Genetics: Early Online, published on May 11, 2017 as 10.1534/genetics.117.200493

Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation.

Julien Jouganous*, Will Long*, Aaron P. Ragsdale* and Simon Gravel*,[1]
*Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada

ABSTRACT Understanding variation in allele frequencies across populations is a central goal of population genetics. Classical models for the distribution of allele frequencies, using forward simulation, coalescent theory, or the diffusion approximation, have been applied extensively for demographic inference, medical study design, and evolutionary studies. Here we propose a tractable model of ordinary differential equations for the evolution of allele frequencies that is closely related to the diffusion approximation but avoids many of its limitations and approximations. We show that the approach is typically faster, more numerically stable, and more easily generalizable than the state-of-the-art software implementation of the diffusion approximation. We present a number of applications to human sequence data, including demographic inference with a five-population joint frequency spectrum and a discussion of the robustness of Out-of-Africa model inference to the choice of modern population.

KEYWORDS Demographic inference; joint allele frequency spectrum; diffusion approximation; moments equations

Introduction

Understanding the role of demography and selection in shaping genetic diversity is a central challenge in population genetics. In recent years, genome sequencing experiments have generated large amounts of data that can be used to test and refine this understanding. Detailed models of human genetic diversity have shed a new light on the origins and history of modern humans and have helped researchers design and interpret biomedical studies.

Present-day genomes are the result of a large number of more or less randomly occurring mutations, recombinations, matings, and deaths. Individual events and their genomic consequences bear limited information about the past. We can, however, learn a lot by statistically integrating information across entire genomes and populations. Classical work has focused on simple summaries of genomic diversity that could be computed analytically, such as the number of pairwise differences between individuals, the number of segregating sites in a sample, or linkage disequilibrium (Crow and Kimura 1970). In recent years, the ability to simulate genomes and the availability of data has led to a wide array of summaries designed to identify different evolutionary forces (Patterson *et al.* 2012; Schiffels and Durbin 2014; Scheinfeldt and Tishkoff 2013). Here we consider the allele frequency spectrum (AFS): We compute the frequency of the derived allele at each site, and build a histogram of the number of sites observed at each frequency.

The AFS is a classical measure of diversity (Kimura 1964) that has seen a surge in popularity recently because of improved computational approaches to estimate it. Back-in-time approaches based on coalescence theory (Excoffier and Foll 2011; Excoffier *et al.* 2013; Kamm *et al.* 2017) can be extremely efficient for neutral simulations and can handle large number of populations, but they often become cumbersome or intractable in models with selection. Forward-in-time approaches tend to be more easily generalized to account for selection. Despite recent progress in discrete whole-population methods (Haller and Messer 2016), most current approaches are based on the diffusion approximation initiated by Fisher and Wright in the early 1930's (Fisher 1930; Wright 1931; Crow and Kimura 1970; Kimura 1964). In this approach, the evolution of the AFS is described as a continuous process through a partial differential equation. Several simulating tools based on this diffusion approximation have been implemented and distributed to the population genetics community (Gutenkunst *et al.* 2009; Lukić and Hey 2012; Lukić *et al.* 2011), among which $\partial a \partial i$ is probably the most commonly used. In the past few years, these numerical tools have been used widely and led to significant results from both historical

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Sunday 7th May, 2017%
[1]McGill University, Department of Human Genetics and Genome Quebec Innovation Centre, 740 Dr Penfield Avenue, Montreal, Quebec H3A 0G1, Canada. Email: simon.gravel@mcGill.ca

and biological points of view (Gutenkunst *et al.* 2009; Gravel *et al.* 2011; Schmutz *et al.* 2014).

Despite these successes, computational cost and stability remains an important limitation. For instance, available software can only handle up to 4 populations simultaneously (*∂a∂i* can handle 3). The time needed to perform individual simulations can be prohibitive when many simulations must be run, for example when performing bootstrap analysis for multiple models. Furthermore, biases inherent to Kingman's coalescent or the diffusion approximation, which were benign in small samples, can become important in contemporary datasets (Bhaskar *et al.* 2014; Spence *et al.* 2016).

Here, we describe a new and efficient method to simulate the AFS that is more efficient and more numerically stable than the state-of-the-art, yet does not require the diffusion approximation. In cases where the diffusion approximation is appropriate, this approach is equivalent to a moment representation of the diffusion equation that is analytically intuitive and numerically tractable. We integrate this moment approach into the *∂a∂i* inference framework to facilitate inference from allele frequency distributions across a broader range of problems than was previously possible.

Method and material

Heterozygosity rate evolution

First consider a large population of N diploid individuals evolving under the neutral Wright-Fisher model. Generations are discrete and individuals from generation $k+1$ receive alleles drawn randomly and with replacement from the parental alleles present in generation k. Because alleles in a diploid individual are inherited and transmitted independently in the neutral case, we can forget about diploid individuals and think of the population as a set of $2N$ haploid samples (or *ploids*, for short). We are interested in the expected number $\phi_n^k(i)$ of sites where the alternate allele is observed exactly i times in a sample of size n at generation k. We neglect correlations between sites and suppose that each locus is transmitted independently.

In this model, heterozygosity is proportional to $\phi_2^k(1)$, the number of heterozygous sites in sample of $n=2$ ploids. Under neutral Wright-Fisher reproduction, this follows the classical recursion

$$\Phi_2^{k+1}(1) = \left(1 - \frac{1}{2N}\right)\Phi_2^k(1). \tag{1}$$

The derivation of this recursion is simple: If two ploids at generation $k+1$ are inherited from the same parental ploid, which happens with probability $\frac{1}{2N}$, they are identical and do not contribute to $\Phi_2^{k+1}(1)$. Otherwise (with probability $1-\frac{1}{2N}$) they are drawn randomly without replacement from the previous generation. The probability that they are different is $\Phi_2^k(1)$, by definition, leading to Equation (1).

Neutral case AFS

We can similarly derive a recursion equation for the neutral AFS for arbitrary n and i. This time we draw n ploids from the previous generation. The probability that a given pair of ploids has the same parent is still $\frac{1}{2N}$. For simplicity, we suppose that at most one ploid pair shares a parent at each generation. The situation with multiple coalescent events per generation, which can occur for large sample sizes (Bhaskar *et al.* 2014), is discussed in Section Drift with multiple coalescences.

If there is no coalescence, the n ploids at generation $k+1$ are copied from n randomly drawn ploids at generation k and the distribution of allele frequencies does not change: $\Phi_n^{k+1} = \Phi_n^k$. If there is a coalescence, the n descendants are copied from a random set of $n-1$ parental haplotypes. The distribution of possible parental frequencies is described by Φ_{n-1}^k. To treat all possible coalescence cases in a unified way, we imagine that we always draw n parental ploids, with the understanding that one selected parental ploid may not leave descendants. By doing this, we can express Φ_n^{k+1} as a simple linear function of Φ_n^k :

$$\Phi_n^{k+1}(i) = \Phi_n^k(i) + \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n^k,$$

where $\tilde{\Delta}_{n,i}$ is a sparse linear operator describing the change in allele frequency due of drift in a single generation. Its coefficients are provided in Appendix Coefficients of the discrete operators.

De novo mutations can change the allele type and therefore also affect the allele frequency. Their effects on the i^{th} entry of the AFS is described by the linear operator $\mathcal{U}_i$:

$$\Phi_n^{k+1}(i) = \Phi_n^k(i) + \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n^k + \mathcal{U}_i\Phi_n^k. \tag{2}$$

As a first approximation and to easily compare to commonly used simulation tools, we consider the infinite-sites model (Kimura 1969), in which the only mutations allowed are at previously invariant sites, and at most one mutation per site can occur. In this model, the mutation term is a source term: $\mathcal{U}_i\Phi_n^k \simeq nu\delta_{i=1}$, with u the genome-wide mutation rate. In the infinite-sites model, we therefore have

$$\Phi_n^{k+1}(i) = \Phi_n^k(i) + \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n^k + nu\delta_{i=1}. \tag{3}$$

A more general form of the mutation operator, accounting for backwards mutations, is discussed in Appendix Finite genome model for mutations.

Equation (3) is formally equivalent to a numerical discretization of the diffusion approximation. However, it avoids two approximations. The first is the approximation of a discrete Wright-Fisher system by a time- and frequency-continuous diffusion. The second is the approximation of a continuous diffusion by a discrete numerical partial differential equation (PDE) solver.

For notational and numerical convenience, we approximate $\Phi_n^{k+1} - \Phi_n^k$ by the time derivative $\dot{\Phi}_n$ and work in continuous time:

$$\dot{\Phi}_n(i) = \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n + nu\delta_{i=1}. \tag{4}$$

Discrete-time results can recovered by interpreting $\dot{\Phi}_n$ as $\Phi_n^{k+1} - \Phi_n^k$ everywhere.

Modeling selection and dominance

To obtain similar recursion equations under selection, we consider a model of selection in which ploids are drawn uniformly from the previous generation, but are accepted with different rates depending on the fitness of the parent. Alleles drawn from a diploid parent are accepted with probability 1 for genotype aa, $1+hs$ for genotype aA, and $1+s$ for genotype AA: $s \leq 0$ is the selection coefficient, and h is the dominance coefficient. If a ploid is not accepted, it must be redrawn at random from the entire population.

In this model, contrary to the neutral case, we may need to draw more than n alleles from generation k to build ϕ_n^{k+1}. In what follows, we assume that $ns \ll 1$, so that at most one allele

is rejected from the sample at each generation. This assumption is also implicit in the diffusion approximation. Higher-order corrections can also be computed.

Assuming $ns \ll 1$, we find

$$\dot{\Phi}_n(i) = \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n + s\tilde{\nabla}_{h,i}\Phi_{n+2} + nu\delta_{i=1}, \tag{5}$$

where $\tilde{\nabla}_{h,n,i}$ is a sparse linear operator describing the change in sample allele frequencies due to selection. Its coefficients are given in Appendix Coefficients of the discrete operators.

To compute Φ_n^{k+1}, Equation (5) requires Φ_{n+2}^k, which itself requires Φ_{n+4}^{k-1}, and so forth. The evolution equation for Φ_n^k is not closed and cannot be solved directly in that form.

Moment closure

To find a closed approximation to Equation (5) and compute $\Phi_n^{k+1}(i)$ directly from $\Phi_n^k(i)$, we want to approximate the larger sample AFS Φ_{n+2}^k in terms of Φ_n^k. Intuitively, this should be possible for n large enough: a sample of size 102 adds little information that is not already contained in a sample of size 100. In fact, the number of variants to be discovered in a sample of size $20n$ can be estimated accurately from a sample of size n by jackknife extrapolation (Gravel and NHLBI GO ESP 2014).

Here we adapt jackknife extrapolation to perform the moment closure approximation. We consider a 3rd order jackknife to express entries of the higher order spectra Φ_{n+1} and Φ_{n+2} as linear combinations of 3 entries of Φ_n. Higher order jackknives would increase accuracy at the cost of computational complexity and stability. Since jackknives provide a linear approximation, we can write $\Phi_{n+2} \simeq \mathcal{J}_{n\to n+2}\Phi_n$ where $\mathcal{J}_{n\to n+2}$ is a sparse linear operator whose coefficients are provided in Appendix Jackknife approximation for moment closure. Under this approximation, the selection term becomes $s\tilde{\nabla}_{h,i}\Phi_{n+2}^k \simeq s\nabla_{h,i}\mathcal{J}_{n\to n+2}\Phi_n = s\hat{\nabla}_{h,n,i}\Phi_n$, where we defined $\hat{\nabla}_{h,n,i} = \tilde{\nabla}_{h,n,i}\mathcal{J}_{n\to n+2}$ as the closed approximation to the selection operator. The resulting system is a closed and linear system of ordinary linear differential equations:

$$\dot{\Phi}_n(i) = \frac{1}{4N}\tilde{\Delta}_{n,i}\Phi_n + s\hat{\nabla}_{h,n,i}\Phi_n + nu\delta_{i=1}. \tag{6}$$

Multiple populations and migrations

In practice, we often want to study the distribution of allele frequencies across multiple populations where mating is more common within than across populations. In such a case, we consider the multidimensional AFS $\Phi_{\mathbf{n}}(\mathbf{i})$ where $\mathbf{n}$ is a vector of sample sizes of the different populations. Thus, $\Phi_{n_1,\cdots,n_p}(i_1,\cdots,i_p)$ is the number of variants that are found i_1 times in population 1, i_2 times in population 2, and so forth.

