Title: *Selection leads to false inferences of introgression using popular methods*

Authors: Megan L. Smith[1,2] and Matthew W. Hahn[2,3]

[1]Department of Biological Sciences, Mississippi State University, Starkville, MS 39762, USA

[2]Department of Biology, Indiana University, Bloomington, IN 47405, USA

[3]Department of Computer Science, Indiana University, Bloomington, IN 47405, USA

Running Title: Selection misleads migration inferences

Key Words: geneflow, migration, background selection, selective sweeps

Correspondence: ms4438@msstate.edu (M.L. Smith)

Abstract

Detecting introgression between closely related populations or species is a fundamental objective in evolutionary biology. Existing methods for detecting migration and inferring migration rates from population genetic data often assume a neutral model of evolution. Growing evidence of the pervasive impact of selection on large portions of the genome across diverse taxa suggests that this assumption is unrealistic in most empirical systems. Further, ignoring selection has previously been shown to negatively impact demographic inferences (e.g., of population size histories). However, the impacts of biologically realistic selection on inferences of migration remain poorly explored. Here, we simulate data under models of background selection, selective sweeps, and adaptive introgression. We show that ignoring selection leads to false inferences of migration in three popularly used methods (fastsimcoal, ∂a∂i, and BPP). Selection results in the rejection of isolation-only models in favor of isolation-with-migration models and leads to elevated estimates of migration rates across methods. Our results suggest that such methods may be unreliable in many empirical systems, such that new methods that are robust to selection will need to be developed.

Article Summary

Detecting migration between closely related populations is a central objective in many evolutionary biology studies. However, popular methods for detecting migration assume a simplified model of evolution. Here, we evaluate the impacts of biologically realistic natural selection, recombination, and mutation on three methods for detecting migration. We find that biological complexity leads to false inferences of migration, suggesting that results should be interpreted with caution and that new methods are needed to make robust inferences of migration across empirical systems.

Introduction

In recent years, as genomic data have become readily available for many taxa, evidence of introgression has accumulated across the tree of life (Mallet *et al.* 2016). A growing interest in understanding the role of introgression in diversification has led to the development of numerous phylogenetic methods for detecting introgression (reviewed in Hibbins and Hahn 2022), but most of these methods cannot detect introgression between sister taxa—only methods that use population genetic data attempt to do this. While detecting introgression between sister taxa is a difficult task, it is of central interest to many researchers. For example, understanding whether closely related taxa exchanged genes during divergence is central to distinguishing among modes of speciation (Payseur and Rieseberg 2016; Roux *et al.* 2016), with evidence of gene flow between closely related taxa being interpreted as a possible signal of sympatric speciation. Characterizing gene flow between sister species is also essential for developing null models in scans for selection (Williamson *et al.* 2005; Nielsen *et al.* 2007; Excoffier *et al.* 2009; Luqman *et al.* 2021). Thus, despite the difficulties of the task, many population genetic methods have been developed (and have been widely applied) to detect gene flow between sister taxa.

Introgression should lead to increased allele-sharing between taxa and increased variance in coalescence times compared to incomplete lineage sorting (ILS) alone, and methods to detect introgression between sister taxa rely on these expectations. Summary-statistic methods aim to detect particular regions of the genome that have introgressed based on the expectation that these regions should be more similar between sister taxa than non-introgressed regions (Joly *et al.* 2009; Geneva *et al.* 2015; Rosenzweig *et al.* 2016). Other approaches focus on comparing models with and without migration and/or estimating genome-wide migration rates. For

example, many site frequency spectrum (SFS)-based methods estimate migration rates and other parameters by finding the parameters that maximize the composite likelihood of the SFS (e.g., Gutenkunst *et al.* 2009; Tellier *et al.* 2011; Excoffier *et al.* 2013), which can be computed using diffusion approximation (e.g., ∂a∂i; Gutenkunst *et al.* 2009) or simulations (e.g., fastsimcoal2; Excoffier *et al.* 2013). Models with and without migration can then be compared based on estimated likelihoods. While powerful, SFS-based approaches do not take advantage of linkage information, and other approaches exist that directly analyze sequence data rather than relying on the SFS as a summary. For example, BPP estimates divergence times and the intensities of introgression events from sequence data under the multispecies coalescent with introgression (MSci) model a using Bayesian Markov chain Monte Carlo (MCMC) approach (Flouri *et al.* 2020). While these methods are powerful and can be highly accurate on simulated datasets, all assume selection does not affect the patterns observed.

Mounting evidence suggests that selection impacts large portions of the genome (reviewed in Cutter and Payseur 2013; Kern and Hahn 2018). Notably, selection can produce genomic signals that mimic demographic processes. For example, linked selection can produce signals that mimic population growth or contraction and can mislead commonly used methods for inferring population size histories (Ewing and Jensen 2016; Schrider *et al.* 2016; Johri *et al.* 2021). Ignoring selection may also pose a substantial problem for methods aiming to detect migration (Cruickshank and Hahn 2014; Mathew and Jensen 2015; Roux *et al.* 2016; Fraïsse *et al.* 2021). Selection leads to increased heterogeneity in levels of divergence among loci by either decreasing (directional selection) or increasing (balancing selection) levels of polymorphism at some loci. Since many methods for detecting introgression rely on these same signals, this can

lead to false inferences of migration (e.g., Cruickshank and Hahn 2014; Roux *et al.* 2016). Despite more widespread acknowledgement of the role of selection and the shortcomings of neutral assumptions in recent years, methods for inferring migration rates that ignore selection are still widely used.

Here, we simulate data under several evolutionary models that include selection, including background selection, selective sweeps, and adaptive introgression. We evaluate the impact of selection on inferences of migration rates in ∂a∂i, fastsimcoal2, and BPP, and show that selection leads to high rates of false inferences of migration. Our results highlight the importance of incorporating selection into tests for migration in the future.

Materials and Methods

Simulations

We simulated two populations that diverged at a set time in the past, T_D, and considered three migration histories: no migration (nomig), a pulse of migration from population 1 to population 2 looking forward in time (p1_p2), and a pulse of migration from population 2 to population 1 (p2_p1). We set all population sizes to 125,000 and considered three divergence times: $T_D = 0.25 \times 4N,\ 1 \times 4N,$ and $4 \times 4N$ (low, medium, high). The time since introgression, T_M, was drawn from a vector $\{0.01 \times T_D, 0.05 \times T_D, 0.10 \times T_D, 0.15 \times T_D, \ldots, 0.9 \times T_D\}$ and the probability of any lineage migrating, p_M, was drawn from a vector $\{0.05, 0.1, 0.15, \ldots, 0.95\}$. We set the per site mutation rate, μ, to 1e-8, and the per site recombination rate, r, to 5e-8, both per generation. To lessen the computational burden of forward-in-time simulations, we scaled all simulations by an order of 100: population sizes were scaled to 1250, and mutation rates,

recombination rates, and selection coefficients (*s*; see below) were all scaled to keep values of $N\mu$, Nr, and Ns constant. Similarly, divergence times and the timing of migration were scaled down by an order of 100. We simulated 10,000 independent 10 kb windows for most conditions (see below for details).

To evaluate the impact of selection on inferences of migration rates, we simulated data under five scenarios in SLiM v4.0.1 (Haller and Messer 2019), overlaying neutral mutations with pyslim v1.0.3 and tskit v0.5.5 (Kelleher et al., 2018; Haller et al., 2019). We considered the following selective scenarios: 1) a neutral model; 2) background selection (BGS); 3) a selective sweep in the ancestor of the two populations; 4) a selective sweep in population 1; and 5) adaptive introgression. For all scenarios except adaptive introgression, we considered all three migration models described above (nomig, p1_p2, and p2_p1). For adaptive introgression, we considered only the p1_p2 model. For all simulations, a neutral burn-in was added via recapitation in pyslim. Each condition is described in detail below:

1) Neutral model: To simulate data in the absence of selection, we overlay all mutations on recorded tree sequences with pyslim.
2) Background selection ("BGS"): To simulate under a model of BGS, we simulated 75% of mutations as deleterious and 25% as selectively neutral. Deleterious mutations had a dominance value of 0.25, corresponding to partially recessive mutations. Selection coefficients for deleterious mutations were drawn from a gamma distribution with a mean and shape of -0.000133 and 0.35 (pre-scaling; mean of -0.0133 post-scaling), corresponding to estimates from *Drosophila* (Huber *et al.* 2017; Schrider 2020). With the population sizes

used in our simulations, this corresponds to a mean $2Ns$ = -33.25. We included a burn-in period in which background selection was acting for 25000 generations prior to population splitting.

