Models of archaic admixture and recent history from two-locus statistics

Aaron P. Ragsdale and Simon Gravel

Department of Human Genetics, McGill University, Montreal, QC, Canada

December 6, 2018

Abstract

We learn about population history and underlying evolutionary biology through patterns of genetic polymorphism. Many approaches to reconstruct evolutionary histories focus on a limited number of informative statistics describing distributions of allele frequencies or patterns of linkage disequilibrium. We show that many commonly used statistics are part of a broad family of two-locus moments whose expectation can be computed jointly and rapidly under a wide range of scenarios, including complex multi-population demographies with continuous migration and admixture events. A full inspection of these statistics reveals that widely used models of human history fail to predict simple patterns of linkage disequilibrium. To jointly capture the information contained in classical and novel statistics, we implemented a tractable likelihood-based inference framework for demographic history. Using this approach, we show that human evolutionary models that include archaic admixture in Africa, Asia, and Europe provide a much better description of patterns of genetic diversity across the human genome. We estimate that individuals in two African populations have $6-8\%$ ancestry through admixture from an unidentified archaic population that diverged from the ancestors of modern humans 500 thousand years ago.

The study of genetic diversity in human populations has shed light on the origins of our species and our spread across the globe. With the growing abundance of sequencing data from contemporary and ancient humans, coupled with archaeological evidence and detailed models of human demography, we continue to refine our understanding of our intricate history. Accurate demographic models also serve as a statistical foundation for the identification of loci under natural selection and the design of biomedical and association studies.

Whole-genome sequencing data are high dimensional and noisy. In order to make inferences of history and biology, we rely on summary statistics of variation across the entire genome and in many sequenced individuals. One such statistic that is commonly used for demographic inference is the distribution of SNP allele frequencies in one or more populations, called the sample or allele frequency spectrum (AFS) (Marth *et al.*, 2004; Gutenkunst *et al.*, 2009; Jouganous *et al.*, 2017; Kamm *et al.*, 2017). AFS-based inference has proven to be a powerful inference approach, yet it assumes independence between SNPs and therefore ignores information contained in correlations between neighboring linked loci, which is also referred to as linkage disequilibrium (LD).

Measures of LD are also informative about demographic history, mutation, recombination, and selection. A separate class of inference methods leverage observed LD across the genome to infer local recombination rates (McVean *et al.*, 2004; Auton and McVean, 2007; Chan *et al.*, 2012; Kamm *et al.*, 2016) and demographic history (Li and Durbin, 2011; Loh *et al.*, 2013; Schiffels and Durbin, 2014; Rogers, 2014).

While two-locus statistics have been extensively studied (Hill and Robertson, 1966, 1968; Karlin and McGregor, 1968; Ohta and Kimura, 1969a,b; Golding, 1984; Ethier and Griffiths, 1990; Hudson, 2001; McVean, 2002; Song and Song, 2007), most of this work focused on a single population at equilibrium demography, precluding their application to realistic demographic scenarios. Recently, approaches for computing the full two-locus sampling distribution for a single population with non-equilibrium demography were developed via the coalescent (Kamm *et al.*, 2016) or a numerical solution to the diffusion approximation (Ragsdale and

Gutenkunst, 2017), allowing for more robust inference of fine-scale recombination rates and single population demographic history. However, there remain significant limitations. Computing the full two-locus haplotype frequency spectrum is computationally expensive, hindering its application to inference problems that require a large number of function evaluations. Alternatively, computationally efficient low-order equations for specific LD statistics have been proposed (Hill and Robertson, 1968; Rogers, 2014), but these have seen limited application and only to single populations.

In this article, we show that the moment system of Hill and Robertson (1968) can be expanded to compute a large family of one- and two-locus statistics with flexible recombination, population size history, and mutation models. Additionally, we show that the system can be extended to multiple populations with continuous migration and discrete admixture events, and that low order statistics can be accurately and efficiently computed for tens of populations with complex demography.

We use this moment system together with likelihood-based optimization to infer multi-population demographic histories. We reexamine how well widely-used models of human demographic history recover observed patterns of polymorphism, and find that these models underestimate LD among rare variants in each population, sometimes by a large amount. The inclusion of archaic admixture in both Eurasian and African populations resolves these differences, and we infer $\sim 6-8\%$ archaic contribution in two populations in Africa. By jointly modeling a wide range of summary statistics across human populations, we can reveal important aspects of our history that are invisible through traditional analyses using individual statistics.

