haller and Messer, 2023) with a non-Wright–Fisher model on a two-dimensional stepping-stone lattice consisting of 20×10 demes. The landscape was bisected into two equal metapopulations (10×10 demes each) that interacted through a central contact zone (Fig. 1A). Each individual carried a 100,000 base pair (bp) diploid genome with a per generation mutation rate of $10^{-7}$ and a per generation recombination rate of $10^{-5}$. Mutations were selectively neutral. Two migration rates governed dispersal between adjacent demes: within each metapopulation, $M_W$ individuals dispersed per adjacent deme per generation, while across the contact zone, $M_B$ individuals dispersed between each pair of adjacent demes from opposing metapopulations per generation. Demes at the landscape edges experienced reduced migration due to having fewer adjacent demes. Each deme maintained a constant carrying capacity of 500 hermaphroditic individuals with a two-generation lifespan. Reproduction occurred during an individual's second generation, when they randomly selected a partner within their deme with whom they produced a single offspring. Following reproduction, the parental generation was removed, and migrants were randomly selected from the offspring cohort for dispersal to adjacent demes.

The simulations progressed through four distinct phases, with Phases I and II serving as the burn-in period (Fig. 1A). Phase I modelled a common ancestral metapopulation with migration rates of $M_W = 4$ individuals per adjacent deme per generation (both within and between metapopulations), approximating panmixia over 100,000 generations. In Phase II, extending for an additional 100,000 generations, initial population divergence was established by reducing $M_B$ to zero while maintaining $M_W$. Phase III simulated secondary contact and

introgressive hybridisation (Fig. 1B, 1C). During this 1,000-generation phase, $M_W$ was fixed at either 4, 2, or 1 individuals per adjacent deme per generation, while $M_B$ was adjusted relative to $M_W$ at ratios of 1:1, 1:4, and 1:8, representing scenarios of neutral diffusion, a moderate barrier to migration, and a strong barrier to migration, respectively (Fig. 1C). Phase IV introduced a subsequent 5,000-generation period of complete isolation, with no further opportunity for gene flow between metapopulations. To account for stochastic variation and ensure statistical robustness, simulations were run for ten replicates for each of the nine combinations of $M_W$ and $M_B$ parameters. Although the timing and duration of phases could influence the resulting patterns of divergence and the extent of gene flow, we kept these parameters constant to focus on the effects of $M_W$ and $M_B$ on our ability to reliably estimate admixture through time. Phase lengths were sufficient to generate allelic diversity and yield variation in gene flow patterns while remaining computationally tractable. These phases were chosen to approximate a typical glacial-cycle scenario in vertebrates, with prolonged isolation during glacial maxima followed by a brief window of secondary contact in subsequent interglacials, such as the early Holocene (Hewitt, 2000).

Alleles that became fixed between metapopulations during Phase II were tracked as population-specific fixed differences. Genotypic sampling was conducted by randomly selecting 10 individuals per deme—1,000 for each of population 1 and population 2, for 2,000 total individuals—at 100-generation intervals throughout Phase III and the initial 2,000 generations of Phase IV, with a final sampling at generation 5,000 of Phase IV. Genotypic data from these samples were exported as variant call files (VCF) for subsequent analyses.

### *2.2 Admixture proportions*

When loci show fixed allelic differences between populations, admixture can be estimated simply by counting the number of alleles inherited from each source population (i.e., a hybrid index). If these loci were indeed fixed at the start of secondary contact and their states are

known, this calculation is straightforward. By contrast, when loci are polymorphic in one or both populations, the ancestry of individual alleles cannot be determined directly, and statistical methods are needed to account for uncertainty regarding the origin of each allele. When considering only two populations, the resulting admixture estimate is equivalent to a hybrid index but is based on allele-frequency differences rather than fixed diagnostic markers (Gompert et al., 2017). We defined true ancestry in our simulations as the proportion of fixed differences (identified at the end of Phase II) inherited from metapopulations P1 and P2. We also estimated ancestry—and thus admixture—using ADMIXTURE, sNMF, and PopCluster. Default parameters were used for all software. Although STRUCTURE is perhaps the most popular program to estimate admixture proportions, the program runs inefficiently on large datasets and was not included here given the number of individual analyses. Given that our simulations explicitly modelled two divergent metapopulations, we specified $K = 2$ for all analyses. Prior to analyses, minor allele count filtering was applied to remove singleton alleles—allelic variants present in only a single allele copy across all sampled individuals—which are uninformative about population structure and can bias admixture estimates (Johnston et al. 2015; Linck and Battey 2019; Ahrens et al. 2021). These filtered data were randomly subset into datasets of 500, 1,000, 2,000, 5,000, and 10,000 variant sites to determine how the number of loci influences the accuracy of admixture estimates. Analyses were run until the specified convergence criterion was met or a maximum of 200 iterations was reached. Criteria were as follows: (i) ADMIXTURE—the increase in log-likelihood between successive iterations falls below $\varepsilon = 10^{-4}$ (Alexander et al., 2009); (ii) sNMF—a stationary condition adapted from the Karush–Kuhn–Tucker (KKT) criteria is satisfied, with the difference between successive iterations below $\varepsilon = 10^{-4}$ (Frichot et al., 2014); (iii) PopCluster—the estimated admixture proportions between iterations have stabilised using the expectation maximisation algorithm (Wang, 2024).

### *2.3 Summary statistics*

We evaluated the performance of ADMIXTURE, sNMF, and PopCluster using two metrics calculated at each sampled time point: average admixture and, to capture the spatial pattern, cline slope ratio. For each metric, we compared the "true" admixture—calculated from alleles fixed at the end of Phase II—with estimated admixture proportions from the three methods.

Average admixture was measured as the mean individual admixture proportion across the 20×10 deme landscape, providing a summary of introgression at metapopulation scale. Admixture accuracy was quantified for each individual as the ratio of estimated to true admixture proportion, where a value of 1 indicates perfect accuracy, values greater than 1 indicate overestimation, and values less than 1 indicate underestimation (Fig. 2A). We also quantified admixture error as the difference between estimated and true admixture proportions, with positive values indicating overestimation and negative values indicating underestimation (Fig. 2A). Considering both measures allows us to distinguish absolute discrepancies in estimation (error) from proportional deviations relative to the true admixture level (accuracy), which may be amplified when true admixture proportions are small. To avoid undefined ratios in the latter case, instances where true admixture was exactly zero were replaced with a small constant ($1 \times 10^{-6}$) in the denominator prior to computing accuracy; such instances were rare and at early time points under low-migration scenarios.

To provide a spatial context for these estimates, we calculated a landscape-level measure of barrier strength, analogous to cline width (Barton, 1983). Because the limited extent of our simulated landscape frequently caused allele frequencies to fall outside the conventional limits used for cline widths (0.2–0.8; Endler 1977), we instead applied the diffusion model of Nagylaki (1976) for stepped clines, which characterises sharp central transitions where migration is locally reduced (Baird and Daley, 2025). Barrier strength (the "slope ratio") was calculated as the ratio of the step in parental ancestry (here, metapopulation 1) between deme columns 10 and 11 to the gradient of metapopulation 1 ancestry across columns 1–10,

using the mean ancestry per column. (Given the symmetry of the landscape, this could also have been done using metapopulation 2.) Accuracy was then quantified as the ratio of estimated to true slope ratio at each sampled time point, providing a direct measure of how well each method recovered the underlying barrier strength (Fig. 2B). Unlike average admixture, we did not calculate error for slope ratio, as absolute differences between two derived spatial ratios lack straightforward biological interpretation.

As a post hoc exploration of why admixture estimation becomes increasingly inaccurate over time, we calculated three summary statistics: Hardy–Weinberg disequilibrium and linkage disequilibrium within metapopulations, and the number of fixed allelic differences between metapopulations. Clustering methods such as ADMIXTURE, sNMF, and PopCluster assume Hardy–Weinberg and linkage equilibrium within clusters and depend on consistent allele-frequency differences between clusters to assign ancestry. We therefore predicted that increasing underestimation of admixture would correspond to the progressive re-establishment of within-metapopulation equilibrium and the erosion of fixed allelic differences between metapopulations. Hardy–Weinberg disequilibrium was quantified as the absolute mean relative deviation of observed from expected heterozygosity across loci ($|H_O - H_{\text{exp}}|/H_{\text{exp}}$), and linkage disequilibrium was measured as the average pairwise $R^2$ across all variants. Both statistics were calculated separately for each metapopulation at each sampled time point using PLINK v1.9 (Chang et al., 2015). Fixed allelic differences were identified between deme column 1 (the outermost column of metapopulation 1) and deme column 20 (the outermost column of metapopulation 2), representing the point of maximum genetic divergence across the landscape.