3) Selective sweep in the ancestral population ("sweep ancestor"): When simulating a selective sweep in the ancestor of the two populations, the selection coefficient was drawn from a uniform (0.001, 0.005) distribution pre-scaling (0.1, 0.5, post-scaling) with a dominance of 1. At generation 1, a single selectively advantageous mutation was introduced into the ancestral population at position 5000 (i.e. the middle of the locus). Then, until generation 1000, we checked whether the mutation had fixed or been lost. If it had been fixed, the two populations split at generation 1000 and the simulation proceeded. If the mutation was lost, we restarted at generation 1 and repeated the procedure until the mutation fixed.

4) Selective sweep in population 1 ("sweep p1"): Immediately after divergence, a selectively advantageous mutation with a selection coefficient and dominance as in the linked ancestor simulation was introduced into population 1 at position 5000. If the mutation was lost before migration between populations began (or before the end of the simulation in the no migration model), we restarted the simulation. The mutation had no fitness effect in population 2.

5) Adaptive introgression ("adaptive int"): For this model, only one migration direction was considered (p1_p2). The selection coefficient and dominance were the same as in the model with a selective sweep in P1, except that the mutation was also advantageous in P2. We did not require that the advantageous mutation actually introgressed in these simulations.

We also evaluated the effects of a more biologically realistic model by including variation in mutation rate, recombination rate, and selection coefficients across genomic segments using the "real BGS-weak CNE" approach described in Schrider (2020). Following Schrider (2020), we used annotation data from the University of California Santa Cruz (UCSC) Table Browser for the *D. melanogaster* genome (release 5 / dm3; Adams *et al.* 2000). We also used the *D. melanogaster* recombination map from Comeron *et al.* (2012). Briefly, each simulated replicate was modelled after a randomly selected genomic region with selection coefficients 10-fold smaller in conserved noncoding elements (CNEs) than in coding regions. We modelled windows based on the *Drosophila* genome. For each 10 kb window, we selected an endpoint (constrained to be a multiple of 10 kb). Windows with >= 75% assembly gaps were not allowed, but otherwise windows were drawn randomly with replacement. The locations of annotated exons and phastCons elements were recorded, and deleterious mutations occurred only at these sites. We used recombination rates drawn from the *Drosophila* recombination map for the selected window. We drew the mutation rate from a uniform (3.445e-9, 3.445e-8) distribution (pre-scaling) for each simulated region. We refer to these simulations as the 'complex genomic architecture' condition below.

We simulated 10,000 10-kb regions for the BGS and neutral conditions and 1,500 10-kb regions for each sweep condition. These simulated regions were used to build datasets for downstream analyses. We sampled 20 diploid individuals (40 chromosomes) total, 10 per population. We then constructed a genotype matrix, discarding sites that were constant in our sample and sites with multiple mutations in the sample. These simulated regions were used to construct test datasets (Table 1) for downstream analyses with fastsimcoal2, ∂a∂i, and BPP. Test datasets for

BGS and neutral conditions consisted of the 10,000 regions simulated under the corresponding condition and model. For the sweep conditions (sweep p1, sweep ancestor, and adaptive introgression), we constructed test datasets by sampling 500 (5%), 1000 (10%), or 1500 (15%) regions simulated under the corresponding sweep condition and the remainder of datasets (9500, 9000, or 8500, respectively) from the corresponding set of simulations under BGS. Sampling was conducted independently to generate datasets for analyses with SFS-based methods (fastsimcoal2 and ∂a∂i) and BPP. For SFS-based methods, we constructed 100 replicate site frequency spectra (SFS) for each condition by sampling a single SNP per region in the test dataset (see below for additional detail). For analyses with BPP, for each condition, we generated a 500-bp alignment for each region included in the test dataset (see below for additional detail).

Comparing models and estimating migration rates in ∂a∂i

To estimate migration rates in ∂a∂i, we constructed SFS for all simulated datasets (54 datasets per divergence time; Table 1). We built 100 replicate SFS for each dataset, sampling a single SNP with replacement from each simulated fragment (10,000 per dataset). When constructing the SFS for ∂a∂i, we did not populate the monomorphic cell.

We estimated migration rates, population sizes, and divergence times using the split_mig model in ∂a∂i v.2.3.0 (Gutenkunst *et al.* 2009). This model includes two populations, with sizes V_1 and V_2 relative to the ancestral population, a divergence time, τ, and two migration rates, M_{12} and M_{21}. We used starting parameter estimates of 0.1 for V_1 and V_2, 0.01 for M_{12} and M_{21}, and 0.5 *f*or τ. We set the lower and upper bounds for v_1 and v_2 to 1e-3 and 5, respectively. The lower and upper bounds for the migration rate parameters were set to 0 and 5, respectively. For τ, the

bounds depended on the divergence time of the simulated dataset being analyzed. The upper and lower bounds were set to 0.005 and 1 for low divergence, 0.02 and 4 for medium divergence, and 0.04 and 16 for high divergence. We perturbed parameters using the perturb_params function in ∂a∂i. Then, we optimized parameters using the BOBYQA algorithm, the default algorithm in ∂a∂i. We performed a maximum of 400 evaluations.

For all datasets without migration, we also estimated parameters by maximizing the likelihood of the same model but with migration rates set to zero. Then, we compared the likelihood of the split_mig model to the likelihood of the model without migration using a likelihood ratio test (LRT). We calculated the test statistic Δ as:

$\Delta = 2 * (ln\left(split_{mig}\right) - ln\left(no_{mig}\right))$. We then computed the *p*-value using a χ^2 distribution with two degrees of freedom, and we rejected the null (no migration) model at a significance level of 0.01.

Comparing models and estimating migration rates in fastsimcoal2

To estimate migration rates in fastsimcoal2, we used the SFS constructed for ∂a∂i, except that we included the number of monomorphic sites. To calculate the number of monomorphic sites, we used the following equation:

$$N_{monomorphic} = \sum_{i=1}^{10000} (10000 - x_i) \times \frac{1}{x_i}$$

where x_i is the number of segregating sites in fragment *i*. This calculation accounts for the fact that we only sampled a single segregating site per fragment.

We estimated migration rates, population sizes, and divergence times in fastimcoal2 v.2.7.0.9 (Excoffier *et al.* 2013) using a model with the same parameterization as used in ∂a∂i. We set the mutation rate to the value used in the simple simulations (1e-8). Note that the units used in fastsimcoal2 differ from those used in ∂a∂i. We set the minimum bounds for migration rates to zero, the minimum bounds for population sizes to 1250, and the minimum bounds for the relative population sizes v_1 and v_2 to 1e-2. The minimum bounds on the divergence time were set to 1250 for low divergence, 5000 for medium divergence, and 10000 for high divergence. We used 100,000 simulations to estimate the expected SFS and performed 40 ECM cycles to estimate parameters. We also compared the migration model to a model with migration rates set to zero using an LRT with two degrees of freedom as described above for ∂a∂i.