Theory and methods

To compute a large set of summary statistics for genetic data, we use mathematical properties of the Wright-Fisher model that are related to the look-down model (Donnelly and Kurtz, 1999a,b). To illustrate this process, we first build intuition through familiar equations from population genetics and then explain how these fit within a larger hierarchy of tractable models.

In this section, we therefore begin with evolution equations for heterozygosity and the frequency spectrum, then turn to recursions for low-order LD statistics and show that the classical Hill-Robertson (1968) system for D^2 can be extended to arbitrary moments of D, multiple populations, and even the full sampling distribution of two-locus haplotypes. Mathematical details and expanded discussion for each result are given in the Appendix. Throughout this article, we assume that human populations can be described, approximately, by a finite number of randomly mating populations.

Motivation: Single site statistics and the allele frequency spectrum

The most basic measure of diversity is the expected heterozygosity $\mathbb{E}[H]$, or the expected number of differences between two haploid copies of the genome. Given $\mathbb{E}[H]$ at time t, population size $N(t)$ and mutation rate u, Wright (1931) showed that enumerating all distinct ways to choose parents among two lineages leads to a recursion for $\mathbb{E}[H]$,

$$\mathbb{E}[H]_{t+1} = \left(1 - \frac{1}{2N(t)}\right)\mathbb{E}[H]_t + 2u. \tag{1}$$

To leading order in $1/N$ and u, two copies of the genome are different if their parents were distinct (which has probability $1-\frac{1}{2N}$) and carried different alleles (which has probability $\mathbb{E}[H]_t$), or if there was a mutation along one of two lineages (which has probability $2u$.)

Heterozygosity is a low order statistic: we only require two copies of the genome to estimate $\mathbb{E}[H]$ genome-wide. More samples provide additional information that can be encoded in the sample AFS Φ_n, the distribution of allele counts within a sample of size n .

A standard forward approach to compute Φ_n involves numerically solving the partial differential equation for the distribution of allele frequencies in the full population and then sampling from this distribution for the given sample size n (e.g. Gutenkunst et al. (2009)). By enumerating mutation events and parental copying probabilities in a sample of size n, Jouganous *et al.* (2017) showed that Equation 1 can be generalized to a recursion for $\{\Phi_n(i)\}_{i=0,\ldots,n}$. $\mathbb{E}[H]$ can be seen as a special case equal to $\Phi_2(1)$, the $i = 1$ bin in the size $n = 2$ frequency spectrum. These recursions can also be derived as moment equations for the diffusion approximation (Evans *et al.*, 2007; Živković *et al.*, 2015; Jouganous *et al.*, 2017).