# 3 Results

We varied rates of within-population migration ($M_W$) and between-population migration ($M_B$) to assess how the extent of admixture influences its detectability over time. The high-

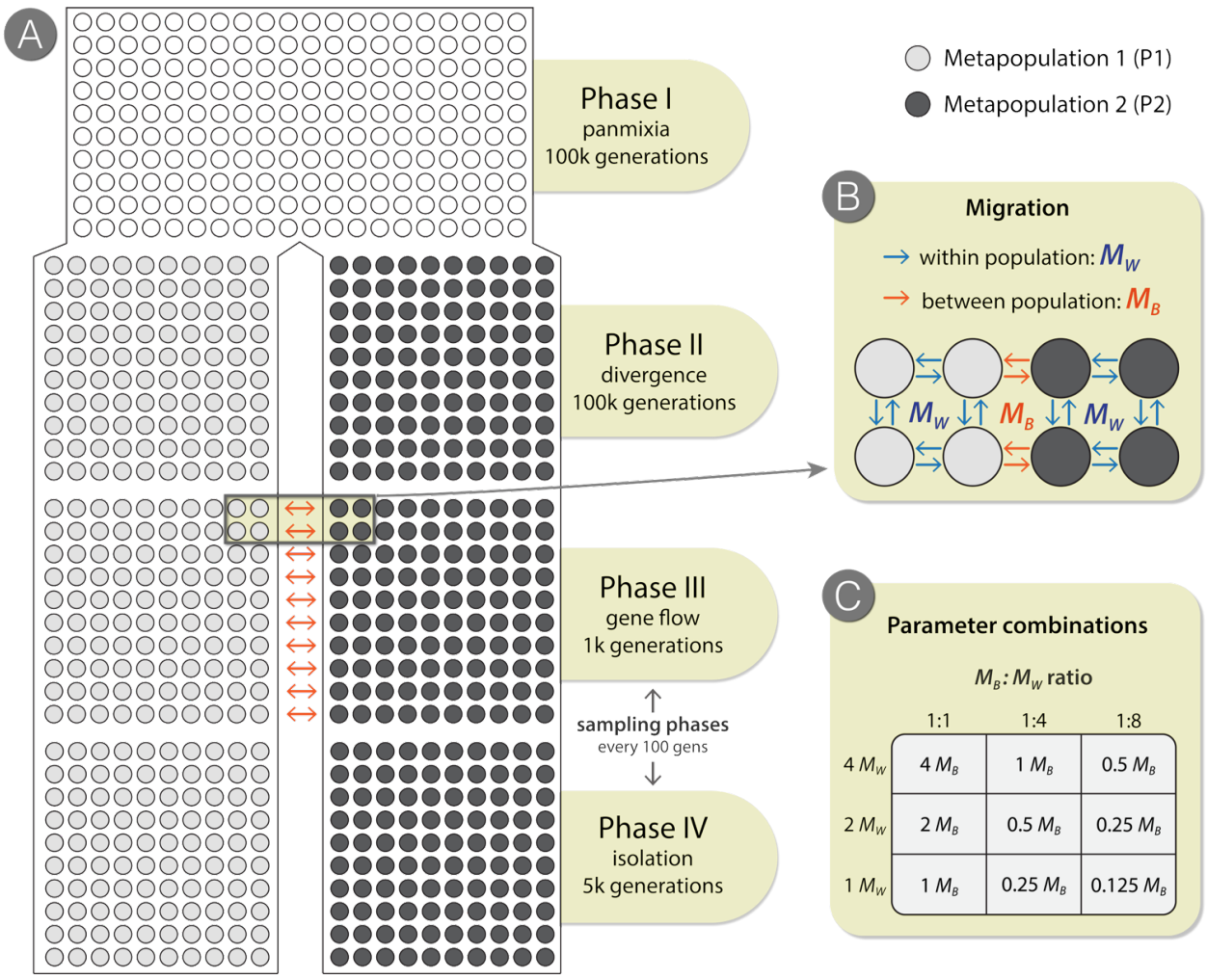


**Figure 1.** Framework and variable parameters for forward-time simulations in SLiM 4. (A) Schematic of the 20×10 deme landscape across four phases. In Phase I, all demes form a single metapopulation for 100,000 generations. Phase II divides the landscape into two metapopulations (P1 and P2), with migration restricted to within each metapopulation for 100,000 generations. Phase III introduces a 1,000-generation pulse of secondary contact, during which gene flow occurs between P1 and P2. Phase IV reinstates complete isolation for 5,000 generations. (B) Illustration of the two migration parameters: within-metapopulation migration ($M_W$) among adjacent demes and between-metapopulation migration ($M_B$) across the central contact zone. (C) The nine combinations of $M_W$ and $M_B$ simulated in this study, each run for 10 replicates.

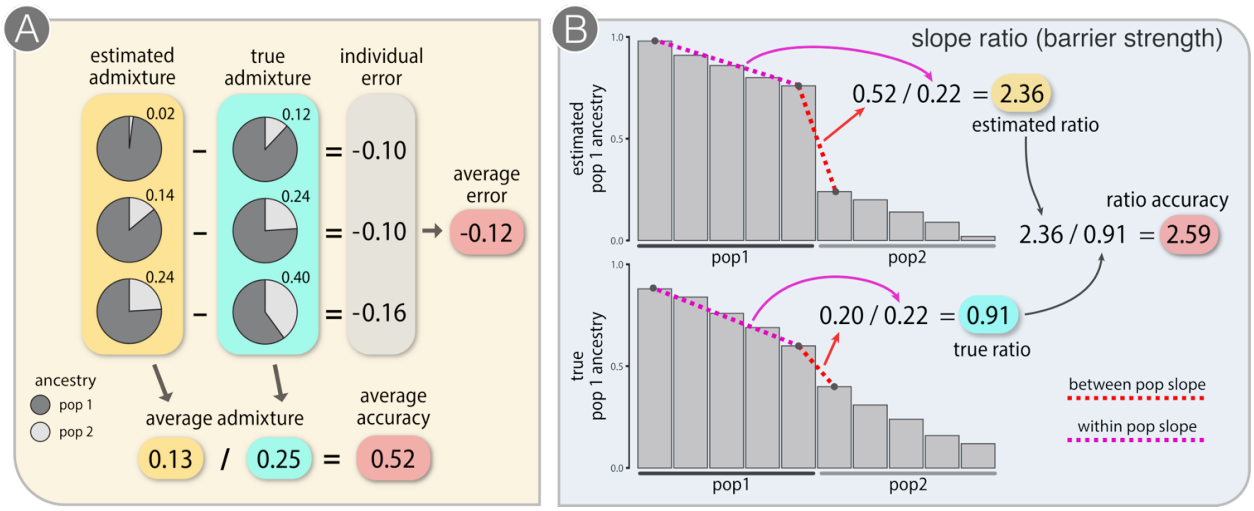


**Figure 2.** Illustration of summary statistics used to evaluate method performance. (A) Average admixture accuracy (ratio of estimated to true admixture proportion) and error (difference between estimated and true admixture proportions). (B) Barrier strength (slope ratio) accuracy, the ratio of estimated to true slope ratio. Note that for (A), calculating accuracy and error is algebraically identical whether computed from population averages or averaged across individuals.

est migration scenario ($M_W = 4$, $M_B = 4$) exhibited the most rapid increase and greatest magnitude of true admixture (Fig. 1), reaching an average admixture proportion of approximately 0.30 (30%) by the close of Phase III, when secondary contact ends ($t = 0$). In contrast, the lowest migration scenario ($M_W = 1$, $M_B = 0.125$) attained only around 0.06 (6%) true average admixture by the close of Phase III. After gene flow ceased, true admixture proportions rapidly stabilised across the landscape, regardless of migration parameters. In all scenarios, the three inference tools systematically underestimated true admixture proportions both during gene flow and after it ended (Figs 3–4, S1A). On average, they recovered only 56% of the true mean admixture proportion by the end of gene flow ($t = 0$), declining to 23% by $t = 1000$ (Fig. 4B). Two thousand generations after gene flow ($t = 2000$), estimated admixture proportions fell below 0.05 across every scenario (Fig. 4A), even when actual admixture remained high (e.g., >0.2). By $t = 2000$ in the high-migration models, ADMIXTURE and PopCluster no longer detected admixture, whereas sNMF retained some signal (Figs 4, S1A), although it continued to underestimate true levels. Patterns of bias differed between proportional accuracy (Fig. 4B) and absolute error (Fig. S2). Proportional

accuracy declined most strongly in low-introgression scenarios (low $M_B$), whereas true admixture was closer to zero and, consequently, small absolute deviations translated into large proportional differences. In contrast, absolute error increased more steeply through time in scenarios with high $M_W$ and $M_B$, reflecting higher true levels of admixture (Fig. S2).