Estimating migration rates with BPP

We estimated introgression probabilities and divergence times in BPP v4.4.0 (Flouri *et al.* 2020). Notably, BPP uses the MSci model, which models an instantaneous introgression event, rather than continuous migration as modelled in our simulations and by ∂a∂i and fastsimcoal2. We first generated sequence alignments for each region equivalent to a 500-bp locus. We generated these alignments from the output matrices generated from our simulations as follows. We used the first 500 base pairs of each of the 10,000 fragments composing the test dataset. For constant sites, we randomly selected a base (‘A’, ‘T’, ‘C’, or ‘G’) to use in the alignment. For segregating sites, we drew two base pairs from (‘A’, ‘T’, ‘C’, ‘G’) at random without replacement and assigned one as the ancestral allele and one as the derived allele. For each simulation condition (Table 1), we created 20 replicate datasets with 500 500-bp loci by sampling loci without replacement from our

simulated fragments. In BPP we fixed the species tree to a two-population tree with introgression allowed in both directions and used an inverse gamma (3, 0.01) prior for the parameters θ and τ. Since the inverse gamma prior is a conjugate prior for θ, this allowed the θ parameters to be integrated out analytically, improving run times (Hey and Nielsen 2007). We allowed mutation rates to vary across loci using the a_mubar, b_mubar, and a_mui priors. We set a_mubar and b_mubar to 0, so that mutation rates were relative. We set a_mui equal to 2, and we used the iid prior. We also used the heredity scalar to allow for variation in effective population sizes across loci. For the heredity scalar, we used a Gamma(4,4) prior. We collected 500,000 samples from the posterior after discarding the first 20,000 samples as burnin and sampling every 2 iterations. Some runs did not finish in 90 or 96 hours, but we collected a minimum of 447,060 samples for all runs. These samples were used to estimate all parameters (using the posterior mean). To assess convergence, we used effective sample size (ESS) values. In order to use the results from BPP to provide a binary determination of the presence of migration, we asked whether the highest posterior density interval (HDI) for migration parameters included 0.

Results

Selection leads to false inferences of migration using ∂a∂i

When combined with a complex genomic architecture including variation in recombination and mutation rates, selection often resulted in false inferences of migration in ∂a∂i using a likelihood ratio test (LRT). For the lowest divergence times, false positive rates were elevated even in the absence of selection and were not heavily impacted by selection (Figure 1a,b, Supporting Figure S1). The rates of rejection ranged from 19% to 33%, when only 1% false positives are expected at this *p*-value. When divergence times were moderate, there was a slight elevation of false

positive rates in the presence of a selective sweep in the ancestral population and a complex genomic architecture (2% rejected; Figure 1c,d). Most strikingly, for the highest divergence times considered here, the isolation-only model was always erroneously rejected in favor of the isolation-with-migration model in the presence of selection and a complex genomic architecture (Figure 1f). Interestingly, it did not seem to matter which type of selection there was (background selection vs. selective sweeps), when selection occurred (ancestrally vs. post-speciation), nor the fraction of all loci affected by sweeps (results for 5% and 10% shown in Supporting Figure S1). This is perhaps unsurprising, as the background data in these cases is drawn from the BGS scenario, which alone causes a false positive rate of 100% for the highest divergence times and the complex genomic architecture. Although the unconstrained model (including migration) should always have a higher likelihood than the constrained model (without migration), this was not always the case. Particularly in the medium divergence case, in ∂a∂i we observed instances in which the constrained model had higher likelihoods, indicating potential issues accurately approximating the likelihood for some datasets (Supporting Figure S2).

Perhaps as expected given the LRT results, the complex genomic architecture in combination with selection resulted in elevated estimates of migration rates in ∂a∂i. As with the LRT, results for the lowest divergence time lack any clear signal related to selection: non-zero rates are observed even in the absence of selection in very recently diverged populations (Figure 2a-b). Again, the most striking impact was seen in the high divergence case when selection was combined with a complex genomic architecture (Figure 2f, Supporting Figures S3). In this case, migration rates were substantially overestimated when the scenario included BGS or a selective

sweep compared to the neutral case (Figure 2f). Results were qualitatively similar whether 5, 10, or 15% of windows in sweep datasets experienced a sweep (Supporting Figures S5-S6). When simulations included migration, migration rate estimates tended to be higher under the complex genomic architecture and were slightly lower in the presence of selection (Supporting Figures S3-S6).

In the absence of migration and selection, ancestral θ was underestimated (Supporting Figure S7). Selection tended to reduce estimates of ancestral θ, while a complex genomic architecture tended to increase estimates of ancestral θ relative to the simple genomic architecture. The presence of migration led to increased estimates of ancestral θ (Supporting Figure S7). We also estimated the size of each population relative to the ancestral population (V_1, V_2, Supporting Figures S8-S9). Notably, especially under the simple genomic architecture, estimates of V_1 and V_2 tended to compensate for mistakes in the estimates of ancestral θ. In other words, when ancestral θ was underestimated, V_1 and V_2 tended to be overestimated, and vice versa. Divergence time estimates were fairly accurate in the absence of migration and selection, although they were somewhat overestimated in the high divergence case (Supporting Figure S10). When combined with a simple genomic architecture, all forms of selection led to overestimates of divergence times in the absence of migration. However, when combined with a complex genomic architecture and moderate or high divergence times, selection either had no effect or led to underestimated divergence times. Divergence times were underestimated in the presence of migration, except for at the lowest divergence times.

Selection leads to false inferences of migration using fastsimcoal2

When combined with a complex genomic architecture including variation in recombination and mutation rates, selection often resulted in false inferences of migration in fastsimcoal2. For the lowest divergence times, we rarely rejected the isolation-only model (one rejection at $p<0.01$ across eight conditions with 100 replicates each) (Figure 3a-b, Supporting Figure S11). When divergence times were medium, false positive rates were elevated across all conditions, even in the absence of selection: rates of rejection ranged from 15% to 35% (Figure 3c-d). For the highest divergence times considered here, the isolation-only model was always erroneously rejected in favor of the isolation-with-migration model in the presence of selection and a complex genomic architecture (Figure 3f). Although the unconstrained model (including migration) should always have a higher likelihood than the constrained model (without migration), as with ∂a∂i this was not always the case. Particularly in the low divergence case, in fastsimcoal2 we observed instances in which the constrained model had higher likelihoods, indicating potential issues accurately approximating the likelihood (Supporting Figure S12).

As with ∂a∂i, the complex genomic architecture in combination with selection resulted in elevated estimates of migration rates in fastsimcoal2 (Figure 4, Supporting Figures S13). Most notably, when divergence times were high, migration rates were substantially overestimated in the presence of a complex genomic architecture and selection relative to the neutral case (Figure 4f). Results were qualitatively similar whether 5, 10, or 15% of windows in sweep datasets experienced a sweep (Supporting Figures S15, S16). When simulations included migration, migration rate estimates tended to be lower under the complex genomic architecture and were slightly elevated in the presence of selection (Supporting Figures S13-S16).

Estimates of the ancestral population sizes were fairly accurate in fastsimcoal2 (Supporting Figure S17). The relative population sizes V_1 and V_2 were overestimated across all models and conditions (Supporting Figures S18, S19). In the absence of selection, migration, and variation in mutation and recombination rates across the genome, divergence time estimates were accurate (Supporting Figure S20a,d,g). Divergence time estimates were higher under the complex genomic architecture compared to the simple genomic architecture, and were reduced in the presence of selection (Supporting Figure S20). Divergence times tended to be underestimated in the presence of migration (Supporting Figure S20).