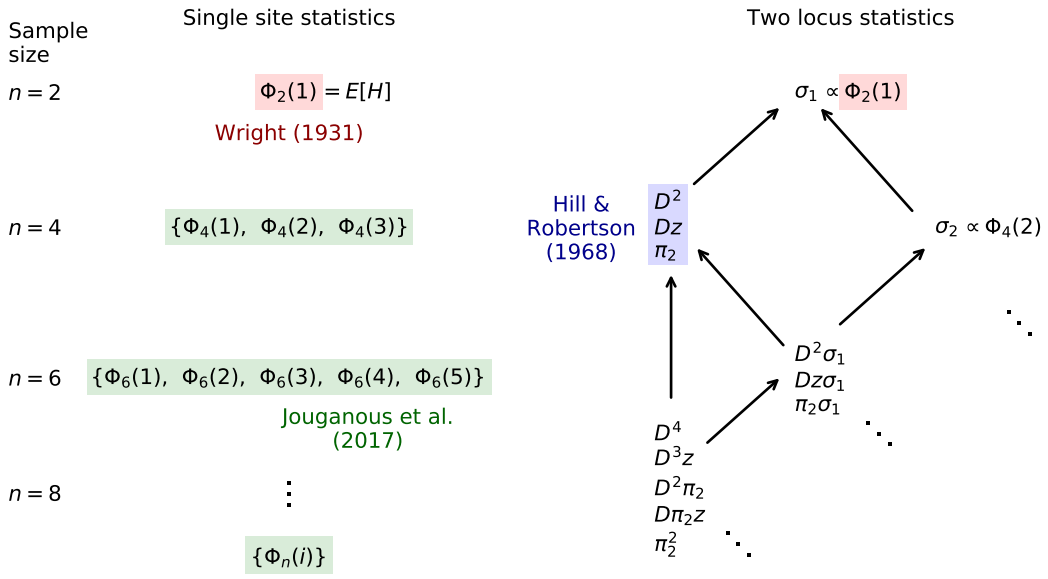

Figure 1: **Hierarchy of even-order moments of Wright-Fisher evolution.** Expected statistics under neutral Wright-Fisher evolution depend on equal or lower-order statistics at the previous generation, allowing for a hierarchy of closed recursion equations. On the left, single-site statistics are represented as the entries in the size-n AFS, Φ_n, and depend only on same-order statistics. On the right, the corresponding two-locus statistics, including the Hill-Robertson system for $\mathbb{E}[D^2]$, which relies on the expected heterozygosity. Closed recursions can be found for any given $\mathbb{E}[D^m]$, giving rise to sparse, linear system of ODEs that may be solved efficiently. We denote $\pi_2 = p(1-p)q(1-q), z = (1-2p)(1-2q)$, and $\sigma_i = p^i(1-p)^i + q^i(1-q)^i$. Arrows indicate dependence of moments and highlighted moments indicate classical recursions. Odd-order moments are shown in Figure A1.

Two-locus statistics

We will use this same intuition for the two-locus theory. First consider the model for two loci that each permit two alleles: alleles A/a at the left locus, and B/b at the right. There are four possible two-locus haplotypes, AB, Ab, aB, and ab, whose frequencies sum to 1 in the population. LD between two loci is measured as the covariance of their allele frequencies:

$$D = f_{AB}f_{ab} - f_{Ab}f_{aB}.$$

In other words, D is the probability of drawing two lineages from the population and observing one lineage of type AB and the other of type ab, minus the probability of observing the two cross types Ab and aB. As such, $\mathbb{E}[D]$ is a two-haplotype statistic, meaning we require just two haploid copies of the genome (or a single phased diploid genome) to estimate D, in the same way that the expected heterozygosity $\mathbb{E}[H]$ is a two-sample statistic of single-site variation.

Moment equations for D and D^2

Enumerating possible copying, recombination, and mutation events for two lineages also leads to a well-known recursion for $\mathbb{E}[D]$ (Hill and Robertson, 1966). The possibility of sharing a common parent from the previous generation leads to the same $\frac{1}{2N}$ decay familiar from Equation 1. $\mathbb{E}[D]$ also decays due to recombination with rate proportional to the probability r of a recombination event between two loci in a given generation. To leading order in r, u, and $\frac{1}{2N}$, we have

$$\mathbb{E}[D]_{t+1} = \left(1 - \frac{1}{2N} - r\right)\mathbb{E}[D]_t. \tag{2}$$

Mutation doesn't contribute to $\mathbb{E}[D]$ because any mutation event is equally likely to contribute positively or negatively to the statistic. As a result, D is expected to be zero across the genome.

However, the second moment $\mathbb{E}[D^2]$ is positive. Hill and Robertson (1968) found a recursion for a triplet of statistics including $\mathbb{E}[D^2]$, which we write as

$$\mathbf{y} = \begin{pmatrix} \mathbb{E}[D^2] \\ \mathbb{E}[D(1-2p)(1-2q)] \\ \mathbb{E}[p(1-p)q(1-q)] \end{pmatrix},$$

where p is the allele frequency of A, and q is the allele frequency of B. The recursion is

$$\mathbf{y}_{t+1} - \mathbf{y}_t = \left(\mathcal{D}_{N(t)} + \mathcal{R}_r\right)\mathbf{y}_t, \tag{3}$$

where $\mathcal{D}$ and $\mathcal{R}$ are matrix operators for drift and recombination, respectively. To leading order in $\frac{1}{2N}$ and r, these take the form

$$\mathcal{D}_{N(t)} = \frac{1}{2N(t)} \begin{pmatrix} -3 & 1 & 1 \\ 4 & -5 & 0 \\ 0 & 1 & -2 \end{pmatrix},$$