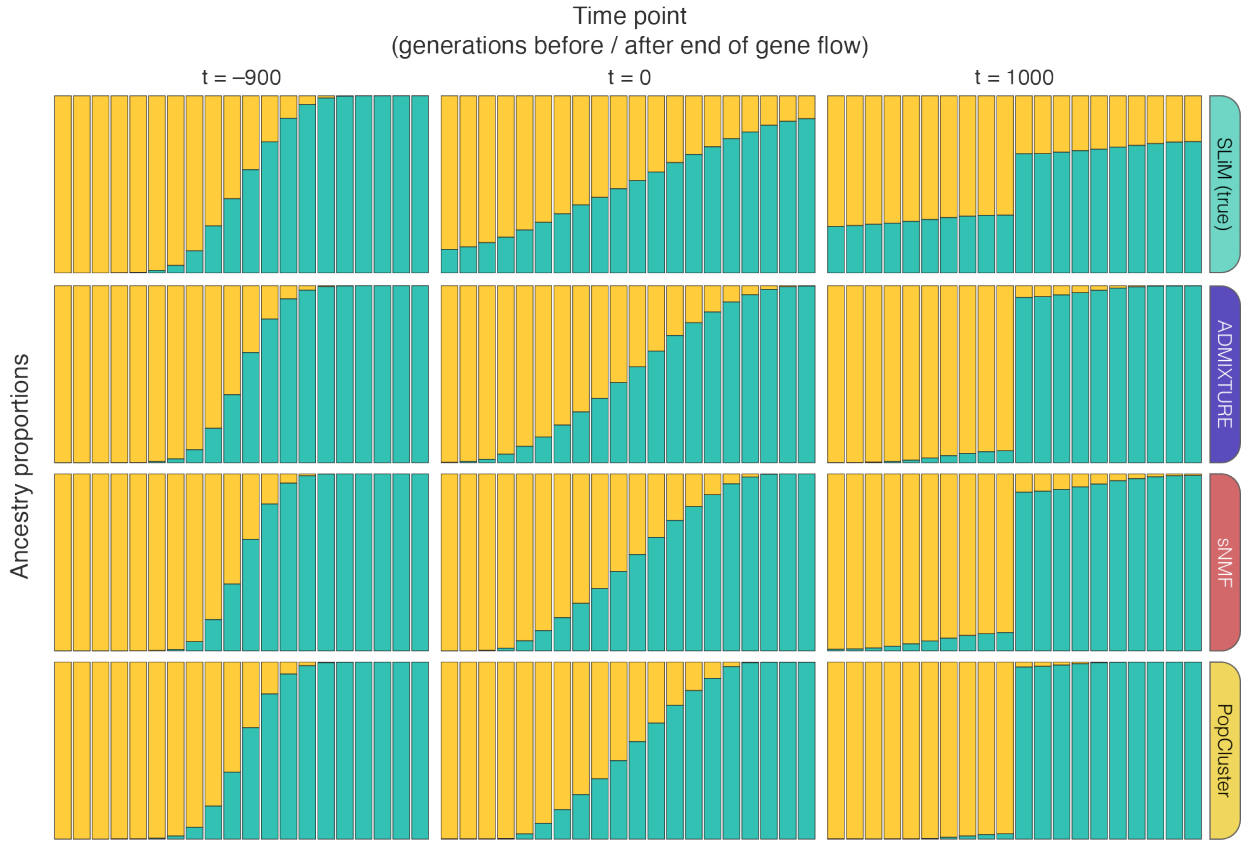


**Figure 3.** Example ancestry proportions at three time points showing true proportions from SLiM simulations (top row) and estimates from ADMIXTURE, sNMF, and PopCluster (rows 2–4, respectively). Time points are 100 generations after gene flow began ($t = -900$; left column), the first generation that gene flow ceased ($t = 0$; middle column), and 1,000 generations after gene flow ceased ($t = 1000$; right column). Results shown are for the highest migration scenario ($M_W = 4$, $M_B = 4$), which produced the greatest magnitude of true admixture and thus the largest absolute underestimation by all three methods (see Figs 4–5), with each value reflecting the mean ancestry proportion averaged across deme rows within each landscape column. Results are based on analyses run using 5,000 variant sites.

We also evaluated how well inference methods captured landscape-scale patterns of gene flow by evaluating accuracy for cline steepness (slope ratio; Fig. 2B). As expected, higher migration rates generated shallower transitions in ancestry, producing lower true slope ratios

(i.e., weaker barrier strength), consistent with more diffuse introgression across the landscape (Fig. 5A). Accuracy for all three methods were relatively stable during the period of gene flow in Phase III ($t = -1000$ to 0; Figs 5B, S1B). Despite this stability, the ratio of estimated to true slope ratio consistently ranged from 1.79±0.17 ($M_W = 4$, $M_B = 4$) to 2.27±0.21 ($M_W = 2$, $M_B = 0.25$), showing that all methods overestimated barrier strength by approximately two-fold throughout secondary contact. Once gene flow ceased (Phase IV, $t = 0$), estimates diverged. ADMIXTURE and PopCluster became progressively less accurate, particularly in high-migration scenarios ($M_W = 4$; Fig. 5B, row 1): ADMIXTURE overestimated slope ratios by an average factor of $6.09 \pm 1.30$, while PopCluster overestimated slope ratio by an average factor of $27.97 \pm 7.34$. In contrast, sNMF maintained a similar level of accuracy in Phase IV ($2.06 \pm 0.36$) to that seen during Phase III, although error did gradually increase. Reported calculations of slope ratios, error, and accuracy in Phase IV were restricted to the first 1,500 generations because estimates of barrier strength became increasingly unstable as the within-population ancestry slopes approached zero.

The number of variant sites included in analyses had only a small effect on estimation error (Fig. S3). PopCluster showed slight improvement with increasing numbers of loci up to $t = 1000$, after which additional sites provided no further reduction in error. In contrast, sNMF and ADMIXTURE exhibited gradually increasing error as the number of variant sites increased. One possibility is that, despite the high recombination rate in our simulations, physical linkage among nearby sites may have introduced estimation bias with increasing numbers of variant sites. Even so, the magnitude of this effect was small, and differences in error were typically only 0.01–0.04 among different numbers of variant sites.