Selection leads to false inferences of migration in BPP

Using BPP, non-zero migration rates were often inferred in the absence of migration (Figure 5; Supporting Figures S21-S23). The highest posterior density interval (HDI) of the two migration parameters, φ-X and φ-Y, rarely contained zero in the low divergence case, regardless of the presence of selection (Figure 5a,b; Supporting Figure S21). In the medium and high divergence cases, BPP still inferred non-zero migration often—the HDI did not include zero 0%-45% of the time (Figure 5c-f; Supporting Figures S21-S23). Particularly in the high divergence case, BPP inferred non-zero migration most often in the presence of selection and a complex genomic architecture (Figure 5c-f; Supporting Figures S21-S23). Although Bayesian methods do not have an equivalent "false positive" rate to frequentist methods, we would not expect such a high proportion of HDIs to not include the true parameter value. As expected given that these HDIs do not overlap zero, migration rate estimates were elevated in the presence of selection and complex genomic architectures, particularly in the high divergence case (Figure 6; Supporting Figures S24). When simulations included migration, estimates of φ-X and φ-Y were generally

higher under the complex genomic architecture, and estimated rates were slightly lower in the presence of selection (Supporting Figures S24-S25). Results were qualitatively similar whether 5, 10, or 15% of sweep datasets experienced a sweep (Supporting Figures S26-S27).

We also evaluated whether there was evidence for a lack of convergence in BPP runs by examining ESS values. We focused on ESS values for the log likelihood, along with the φ-*X* and φ-*Y* parameters. In the low divergence case, ESS values for the φ parameters were often low, indicating a lack of convergence, particularly under the complex genomic architecture; however, ESS values were generally greater than 200 in the medium and high divergence cases (Supporting Figure S28). Under some conditions, there was evidence of a correlation between ESS values and parameter estimates (e.g., for φ-*X* under a neutral model with a complex genomic architecture, Supporting Figure S29). This suggests that in some (but not all) instances, assessing convergence may allow researchers to identify problematic cases.

In the absence of selection and migration, divergence times were overestimated when divergence times were medium or low and accurately estimated when divergence times were high (Supporting Figure S30). Divergence time estimates were reduced in the presence of selection, and were elevated under the complex genomic architecture relative to the simple genomic architecture. The presence of migration led to underestimates of divergence times when divergence times were high.

Discussion

Our results paint a concerning picture for popular methods currently used to estimate migration rates between sister populations or species. The three approaches tested (fastsimcoal, ∂a∂i, and BPP) all showed high rates of misleading results in the presence of selection (Figures 1, 3, 5). This was particularly true under more realistic models that included variation in recombination and mutation rates, as well as in the strength of selection. Given that large portions of the genomes of many species are impacted by selection (Begun *et al.* 2007; McVicker *et al.* 2009; Sella *et al.* 2009; Langley *et al.* 2012; Corbett-Detig *et al.* 2015; Phung *et al.* 2016; Pouyet *et al.* 2018) and that variation in mutation and recombination rates across the genome is the norm, these results suggest that many inferences of introgression may be artifacts that do not reflect biological reality.

Numerous studies have found that ignoring natural selection can negatively impact different types of demographic inferences. Selection can lead to false inference of population size changes (e.g., Ewing and Jensen 2016; Schrider *et al.* 2016; Johri *et al.* 2021) and several studies have suggested that selection can also mislead inferences of migration (e.g., Cruickshank and Hahn 2014; Mathew and Jensen 2015; Roux *et al.* 2016). To account for the effects of selection, several approaches for inferring migration have been developed that allow for heterogeneous effects of selection among loci (Sousa *et al.* 2013; Roux *et al.* 2016; Sethuraman *et al.* 2019; Fraïsse *et al.* 2021). Although these methods vary in the types of inferences that can be made—from locus-specific migration rates to genome-wide migration rates—they all model selection by allowing for variation in θ among loci. BPP also allows for variation in rates across loci, either by using the a rate multiplier for θ and τ or for θ alone (Flouri *et al.* 2020). In our analyses, we used the both rate multipliers and still found that selection led to false inferences of migration. It

is not clear that the results using any of these similar methods would differ qualitatively from those shown in Figure 5 (to our knowledge, none have been tested against a no-migration scenario with selection).

Approaches that allow θ to vary among loci assume that the only effect of selection is to re-scale the height of coalescent genealogies, with no change in the relative length of branches. However, selection will skew the allele frequency of linked neutral variants, whether due to advantageous mutations ("hitchhiking"; Aguade *et al.* 1989; Braverman *et al.* 1995) or deleterious mutations ("background selection"; Zeng and Charlesworth 2011; Zeng 2013; Cvijović *et al.* 2018); non-equilibrium population histories can make the effect of background selection on the allele frequency spectrum even stronger (Torres *et al.* 2018). These effects are likely to be driving the false inferences of migration found in all three methods tested here, with shifts in the allele frequency spectrum—or more subtle effects of selection, such as increased variance in coalescence times—causing generally more complex models to be favored over simpler models that do not include migration. Importantly, the effects of linked selection extend well beyond exons of protein-coding genes. The "neutral" assumption made by ∂a∂i, fastsimcoal2, and BPP is not that the sequences analyzed be free from direct (negative) selection, but rather that the genealogical dynamics at every locus follow neutral expectations. This means that non-coding, non-functional regions of the genome that are free from constraint can be affected by exactly the sort of linked selection that drives false inferences here. Therefore, obtaining similar estimates of migration in coding and non-coding datasets is likely to be driven by the near-universal effects of linked selection across a genome, rather than by a lack of selective effects on protein-coding genes.

Phylogenetic methods for inferring gene flow are much more robust to assumptions about selection, largely because they often depend on asymmetries in tree topologies (Hibbins and Hahn 2022). However, the dependence on tree asymmetry also means that they cannot be used to detect gene flow between sister lineages. So what is the way forward? Our results, along with previous studies, highlight several potential possibilities. Statistical-learning approaches are highly flexible (Schrider and Kern 2018), and one path forward involves training these algorithms under appropriate models that incorporate selection. One such approach has successfully used approximate Bayesian computation to jointly estimate DFEs and population size histories (Johri *et al.* 2020). Beyond training statistical learning algorithms on more realistic training data, techniques for domain adaptation—a subfield of machine learning that aims to adapt an algorithm trained on the source domain (e.g., on simulations under a model of interest) to the target domain (i.e., empirical data; reviewed in Wilson and Cook 2020) offer a promising path forward to accommodating complex biological realities in population genomics (e.g., Mo and Siepel 2023). Regardless of which methods are used, accurate inferences of demographic histories will have to include the complexities introduced by selection (Hahn 2008; Sheehan and Song 2016; Johri *et al.* 2020).

Accurate inferences of introgression histories are important because they can tell us about modes of speciation. A large number of studies supporting gene flow between closely related populations have been interpreted as lending support to speciation-with-gene-flow models; our results highlight that caution is warranted in these interpretations (cf. Cruickshank and Hahn, 2014), though there are other reasons to exercise caution as well (Yang *et al.* 2017). Although

many estimates were non-zero, migration rates estimated for datasets generated under models including selection were low absolutely. In fastsimcoal2, estimated migration rates per generation were on the order of 10^{-9}, in ∂a∂i, $2 \times N_{ref} \times m_{ij}$ was on the order of 0.002 (~10^{-9} migrants per generation), and in BPP the weight of the hybrid edge was on the order of 0.002 (Figures 2, 4, 6). Moving forward, we recommend that inferences of migration between sister populations made without considering selection be interpreted with caution, particularly when inferred rates of migration are low. Importantly, we do not believe that the results found here with a limited set of selective scenarios and a limited set of introgression histories can fully describe the effects of selection, mutation, and recombination on inaccurate demographic inferences. Inferences about the presence, direction, and timing of introgression (including whether speciation and introgression occur at the same time—i.e. homoploid hybrid speciation) may all be affected by models that ignore natural selection and complex genomic architectures. We hope that new methods can be developed to overcome these obstacles.

Data Availability

Simulated data formatted for various programs are available on Figshare (DOI: 10.6084/m9.figshare.24354277). All scripts are available on GitHub (https://meganlsmith.github.io/selectionandmigration/).

Acknowledgements

This work was supported by a National Science Foundation (NSF) postdoctoral fellowship to MLS (DBI-2009989) and NSF grant to MWH (DBI-2146866).

Tables

Table 1: Simulation conditions considered in this study.