and

$$\mathcal{R}_r = r \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The three statistics in the Hill-Robertson system have a natural interpretation. $\mathbb{E}[D^2]$ is the variance of D and has received plenty of attention over the years. The second statistic includes a term $z = (1-2p)(1-2q)$ whose magnitude is largest when there are rare alleles at both loci. Thus $\mathbb{E}[D(1-2p)(1-2q)] = \mathbb{E}[Dz]$ is a measure of LD weighted towards rare variants. In other words, pairs of rare variants in high LD contribute sizably to this statistic. $\mathbb{E}[\pi_2] = \mathbb{E}[p(1-p)q(1-q)]$ is the joint heterozygosity across pairs of SNPs. If we sample four haplotypes from the population, this is the probability that the first pair differ at the left locus, and the second pair differ at the right locus.

The applications in this article focus on generalizing the Hill-Robertson equations to multi-population settings. However, we first outline generalizations to high-order moments and non-neutral evolution, leaving theoretical developments and simulations to the Appendix.

Generalizing to higher moments of D

The existence of tractable higher-order moment equations for one-locus statistics (Jouganous *et al.*, 2017) suggests the existence of a similar high-order system for two-locus statistics. Higher moments of D provide additional information about the distribution of two-locus haplotypes. Appendix A.1 shows that the Hill-Robertson system can be extended to compute any moment of D, and presents recursions for those systems of arbitrary order D^m that closes under drift, recombination, and mutation.

This family of recursion equations takes a form similar to the D^2 system: the evolution of $\mathbb{E}[D^m]$ requires $\mathbb{E}[D^{m-1}z]$ and $\mathbb{E}[D^{m-2}\pi_2]$, with each of those terms depending on additional terms of the same order and smaller orders (Figure 1). For any order m, Appendix A.4.1 shows the system closes and forms a hierarchy of moment equations, in that the D^m recursion contains the D^{m-2} system, which itself contains the D^{m-4} system, and so on (Figure A1). Just as the Wright equation for heterozygosity generalizes naturally to equations for the more informative distribution of allele frequency (Jouganous *et al.*, 2017), the Hill and Robertson equations for $E[D]$ and $E[D^2]$ generalize to informative higher-order LD statistics.

Generalizing to arbitrary two-locus haplotype distribution

Given the analogy between the frequency spectrum and the Hill-Robertson equations, it is natural to study the connection between the moment equations for D^n and the evolution of the two-locus haplotype frequency distribution $\Psi_n(f_{AB}, f_{Ab}, f_{aB}, f_{ab})$.