Using the same method as for the single population case, we can derive a system of ordinary equations for the joint AFS $\Phi_{\mathbf{n}}$ (see Appendix Extension to multiple dimensions). Because of migration, however, the evolution of $\Phi_{\mathbf{n}}$ depends on $\Phi_{\mathbf{n}'}$ for $\mathbf{n}' \neq \mathbf{n}$: When creating n_i ploids in population i and n_j in population j, we may occasionally draw e.g. $n_i + 1$ ploids from i and $n_j - 1$ from j. Thus evolution equations for $\Phi_{\mathbf{n}'}$ are coupled for all $\mathbf{n}'$ such that $n'_j + n'_k = n_j + n_k$. We could therefore write evolution equations that are closed on the space of all $\Phi_{\mathbf{n}'}$. However, this is numerically costly. For weak migration, with $n_k m_{jk} \ll 1$ for all j,k, we can again use the jackknife to write uncoupled approximations to these equations. A simple way of achieving this is to write the evolution equation for $\Phi_{\mathbf{n}}$ in terms of the slightly larger AFS $\Phi_{\hat{\mathbf{n}}}$ with $\hat{\mathbf{n}}_j = \mathbf{n} + \mathbf{e}_j$ and $\mathbf{e}_j$ a unit vector with a unit value for the jth population. We then have

$$\begin{aligned}\dot{\Phi}_{\mathbf{n}}(\mathbf{i}) = \sum_{j=1}^{p}\Bigg[&\frac{1}{4N_j}\tilde{\Delta}_{n_j,i_j}\Phi_{\mathbf{n}} + s_j\hat{\nabla}_{j;h_j,n_j,i_j}\Phi_{\mathbf{n}} \\ &+ \sum_{k\neq j} m_{jk}\mathcal{M}_{jk}\Phi_{\hat{\mathbf{n}}_j} + n_j u\delta_{\mathbf{i}=\mathbf{e}_j}\Bigg],\end{aligned} \tag{7}$$

where $\tilde{\Delta}_{n_j,i_j}$ is the drift operator and $\hat{\nabla}_{j;h_j,n_j,i_j}$ is the selection operator for population j, $\mathcal{M}_{jk}$ is the linear migration operator from populations j to population k, and the migration rate m_{jk} is the proportion of ploids in population k whose parent was in population j. Parameters for the migration operator $\mathcal{M}_{jk}$ are provided in Appendix Coefficients of the discrete operators.

We can then use a jackknife approximation as above to write

$$\Phi_{\hat{\mathbf{n}}_j} \simeq \mathcal{J}_{\mathbf{n}\to\hat{\mathbf{n}}_j}\Phi_{\mathbf{n}},$$

where $\mathcal{J}_{\mathbf{n}\to\hat{\mathbf{n}}_j}$ is a sparse and linear jackknife operator (The code distributed with this article uses a uses a slightly more complex jackknife formulation described in Appendix Jackknife approximation for moment closure). We then obtain a closed version of the migration operator, $\hat{\mathcal{M}}_{jk} = \mathcal{M}_{jk}\mathcal{J}_{\mathbf{n}\to\hat{\mathbf{n}}_j}$, leading to a closed-form evolution equation for $\Phi_{\mathbf{n}}$:

$$\begin{aligned}\dot{\Phi}_{\mathbf{n}}(\mathbf{i}) = \sum_{j=1}^{p}\Bigg[&\frac{1}{4N_j}\tilde{\Delta}_{n_j,i_j}\Phi_{\mathbf{n}} + s_j\hat{\nabla}_{h_j,n_j,i_j}\Phi_{\mathbf{n}} \\ &+ \sum_{k\neq j} m_{jk}\hat{\mathcal{M}}_{jk}\Phi_{\mathbf{n}} + n_j u\delta_{\mathbf{i}=\mathbf{e}_j}\Bigg].\end{aligned} \tag{8}$$

So far, we supposed that migration was weak enough that the expected number of migrants per generation in our sample was less than 1. However, many populations have received many migrants over few generations. This is the case, for example, in many human populations in the Americas.

In such cases, we compute the resulting AFS directly without using the jackknife approximation. Because migration is a stochastic process, the number of migrant lineages varies across loci. A given sample of 10 ploids may derive 100% ancestry from population 1 at one locus, and 100% from population 2 at another. In the independent-sites model and two-way admixture, the distribution of lineages that trace back to one population is binomial: The expected frequency spectrum is an integral over the possible numbers of migrating lineages. To generate n admixed lineages, we can therefore need as many as n lineages from each of the source populations. We explain in Appendix Strong migration and admixture how the number of source population lineages can be chosen to maximize computational efficiency.

Implementation and performance

We developed a python library, called *Moments*, to simulate multidimensional AFS and infer demographic history using the method described above. The linear set of ordinary differential equations is integrated using an implicit Crank-Nicolson scheme which is well suited to diffusion problems. For three and more populations problems, we use alternating direction implicit methods to speed up computations (Baolin and Wenzhi 1994). From the user perspective, the library is very similar to $\partial a \partial i$, since we reused the $\partial a \partial i$ software architecture, including the user interface and data handling methods. We added

a number of new convenience features, described in Appendix Additional features, but the most important differences are performance and generality.

Whereas *∂a∂i* could model up to three populations and *Multipop* (Lukić and Hey 2012) up to four, *Moments* can handle models with up to 5 populations with selection, migrations and population splits. The run-time is still exponential in the number of populations, so the direct computation of large sample sizes for more than three populations remains challenging.

To provide a fair comparison between *Moments* and *∂a∂i* for a comparable problem size, we need to consider both accuracy and computation time. *∂a∂i* users can chose the number of grid points used for solving the diffusion equation. A large number of points results in slower but more accurate integration. We used the recommended strategy in *∂a∂i*, namely using three different grid sizes and performing Richardson extrapolation.

Moments also presents a tradeoff between speed and accuracy because the moment closure approximation improves with increasing n. *Moments* can therefore be made more accurate by computing the AFS for n' larger than the desired n, and subsampling to n after the integration. For the simulations below, we simply integrated with the desired n.

Table 1 shows comparisons between the two methods in several test cases. A more extensive set of test cases is provided in the Appendix Benchmarks.

In the case of neutral equilibrium with constant population size, we can compare our computations to the exact solution. For more complex cases, where analytical results are unavailable, we use *∂a∂i* with a very fine frequency grid and use the spectrum thus obtained as a reference to compare our method and *∂a∂i*. Specifically, the reference AFS is computed with third order Richardson extrapolations using fine grids — 30, 35 and 40 times the sample size for 1D and 2D cases and 5, 6 and 8 times the sample size for 3D simulations. Using this reference as the true AFS can induce a bias in favor of *∂a∂i*.

We consider several configurations with up to 3 dimensions including selection, migration and non constant populations sizes. We used 30 samples per population. The first 5 cases are simple integrations starting from a null spectrum: a single population without selection until the equilibrium is reached, two isolated populations under selection and three populations with selection and migrations. The sixth model is the demographic model of Out of Africa expansion described in (Gravel *et al.* 2011). Details of the models, and additional benchmarks, are provided in Appendix Benchmarks.

We report the execution time, the mean relative error ε_r compared to the 'true spectrum', and the difference in likelihood $\Delta LL = \log(L(truth, truth)) - \log(L(truth, test))$, where $L(x, y)$ is the probability to observe the AFS x assuming that the expected AFS is y. Here 'truth' is reference AFS and 'test' is the AFS computed using each software with regular settings. The likelihood is computed as a product of Poisson likelihoods, one for each frequency spectrum entry (Gutenkunst *et al.* 2009). Finally, we report the number of biologically impossible negative AFS entries.

In the given examples, *Moments* performs better than *∂a∂i* for most metrics. In the more general set of benchmarks provided in Appendix Benchmarks, we find a few instances where *∂a∂i* outperforms *moments*. Overall, we find that *Moments* is particularly efficient for integrations over long time periods, and rarely predicts negative frequencies.

For 4 and 5 populations models, we could not compare *Moments* to *∂a∂i*, since *∂a∂i* does not allow for such high-dimensional simulations. Similarly, comparison with *∂a∂i* was not possible when modelling reversible mutations in a finite genome. Figure 1 shows the predictions of *Moments* in two finite-genome models with a constant-size population, where analytical solutions are available for comparison.

Demographic model	method	exec time(s)	mean(ε_r)	Δ LL	# < 0
Neutral equilibrium 1D	*∂a∂i*	3.6e-02	1.3e-02	**1.3e-05**	0
	Moments	**1.5e-03**	**9.7e-03**	4.6e-05	0
Selection 2D	*∂a∂i*	1.7e-01	1.3e-02	2.9e-03	0
$T = 1.0$	*Moments*	**2.1e-02**	**9.1e-04**	**-1.4e-03**	0
Selection 2D	*∂a∂i*	6.2e-01	5.5e-03	9.3e-03	0
$T = 5.0$	*Moments*	**7.5e-02**	**9.2e-04**	**-6.5e-03**	0
Selection, migration 3D	*∂a∂i*	8.2e+01	4.2e-02	6.2e-04	0
$T = 1.0$	*Moments*	**5.0**	**1.5e-02**	**1.7e-04**	0
Selection, migration 3D	*∂a∂i*	4.0e+02	4.2e-02	3.4e-03	12
$T = 5.0$	*Moments*	**2.5e+01**	**1.4e-02**	**2.8e-04**	0
Out of Africa 3D	*∂a∂i*	5.7	1.5e+02	2.5e-04	0
	Moments	**4.1**	**1.6e-02**	**8.1e-05**	0

Table 1 Performance comparisons between *∂a∂i* and *Moments* on several scenarios (30 samples per population). The full bench is provided in Appendix . For *∂a∂i* simulations we used Richardson extrapolation to improve convergence. The time T provided is the simulation time in genetic units.

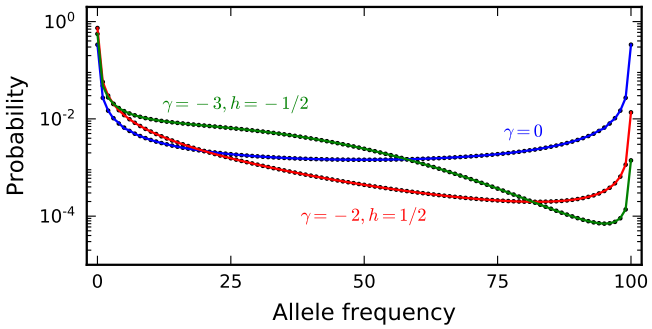


Figure 1 *Moments* predictions for finite-genome frequency spectrum for a population of constant size $N = 10,000$ with reversible mutation rates $u = v = 2 \times 10^{-6}$ (dots), compared to theoretical expectations obtained by integrating Eqs 5.70 and 5.72 in (Ewens 2004) (solid lines).

Application to data

3 populations Out of Africa

We used our method to update the Out of Africa expansion model described in (Gravel *et al.* 2011; Gutenkunst *et al.* 2009). These models are used in a wide variety of medical and evolutionary applications.

We used autosomal synonymous sequence data from the publicly available Thousand Genome Project (1000G) (1000 Genomes Project 2015; Sudmant *et al.* 2015). We first computed the joint AFS for three populations: Yoruba individuals in Ibadan, Nigeria (YRI); Utah residents with northern and western European ancestry (CEU); and Han Chinese from Beijin (CHB). We restricted our analysis to 80 ploids for each population and fit the same 13-parameter demographic model as in (Gutenkunst *et al.* 2009; Gravel *et al.* 2011). However, whereas previous studies had access to capture data for a subset of the exome, here we

had access to the entire high-coverage exome data. Moreover, we updated the mutation coefficient and the generation time to more realistic values: respectively $\mu = 1.44 \times 10^{-8}$ (Gravel *et al.* 2013) and $T_g = 29y$ (Tremblay and Vézina 2000). The best-fit model is represented in Figure 2.

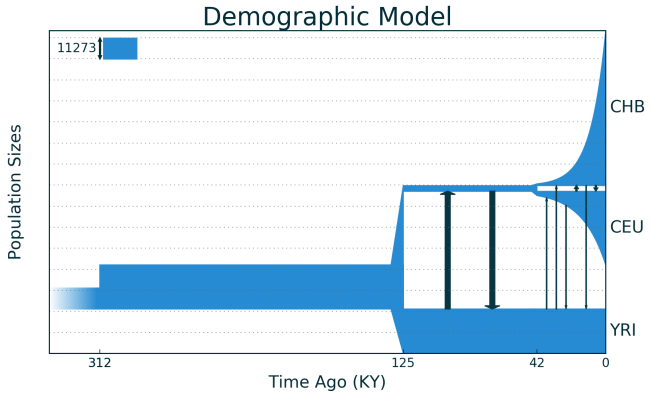


Figure 2 Out of Africa expansion.

We used a Powell optimizer to find maximum likelihood parameters (see Appendix Additional features). Confidence intervals were obtained using recently introduced approaches for uncertainty estimation for composite likelihood (Coffman *et al.* 2016). Maximum-likelihood parameters are presented in Table 2.