At metapopulation scale, Hardy–Weinberg disequilibrium (Fig. S4) and linkage disequilibrium (Fig. S5) were highest shortly after the onset of gene flow (Phase III) in high migration scenarios, and peaked toward the end of Phase III in low migration scenarios. As gene flow continued, disequilibrium weakened, declining more rapidly still after gene flow ended. Two thousand generations after gene flow had ended ($t = 2000$), linkage disequilibrium was

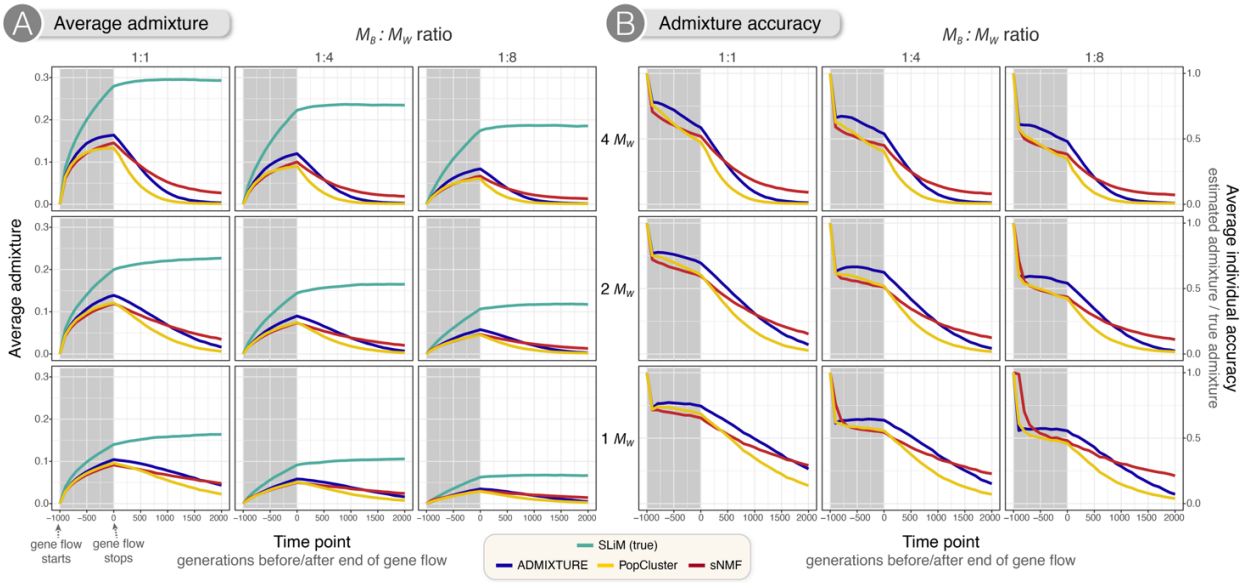


**Figure 4.** True and estimated average admixture proportions (A) and proportional accuracy of admixture estimates (B) across the sampling period, for ADMIXTURE, sNMF, and PopCluster (coloured lines), alongside the true admixture proportion derived directly from SLiM simulations (green line; panel A only). Average admixture is the mean individual admixture proportion across the deme landscape, providing a summary of introgression at metapopulation scale. Accuracy (B) is the ratio of estimated to true admixture proportion, where a value of 1 indicates perfect agreement, values below 1 indicate underestimation, and values above 1 indicate overestimation. Across all scenarios, all three methods consistently underestimated true admixture. The shaded region indicates Phase III, during which gene flow occurred ($t = -1000$ to 0); the unshaded region represents the first 2,000 generations of Phase IV, during which populations were in complete isolation. Rows indicate the level of within-metapopulation migration ($M_W$) and columns indicate the ratio of between- to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses of 5,000 variant sites.

negligible in all scenarios. Hardy–Weinberg disequilibrium remained modest (∼0.1) and, as expected, was greater in metapopulations with lower $M_W$. The number of fixed allelic differences between opposite ends of the landscape also declined rapidly (Fig. S6), with no fixed differences remaining by the end of Phase III when $M_W = 4$, and some persisting through Phase IV under low $M_W$ and $M_B$. Together, these patterns show that declines in within-metapopulation disequilibrium coincide with the loss of allele-frequency differences between metapopulations, consistent with the progressive erosion of the signal on which clustering methods depend.

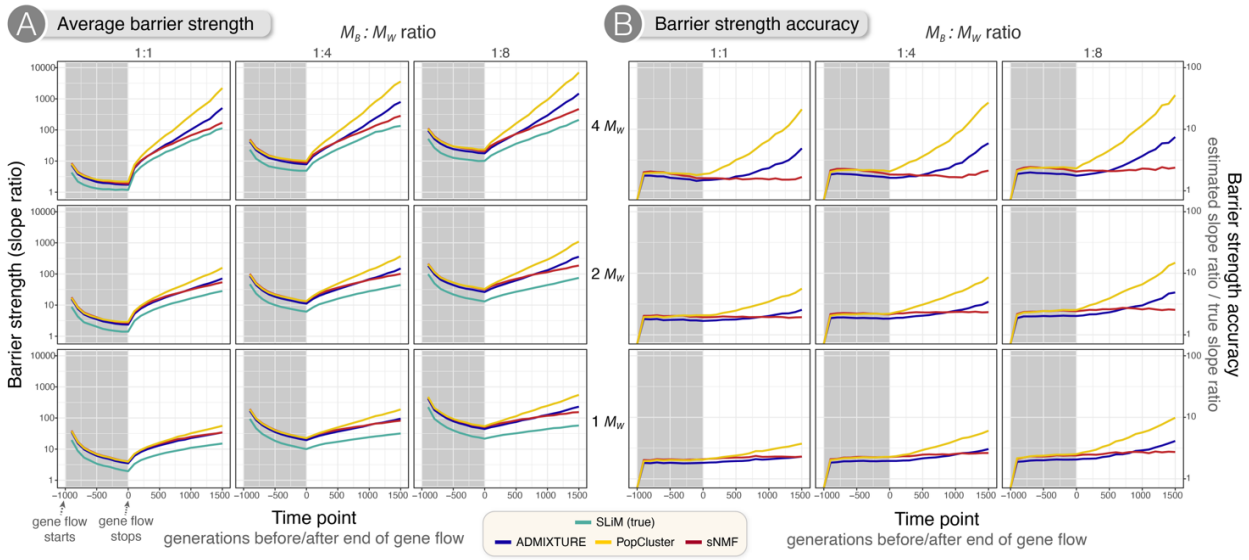


**Figure 5.** True and estimated barrier strength (slope ratio) across the sampling period (A) and proportional accuracy of slope ratio estimates (B), for ADMIXTURE, sNMF, and PopCluster (coloured lines), alongside the true slope ratio derived directly from SLiM simulation ancestry tracking (green line; panel A only). The slope ratio is the ratio of the step in metapopulation 1 ancestry between deme columns 10 and 11 to the gradient of metapopulation 1 ancestry across columns 1–10, providing a landscape-scale measure of barrier strength (shown on a $\log_{10}$ scale). Accuracy (B) is the ratio of estimated to true slope ratio, where a value of 1 indicates perfect agreement; in all cases, estimated slope ratios exceeded 1, indicating systematic overestimation of barrier strength. The shaded region indicates Phase III, during which gene flow occurred ($t = -1000$ to 0); the unshaded region represents the first 1,500 generations of Phase IV, during which populations were in complete isolation, with later time points excluded due to increasing instability in slope ratio estimates as within-population ancestry gradients approached zero. Rows indicate the level of within-metapopulation migration ($M_W$) and columns indicate the ratio of between-metapopulation to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses of 5,000 variant sites.

# 4 Discussion

Our simulations demonstrate that clustering methods become increasingly unreliable at detecting admixture as the period of isolation following gene flow lengthens, regardless of whether gene flow was limited or extensive. Admixture was modestly but increasingly underestimated during active gene flow after secondary contact, but accuracy declined sharply once populations returned to isolation—falling to less than a quarter of the true value within 1,000 generations and becoming effectively undetectable within 2,000 generations, even where true admixture remained substantial. These results extend recent findings documenting bias and estimation error in clustering methods (Pang and Zhang 2025; Wiens and Colella 2025). Critically, whereas prior simulation studies have not combined explicit spatial structure with temporal tracking across a range of dispersal rates, our metapopulation design allowed us to disentangle the effects of within- and between-metapopulation dispersal on both the establishment and decay of admixture signal. While clustering approaches remain powerful tools for detecting contemporary substructure, these findings add to a growing body of evidence that their outputs require cautious interpretation (Latch et al. 2006; Wang 2017; Bradburd et al. 2018; Lawson et al. 2018; Pang and Zhang 2025; Wiens and Colella 2025) and underscore that the absence of inferred admixture at a contact zone should not be interpreted as evidence of long-term isolation.

To better understand why such underestimation arises, we examined how within-population disequilibrium and between-population allelic differences changed through time. Hardy–Weinberg and linkage disequilibrium were initially elevated following secondary contact but weakened steadily, then declined more rapidly once gene flow ceased. At the same time, fixed allelic differences disappeared quickly in high-migration scenarios and persisted only when $M_W$ and $M_B$ were low. The correspondence between declining within-population disequilibrium and the loss of allelic differences between metapopulations indicates that recombination and gene flow progressively erode the allele-frequency contrasts on which

ancestry inference depends. As this structure decays, admixture estimates drift toward zero even when the true extent of admixture is high.