Genomic Architecture	Model	Condition
simple	no migration	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)
	p1_p2	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)
		Adaptive Introgression (5, 10, 15%)
	p2_p1	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)
complex	no migration	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)
	p1_p2	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)
		Adaptive Introgression (5, 10, 15%)
	p2_p1	Neutral
		BGS
		Sweep P1 (5, 10, 15%)
		Sweep Ancestor (5, 10, 15%)

Figures

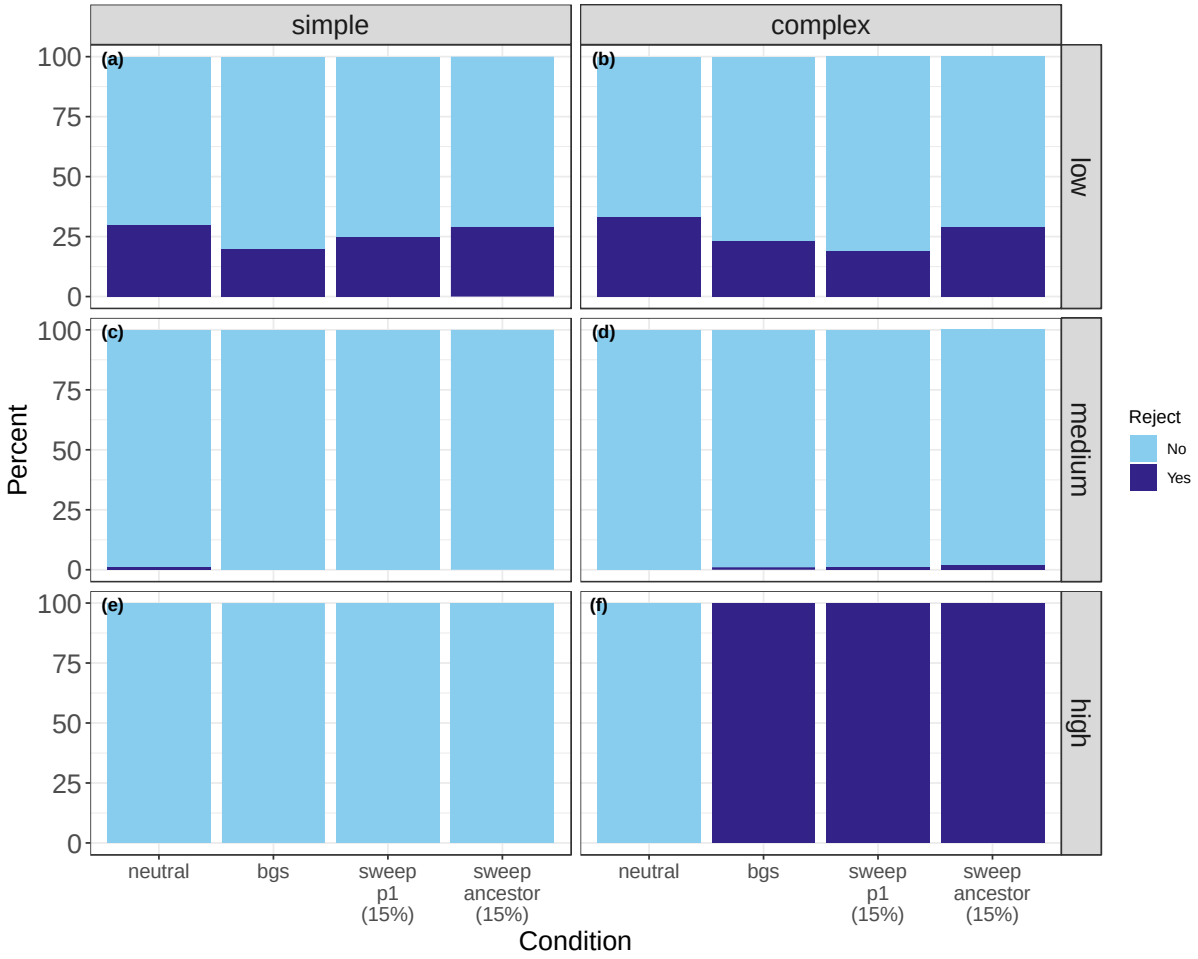

Figure 1. Results of the LRT in ∂a∂i. Light blue indicates cases where we failed to reject the isolation-only model, and dark blue indicates cases where we rejected the isolation-only model in favor of the isolation-with-migration model. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

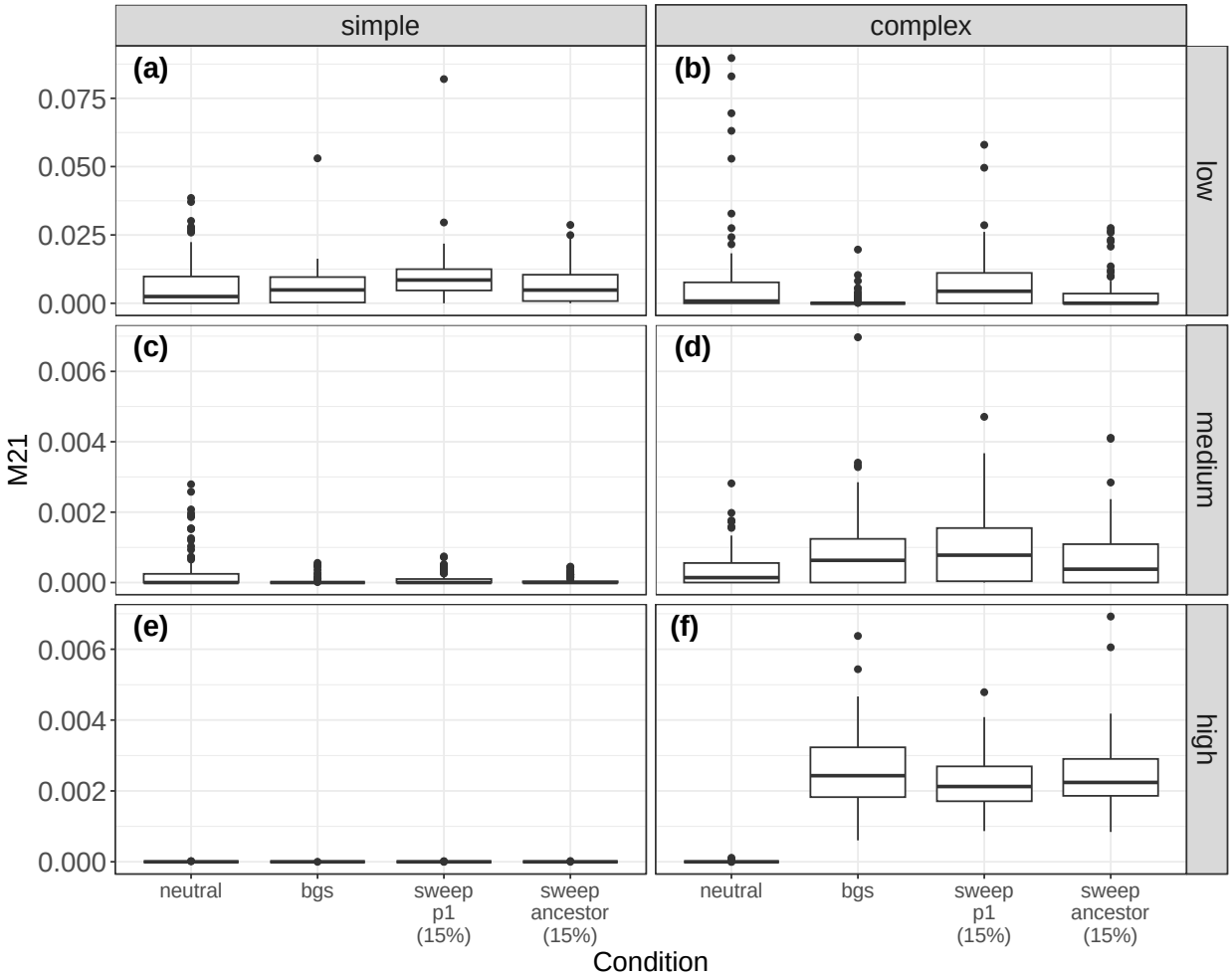

Figure 2. Estimates of M_{21} in ∂a∂i for datasets simulated without migration. Estimates are in units of $2 \times N_{ref} \times m_{ij}$. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