Parameter	Estimate	95% CI
N_A	11273	-
N_{AF}	23721	23033 - 24409
N_B	3104	2882 - 3326
N_{EU0}	2271	2091 - 2452
$r_{EU}(\%)$	0.196	0.189 - 0.203
N_{AS0}	924	864 - 984
$r_{AS}(\%)$	0.309	0.30 - 0.318
$m_{AF-B}(\times 10^{-5})$	15.8	14.7 - 16.8
$m_{AF-EU}(\times 10^{-5})$	1.10	0.99 - 1.22
$m_{AF-AS}(\times 10^{-5})$	0.48	0.42 - 0.55
$m_{EU-AS}(\times 10^{-5})$	4.19	3.86 - 4.51
$T_{AF}(kya)$	312	259 - 365
$T_B(kya)$	125	110 - 139
$T_{EU-AS}(kya)$	42.3	40.6 - 44.0

Table 2 Parameters estimates inferred with the likelihood approach. For the mutation rate and the generation time, we used respectively: $\mu = 1.44 \times 10^{-8}$ (Gravel *et al.* 2013) and $T_g = 29y$ (Tremblay and Vézina 2000). We used 80 samples per population, performed the inference in genetic units, and scaled the parameters so that N_A matches the point estimate from (Gravel *et al.* 2011). Confidence intervals are obtained by bootstrap over genic regions. Parameters are defined in Reference (Gravel *et al.* 2011)

Validation with different trios of populations

Because of data and computational limitations, most previous studies have only considered a single triplet of populations for estimating an Out-Of-Africa model. We investigated whether other populations would yield consistent results. Here we inferred the parameters with three different trios incorporating data from Luhya from Kenya (LWK), British (GBR) and Kinh Vietnamese (KHV) populations in addition to the three original populations we used. Inference results are presented in Table 3.

Parameter	YRI CHB GBR	YRI KHV GBR	LWK CHB CEU
N_A	11273	11273	11273
N_{AF}	24486	23908	29034
N_B	3034	2986	2746
N_{EU0}	2587	2485	2141
$r_{EU}(\%)$	0.17	0.17	0.21
N_{AS0}	958	951	859
$r_{AS}(\%)$	0.30	0.30	0.33
$m_{AF-B}(\times 10^{-5})$	15.6	16.0	17.7
$m_{AF-EU}(\times 10^{-5})$	1.00	1.02	1.50
$m_{AF-AS}(\times 10^{-5})$	0.48	0.60	0.63
$m_{EU-AS}(\times 10^{-5})$	3.99	4.46	4.43
$T_{AF}(kya)$	349	313	218
$T_B(kya)$	121	120	99
$T_{EU-AS}(kya)$	44	43	40

Table 3 Parameters estimates inferred for the Out of Africa with different trios of African, European and Asian populations from the 1000 Genomes dataset. For the mutation rate and the generation time, we used respectively: $\mu = 1.44 \times 10^{-8}$ (Gravel *et al.* 2013) and $T_g = 29y$ (Tremblay and Vézina 2000). We used 80 samples per population. We performed the inference in genetic units, and scaled the parameters so that N_A matches the point estimate from (Gravel *et al.* 2011).

The inference is robust to permutation of European and Asian populations, in the sense that inferred parameters across triplets are consistent with the confidence intervals obtained in the YRI-CEU-CHB triplet. However, changing the African population from YRI to LWK changes parameters beyond the confidence intervals. This is because the reported confidence intervals take into account the finite nature of the genome, but not the variation across populations. Parameters inferred from different triplets are not wildly different, reflecting the shared or similar histories of some of the populations, but quantitative details can vary substantially. Perhaps the most interesting difference is the more recent inferred split between between LWK and Eurasians, compared to the split between YRI and Eurasians, supporting either population structure within Africa prior to the Out-of-Africa population (with LWK ancestors more closely related to the Out-of-Africa population), or interaction between LWK and Eurasians after the OOA event. This naturally begs for more detailed modelling of the relationships between African popula-

tions in the 1000 Genome project. This would however require a more thorough discussion of the historical and archeological context, and is outside the scope of this manuscript.

Out of Africa model with 4 and 5 populations

As *moments* handles up to five populations, we can simulate more complex models than the Out of Africa model described in (Gutenkunst *et al.* 2009). In this Section, we consider a model of Out of Africa expansion with four populations including one additional Asian population, the Japanese from Tokyo (JPT). The model is identical to that of Section 3 populations Out of Africa with an additional population split in Asia leading to the Japanese population (see Figure 3 and Table 4).

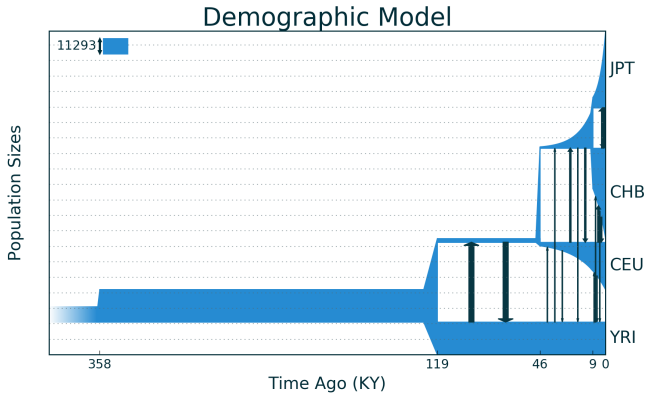


Figure 3 Out of Africa expansion model with four populations.

We also considered a model with five populations by adding a third Asian population, the Kinh Vietnamese (KHV). To reduce computational time and facilitate comparison with the 4-population model, we fixed the parameters inferred from the four population model and inferred the new parameters (see Table 4) introduced by the split which gives rise to the KHV population.

Parameter	Estimate
N_A	11293
N_{AF}	23721
N_B	2831
N_{EU0}	2512
$r_{EU}(\%)$	0.16
N_{AS0}	1019
$r_{AS}(\%)$	0.26
N_{KHV0}	2356
$r_{KHV}(\%)$	17.4
N_{JP0}	4384
$r_{JP}(\%)$	1.29
$m_{AF-B}(\times 10^{-5})$	16.8
$m_{AF-EU}(\times 10^{-5})$	1.14
$m_{AF-AS}(\times 10^{-5})$	0.56
$m_{EU-AS}(\times 10^{-5})$	4.75
$m_{CH-KHV}(\times 10^{-5})$	21.3
$m_{CH-JP}(\times 10^{-5})$	3.3
$T_{AF}(kya)$	357
$T_B(kya)$	119
$T_{EU-AS}(kya)$	46
$T_{CH-KHV}(kya)$	9.8
$T_{CH-JP}(kya)$	9

Table 4 Maximum likelihood parameters estimates for the 4 and 5 populations models. We used 30 samples per population. Parameters in black were inferred using the 4-population model. They were held fixed when inferring the remaining parameters in the 5 population model

Discussion

Connections with other models

Diffusion approximation A standard approach to simulate AFS evolution relies on a continuous approximation to the Wright-Fisher process. This approximation leads to an advection-diffusion equation describing the evolution of the population allele frequencies density $\phi(x,t)$ (*i.e.* the expected proportion of alternate ploids at frequency $x \in [0,1]$ in the population and at time t).

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} \simeq & \frac{1}{4N}\Delta x(1-x)\phi(x,t) \\ & - s\nabla\cdot\left(h+(1-2h)x\right)x(1-x)\phi(x,t) \\ & + 2Nu\delta(x-\frac{1}{2N}).\end{aligned} \tag{9}$$

The software $\partial a\partial i$ uses a finite difference scheme to solve this equation and approximate the distribution ϕ. The expected AFS entries are then computed via the integral:

$$\Phi_n(i) = \int_0^1 \binom{n}{i} x^i (1-x)^{n-i}\phi dx.$$

The AFS entries, which we had computed without reference to an underlying continuous model, can thus be seen as moments of the distribution ϕ computed in a non-canonical basis of polynomials. The idea of deriving evolution equations for the classical moments $\mu_k = \int_0^1 x^k\phi dx$ was proposed in (Evans *et al.* 2007). Unfortunately, obtaining the allele frequency distribution $\Phi_n(k)$ from the μ_k requires a sum with large fluctuating terms, making it impractical for large sample sizes. Evans et al. pointed out that writing an equation in terms of the $\Phi_n(k)$, as was done here, might be preferable, but apparently did not realize that these equations could be brought to a simple form that avoids alternating coefficients.

In fact, we show in Appendix One-dimensional case that by computing a moment equation for the diffusion equation, we obtain precisely Equation (5), which is not more complex than the original diffusion equation and is, as we have seen, more numerically stable. Furthermore, the lack of closure of the evolution equation with selection is easily addressed by a jackknife procedure, at least for the weak selective values consistent with the diffusion approximation.

Despite the alternating sum problem, Živković *et al.* (2015) did show that the Evans et al. equations for the μ_k could be used to compute the frequency spectrum, at least for a single population with piece-wise constant size and for up to 200 ploids. They also proposed a truncation approach to resolve the moment closure problem. This truncation approach would not be practical in the moment representation of Equation (5), as setting high-order terms to zero would amount to neglecting selection altogether.

In short, the Evans et al. equations and the ones presented here are algebraically equivalent. The main advantages of the present formulation is numerical stability, ease of generalization, and the availability of the moment closure approach.

Particle models The existence of closed evolution equations for the AFS in a sample of size n suggests that we can learn about the evolution of a large population by tracking only n lineages forward in time. Consider for example Equation (3), describing neutral evolution for a large Wright-Fisher population. Equation (3) uses the fact that n samples are either copied from n distinct parents, or from $n-1$ distinct parents with one parent contributing to two offspring. The evolution of the expected sample frequency spectrum can therefore be emulated as a Moran model for a population of size n with a suitably scaled replacement rate, an interpretation followed in Kamm *et al.* (2017). A more general theoretical discussion of the relationship between effective 'particle' models and continuous models can be found in Donnelly and Kurtz (Donnelly and Kurtz 1999).

Coalescent models Three related methods for estimating the neutral, multi-population allele frequency distribution have been presented recently (Kamm *et al.* 2017; De Maio *et al.* 2015; Bryant *et al.* 2012). These use a transition matrix similar to the one computed here for the single population, neutral, large-N limit. Rather than directly integrating a high-dimensional diffusion-like equation, these methods compute transition matrices for entire epochs during which populations are assumed to be isolated. They then combine these transition matrices using dynamic programming approaches, such as Felsenstein's tree-peeling algorithm. These approaches can handle many populations because they can compute a subset of the entries of the AFS, reducing the dimensional curse affecting *∂a∂i* and *moments*. However, they handle neither selection nor continuous migration. We see no reason why the jackknife-based transition matrices, which we introduced here to simulate the evolution of allele frequencies forward in time, could not also be used to introduce selection, multiple coalescence, and complex mutation models to the coalescent-based dynamic programming framework.

Conclusion

We described a highly accurate and numerically robust approach to simulate the evolution of allele frequency distributions over time in a discrete Wright-Fisher model. On a practical level, this approach can be used wherever the diffusion approximation is applicable, in which case it typically provides faster, more accurate, and more robust solutions. Our software implementation will provide an easy transition to *∂a∂i* users: data handling and ancillary methods for *moments* are largely copied from open-source code from *∂a∂i*, model specification is simplified, and the number of adjustable parameters is reduced.

The applications of our solver to data from the 1000 Genomes project largely recapitulate and refine worldwide models of genetic diversity. Within Asia, we found that the best model involved a split between Han Chinese, Japanese, and Kinh Vietnamese approximately 9000 years ago. Importantly, we found the choice of a 'representative' population in Out-of-Africa models can substantially affect the inferred parameters, even ones not directly involving the population. Unsurprisingly, the choice of the African population has the largest impact on the inferred demography, emphasizing that previous OOA models may be applicable to many Eurasian populations, but not to other African populations. Building models including more than one African population will likely provide much more information about Human ancestry both in Africa and across the world.

Acknowledgements

We thank Ryan Gutenkunst and Yun Song for useful discussions. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program, CIHR award MOP 136855, and NSERC Discovery grant (to SG), and Google Summer of Code (to WL).

Literature Cited

1000 Genomes Project, 2015 A global reference for human genetic variation. Nature **526**: 68–74.

Baolin, Z. and L. Wenzhi, 1994 On alternating segment Crank-Nicolson scheme. Parallel Computing **20**: 897–902.

Bhaskar, A., A. G. Clark, and Y. S. Song, 2014 Distortion of genealogical properties when the sample is very large. Proc Natl Acad Sci USA **111**: 2385–2390.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury, 2012 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular biology and evolution **29**: 1917–1932.

Coffman, A. J., P. H. Hsieh, S. Gravel, and R. N. Gutenkunst, 2016 Computationally efficient composite likelihood statistics for demographic inference. Molecular Biology and Evolution **33**: 591–593.

Crow, J. F. and M. Kimura, 1970 *An introduction to population genetics theory*. Harper and Row.

De Maio, N., D. Schrempf, and C. Kosiol, 2015 PoMo: an allele frequency-based approach for species tree estimation. Syst. Biol. **64**: 1018–1031.

Donnelly, P. and T. G. Kurtz, 1999 Genealogical processes for fleming-viot models with selection and recombination. Annals of Applied Probability **9**: 1091–1148.