The three methods we evaluated—ADMIXTURE, sNMF, and PopCluster—displayed broadly similar temporal patterns but differed subtly with respect to error and accuracy. ADMIXTURE performed best during and shortly after the period of gene flow, maintaining reasonable estimates for roughly 1,000 generations after migration ceased, whereas sNMF was more accurate at the end of the sampling period. PopCluster consistently underperformed, albeit modestly, relative to both other methods across all scenarios, in both individual-level admixture estimation and spatial cline recovery. In the highest-migration model ($M_W = 4$, $M_B = 4$), all three methods underestimated admixture by an absolute margin of approximately 0.25–0.30 after 1,000 generations of isolation, whereas in the lowest-migration model ($M_W = 1$, $M_B = 0.125$), the bias was much smaller (around 0.05). This contrast arises because high $M_W$ and $M_B$ scenarios produce levels of true admixture, leaving more introgressed ancestry available to be underestimated. By comparison, true admixture in low-introgression scenarios is close to zero and, although proportional accuracy appears poor, the underlying absolute discrepancies are small. Thus, the greatest absolute errors occurred in scenarios with the most extensive introgression, reinforcing the conclusion that clustering methods struggle most under strong historical gene flow.

The practical implications of these findings are considerable. First and foremost, the absence of inferred admixture in contemporary samples should not be interpreted as conclusive evidence against past gene flow. Populations that have experienced extended isolation following introgression may sometimes appear genetically "pure" despite retaining a substantial proportion of introgressed alleles. Reliably estimating admixture depends not only on biological characteristics such as dispersal ability, time since introgression, and population or range size, but also on the geographic scale of sampling relative to the zone of admixture (Bradburd et al. 2018; Duranton et al. 2019). Including more geographically distant "parental" populations, where population-wide mixing remains incomplete, would likely improve the

ability of clustering methods to detect admixture (Ravagni et al., 2021). Even so, biological context remains critical—in species with small ranges, no amount of spatially broad sampling can recover historical admixture once introgressed alleles have become widespread and reached equilibrium. This is especially consequential for conservation genetics, where failing to detect historical admixture may lead to overestimation of genetic distinctiveness and misinformed management or listing decisions. Thus, our findings suggest that clustering approaches should be used with care, particularly when gene flow is expected to have occurred long ago or in species with high dispersal, and that methods explicitly modelling historical gene flow—such as coalescent-based frameworks (e.g., Flouri et al. 2023)—might provide valuable complementary insight in such cases.

While our simulations are powerful for understanding how the signal of admixture can decay when using clustering-based methods, they do have many limitations. Our simplified demographic model—featuring just two populations with a single pulse of secondary contact—will not fully capture the complexities of many natural populations, where multiple admixture events, more than two ancestral populations, continuous gene flow, or variable population sizes can be common. In such cases, accurately estimating admixture is likely to be even more challenging. This difficulty is compounded further by the fact that accurately specifying $K$—the number of ancestral populations—is itself non-trivial and can introduce further error; $K$ is a statistical convenience rather than a biological reality, and methods for choosing it remain inconsistent and prone to over- or under-splitting populations (Lawson et al., 2018). Admixture estimates may remain accurate for longer in organisms with very large distributions relative to dispersal—with correspondingly extensive geographic sampling—as introgressed alleles require more time to disperse across the landscape; however, in extremely large ranges, isolation by distance may introduce new biases to admixture estimates (Bradburd et al. 2013; Wiens and Colella 2025). These factors underscore the need for caution when extrapolating simulation results to empirical data. Importantly, we do not suggest that all empirical studies will underestimate admixture: our simulations capture only a small sub-

set of the parameter space of real populations, and outcomes when using real data will depend strongly on factors such as the true number of population clusters, sampling design, demographic history, and model assumptions not explored here. Nonetheless, recognising the temporal limitations of clustering-based methods remains essential for reconstructing population history accurately and for drawing appropriately nuanced conclusions.

## Acknowledgements

We thank Peta Hill for advice and discussion, Ben Haller for his assistance in optimising our SLiM code, and members of the ANU PhyloMacro Snack Group for providing constructive feedback (and snacks).

## Funding

This work was supported by the Australian National University's Future Research Talent scholarship programme (awarded to J.S.) and an Australian Research Council Discovery Project (DP210102267).

## References

Ahrens, C. W., Jordan, R., Bragg, J., Harrison, P. A., Hopley, T., Bothwell, H., Murray, K., Steane, D. A., Whale, J. W., Byrne, M., Andrew, R., and Rymer, P. D. (2021). Regarding the F-word: The effects of data filtering on inferred genotype-environment associations. *Molecular Ecology Resources*, 21(5):1460–1474.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9):1655–1664.

Anderson, E. and Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, 8(4):378–388.

Arnold, M. L. and Kunte, K. (2017). Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends in Ecology & Evolution*, 32(8):601–611.

Baird, S. J. E. and Daley, N. (2025). The shapes of clines and wavefronts. *Molecular Ecology*, 34:e70109.

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

Bradburd, G. S., Coop, G. M., and Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1):33–52.

Bradburd, G. S., Ralph, P. L., and Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, 67(11):3258–3273.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1).

Duranton, M., Bonhomme, F., and Gagnaire, P. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications*, 12(9):1743–1756.

Endler, J. A. (1977). *Geographic Variation, Speciation, and Clines*, volume 10 of *Monographs in Population Biology*. Princeton University Press.

Flouri, T., Jiao, X., Huang, J., Rannala, B., and Yang, Z. (2023). Efficient Bayesian inference under the multispecies coalescent with migration. *Proceedings of the National Academy of Sciences*, 120(44):e2310708120.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4):973–983.

Gompert, Z., Mandeville, E. G., and Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, 48(1):207–229.

Haller, B. C. and Messer, P. W. (2023). SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist*, 201(5):E127–E139.

Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18):4606–4618.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789):907–913.

Johnston, H. R., Hu, Y., and Cutler, D. J. (2015). Population genetics identifies challenges in analyzing rare variants. *Genetic Epidemiology*, 39(3):145–148.

Latch, E. K., Dharmarajan, G., Glaubitz, J. C., and Rhodes, O. E. (2006). Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, 7(2):295–302.

Lawson, D. J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9:3258.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183–189.

Linck, E. and Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3):639–647.

Nagylaki, T. (1976). Clines with variable migration. *Genetics*, 83(4):867–886.

Pang, X.-X. and Zhang, D.-Y. (2025). A cautionary note on using STRUCTURE to detect hybridization in a phylogenetic context. *Journal of Systematics and Evolution*, 63(6):1560–1576.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.

Ravagni, S., Sanchez-Donoso, I., and Vilà, C. (2021). Biased assessment of ongoing admixture using STRUCTURE in the absence of reference samples. *Molecular Ecology Resources*, 21(3):677–689.

Rhymer, J. M. and Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, 27(1):83–109.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., and Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602):2381–2385.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19(4):198–207.

Suarez-Gonzalez, A., Lexer, C., and Cronk, Q. C. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, 14(3):20170688.

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., and Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7):892–908.

Wang, J. (2017). The computer program STRUCTURE for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources*, 17(5):981–990.

Wang, J. (2022). Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity*, 129:79–92.

Wang, J. (2024). PopCluster: A population genetics model-based toolset for simulating, inferring and visualising individual admixture and population structure. *Molecular Ecology Resources*.

Wiens, B. J. and Colella, J. P. (2025). That's not a hybrid: How to distinguish patterns of admixture and isolation by distance. *Molecular Ecology Resources*, 25:e14039.

Wogan, G. O. U., Yuan, M. L., Mahler, D. L., and Wang, I. J. (2023). Hybridization and transgressive evolution generate diversity in an adaptive radiation of *Anolis* lizards. *Systematic Biology*, 72(4):874–884.