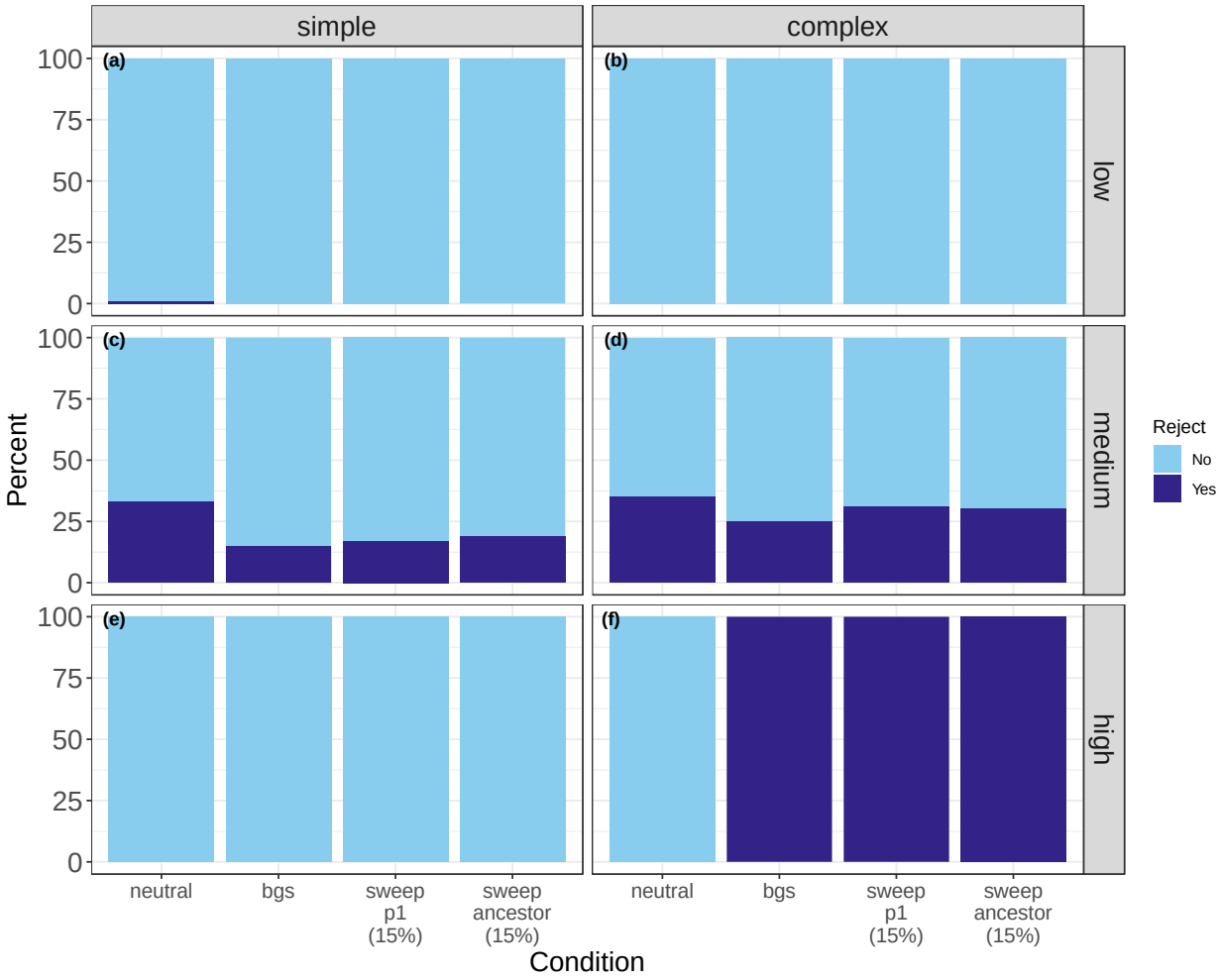

Figure 3. Results of the LRT in fastsimcoal2. Light blue indicates cases where we failed to reject the isolation-only model, and dark blue indicates cases where we rejected the isolation-only model in favor of the isolation-with-migration model. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

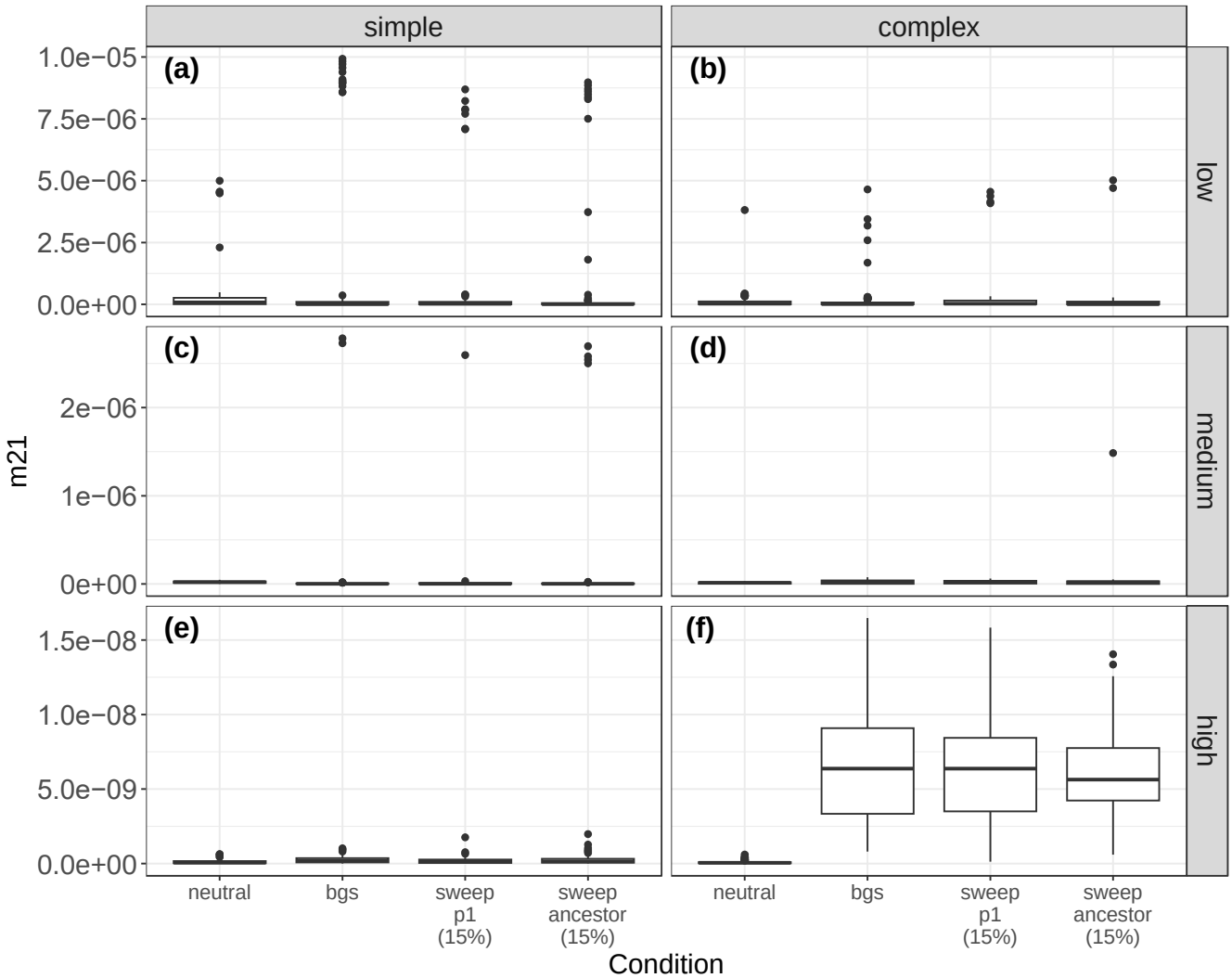

Figure 4. Estimates of m_{21} in fastsimcoal2 for datasets simulated without migration. The rate m_{ij} is the probability of any gene moving from population *i* to population *j* backwards in time each generation. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