Evans, S. N., Y. Shvets, and M. Slatkin, 2007 Non-equilibrium theory of the allele frequency spectrum. Theoretical population biology **71**: 109–119.

Ewens, W. J., 2004 *Mathematical Population Genetics*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY.

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

Excoffier, L. and M. Foll, 2011 Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics **27**: 1332–1334.

Fisher, R., 1930 The distribution of gene ratios for rare mutations. Proc. R. Soc. Edinb. **50**: 205–220.

Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, *et al.*, 2011 Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences **108**: 11983–11988.

Gravel, S. and NHLBI GO ESP, 2014 Predicting discovery rates of genomic features. Genetics **197**: 601–610.
Gravel, S., F. Zakharia, A. Moreno-Estrada, J. K. Byrnes, M. Muzzio, *et al.*, 2013 Reconstructing native american migrations from whole-genome and whole-exome data. PLoS Genet **9**: e1004023.
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 Genet **5**: e1000695.
Haller, B. C. and P. W. Messer, 2016 Slim 2: Flexible, interactive forward genetic simulations. Molecular Biology and Evolution **34**: 230–240.
Kamm, J. A., J. Terhorst, and Y. S. Song, 2017 Efficient computation of the joint sample frequency spectra for multiple populations. Journal of Computational and Graphical Statistics **26**: 182–194.
Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput Biol **12**: e1004842.
Kimura, M., 1964 Diffusion models in population genetics. Journal of Applied Probability **1**: 177–232.
Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61**: 893.
Kuehn, C., 2016 Moment closure - a brief review. In *Control of Self-Organizing Nonlinear Systems*, pp. 253–271, Springer.
Libby, W. F., E. C. Anderson, and J. R. Arnold, 1949 Age determination by radiocarbon content: world-wide assay of natural radiocarbon. Science **109**: 227–228.
Lukić, S. and J. Hey, 2012 Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-africa expansion. Genetics **192**: 619–639.
Lukić, S., J. Hey, and K. Chen, 2011 Non-equilibrium allele frequency spectra via spectral methods. Theoretical population biology **79**: 203–219.
Mendelson, A. F., M. A. Zuluaga, B. F. Hutton, and S. Ourselin, 2016 What is the distribution of the number of unique original items in a bootstrap sample? ArXiv e-prints .
Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, *et al.*, 2012 Ancient admixture in human history. Genetics **192**: 1065–1093.
Scheinfeldt, L. B. and S. A. Tishkoff, 2013 Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet **14**: 692–702.
Schiffels, S. and R. Durbin, 2014 Inferring human population size and separation history from multiple genome sequences. Nat. Genet. **46**: 919–925.
Schmutz, J., P. E. McClean, S. Mamidi, G. A. Wu, S. B. Cannon, *et al.*, 2014 A reference genome for common bean and genome-wide analysis of dual domestications. Nature genetics **46**: 707–713.
Spence, J. P., J. A. Kamm, and Y. S. Song, 2016 The site frequency spectrum for general coalescents. Genetics **202**: 1549–1561.
Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, *et al.*, 2015 An integrated map of structural variation in 2,504 human genomes. Nature **526**: 75–81.
Tremblay, M. and H. Vézina, 2000 New estimates of intergenerational time intervals for the calculation of age and origins of mutations. The American Journal of Human Genetics **66**: 651–658.
Valladas, H., J. Clottes, J.-M. Geneste, M. A. Garcia, M. Arnold, *et al.*, 2001 Palaeolithic paintings: evolution of prehistoric cave art. Nature **413**: 479–479.
Walter, R. C., 1994 Age of Lucy and the First Family: single-crystal 40Ar/39Ar dating of the Denen Dora and lower Kada Hadar members of the Hadar Formation, Ethiopia. Geology **22**: 6–10.
Wright, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.
Živković, D., M. Steinrücken, Y. S. Song, and W. Stephan, 2015 Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. Genetics pp. genetics–115.

Coefficients of the discrete operators

In this section, we first present the full form of the discrete drift, selection, and migration operators involved in equation (7) in the simplest model where selection, migration, and drift are treated linearly, and then present an outline of the derivation.

Remember that the ODE for an AFS in p populations, with N_j diploid individuals in population j and n_j sequenced ploids per population, is

$$\dot{\Phi}_{\mathbf{n}}(\mathbf{i}) = \sum_{j=1}^{p} \Big[\frac{1}{4N_j} \tilde{\Delta}_{n_j,i_j} \Phi_{\mathbf{n}} + s_j \tilde{\nabla}_{h_j,n_j,i_j} \Phi_{\mathbf{n}+2\mathbf{e_j}} + \sum_{k \neq j} m_{jk} \mathcal{M}_{jk} \Phi_{\hat{\mathbf{n}}_\mathbf{j}} + n_j u \delta_{\mathbf{i}=\mathbf{e_j}} \Big], \tag{10}$$

where $\hat{\mathbf{n}}_j = \mathbf{n} + \mathbf{e}_j$ is the parental sample size required to reach sample size $\mathbf{n}$ after migration from j to k, and $\mathbf{e}_j$ is the j^{th} vector of the canonical basis of $\mathbb{R}^p$, $\mathbf{e}_j(i) = \delta_{ij}$.

The drift operator $\tilde{\Delta}_{n_j,i_j}$ has the form

$$\tilde{\Delta}_{n_j,i_j}\Phi_{\mathbf{n}}(\mathbf{i}) = \Big[(i_j-1)(n_j-i_j+1)\Phi_{\mathbf{n}}(\mathbf{i}-\mathbf{e_j}) - 2i_j(n_j-i_j)\Phi_{\mathbf{n}}(\mathbf{i}) + (n_j-i_j-1)(i_j+1)\Phi_{\mathbf{n}}(\mathbf{i}+\mathbf{e_j}) \Big], \tag{11}$$

and is closely related to transition matrices derived in (Evans *et al.* 2007; Spence *et al.* 2016). The selection operator $\tilde{\nabla}_{h_j,n_j,i_j}$ is defined by

$$\begin{aligned}\tilde{\nabla}_{h_j,n_j,i_j}\Phi_{\mathbf{n}+2\mathbf{e_i}}(\mathbf{i}) =& \frac{h_j}{n_j+1}\Big[i_j(n_j+1-i_j)\Phi_{\mathbf{n}+\mathbf{e_i}}(\mathbf{i}) \\ &- (n_j-i_j)(i_j+1)\Phi_{\mathbf{n}+\mathbf{e_i}}(\mathbf{i}+\mathbf{e_j})\Big] \\ &+ \frac{(1-2h_j)(i_j+1)}{(n_j+1)(n_j+2)} \\ &\times \Big[i_j(n_j+1-i_j)\Phi_{\mathbf{n}+2\mathbf{e_i}}(\mathbf{i}+\mathbf{e_j}) \\ &- (n_j-i_j)(i_j+2)\Phi_{\mathbf{n}+2\mathbf{e_i}}(\mathbf{i}+2\mathbf{e_j})\Big]\end{aligned} \tag{12}$$

and the migration operator $\mathcal{M}_{jk}$ is

$$\begin{aligned}\mathcal{M}_{jk}\Phi_{\hat{\mathbf{n}}_\mathbf{j}} =& \frac{1}{n_j+1}\Big((i_j+1)(n_k-i_k+1)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(\mathbf{i}+\mathbf{e_j}-\mathbf{e_k}) \\ &- (i_j+1)(n_k-i_k)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(\mathbf{i}+\mathbf{e_j})\Big) \\ &+ (n_j-i_j)(i_k+1)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(\mathbf{i}+\mathbf{e_j}) - i_k(n_j-i_j)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(\mathbf{i}).\end{aligned} \tag{13}$$

Derivation for the drift operator

For simplicity, we derive the drift and selection operators in one dimension, since the multi-dimension case is a straightforward generalization with a heavier notation burden. The probability of observing a single two-way coalescence in a sample of size n is

$$P_{n,N}(1 \to 2) = \frac{1}{2N}\binom{n}{2}\left(1 - \frac{1}{2N}\right)^{n-2}(1 - P_{n-2,N-1}), \quad (14)$$

where $\frac{1}{2N}$ is the probability that a given pair of lineages coalesces, $\binom{n}{2}$ is the number of distinct pairs of lineages that can coalesce, $(1-1/2N)^{n-2}$ is the probability that there is no additional coalescence to the same ploid, and $1 - P_{n-2,N-1}$ is the probability that none of the remaining $n-2$ lineage coalesce to the remaining $N-1$ ancestors. If we only keep leading order terms in $\frac{1}{N}$, we find the classical large-population limit

$$P_{n,N}(1 \to 2) = \frac{1}{2N}\binom{n}{2}.$$

Given a single two-way coalescence, we then want to compute the frequency distribution in the offspring given the distribution in the parental generation. We imagine that the n descendent ploids are copied from n parental ploids, with one parental ploid drawn twice, and one parental ploid never drawn. Since the n parental ploids were drawn independently, their expected allele frequency distribution corresponds to the expected frequency in the parental generation. To compute the change in expected allele frequency between parental and offspring generation, we must compute the probability of allele frequency changes between parents and offsprings.

The probability of observing a transition from parental frequency $i+1$ to descendent frequency i after a two-way coalescence is the probability that the twice-selected ploid carries the reference genotype (which we represent as ), and the unselected ploid is non-reference (which we represent as •):

$$P^{1\to 2}(i+1, i) = \frac{(i+1)(n-i-1)}{n(n-1)}\delta_{0\le i\le n-1}.$$

Similarly, the gain of an alternate ploid results from event  and has probability

$$P^{1\to 2}(i-1, i) = \frac{(i-1)(n-i+1)}{n(n-1)}\delta_{1\le i\le n}.$$

The probability of transitioning from i alternate ploids to $i-1$ (via ) or $i+1$ (via) is

$$P^{1\to 2}(i, *) \equiv P^{1\to 2}(i, i-1) + P^{1\to 2}(i, i+1) = 2\frac{i(n-i)}{n(n-1)}.$$

We can finally compute the drift operator for a single two-way coalescence:

$$\begin{aligned}\frac{1}{4N}\tilde{\Delta}^{1\to 2}_{n,i}\Phi_n(i) =& P_{n,N}(1 \to 2) \\ &\times \Big(\Phi_n(i-1)P^{1\to 2}(i-1, i) \\ &+ \Phi_n(i+1)P^{1\to 2}(i+1, i) \\ &- \Phi_n(i)P^{1\to 2}(i, *)\Big),\end{aligned} \quad (15)$$

which simplifies to Equation (11).

Derivation of the selection operator

Here we derive a one-dimensional version of Equation (12). The derivation is elementary but a bit cumbersome. See Appendix ODEs on the moments for a different derivation that relies on the diffusion approximation.

We consider the action of selection on a single transmission, when a new ploid is drawn from a diploid individual. We derive the results in the context of negative selection ($s < 0$ and $sh < 0$) where we can think of selection as eliminating a proportion of the neutral transmissions. A transmission is eliminated by selection with probability $-sh$ if the parent is a heterozygote, and with probability $-s$ if the parent is an alternate homozygote. For bookkeeping, we write the state of the parent as  , with the empty circle representing a reference allele, solid circle representing the alternate allele, and the vertical line representing the putative initial selected ploid. When selection acts, this ploid is replaced by a new ploid, taken at random from the parental population. Since we assume that at most one selective event occurs per generation, we do not need to know the diploid state of this replacement ploid. In the selective event labeled as  , the transmission of an alternate allele from a homozygous ancestor was eliminated and replaced by transmission from a reference allele. To generate n descendent ploids with one selective event, there have been a total of $n+2$ relevant ploids: the n selected initially, the diploid companion of the rejected ploid and the replacement ploid. We want to compute the change in allele frequency caused by the selection process relative to the neutral transmission of the n selected ploids. We write $S_{\text{  }}(i+1 \to i)$ for the probability that the n initial parental ploids included $i+1$ alternate alleles, but the offspring ploids include i because of selective process  . This is proportional to the number $\Phi_{n+2}(i+2)$ of loci having the required $i+2$ alternate alleles in a sample of size $n+2$, the probability of selecting the correct triplet of ploids (namely the hypergeometric $\frac{\binom{i+2}{2}\binom{n-i}{1}}{\binom{n+2}{3}}$ for the correct choice of three ploids times $\frac{1}{3}$ for the correct order), the probability that the selection event occurs for one transmission ($-s$), and the number of transmissions where selection can act (n). Putting all these together, we get

$$\begin{aligned}
S_{\text{  }}(i+1 \to i) &= -\frac{ns}{3}\left(\Phi_{n+2}(i+2)\frac{\binom{i+2}{2}\binom{n-i}{1}}{\binom{n+2}{3}}\right), \\
S_{\text{  }}(i+1 \to i) &= -\frac{nsh}{3}\left(\Phi_{n+2}(i+1)\frac{\binom{i+1}{1}\binom{n-i+1}{2}}{\binom{n+2}{3}}\right), \\
S_{\text{  }}(i-1 \to i) &= -\frac{nsh}{3}\left(\Phi_{n+2}(i+1)\frac{\binom{i+1}{2}\binom{n-i+1}{1}}{\binom{n+2}{3}}\right), \\
S_{\text{}}(i \to i-1) &= -\frac{ns}{3}\left(\Phi_{n+2}(i+1)\frac{\binom{i+1}{2}\binom{n-i+1}{1}}{\binom{n+2}{3}}\right), \\
S_{\text{  }}(i \to i-1) &= -\frac{nsh}{3}\left(\Phi_{n+2}(i)\frac{\binom{i}{1}\binom{n-i+2}{2}}{\binom{n+2}{3}}\right), \\
S_{\text{}}(i \to i+1) &= -\frac{nsh}{3}\left(\Phi_{n+2}(i+2)\frac{\binom{i+2}{2}\binom{n-i}{1}}{\binom{n+2}{3}}\right).
\end{aligned} \quad (16)$$

The first three terms, of the form $S_{\cdot}(\cdot \to i)$, describe increases to the number of alleles at frequency i. The last three terms,

of the form $S_{\cdot}(i \to \cdot)$, contribute to a decrease relative to the neutral case: Their contribution to the evolution equation will have opposite signs. Other combinations of ploids either do not change the frequency spectrum (e.g., ), or have zero probability of happening in our selective model ().