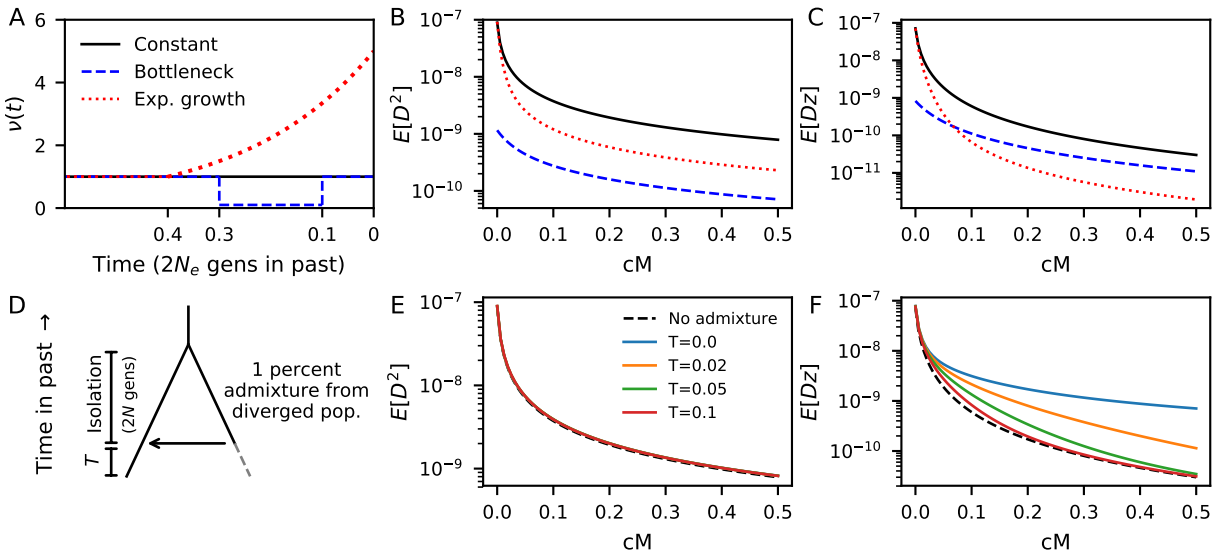

Figure 2: **LD curves are sensitive to demography.** (A-C) The statistics in the Hill-Robertson system, including the decay of linkage disequilibrium, are sensitive to demography. Here we compute expected decay curves for three distinct population size histories: constant size, a recent bottleneck and recovery, and recent exponential growth. Both the amplitude and shape of the LD curves differ between demographic models for (B) $\mathbb{E}[D^2]$ and (C) $\mathbb{E}[Dz] = \mathbb{E}[D(1-2p)(1-2q)]$. (D) To illustrate the effect of admixture on LD curves, we consider two populations in isolation for $2N$ generations, followed by an admixture event where the focal population receives 1% of lineages from the diverged population. (E) $\mathbb{E}[D^2]$ curves are largely unaffected by this low level of admixture. (F) However, $\mathbb{E}[Dz]$ is immediately and strongly elevated following admixture, and remains significantly elevated for prolonged time T since the admixture event.

While classical approaches for computing Ψ_n (Golding, 1984; Hudson, 2001) were limited to neutrality and steady-state demography, recent coalescent and diffusion developments allow for Ψ_n to be computed under non-equilibrium demography and selection (Kamm *et al.*, 2016; Ragsdale and Gutenkunst, 2017). These approaches are computationally expensive and limited to one population, as Ψ_n has size $\frac{(n+1)(n+2)(n+3)}{6}$, and the P-population distribution grows asymptotically as n^{3P}.

Generalizing the approach of Jouganous *et al.* (2017), we can write a recursion equation on the entries of Ψ_n under drift, mutation, recombination, and selection at one or both loci (Appendix A.3). As expected, this recursion does not close under selection: to find Ψ_n at time $t+1$, we require Ψ_{n+1} and Ψ_{n+2} at time t. It also does not close under recombination, requiring a closure approximation. Using the same closure strategy for selection and recombination, however, we can approximate the entries of Ψ_{n+1} and Ψ_{n+2} as linear combinations of entries in Ψ_n and obtain a closed equation. This approach provides accurate approximation for moderate n under recombination and selection (Appendix A.3.5) that represent a 10 to 100-fold speedup over the numerical PDE implementation in Ragsdale and Gutenkunst (2017) (Table A1). However, closure is inaccurate for small n.

By contrast to the full two-locus model, equations for moments of D close under recombination because the symmetric combination of haplotype frequencies that define D ensures the cancellation of higher-order terms (Appendix A.1.2). This makes the moments of D particularly suitable for rapid computation of low-order statistics over a large number of populations.

The Hill-Robertson system does not close, however, if one or both loci are under selection. Appendix A.1.4 considers a model where one of the two loci is under additive selection. We derive recursion equations for terms in the D^2 system and describe the moment hierarchy and a closure approximation, though we leave its development to future work. In the following we focus on neutral evolution.

Multiple populations

While a large body of work exists for computing expected LD in a single population, little progress has been made toward extending these models to multiple populations. Forward equations for the full two-locus sampling distribution become computationally intractable beyond just a single population, even with the moment-based approach described above. Here, we extend the Hill-Robertson system to any number of populations, allowing for population splits, admixture, and continuous migration.

Motivation: Heterozygosity across populations

To motivate our derivation of the multi-population Hill-Robertson system and provide intuition, we begin with a model for heterozygosity across populations with migration. With two populations we consider the cross-population heterozygosity, $\mathbb{E}[H_{12}] = \mathbb{E}[p_1(1-p_2)] + \mathbb{E}[p_2(1-p_1)]$, the probability that two lineages, one drawn from each population, differ by state. At the time of split between populations 1 and 2, $\mathbb{E}[H_1] = \mathbb{E}[H_2] = \mathbb{E}[H_{12}]$. Because coalescence between lineages in different populations is unlikely, $\mathbb{E}[H_{12}]$ is not directly affected by drift. In the absence of migration, this statistic increases linearly with the mutation rate over time (Figure A2).