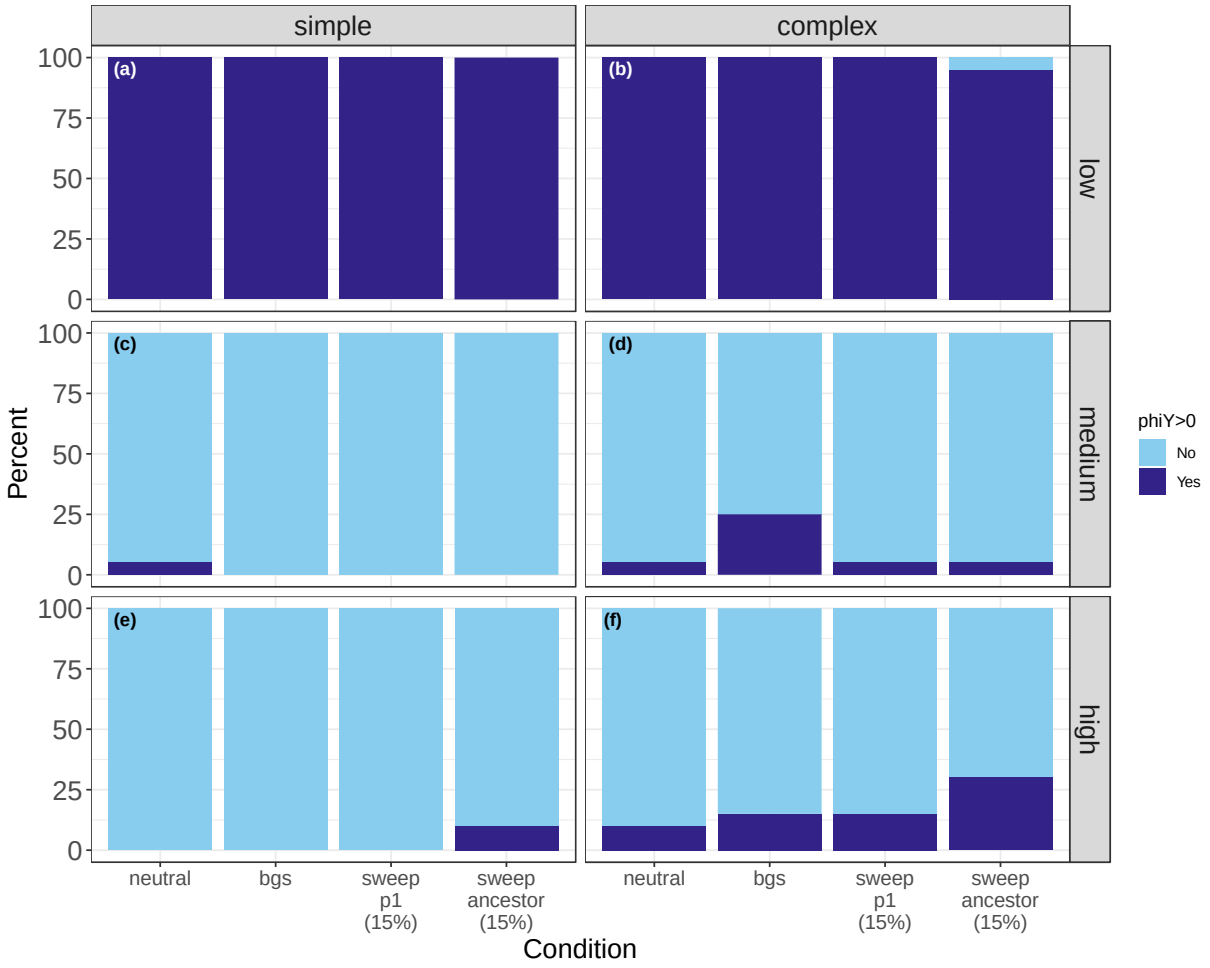

Figure 5. The proportions of replicates with 95% highest posterior density intervals for φ-*Y* that do not include zero in the absence of migration in BPP. Light blue indicates cases where the HDI includes zero, and dark blue indicates cases where the HDI does not include zero. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

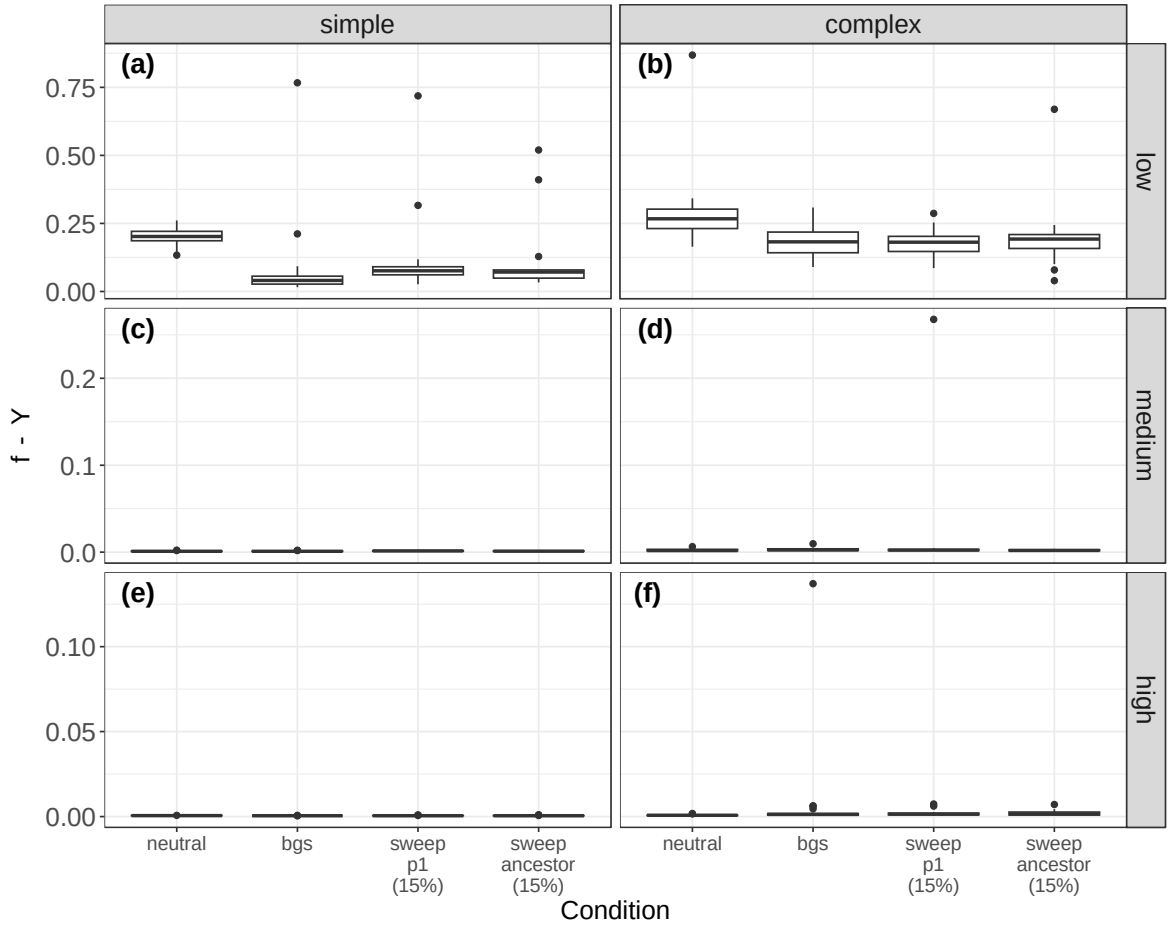

Figure 6. Mean posterior estimates of φ-*Y* in BPP for models without migration. . φ-*Y* is the weight of the introgression edge Y in the MSci model. a) results from the low divergence time scenario with a simple genomic architecture; b) results from the low divergence time scenario with a complex genomic architecture; c) results from the medium divergence time scenario with a simple genomic architecture; d) results from the medium divergence time scenario with a complex genomic architecture; e) results from the high divergence time scenario with a simple genomic architecture; f) results from the high divergence time scenario with a complex genomic architecture.