Under an additive model, the diploid companion to the selected allele plays no role and the evolution should only depend on Φ_{n+1} rather than Φ_{n+2}. To make this explicit, we collect terms into an additive term, proportional to h, and a dominance term, proportional to $(1-2h)$. The dominance term thus vanishes under genic selection ($h = \frac{1}{2}$). Its coefficient can be obtained by setting $h = 0$ in the list of contributions from Equation (16). The dominance term reads

$$(1-2h)\left(S_{\text{  }}(i+1 \to i) - S_{\text{  }}(i \to i-1)\right).$$

The additive term contains the remaining terms from (16). We use the fact that $hS_{\text{  }}(i+1 \to i) = S_{\text{  }}(i \to i+1)$ and $S_{\text{  }}(i-1 \to i) = hS_{\text{  }}(i \to i-1)$, to simplify the additive term:

$$\begin{aligned}
&S_{\text{  }}(i+1 \to i) + S_{\text{  }}(i-1 \to i) - S_{\text{  }}(i \to i-1) \\
&\quad - S_{\text{  }}(i \to i+1) \\
&\quad + 2h\left(S_{\text{  }}(i+1 \to i) - S_{\text{  }}(i \to i-1)\right) \\
&= S_{\text{  }}(i+1 \to i) - S_{\text{  }}(i-1 \to i) \\
&\quad - S_{\text{  }}(i \to i-1) + S_{\text{  }}(i \to i+1).
\end{aligned} \tag{17}$$

We combine the first and last term using the downsampling formula

$$\Phi_{n+1}(i+1) = \frac{i+2}{n+2}\Phi_{n+2}(i+2) + \frac{n-i+1}{n+2}\Phi_{n+2}(i+1)$$

and similarly combine the middle two terms using

$$\Phi_{n+1}(i) = \frac{i+1}{n+2}\Phi_{n+2}(i+1) + \frac{n-i+2}{n+2}\Phi_{n+2}(i),$$

to reach Equation (12).

Derivation of the migration operator

We derive the migration operator for populations j and k, with m_{jk} the probability that a ploid in population k has a parent from population j. We consider a sample of n_j ploids from population 1 and n_k from population k and suppose that the migration rate is small enough that at most one migration occurs per site per generation. Here we consider parental sample sizes $\hat{\mathbf{n}} = (n_j+1, n_k)$, to account for the fact that a lineage from j may be needed to produce the n_k offspring.

In this case there are only two ploids involved in the migration, the replaced and the replacement ploids, and two possible configurations that result in changes in the allele frequency,  and  , where circles represent ploids: a dark circle represents an alternate allele, and the ploid on the left is the migrant replacing the ploid to the right. If $M_{\text{}}((i_j, i_k-1) \to (i_j, i_k))$ is the rate of increase of the number of loci with alternate allele counts (i_j, i_k) because of process , we have two rates of increase:

$$\begin{aligned}
M_{\text{}}((i_j+1, i_k-1) \to (i_j, i_k)) &= m_{jk}n_k\Big(\frac{i_j+1}{n_j+1} \\
&\quad \times \frac{n_k - i_k + 1}{n_k}\Phi_{\hat{\mathbf{n}}}(i_j+1, i_k-1)\Big) \\
M_{\text{}}((i_j, i_k+1) \to (i_j, i_k)) &= m_{jk}n_k\Big(\frac{n_j+1-i_j}{n_j+1} \\
&\quad \times \frac{i_k+1}{n_k}\Phi_{\hat{\mathbf{n}}}(i_j, i_k+1)\Big),
\end{aligned} \tag{18}$$

and two corresponding rates of decrease:

$$\begin{aligned}
M_{\text{}}((i_j+1, i_k) \to (i_j, i_k+1)) &= m_{jk}n_k\Big(\frac{i_j+1}{n_j+1}\frac{n_k-i_k}{n_k}\Phi_{\hat{\mathbf{n}}}(i_j+1, i_k)\Big), \\
M_{\text{}}((i_j, i_k) \to (i_j, i_k-1)) &= m_{jk}n_k\Big(\frac{n_j+1-i_j}{n_j+1}\frac{i_k}{n_k}\Phi_{\hat{\mathbf{n}}}(i_j, i_k)\Big).
\end{aligned} \tag{19}$$

(The indices (i_j+1, i_k) in $M_{\text{}}((i_j+1, i_k) \to (i_j, i_k+1))$ represent the allele frequency in parental sample of size $\hat{\mathbf{n}}$. The allele frequency would have been (i_j, i_k) after discarding the extra lineage if there had been no migration.)

We can combine these to get the transition rate due to migration:

$$\begin{aligned}
\mathcal{M}_{jk}\Phi_{\hat{\mathbf{n}}_\mathbf{j}} = \frac{1}{n_j+1}\Big(&(i_j+1)(n_k-i_k+1)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(i_j+1, i_k-1) \\
&- (i_j+1)(n_k-i_k)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(i_j+1, i_k) \\
&+ (n_j-i_j+1)(i_k+1)\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(i_j, i_k+1) \\
&- (n_j-i_j+1)i_k\Phi_{\hat{\mathbf{n}}_\mathbf{j}}(i_j, i_k)\Big).
\end{aligned} \tag{20}$$

The code in *moments* uses a slightly different equation that can be derived from this by the application of the downsampling formulas, namely

$$\begin{aligned}
\mathcal{M}_{jk}\Phi_{\hat{\mathbf{n}}_\mathbf{j}} =&\frac{n_k(i_j+1)}{n_j+1}\left(\Phi_{\mathbf{n}+\mathbf{e}_\mathbf{j}-\mathbf{e}_\mathbf{k}}(\mathbf{i}+\mathbf{e}_\mathbf{j}-\mathbf{e}_\mathbf{k}) - \Phi_{\mathbf{n}+\mathbf{e}_\mathbf{j}-\mathbf{e}_\mathbf{k}}(\mathbf{i}+\mathbf{e}_\mathbf{j})\right) \\
&- i_k\Phi_\mathbf{n}(\mathbf{i}) + (i_k+1)\Phi_\mathbf{n}(\mathbf{i}+\mathbf{e}_\mathbf{k}),
\end{aligned} \tag{21}$$

This makes it possible to only use the jackknife on the first two terms.

Generalized forms of the operators

Finite genome model for mutations

To facilitate comparison with previous work, we focused in this work on the infinite-sites model, assuming that mutations occur at previously invariant loci and neglecting back mutations. However, finite genome and back mutations are easily accommodated by the present model. In this case, the single population mutation term $\mathcal{U}_i\Phi_n$ is:

$$\begin{aligned}
\mathcal{U}_i\Phi_n =&[\mu(n-i+1)\Phi_n(i-1) - \nu i\Phi_n(i)]\delta_{i>0} \\
&+ [\nu(i+1)\Phi_n(i+1) - \mu(n-i)\Phi_n(i)]\delta_{i<n},
\end{aligned}$$

where μ and ν are the forward and backward mutation rates per base pair.

Drift with multiple coalescences

The standard diffusion approximation and Kingman's coalescent model neglect the possibility that multiple coalescences may occur during the same generation in the history of a sample. We used a similar approximation to compute the drift operator in Appendix Coefficients of the discrete operators. Even though multiple coalescences are indeed rare for small sample sizes and large populations, they can have a measurable effect for large sample sizes (Kamm *et al.* 2017).

Here we consider the next-order correction to the drift operator by accounting for three-way and double two-way coalescences (Figure 4), each contributing corrections of order $\frac{1}{N^2}$.

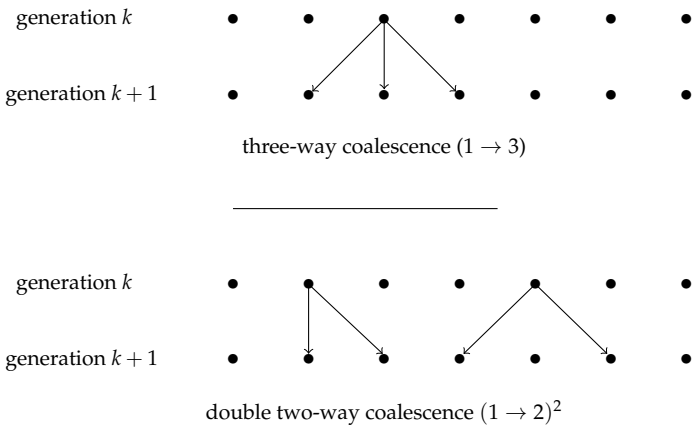


Figure 4 Multiple coalescences.

We can decompose the drift operator $\Delta_{n,i}$ in contributions from two-way, three-way, and double two-way coalescence

$$\tilde{\Delta}_{n,i} = \tilde{\Delta}^{1\to 2}_{n,i} + \tilde{\Delta}^{1\to 3}_{n,i} + \tilde{\Delta}^{(1\to 2)^2}_{n,i}.$$

The computation of individual terms, outlined below, is cumbersome but elementary. The key result is that the rate of change in $\Phi_n(i)$ is a linear combination of the $\Phi_n(j)$ with j ranging from $i-2$ to $i+2$.

Single two-way coalescent The two-way coalescent contribution is also described by Equation (15), but if we consider multiple coalescences we must also account for corrections to $P_{n,N}(1 \to 2)$ of order $\frac{1}{N^2}$, i.e.,

$$P_{n,N}(1\to 2) = \frac{1}{2N}\binom{n}{2}\left(1 - \frac{n-2}{2N} - \frac{1}{2N}\binom{n-2}{2}\right) + O\left(\frac{1}{N^3}\right).$$

Single three-way coalescent Similarly, the probability of a three-way coalescent is, to leading order in N, is

$$P_n(1\to 3) = \frac{\binom{n}{3}}{4N^2} + O\left(\frac{1}{N^3}\right),$$

and we must now compute 5 transition probabilities.

For example the transition probability from $i+2$ to i is,

$$P^{1\to 3}(i+2,i) = \frac{1}{3}H_{n,i+2,3}(2) = \frac{1}{3}\frac{\binom{n-(i+2)}{1}\binom{i+2}{2}}{\binom{n}{3}}.$$

Here $H_{n,k,j}(i) = \frac{\binom{i}{i}\binom{n-j}{j-i}}{\binom{n}{j}}$ is the hypergeometric distribution with j trials and i successes, sampling from a finite population of size n with k successes. To derive this result, we note that a triple coalescence can create a transition from $i+2$ to i if it involves three parental ploids: one (reference) chosen three times, and two (alternate) never chosen:  . The probability of this happening is the product of the probability of drawing two alternate ploids in a sample of three (i.e., $H_{n,i+2,3}(2)$), times the probability that we picked the reference allele to coalesce (i.e., $1/3$).

Similarly, we have contributions from  , , and

$$P^{1\to 3}(i+1,i) = \frac{2}{3}\frac{\binom{n-(i+1)}{1}\binom{i+1}{2}}{\binom{n}{3}},$$

$$P^{1\to 3}(i-1,i) = \frac{2}{3}\frac{\binom{n-(i-1)}{1}\binom{i-1}{2}}{\binom{n}{3}},$$

$$P^{1\to 3}(i-2,i) = \frac{1}{3}\frac{\binom{n-(i-2)}{1}\binom{i-2}{2}}{\binom{n}{3}}.$$

Finally we want the probability of changing from ancestral frequency i to any other offspring frequency. Because the only way to maintain the number of reference ploids under a triple coalescence is to pick either three reference ploids () or three non-reference ploids (), we have

$$\begin{aligned} P^{1\to 3}(i,*) \equiv& P^{1\to 3}(i,i-2) + P^{1\to 3}(i,i-1) \\ &+ P^{1\to 3}(i,i+1) + P^{1\to 3}(i,i+2) \\ =& 1 - H_{n,i+2,3}(0) - H_{n,i+2,3}(3) \\ =& 1 - \frac{\binom{i}{3} + \binom{n-i}{3}}{\binom{n}{3}}. \end{aligned} \tag{22}$$

We finally have the corresponding drift operator

$$\begin{aligned} \frac{1}{4N}\tilde{\Delta}^{1\to 3}_{n,i}\Phi_n(i) =& P_n(1\to 3)\sum_{j\in\mathcal{I}} P^{1\to 3}(j,i)\Phi_n(j) \\ &- P^{1\to 3}(i,*)\Phi_n(i), \end{aligned} \tag{23}$$

with $\mathcal{I} = \{i-2, i-1, i+1, i+2\}$.