# Supplementary Figures

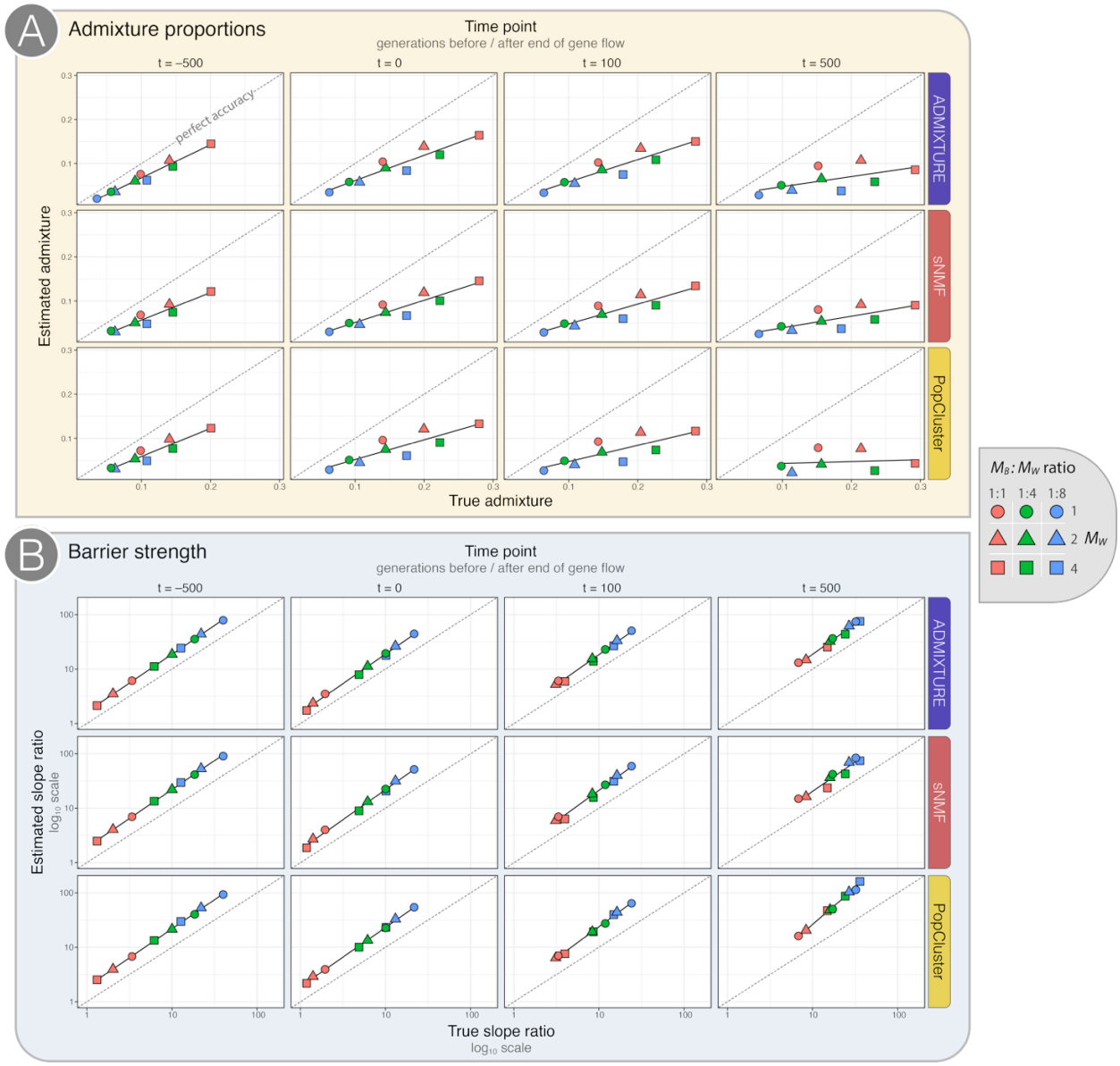


**Figure S1.** Scatter plots of true versus estimated values for (A) average admixture and (B) barrier strength (slope ratio) at four sampled time points (columns) for each of the three inference methods (rows). True values are derived from SLiM simulations and plotted on the $x$-axis; estimated values from each method are on the $y$-axis. The dashed diagonal line indicates perfect agreement between estimated and true values. Point shapes indicate the level of within-metapopulation migration ($M_W$) and colours indicate the ratio of between-metapopulation migration to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses run using 5,000 variant sites.

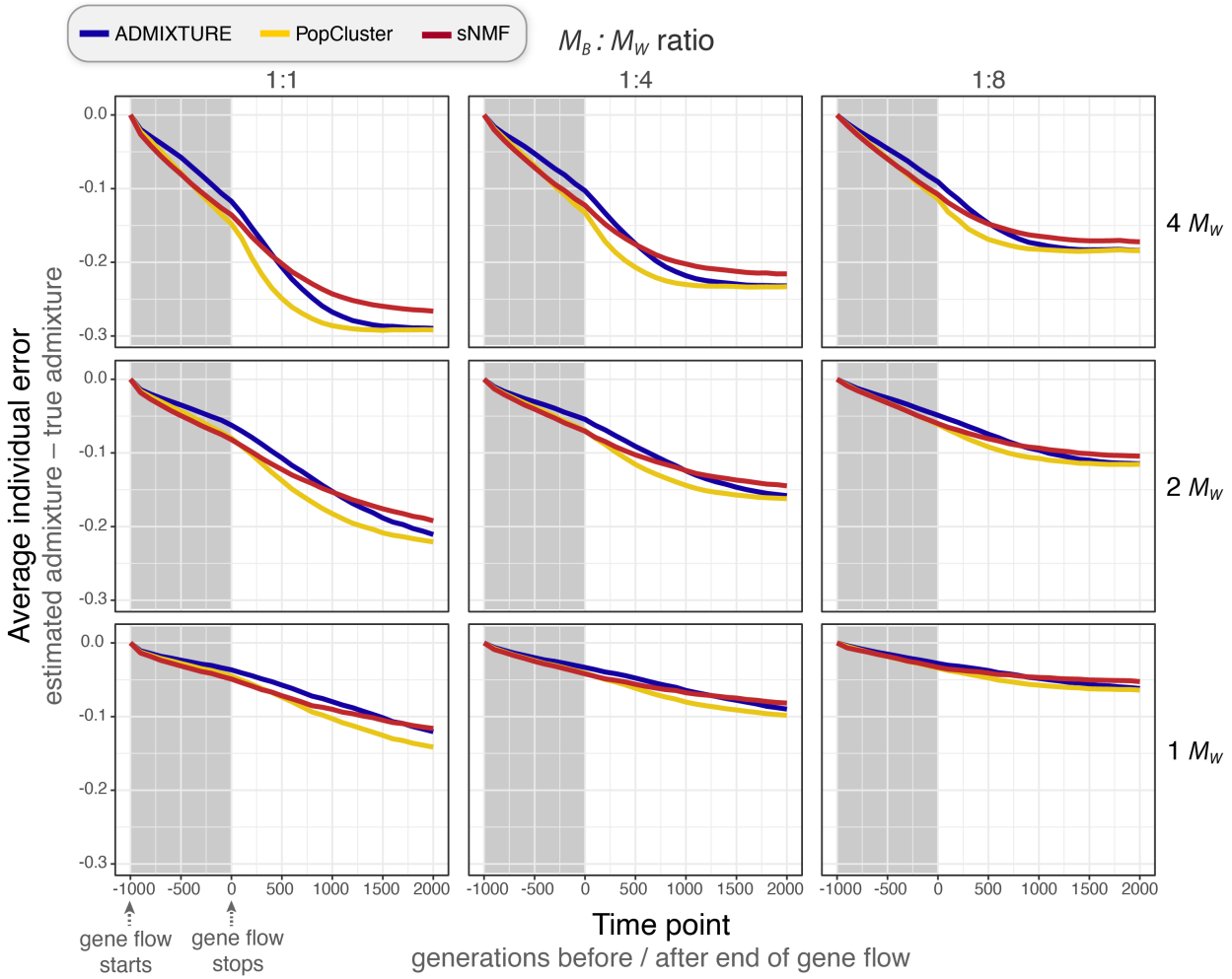


**Figure S2.** Absolute error of average admixture estimates across the sampling period, for ADMIXTURE, sNMF, and PopCluster (coloured lines). Error is the difference between estimated and true admixture proportions, where negative values indicate underestimation. Across all scenarios, all three methods consistently underestimated true admixture. The shaded region indicates Phase III, during which gene flow occurred ($t = -1000$ to 0); the unshaded region represents the first 2,000 generations of Phase IV that were sampled, during which populations were in complete isolation. Rows indicate the level of within-metapopulation migration ($M_W$) and columns indicate the ratio of between-metapopulation to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses of 5,000 variant sites.