References

Adams M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science 287: 2185–2195.

Aguade M., N. Miyashita, and C. H. Langley, 1989 Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. Genetics 122: 607–615.

Begun D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh, *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biology 5: e310.

Braverman J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140: 783–796.

Comeron J. M., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genetics 8: e1002905.

Corbett-Detig R. B., D. L. Hartl, and T. B. Sackton, 2015 Natural selection constrains neutral diversity across a wide range of species. PLoS Biology 13: e1002112.

Cruickshank T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23: 3133–3157.

Cutter A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. Nature Reviews Genetics 14: 262–274.

Cvijović I., B. H. Good, and M. M. Desai, 2018 The effect of strong purifying selection on genetic diversity. Genetics 209: 1235–1278.

Ewing G. B., and J. D. Jensen, 2016 The consequences of not accounting for background selection in demographic inference. Molecular Ecology 25: 135–141.

Excoffier L., T. Hofer, and M. Foll, 2009 Detecting loci under selection in a hierarchically structured population. Heredity 103: 285–298.

Excoffier L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll, 2013 Robust demographic inference from genomic and SNP data. PLoS Genetics 9: e1003905.

Flouri T., X. Jiao, B. Rannala, and Z. Yang, 2020 A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Molecular Biology and Evolution 37: 1211–1223.

Fraïsse C., I. Popovic, C. Mazoyer, B. Spataro, S. Delmotte, *et al.*, 2021 DILS: Demographic inferences with linked selection by using ABC. Molecular Ecology Resources 21: 2629–2644.

Geneva A. J., C. A. Muirhead, S. B. Kingan, and D. Garrigan, 2015 A new method to scan genomes for introgression in a secondary contact model. PLoS One 10: e0118621.

Gutenkunst R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics 5: e1000695.

Hahn M. W., 2008 Toward a selection theory of molecular evolution. Evolution 62: 255–265.

Haller B. C., and P. W. Messer, 2019 SLiM 3: forward genetic simulations beyond the Wright–Fisher model. Molecular Biology and Evolution 36: 632–637.

Hey J., and R. Nielsen, 2007 Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Sciences 104: 2785–2790.

Hibbins M. S., and M. W. Hahn, 2022 Phylogenomic approaches to detecting and characterizing introgression. Genetics 220: iyab173.

Huber C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller, 2017 Determining the factors driving selective effects of new nonsynonymous mutations. Proceedings of the National Academy of Sciences 114: 4465–4470.

Johri P., B. Charlesworth, and J. D. Jensen, 2020 Towards an Evolutionarily Appropriate Null Model: Jointly Inferring Demography and Purifying Selection. Genetics 215: 173–192.

Johri P., K. Riall, H. Becher, L. Excoffier, B. Charlesworth, *et al.*, 2021 The impact of purifying and background selection on the inference of population history: problems and prospects. Molecular Biology and Evolution 38: 2986–3003.

Joly S., P. A. McLenachan, and P. J. Lockhart, 2009 A statistical approach for distinguishing hybridization and incomplete lineage sorting. The American Naturalist 174: E54–E70.

Kern A. D., and M. W. Hahn, 2018 The neutral theory in light of natural selection. Molecular Biology and Evolution 35: 1366–1371.

Langley C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider, *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192: 533–598.

Luqman H., A. Widmer, S. Fior, and D. Wegmann, 2021 Identifying loci under selection via explicit demographic models. Molecular Ecology Resources 21: 2719–2737.

Mallet J., N. Besansky, and M. W. Hahn, 2016 How reticulated are species? BioEssays 38: 140–149.

Mathew L. A., and J. D. Jensen, 2015 Evaluating the ability of the pairwise joint site frequency spectrum to co-estimate selection and demography. Frontiers in Genetics 6: 268.

McVicker G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genetics 5: e1000471.

Mo Z., and A. Siepel, 2023 Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data. bioRxiv. https://doi.org/10.1101/2023.03.01.529396

Nielsen R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark, 2007 Recent and ongoing selection in the human genome. Nature Reviews Genetics 8: 857–868.

Payseur B. A., and L. H. Rieseberg, 2016 A genomic perspective on hybridization and speciation. Molecular Ecology 25: 2337–2360.

Phung T. N., C. D. Huber, and K. E. Lohmueller, 2016 Determining the effect of natural selection on linked neutral divergence across species. PLoS Genetics 12: e1006199.

Pouyet F., S. Aeschbacher, A. Thiéry, and L. Excoffier, 2018 Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. eLife 7: e36317.

Rosenzweig B. K., J. B. Pease, N. J. Besansky, and M. W. Hahn, 2016 Powerful methods for detecting introgressed regions from population genomic data. Molecular Ecology 25: 2387–2397.

Roux C., C. Fraisse, J. Romiguier, Y. Anciaux, N. Galtier, *et al.*, 2016 Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biology 14: e2000234.

Schrider D. R., A. G. Shanku, and A. D. Kern, 2016 Effects of linked selective sweeps on demographic inference and model selection. Genetics 204: 1207–1223.

Schrider D. R., and A. D. Kern, 2018 Supervised machine learning for population genetics: a new paradigm. Trends in Genetics 34: 301–312.

Schrider D. R., 2020 Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. Genetics 216: 499–519.

Sella G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the *Drosophila* genome? PLoS Genetics 5: e1000495.

Sethuraman A., V. Sousa, and J. Hey, 2019 Model-based assessments of differential introgression and linked natural selection during divergence and speciation. bioRxiv. https://doi.org/10.1101/786038

Sheehan S., and Y. S. Song, 2016 Deep learning for population genetic inference. PLoS Computational Biology 12: e1004845.

Sousa V. C., M. Carneiro, N. Ferrand, and J. Hey, 2013 Identifying loci under selection against gene flow in isolation-with-migration models. Genetics 194: 211–233. https://doi.org/10.1534/genetics.113.149211

Tellier A., P. Pfaffelhuber, B. Haubold, L. Naduvilezhath, L. E. Rose, *et al.*, 2011 Estimating parameters of speciation models based on refined summaries of the joint site-frequency spectrum. PLoS One 6: e18155.

Torres R., Z. A. Szpiech, and R. D. Hernandez, 2018 Human demographic history has amplified the effects of background selection across the genome. PLoS Genetics 14: e1007387.

Williamson S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, *et al.*, 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proceedings of the National Academy of Sciences 102: 7882–7887.

Wilson G., and D. J. Cook, 2020 A survey of unsupervised deep domain adaptation. ACM Transactions on Intelligent Systems and Technology (TIST) 11: 1–46.

Yang M., Z. He, S. Shi, and C.-I. Wu, 2017 Can genomic data alone tell us whether speciation happened with gene flow? Molecular Ecology 26: 2845–2849.

Zeng K., and B. Charlesworth, 2011 The joint effects of background selection and genetic recombination on local gene genealogies. Genetics 189: 251–266.

Zeng K., 2013 A coalescent model of background selection with recombination, demography and variation in selection coefficients. Heredity 110: 363–371.