Double two-way coalescent The probability of drawing a double coalescence is

$$P_n((1\to 2)^2) = \frac{3}{4N^2}\binom{n}{4} + O\left(\frac{1}{N^3}\right).$$

Here we need to consider four parental ploids with two coalescences and two not selected. We have a term corresponding to 

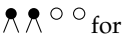
$$P^{(1\to 2)^2}(i+2,i) = \frac{H_{n,i+2,4}(2)}{\binom{4}{2}}.$$

Similarly, and contribute to

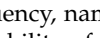
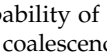
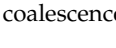
$$P^{(1\to 2)^2}(i+1,i) = \frac{H_{n,i+1,4}(1)}{2} + \frac{H_{n,i+1,4}(3)}{2}.$$

We then have the complementary contributions  and to

$$P^{(1\to 2)^2}(i-1,i) = \frac{H_{n,i-1,4}(1)}{2} + \frac{H_{n,i-1,4}(3)}{2}$$

and  for

$$P^{(1\to 2)^2}(i-2,i) = \frac{H_{n,i+2,4}(2)}{\binom{4}{2}}.$$

There are now three possibilities for not changing the allele frequency, namely  ;  ; and  , so that the probability of changing from frequency i given a double two-way coalescence is

$$P^{(1\to 2)^2}(i,*) = 1 - H_{n,i,4}(0) - \frac{2H_{n,i,4}(2)}{3} - H_{n,i,4}(4).$$

The drift operator is therefore

$$\frac{1}{4N}\tilde{\Delta}^{(1\to 2)^2}_{n,i}\Phi_n(i) = P_n((1\to 2)^2)\Big(- P^{(1\to 2)^2}(i,*)\Phi_n(i) + \sum_{j\in\mathcal{I}} P^{(1\to 2)^2}(j,i)\Phi_n(j)\Big). \tag{24}$$

Changes to $\Phi_n(i)$ caused by drift are a simple linear function of the $\Phi_n(j)$ for $j \in i-2,...,i+2$. This operator is sparse for large n and easy to compute numerically. Higher-order corrections in $\frac{1}{N}$ would simply involve more terms and a progressively denser linear operator.

Strong migration and admixture

To compute the frequency spectrum $\phi^{k+1}_{n_1,n_2}$ after admixture, we need to consider the origin of n_1+n_2 lineages. Under strong recent bidirectional migration, there are loci at which all n_1+n_2 lineages come from population 1, and other loci at which all lineages come from population 2. Thus $\phi^{k+1}_{n_1,n_2}$ depends on the set $\{\phi^k_{\tilde{\mathbf{n}}}\}$ for all $\tilde{\mathbf{n}}$ containing n_1+n_2 lineages. This information is contained in the frequency spectrum $\phi^k_{n_1+n_2,n_1+n_2}$.

We can therefore write

$$\phi^{k+1}_{n_1,n_2}(i,j) = \sum_{IJ} P(I,J\to i,j)\phi^k_{n_1+n_2,n_1+n_2}(I,J),$$

where $P(I,J\to i,j)$ is the probability of observing alternate allele counts (i,j) when each allele is drawn independently and without replacement from the two ancestral samples with frequency (I,J). The direct computation of $P(I,J\to i,j)$, summing over the possible inheritance patterns, is straightforward but computationally demanding for large sample sizes or high-dimensional applications. Here we describe two alternatives implemented in *moments*.

Exact dynamic programming

One alternative is to use dynamic programming: We can add an additional population to the frequency spectrum, and define a function that migrates a single lineage from populations 1 or 2 into population 3. For example, if we want to implement migration from population 1 into population 2 with migration rate m_{12}, we'd need to start with a frequency spectrum of dimension (n_1+n_2,n_2). After creating an additional dimension, the spectrum has dimension $(n_1+n_2,n_2,0)$. Then we'd iteratively extrude migrant lineages, creating spectra of dimension $(n_1+n_2-1,n_2-1,1)$, then $(n_1+n_2-2,n_2-2,2)$, until we reach $(n_1,0,n_2)$, at which point we can simply discard the empty population 2.

The downside of this approach is that it requires the creation of a new population, and $o(n^4)$ operations.

Approximate dynamic programming

To further speed this up, we used a dynamic programming approximation that speeds up computation to $o(n^3)$, and further reduces the number of lineages that must be tracked prior to the admixture event.

The general idea is to first perform migrations in place one ploid at a time, allowing later migrants to replace earlier ones. As we will see, the end result is not the desired frequency spectrum under the conventional admixture model which does not allow for replacement among migrant lineages. However, we will show how we can compute and correct for this difference.

To simulate one-way admixture from population 1 into population 2 at rate m_{12}, we therefore first allow one lineage at a time from population 1 to migrate into population 2. If $M_1\phi_{n_1,n_2}$ is the frequency spectrum resulting from migrating one lineage from population 1 to population 2, starting from ϕ_{n_1,n_2} and changing the sample size from (n_1,n_2) to (n_1-1,n_2+1), the (i,j)th element of the resulting frequency spectrum is

$$(M_1\phi_{n_1,n_2})(i,j) = \frac{i+1}{n_1}\phi_{n_1,n_2}(i+1,j-1) + \frac{n_1-i}{n_1}\phi_{n_1,n_2}(i,j). \tag{25}$$

Starting from a frequency spectrum $\phi_{n_1+\nu,n_2}$, we sequentially compute the spectra $M_1^R\phi_{n_1+\nu,n_2}$, resulting from $R=0,1,2,\ldots,\nu$ migrations. Using projection operator P_{n_1,n_2} to a sample size (n_1,n_2), we obtain frequency spectra $\psi_R = P_{n_1,n_2}M_1^R\phi_{n_1+\nu,n_2}$. The parameter ν is chosen so that all plausible migrant numbers are covered, as explained below. The ψ_R do not correspond to the desired frequency distribution after bulk admixture, because the number of migrant alleles per site does not have the desired binomial distribution. Our goal is then to choose a linear combination $\sum_R w_R\psi_R$ such that

$$\sum_R w_R\psi_R \simeq \phi^{k+1}_{n_1,n_2}. \tag{26}$$

In the standard infinite-sites model, the number of replaced lineages in the admixed sample follows the binomial distribution $B(n,m_{12})$. We can therefore write the frequency spectrum $\phi^{k+1}_{n_1,n_2}$ as the sum $\phi^{k+1}_{n_1,n_2} = \sum_{j=1}^n B(n_2,m_{12})_j\xi_j$, where ξ_j is the frequency spectrum that would result if exactly j lineages were replaced through migration.

To identify the set of w_R that satisfy Equation (26), we also express ψ_R as a linear combination of the ξ_j. The probability Γ_{jR} of observing j net replacements after R sequential replacements is easily computed numerically. This leaves us with $\psi_R = \sum_j \Gamma_{jR}\xi_j$. If we express both sides of Equation (26) in terms of the ξ_j, we find

$$\sum_j\sum_R \Gamma_{jR}w_R\xi_j \simeq \sum_{j=1}^n B(n_2,m_{12})_j\xi_j.$$

We would then like to choose the w_R such that the coefficients are equal for each j:

$$\sum_{R=1}^{\nu}\Gamma_{jR}w_R = B(n_2,m_{12})_j. \tag{27}$$

Provided that ν is large enough, this linear problem admits solutions for w_R. Unfortunately, exact solutions to this equation are prone to strong oscillations in the w_R, which lead to numerical instabilities. Rather than seeking exact but oscillating solutions to Equation (27), we seek non-negative solutions that minimize

the squared error $\sum_j(\sum_{R=1}^{\nu}\Gamma_{jR}w_R - B(n_2,m_{12})_j)^2$ using the active set method implemented in the scipy.optimize.nnls routine. Assuming that we found a solution with an acceptably small error, we finally use Equation (26) to compute the frequency spectrum.

To choose ν, we note that ν sequential replacements give an average of $n(1-(1-1/n)^{\nu})$ net replacements once multiple replacements of the same lineage have been accounted for (e.g., (Mendelson *et al.* 2016)). We find that we get accurate results if we have enough net replaced lineages to cover most likely cases in binomial sampling: $n_2(1-(1-1/n_2)^{\nu}) > m_{12}n_2 + 3\sqrt{n_2 m_{12}(1-m_{12})}$.

We compared the exact and approximate dynamic programming approaches to simulate 50% admixture from population 1 into population 2 with target sizes $n_1 = n_2 = 200$. The exact approach required 400 lineages from population 1 and took 70 seconds on a 3.5 GHz processor. The approximate approach used 385 lineages from population 1 and took 5 seconds on the same processor. The mean squared difference to the exact distribution of allele frequencies was $1\times10^{-}13$, and the largest *relative* error for an entry containing more than one millionth of counts was $6\times10^{-}6$. The approximate approach is therefore much more economical in computational time and the required number of lineages from the source populations.

ODEs on the moments

One-dimensional case

In this section, we show how the system of ordinary differential equations (7) derived in the main text can also be derived by integration by parts from the diffusion approximation to the Wright-Fisher process. Under the diffusion approximation, the evolution of the single-population allele frequencies density $\phi(x,t)$ follows (e.g., (Gutenkunst *et al.* 2009)):

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} \simeq& \frac{1}{4N}\frac{\partial^2}{\partial x^2}x(1-x)\phi(x,t)\\ &-s\frac{\partial}{\partial x}\left(h+(1-2h)x\right)x(1-x)\phi(x,t)\\ &+2Nu\delta(x-\frac{1}{2N}).\end{aligned} \tag{28}$$

The first term of the right hand side is a diffusion term modeling the effect of genetic drift and the second term is a transport term that accounts for selection. Finally, the source term $2Nu\delta(x-\frac{1}{2N})$ models the mutation process under the infinite site assumption.

We are interested in the expected sample frequency spectra $\Phi_n(i)$. We only consider $1\le i\le n-1$, because $\Phi_n(0) = \Phi_n(n) = \infty$ in the infinite-sites model. The expected sample frequencies can be obtained by projecting the allele frequency density ϕ on weights $w_{n,i}(x) = \binom{n}{i}x^i(1-x)^{n-i}$:

$$\Phi_n(i) = \int_0^1 w_{n,i}(x)\phi dx.$$

The time derivative of $\Phi_n(i)$ is given by:

$$\frac{\partial\Phi_n(i)}{\partial t} = \int_0^1 w_{n,i}\dot{\phi}dx.$$

The delta function source term in Equation (28) is integrated easily. The leading-order term in $\frac{1}{N}$ is simply $nu\delta_{i=1}$: this is the rate at which mutations appear on n lineages over one generation:

$$\begin{aligned}\frac{\partial\Phi_n(i)}{\partial t} =& nu\delta_{i=1} + \frac{1}{4N}\int_0^1 w_i\frac{\partial^2}{\partial x^2}\left(x(1-x)\phi(x,t)\right)dx\\ &-s\int_0^1 w_i\frac{\partial}{\partial x}\left(\left(h+(1-2h)x\right)x(1-x)\phi(x,t)\right)dx.\end{aligned} \tag{29}$$

drift term: We want to rewrite the drift term of equation (29) in terms of the moments $\Phi_n(i)$. To this end, we integrate by parts:

$$\begin{aligned}\text{drift term} =& \frac{1}{4N}\int_0^1 w_i\frac{\partial^2}{\partial x^2}\left(x(1-x)\phi(x,t)\right)dx\\ =& \frac{1}{4N}\left[w_i\frac{\partial}{\partial x}\left(x(1-x)\phi(x,t)\right)\right]_0^1\\ &-\frac{1}{4N}\int_0^1\frac{\partial w_i}{\partial x}\frac{\partial}{\partial x}\left(x(1-x)\phi(x,t)\right)dx.\end{aligned}$$

As $w_i(0) = w_i(1) = 0$ we assume that the first term is 0. We can integrate by parts the second term:

$$\begin{aligned}\text{drift term} =& -\frac{1}{4N}\int_0^1\frac{\partial w_i}{\partial x}\frac{\partial}{\partial x}\left(x(1-x)\phi(x,t)\right)dx\\ =& -\frac{1}{4N}\left[\frac{\partial w_i}{\partial x}\times x(1-x)\phi(x,t)\right]_0^1\\ &+\frac{1}{4N}\int_0^1\frac{\partial^2 w_i}{\partial x^2}x(1-x)\phi(x,t)dx.\end{aligned}$$