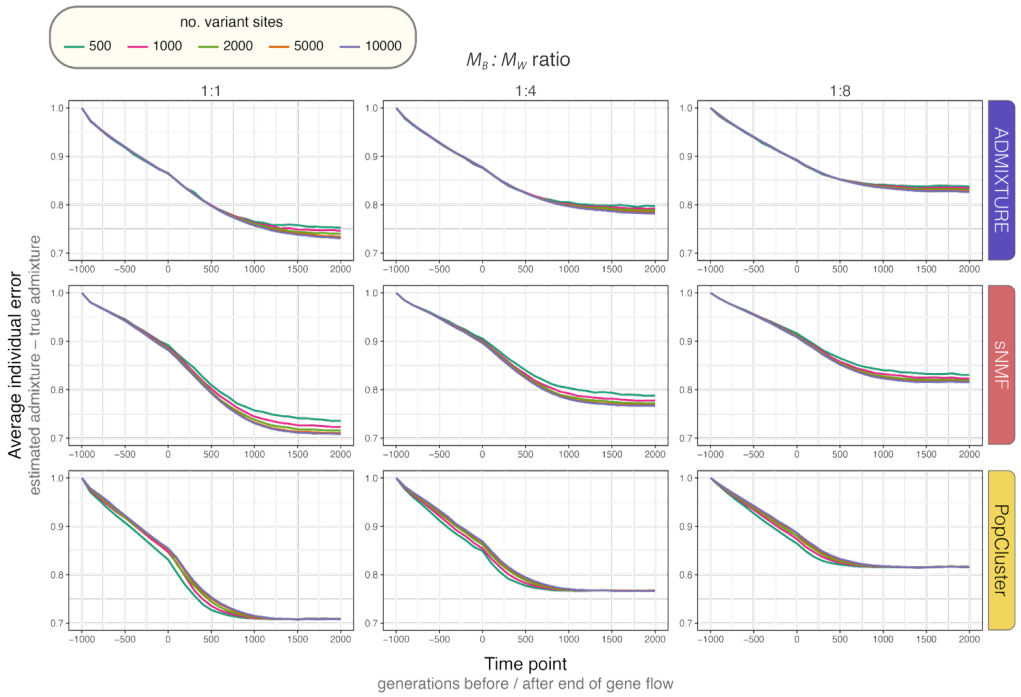


**Figure S3.** Examples showing the influence of the number of variant sites on average admixture error through time. Error is the difference between estimated and true admixture proportions, where negative values indicate underestimation (in all cases here). Rows indicate the inference method (ADMIXTURE, sNMF, and PopCluster) and coloured lines represent different numbers of variant sites (500, 1,000, 2,000, 5,000, and 10,000). All panels show results for $M_W = 4$, the within-metapopulation migration level at which error is generally greatest (see Fig. 5). Columns indicate the ratio of between-metapopulation to within-metapopulation migration ($M_B$:$M_W$). The shaded region indicates Phase III, during which gene flow occurred; the unshaded region represents the first 2,000 generations of Phase IV that were sampled, during which populations were in complete isolation.

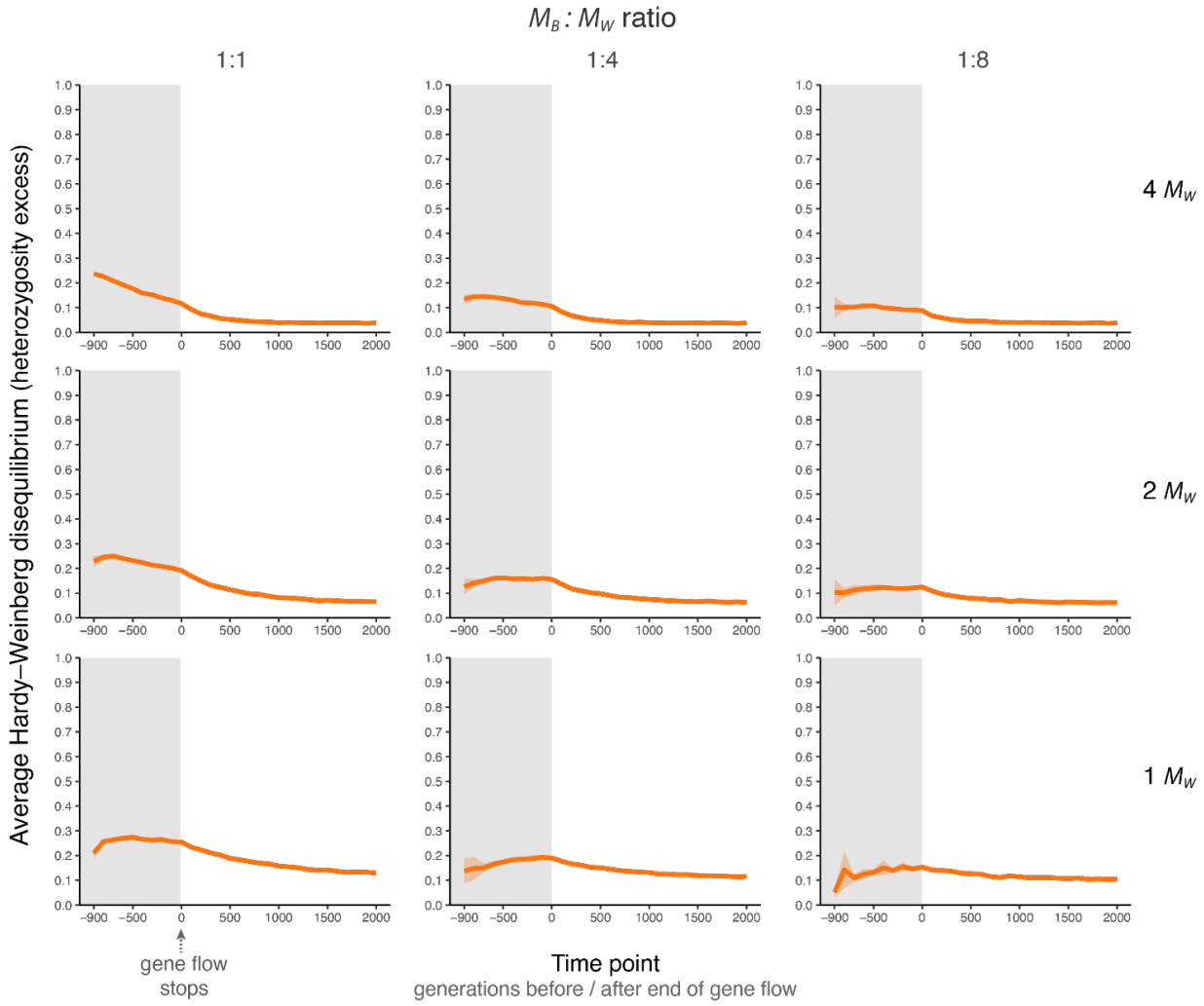


**Figure S4.** Average Hardy–Weinberg disequilibrium (deviation of observed from expected heterozygosity) within metapopulation 1 across the sampling period (metapopulation 2 is omitted given the simulation's spatial symmetry). Rows indicate within-metapopulation migration ($M_W$); columns indicate the ratio of between- to within-metapopulation migration ($M_B$:$M_W$). The shaded region (Phase III) marks the period of gene flow; the unshaded region shows the first 2,000 generations of complete isolation (Phase IV). Results based on 5,000 variant sites.