As we consider a finite genome, $\phi(0,t)$ and $\phi(1,t)$ are finite and, once again, the term in square brackets is zero. Moreover, we have:

$$\begin{aligned}\frac{\partial^2 w_i}{\partial x^2} =& \binom{n}{i}\Big[i(i-1)x^{i-2}(i-x)^{n-i} - 2i(n-i)x^{i-1}(1-x)^{n-i-1}\\ &+(n-i)(n-i-1)x^i(1-x)^{n-i-2}\Big]\end{aligned}$$

Thus,

$$\begin{aligned}\text{drift term} =& \frac{1}{4N}\int_0^1\frac{\partial^2 w_i}{\partial x^2}x(1-x)\phi(x,t)dx\\ =& \frac{1}{4N}\Big[\int_0^1\binom{n}{i}i(i-1)x^{i-1}(i-x)^{n-i+1}\phi(x,t)dx\\ &-2\int_0^1\binom{n}{i}i(n-i)x^i(1-x)^{n-i}\phi(x,t)dx\\ &+\int_0^1\binom{n}{i}(n-i)(n-i-1)x^{i+1}(1-x)^{n-i-1}\phi(x,t)dx\Big].\end{aligned}$$

Rearranging this expression, we can write it in terms of the $\Phi_n(i)$:

$$\begin{aligned}\text{drift term} =& \frac{1}{4N}\left[(i-1)(n-i+1)\Phi_n(i-1)\delta_{i\ge2}\right.\\ &-2i(n-i)\Phi_n(i)\delta_{1\le i\le n-1}\\ &\left.+(n-i-1)(i+1)\Phi_n(i+1)\delta_{i\le n-2}\right],\end{aligned}$$

which corresponds to the one-dimensional version of equation (11).

selection term: Here again, we integrate by parts the selection term:

$$\begin{aligned}
&-s\int_0^1 w_i \frac{\partial}{\partial x}\left((h+(1-2h)x)\,x(1-x)\phi(x,t)\right)dx \\
&= -s\left[w_i\left(h+(1-2h)x\right)x(1-x)\phi(x,t)\right]_0^1 \\
&\quad + s\int_0^1 \frac{\partial w_i}{\partial x}\left(h+(1-2h)x\right)x(1-x)\phi(x,t)dx.
\end{aligned}$$

We can again assume that the term in square brackets is zero, we rearrange the other integral term in order to write it in terms of the $\Phi_n(i)$. We finally get:

$$\begin{aligned}
\text{selection term} =& \frac{sh}{n+1}\Big[i(n+1-i)\Phi_{n+1}(i) \\
&\qquad -(n-i)(i+1)\Phi_{n+1}(i+1)\Big] \\
&+\frac{s(1-2h)(i+1)}{(n+1)(n+2)}\Big[i(n+1-i)\Phi_{n+2}(i+1) \\
&\qquad -(n-i)(i+2)\Phi_{n+2}(i+2)\Big],
\end{aligned}$$

which corresponds to the one-dimensional version of equation (12).

Bringing these expressions together, we get a system of ordinary differential equations on the $\Phi_n(i)$:

$$\begin{aligned}
\frac{\partial\Phi_n(i)}{\partial t} =& nu\delta_{i=1} + \frac{1}{4N}\Big[(i-1)(n-i+1)\Phi_n(i-1)\delta_{i\geq 2} \\
&\qquad -2i(n-i)\Phi_n(i)\delta_{1\leq i\leq n-1} \\
&\qquad +(n-i-1)(i+1)\Phi_n(i+1)\delta_{i\leq n-2}\Big] \\
&+\frac{sh}{n+1}\Big[i(n+1-i)\Phi_{n+1}(i) \\
&\qquad -(n-i)(i+1)\Phi_{n+1}(i+1)\Big] \\
&+\frac{s(1-2h)(i+1)}{(n+1)(n+2)}\Big[i(n+1-i)\Phi_{n+2}(i+1) \\
&\qquad -(n-i)(i+2)\Phi_{n+2}(i+2)\Big]
\end{aligned} \tag{30}$$

This is the one-dimensional, continuous time equivalent to equation (10).

Extension to multiple dimensions

The diffusion model can be generalized to multiple population studies. The drift, selection and mutation terms are directly derived from the unidimensional case. We just need to add the effect of the migrations between the populations. The equation on the joint frequency spectrum is:

$$\begin{aligned}
\frac{\partial\phi(\mathbf{x},t)}{\partial t} \simeq& \sum_{j=1}^{p}\Bigg[\frac{1}{4N_j}\frac{\partial^2}{\partial x_j^2}\Big(x_j(1-x_j)\phi(\mathbf{x},t)\Big) \\
&-s_j\frac{\partial}{\partial x_j}\Big(\Big(h_j+(1-2h_j)x_j\Big)x_j(1-x_j)\phi(\mathbf{x},t)\Big) \\
&-\sum_{k\neq j} m_{jk}\frac{\partial}{\partial x_k}\Big((x_j-x_k)\phi(\mathbf{x},t)\Big) \\
&+2N_j u\delta(x_j-\frac{1}{2N_j})\Pi_{k\neq j}\delta(x_k)\Bigg].
\end{aligned} \tag{31}$$

Now $\mathbf{x}$ is a vector: $\mathbf{x}=[x_1,\cdots,x_p]\in[0;1]^p$. It is the same for the parameters $\mathbf{s}$ and $\mathbf{h}$ and N_j is the reference population size for population j. The term $-\sum_{j=1}^{p}\sum_{k\neq j} m_{jk}\frac{\partial}{\partial x_k}\Big((x_j-x_k)\phi(\mathbf{x},t)\Big)$ accounts for the migrations where m_{jk} is the proportion of individuals in population k whose parents were born in population j.

We are interested in the multi dimensional statistics generalizing the $\Phi_n(i)$:

$$\Phi_{\mathbf{n}}(\mathbf{i}) = \int\prod_{j=1}^{p}\binom{n_j}{i_j}x_j^{i_j}(1-x_j)^{n_j-i_j}dx_j\phi(\mathbf{x}),$$

where $\mathbf{n}$ and $\mathbf{i}$ are vectors. The calculations are a bit more tedious but we can do the same as for the single population case: integrating by parts each term and writing it in terms of the $\Phi_{\mathbf{n}}(\mathbf{i})$, we recover Equation (10).

Jackknife approximation for moment closure

The evolution equations for Φ_n involves higher order terms, such as Φ_{n+1} or Φ_{n+2}, which leads to a moment closure problem. To tackle this, we use a jackknife approach to approximate higher-order terms as linear functions of Φ_n:

$$\Phi_{n+1}(i)\simeq\sum_{k_i\in I_i}\pi_{k_i}\Phi_n(k_i),$$

where I_i is a set of indices chosen so that the frequency k_i/n is close to the target frequency $i/(n+1)$.

Given a set I_i, we choose the coefficients π_{k_i} so that the jackknife is exact for a given parameterized family of functions Φ_{n+1}.

The 'order' of the jackknife refers to the number of terms in I_i: the more terms, the more general the family of functions can be, but the more numerically unstable the jackknife becomes.

We focus here on the order 3 jackknife for the terms $\Phi_{n+1}(i)$ and we will use the following approximation:

$$\forall i\in[1;n],\Phi_{n+1}(i)=\alpha_i\Phi_n(i'-1)+\beta_i\Phi_n(i')+\gamma_i\Phi_n(i'+1),$$

where the index i' is chosen so that the frequency $\frac{i'}{n}$ is as close as possible to $\frac{i}{n+1}$ and to satisfy the following boundary conditions:

$$\begin{aligned}
i'-1&\geq 1,\\
i'+1&\leq n-1.
\end{aligned}$$

We choose the approximation so that the interpolation is exact for quadratic allele frequency distribution:

$$\phi(x)=a+bx+cx^2.$$

Even though this seems like a drastic approximation, the jackknife only requires the quadratic approximation to hold locally: the parameters a, b, and c will be chosen independently for each value of i.

The moments become:

$$\Phi_n(i)=\int_0^1\binom{n}{i}x_i(1-x)^{n-i}\left[a+bx+cx^2\right]dx.$$

The integral can be performed analytically to yield:

$$\Phi_n(i)=\frac{a(n+2)(n+3)+(i+1)(c(i+2)+b(n+3))}{(n+1)(n+2)(n+3)}. \tag{32}$$

We then look for the jackknife coefficients α_i, β_i and γ_i that cancel the approximation error:

$$\Phi_{n+1}(i) - \alpha_i\Phi_n(i'-1) - \beta_i\Phi_n(i') - \gamma_i\Phi_n(i'+1) = 0.$$

We replace the Φ by their expressions under Equation (32):

$$\begin{aligned}
&\frac{a(n+3)(n+4) + (i+1)(c(i+2)+b(n+4))}{(n+4)} \\
&-\alpha_i\frac{a(n+2)(n+3) + i'(c(i'+1)+b(n+3))}{(n+1)} \\
&-\beta_i\frac{a(n+2)(n+3) + (i'+1)(c(i'+2)+b(n+3))}{(n+1)} \\
&-\gamma_i\frac{a(n+2)(n+3) + (i'+2)(c(i'+3)+b(n+3))}{(n+1)} = 0.
\end{aligned}$$

This equation should hold for all values of (a, b, c), so we can use the particular values $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ and write the corresponding system of equations:

$$\begin{cases}
\dfrac{(n+1)}{n+2} &= \alpha_i + \beta_i + \gamma_i, \\
\dfrac{(n+1)(i+1)}{n+3} &= \alpha_i i' + \beta_i(i'+1) + \gamma_i(i'+2), \\
\dfrac{(n+1)(i+1)(i+2))}{n+4} &= \alpha_i i'(i'+1) + \beta_i(i'+1)(i'+2) \\
&\quad + \gamma_i(i'+2)(i'+3).
\end{cases}$$

We can compute the jackknife coefficients α_i, β_i and γ_i by solving the previous linear system. We finally get:

$$\begin{cases}
\alpha_i = \dfrac{Q_\alpha}{2(n+2)(n+3)(n+4)}, \\
\beta_i = \dfrac{Q_\beta}{(n+2)(n+3)(n+4)}, \\
\gamma_i = \dfrac{Q_\gamma}{2(n+2)(n+3)(n+4)}.
\end{cases}$$

with

$$\begin{aligned}
Q_\alpha &= (n+1)\Big[4 + i^2(6+5n+n^2) - i(14+9n+n^2) \\
&\qquad - (n+4)(2i(n+2) - n - 5)i' + (n^2+7n+12)i'^2\Big] \\
Q_\beta &= (n+1)\Big[(i+1)(n+2)(i(n+3) - n - 6) \\
&\qquad - 2(n+4)(i(n+2)-1)i' + (n^2+7n+12)i'^2\Big] \\
Q_\gamma &= (n+1)\Big[(i+1)(n+2)(i(n+3) - 2) \\
&\qquad - (n+4)(2i(n+2) + n + 1)i' + (n^2+7n+12)i'^2\Big]
\end{aligned}$$

The same strategy is used to approximate $\Phi_{n+2}(i)$ as a linear combination of the Φ_n entries. We use the same formulation for multidimensional simulations as the moment closure problem can be addressed separately in each dimension. For instance, in the 2-dimensional case, the selection in the first population will involve the higher order term $\Phi_{n_1+1,n_2}(i_1, i_2)$ that we will approximate as follows:

$$\begin{aligned}
\Phi_{n_1+1,n_2}(i_1,i_2) =& \alpha\Phi_{n_1,n_2}(i_1'-1,i_2) + \beta\Phi_{n_1,n_2}(i_1',i_2) \\
&+ \gamma\Phi_{n_1,n_2}(i_1'+1,i_2).
\end{aligned}$$

Benchmarks

In this section we present an extensive set of test cases to compare computational performances of *Moments* and *∂a∂i* in terms of speed and accuracy. Refer to paragraph Implementation and performance for more details about the metrics used for the comparisons. The results on test cases with 30, 80 and 200 ploids are respectively given in Tables 5, 6 and 7. A more explicit description of the test cases is given in Table 8. Unless otherwise specified, populations sizes grow linearly: $N(t) = 1 + 0.01t$ with t the time in genetic units. For the exponential growth, we used the function $N(t) = 10^t$. When dealing with several populations, we used the same selection and migration parameters as well as the same population growth functions for all the populations, except for the Out-of-Africa model.

Additional features

Moments is based on *∂a∂i*'s interface but incorporates a new computation engine based on the method described in this paper. In addition, we added a few convenience features and optimizations. We developed a model plotting module that returns a schematic representation of a given demographic model. This is not only convenient for presenting results, but also proved very useful in troubleshooting. Figure 2 has been generated with this module. Moreover, it is now possible to directly import data and extract the allele frequency spectrum from a vcf file and to generate bootstrapped frequency spectra from the original data set. Finally, we implemented the Powell method for likelihood optimization, which we found to be more robust than the gradient and simplex methods used previously in *∂a∂i*, consistent with the findings of (Excoffier *et al.* 2013).