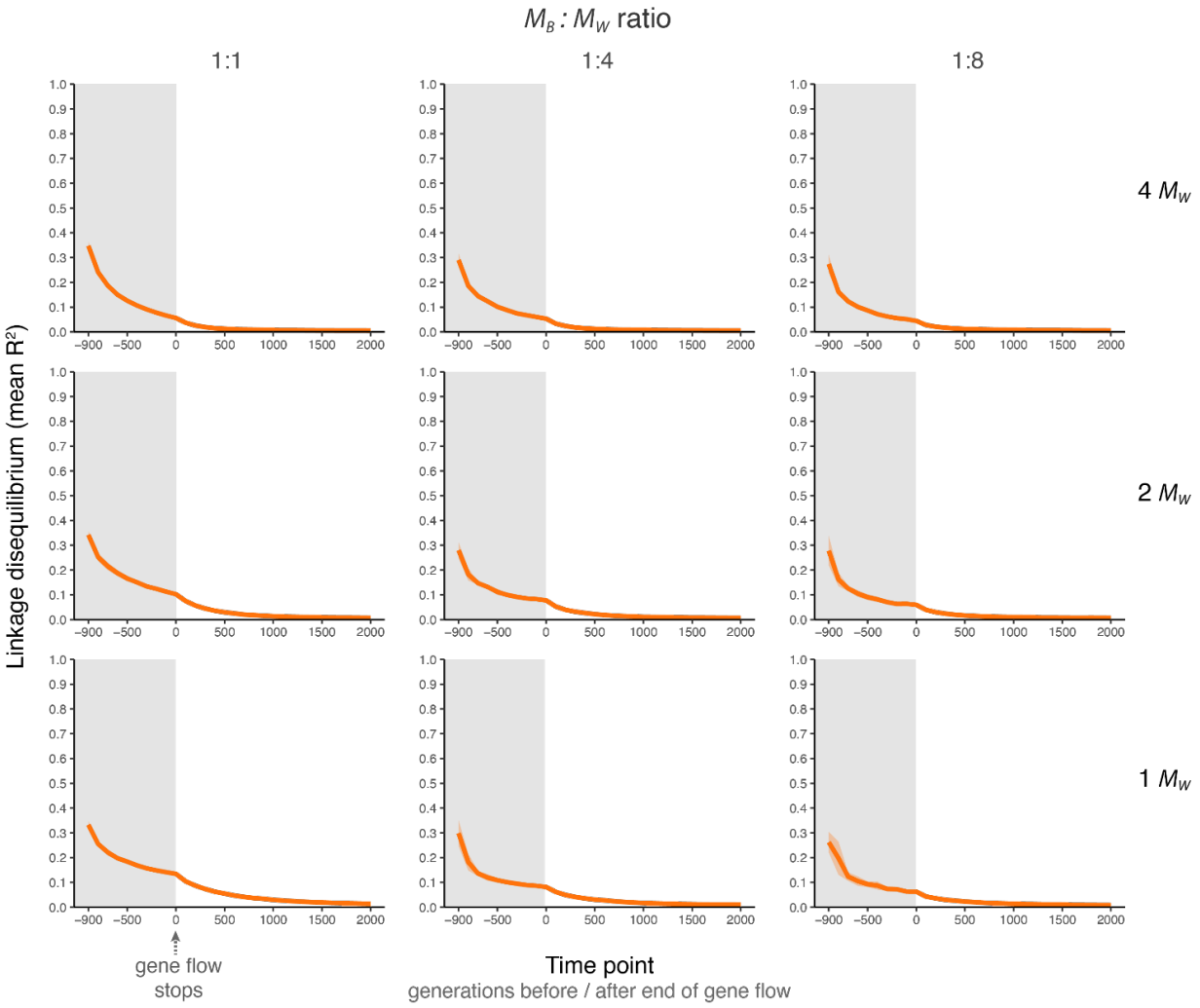


**Figure S5.** Average linkage disequilibrium (LD) across the sampling period, quantified using the $R^2$ metric and calculated for metapopulation 1 only (metapopulation 2 is omitted given the simulation's spatial symmetry). Rows indicate the level of within-metapopulation migration ($M_W$) and columns indicate the ratio of between-metapopulation to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses of 5,000 variant sites. The shaded region indicates Phase III, during which gene flow occurred; the unshaded region represents the first 2,000 generations of Phase IV that were sampled, during which populations were in complete isolation.

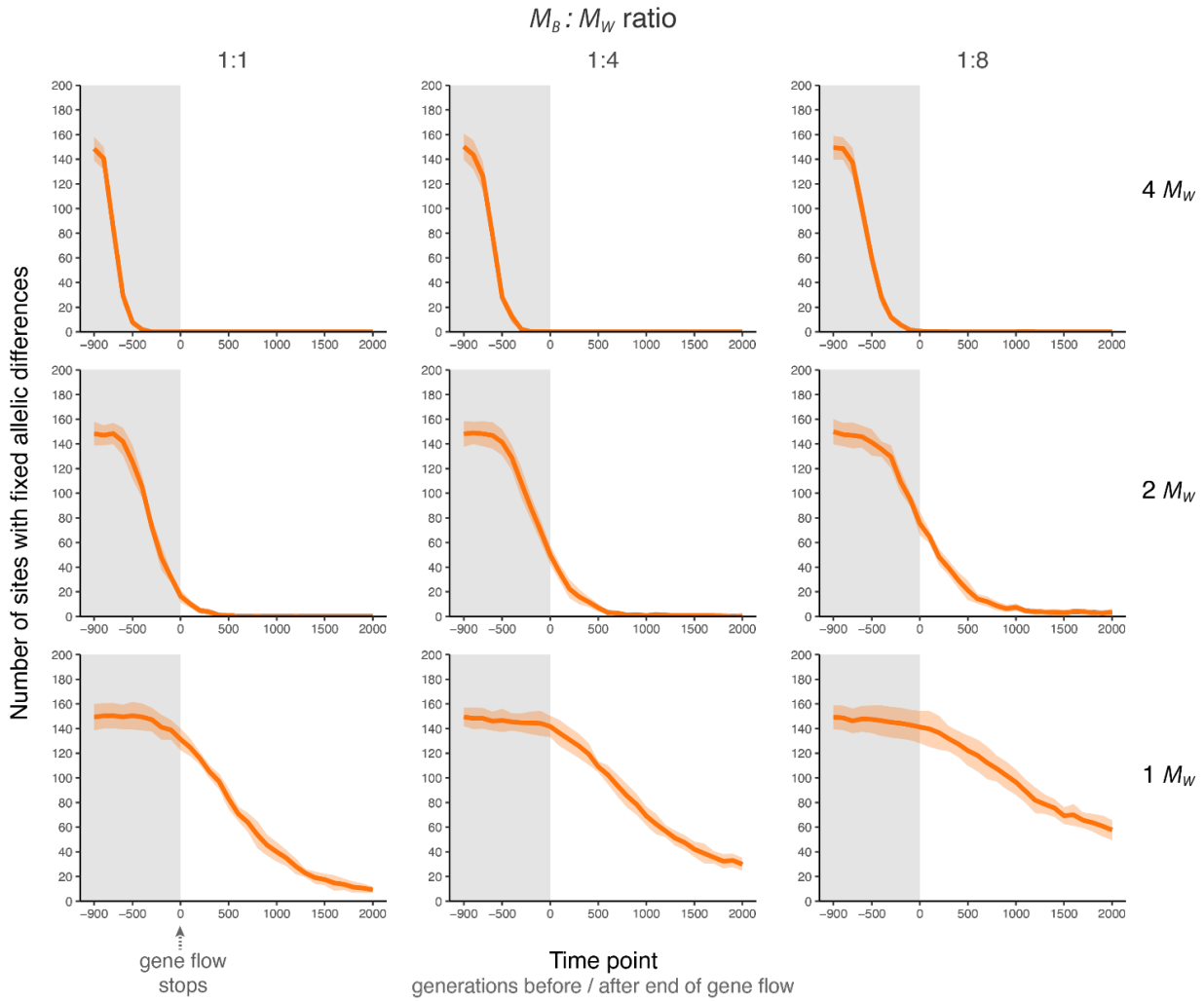


**Figure S6.** Mean number of fixed allelic differences between the most geographically isolated deme columns (see Fig. 1) across the sampling period. Fixed differences were identified between deme column 1 (the outermost column of metapopulation 1) and deme column 20 (the outermost column of metapopulation 2), representing the maximum divergence across the landscape. Rows indicate the three levels of within-metapopulation migration ($M_W$) and columns indicate the ratio of between-metapopulation to within-metapopulation migration ($M_B$:$M_W$). Results are based on analyses of 5,000 variant sites. The shaded region indicates Phase III, during which gene flow occurred; the unshaded region represents the first 2,000 generations of Phase IV that were sampled, during which populations were in complete isolation.