Demographic model	method	exec time (s)	mean(ε_r)	Δ LL	# < 0
Neutral equilibrium 1D	*∂a∂i*	3.6e-02	1.3e-02	**1.5e-05**	0
	Moments	**1.5e-03**	**9.7e-03**	4.6e-05	0
Neutral 1D,	*∂a∂i*	3.6e-02	**3.7e-03**	**4.2e-06**	0
T = 1.0	*Moments*	**1.2e-03**	1.4e-02	3.2e-04	0
Neutral 1D,	*∂a∂i*	1.4e-01	2.5e-03	1.2e-05	0
T = 5.0	*Moments*	**4.6e-03**	**2.8e-04**	**4.8e-08**	0
Selection 1D,	*∂a∂i*	3.7e-02	**6.5e-03**	**1.1e-06**	0
T = 1.0	*Moments*	**8.2e-03**	1.5e-02	3.7e-04	0
Selection 1D,	*∂a∂i*	1.5e-01	5.6e-03	3.4e-06	0
T = 5.0	*Moments*	**3.9e-02**	**9.2e-04**	**5.5e-07**	0
Neutral equilibrium 2D	*∂a∂i*	2.8e-01	1.3e-02	3.1e-03	0
	Moments	**4.2e-02**	**9.7e-03**	**9.3e-05**	0
Selection 2D,	*∂a∂i*	1.7e-01	1.3e-02	2.9e-03	0
T = 1.0	*Moments*	**2.1e-02**	**9.1e-04**	**-1.4e-03**	0
Selection 2D,	*∂a∂i*	6.2e-01	5.5e-03	9.3e-03	0
T = 5.0	*Moments*	**7.5e-02**	**9.2e-04**	**-6.5e-03**	0
Selection 2D,	*∂a∂i*	1.7e-01	**5.5e-03**	4.7e-04	0
T = 1.0, neutral fs0	*Moments*	**1.7e-02**	1.7e-02	**2.6e-04**	0
Selection, migration 2D,	*∂a∂i*	9.3e-01	8.1e-03	1.0e-04	0
T = 1.0	*Moments*	**2.2e-01**	**3.7e-03**	**2.1e-05**	0
Selection, migration 2D,	*∂a∂i*	4.4	8.4e-03	4.1e-04	0
T = 5.0	*Moments*	**1.0**	**2.5e-03**	**2.1e-05**	0
Selection, migration 2D,	*∂a∂i*	9.1e-01	5.2e-03	1.3e-04	0
T = 1.0, neutral fs0	*Moments*	**2.2e-01**	**1.5e-03**	**1.5e-05**	0
Fast growth 2D,	*∂a∂i*	9.6e-01	2.6e-01	**7.1e-04**	16
T = 1.0	*Moments*	**2.2e-01**	**6.3e-02**	8.0e-04	0
YRI-CEU 2D	*∂a∂i*	**1.3e-01**	**3.6e-03**	**3.4e-05**	0
	Moments	1.9e-01	8.3e-03	1.5e-04	0
Neutral equilibrium 3D	*∂a∂i*	1.6e+01	1.3e-02	9.1e-03	0
	Moments	**4.7e-01**	**9.7e-03**	**1.4e-04**	0
Selection 3D,	*∂a∂i*	6.6	7.4e-01	1.8e-03	0
T = 1.0	*Moments*	**1.2e-01**	**4.4e-05**	**-5.3e-02**	0
Selection 3D,	*∂a∂i*	2.7e+01	7.4e-01	6.2e-03	0
T = 5.0	*Moments*	**5.8e-01**	**8.8e-06**	**-1.7e-01**	0
Selection 3D,	*∂a∂i*	6.3	**3.6e-03**	1.8e-03	0
T = 1.0, neutral fs0	*Moments*	**1.2e-01**	2.2e-02	**1.5e-03**	0
Selection, migration 3D,	*∂a∂i*	8.2e+01	4.2e-02	6.2e-04	0
T = 1.0	*Moments*	**5.0**	**1.5e-02**	**1.7e-04**	0
Selection, migration 3D,	*∂a∂i*	4.0e+02	4.2e-02	3.4e-03	12
T = 5.0	*Moments*	**2.5e+01**	**1.4e-02**	**2.8e-04**	0
Selection, migration 3D,	*∂a∂i*	7.8e+01	3.0e-02	5.9e-04	3
T = 1.0 , neutral fs0	*Moments*	**5.0**	**4.1e-03**	**8.6e-05**	0
Fast growth 3D,	*∂a∂i*	7.9e+01	3.5e+02	3.1e-01	7021
T = 1.0	*Moments*	**5.0**	**7.1e-02**	**3.7e-03**	0
Out of Africa 3D	*∂a∂i*	5.7	3.1e-02	2.5e-04	0
	Moments	**4.1**	**1.6e-02**	**8.1e-05**	0

Table 5 Performance comparisons between *∂a∂i* and *Moments* on several scenarios (30 samples per population). For *∂a∂i* simulations we used Richardson extrapolation to improve convergence. The time T provided is the simulation time in genetic units.

Demographic model	method	exec time (s)	mean(ε_r)	Δ LL	< 0 entries
Neutral equilibrium 1D	*∂a∂i*	6.9e-02	1.1e-02	**5.4e-05**	0
	Moments	**2.0e-03**	**1.0e-02**	7.8e-05	0
Neutral 1D,	*∂a∂i*	5.3e-02	**4.1e-04**	**1.1e-07**	0
T = 1.0	*Moments*	**1.4e-03**	1.9e-02	5.8e-03	0
Neutral 1D,	*∂a∂i*	1.6e-01	**2.7e-04**	**2.5e-07**	0
T = 5.0	*Moments*	**4.9e-03**	4.6e-04	1.2e-05	0
Selection 1D,	*∂a∂i*	5.3e-02	**7.2e-04**	**4.9e-08**	0
T = 1.0	*Moments*	**8.7e-03**	1.9e-02	5.9e-03	0
Selection 1D,	*∂a∂i*	1.7e-01	6.2e-04	**9.8e-08**	0
T = 5.0	*Moments*	**3.6e-02**	**5.1e-04**	1.3e-05	0
Neutral equilibrium 2D	*∂a∂i*	1.4	1.0e-02	4.0e-03	15
	Moments	**8.9e-02**	**1.0e-02**	**1.6e-04**	0
Selection 2D,	*∂a∂i*	7.7e-01	1.0e-02	**3.6e-03**	0
T = 1.0	*Moments*	**2.5e-02**	**4.6e-04**	1.0e-02	0
Selection 2D,	*∂a∂i*	2.4	**4.5e-04**	9.1e-03	0
T = 5.0	*Moments*	**1.1e-01**	6.8e-04	**-3.9e-03**	0
Selection, migration 2D,	*∂a∂i*	3.5	**1.5e-03**	**1.6e-06**	0
T = 1.0	*Moments*	**2.4**	2.5e-03	5.5e-06	0
Selection, migration 2D,	*∂a∂i*	1.6e+01	5.4e-03	6.3e-06	0
T = 5.0	*Moments*	**1.1e+01**	**4.7e-03**	**4.7e-06**	0
Fast growth 2D,	*∂a∂i*	3.6	9.6e-02	**2.6e-05**	133
T = 1.0	*Moments*	**5.1**	**6.1e-02**	3.5e-05	0
YRI-CEU 2D	*∂a∂i*	**6.0e-01**	**1.1e-03**	**1.1e-06**	0
	Moments	1.6	9.7e-03	2.1e-04	0
Neutral equilibrium 3D	*∂a∂i*	2.0e+02	1.0e-02	1.2e-02	113549
	Moments	**4.4**	**1.0e-02**	**2.3e-04**	0
Selection 3D,	*∂a∂i*	8.0e+01	9.2e-01	6.5e-03	0
T = 1.0	*Moments*	**9.6e-01**	**8.7e-06**	**-7.9e-03**	0
Selection 3D,	*∂a∂i*	2.8e+02	9.6e-01	1.7e-02	0
T = 5.0	*Moments*	**4.8**	**1.3e-06**	**-6.3e-02**	0
Selection, migration 3D,	*∂a∂i*	8.4e+02	6.4e-03	**9.3e-06**	126
T = 1.0	*Moments*	**1.6e+02**	**4.5e-03**	2.7e-04	0
Fast growth 3D,	*∂a∂i*	8.3e+02	1.8e-01	**5.0e-05**	62559
T = 1.0	*Moments*	**1.6e+02**	**2.1e-02**	2.5e-04	0
Out of Africa 3D	*∂a∂i*	**5.7e+01**	**5.9e-03**	**6.9e-06**	102
	Moments	1.1e+02	1.9e-02	1.3e-04	0

Table 6 Performance comparisons between *∂a∂i* and *Moments* on several scenarios (80 samples per population). For *∂a∂i* simulations we used Richardson extrapolation to improve convergence. The time T provided is the simulation time in genetic units.

Demographic model	method	exec time (s)	mean(ε_r)	Δ LL	< 0 entries
Neutral equilibrium 1D	*∂a∂i*	1.15e-01	**1.03e-02**	**6.78e-05**	0
	Moments	**3.20e-03**	1.06e-02	9.62e-04	0
Neutral 1D,	*∂a∂i*	1.16e-01	**3.58e-05**	**4.41e-09**	0
T = 1.0	*Moments*	**1.81e-03**	2.04e-02	3.08e-02	0
Neutral 1D,	*∂a∂i*	3.38e-01	**2.35e-05**	**6.49e-09**	0
T = 5.0	*Moments*	**7.83e-03**	8.32e-04	6.73e-04	0
Selection 1D,	*∂a∂i*	1.08e-01	**6.45e-05**	**3.64e-09**	0
T = 1.0	*Moments*	**1.50e-02**	2.05e-02	3.17e-02	0
Selection 1D,	*∂a∂i*	2.69e-01	**5.53e-05**	**4.71e-09**	0
T = 5.0	*Moments*	**4.02e-02**	8.41e-04	7.41e-04	0
Neutral equilibrium 2D	*∂a∂i*	8.37	**1.02e-02**	4.80e-03	13934
	Moments	**3.51e-01**	1.06e-02	**1.92e-03**	0
Selection 2D,	*∂a∂i*	4.69	1.02e-02	**4.06e-03**	0
T = 1.0	*Moments*	**5.08e-02**	**2.04e-04**	6.16e-02	0
Selection 2D,	*∂a∂i*	1.29e+01	**8.72e-04**	8.91e-03	0
T = 5.0	*Moments*	**2.27e-01**	1.77e-03	**-2.42e-03**	0
Selection, migration 2D,	*∂a∂i*	**1.84e+01**	**1.52e-03**	**4.71e-07**	0
T = 1.0	*Moments*	3.31e+01	3.67e-03	1.74e-05	0
Selection, migration 2D,	*∂a∂i*	**8.24e+01**	**8.23e-03**	**4.89e-06**	18
T = 5.0	*Moments*	1.58e+02	8.68e-03	1.14e-05	0
Fast growth 2D,	*∂a∂i*	**1.84e+01**	**1.56e-02**	**5.12e-06**	1318
T = 1.0	*Moments*	6.65e+02	6.11e-02	2.01e-05	0
YRI-CEU 2D	*∂a∂i*	**3.67**	**3.41e-03**	**2.50e-06**	0
	Moments	1.54e+01	1.27e-02	2.53e-04	0
Neutral equilibrium 3D	*∂a∂i*	3.71e+03	**1.02e-02**	1.42e-02	6325463
	Moments	**4.38e+01**	1.06e-02	**2.89e-03**	0

Table 7 Performance comparisons between *∂a∂i* and *Moments* on several scenarios (200 samples per population). For *∂a∂i* simulations we used Richardson extrapolation to improve convergence. The time T provided is the simulation time in genetic units.

Demographic model	Description
Neutral equilibrium kD	k constant size population(s), neutral simulation up to equilibrium (T=5.0)
Neutral kD	k growing population(s), neutral simulation.
Selection kD	k growing population(s) with selection: $Ns = 1$ and $h = 0.7$.
Selection kD, neutral fs0	k growing population(s) starting at equilibrium with selection: $Ns = 1$ and $h = 0.7$.
Selection, migration kD	k growing populations with selection and migrations: $Ns = 1$, $h = 0.7$ and $m = 2.0$.
Selection, migration kD, neutral fs0	k growing populations starting at equilibrium, with selection and migrations: $Ns = 1$, $h = 0.7$ and $m = 2.0$.
Fast growth kD	k exponentially growing populations with selection and migrations: $Ns = 1$, $h = 0.7$ and $m = 2.0$.
YRI-CEU	2 populations Out of Africa expansion model, test case given in *∂a∂i*.
Out of Africa 3D	3 populations Out of Africa expansion model as described in Gravel *et al.* (2011); Gutenkunst *et al.* (2009)

Table 8 Test cases descriptions.