With migration, the evolution of $\mathbb{E}[H_{12}]$ depends also on $\mathbb{E}[H_1]$ and $\mathbb{E}[H_2]$. We define the migration rate m_{12} to be the probability that a lineage from population 1 replaces a given lineage in population 2 in one generation. If the lineage drawn from population 2 is a new migrant from population 1 (with probability m_{12}), then $\mathbb{E}[H_{12}]_{t+1}$ is equal to $\mathbb{E}[H_1]_t$ (and similarly, equal to $\mathbb{E}[H_2]_t$ with probability m_{21} for a migration event in the opposite direction). To leading order in m_{ij}, we have

$$\mathbb{E}[H_{12}]_{t+1} = m_{12}\mathbb{E}[H_1]_t + m_{21}\mathbb{E}[H_2]_t + (1 - m_{12} - m_{21})\mathbb{E}[H_{12}]_t.$$

Similar intuition leads to recursions for $\mathbb{E}[H_1]$ and $\mathbb{E}[H_2]$ under migration, and this system easily extends to more than two populations.

The Hill-Robertson system with migration

We take the same approach to determine transition probabilities in the multi-population Hill-Robertson system. Suppose that at some time, a population splits into two populations. At the time of the split, expected two-locus statistics (D^2, Dz, π_2) in each population are each equal to those in the parental population at the time of split (Appendix A.2.1). Additionally, the covariance of D between the two populations, $\mathbb{E}[D_1 D_2]$, is initially equal to $\mathbb{E}[D^2]$ in the parental population. In the absence of migration, Hill-Robertson statistics in each population evolve according to Equation 3, and

$$\mathbb{E}[D_1 D_2]_{t+1} = \left(1 - \frac{1}{2N_1(t)} - \frac{1}{2N_2(t)} - 2r\right)\mathbb{E}[D_1 D_2]_t. \tag{4}$$

With migration, additional moments are needed to obtain a closed system. These additional terms take the same general form as the original terms in the Hill-Robertson system, but include cross-population statistics, analogous to H_{12} in the heterozygosity model with migration. We denote this basis $\mathbf{z}$,

$$\mathbf{z} = \begin{pmatrix} D_i D_j \\ D_i z_{j,k} \\ \pi_2(i,j;k,l) \\ H_{i,j} \end{pmatrix}, 1 \le i,j,k \le P \tag{5}$$

where P is the number of populations, and we slightly abuse notation so that $D_i D_j$ stands in for all index permutations (D_1^2, D_2^2, and $D_1 D_2$ in the two-populations case). We derive transition probabilities under continuous migration in Appendix A.2.2 leading to the closed recursion,

$$\mathbf{z}_{t+1} - \mathbf{z}_t = \left(\mathcal{D}_{\mathbf{N}(t)} + \mathcal{M}_{\mathbf{m}} + \mathcal{R}_r + \mathcal{U}_u\right)\mathbf{z}_t, \tag{6}$$

where $\mathcal{D}$, $\mathcal{M}$, $\mathcal{R}$, and $\mathcal{U}$ are sparse matrices for drift, migration, recombination and mutation that depend on the number of populations, population sizes $\mathbf{N}(t)$, and migration rates $\mathbf{m}$.

Admixture

Patterns of LD are sensitive to migration and admixture events, and low order LD statistics are commonly used to infer the parameters of admixture events (Moorjani *et al.*, 2011; Loh *et al.*, 2013). A well-known result (Cavalli-Sforza and Bodmer, 1971) shows that $\mathbb{E}[D]$ in an admixed population can be nonzero even when D is zero in both parental populations, if frequencies have changed appreciably between the two populations. This is seen by enumerating all possible combinations of haplotype sampling when a fraction f of lineages were contributed by population 1, and $1-f$ by population 2 (Appendix A.2.3). Right after admixture, the expectation $E[D_{\text{adm}}]$ in the admixed population is

$$\mathbb{E}[D_{\text{adm}}] = f\mathbb{E}[D_1] + (1-f)\mathbb{E}[D_2] + f(1-f)\mathbb{E}[\delta], \tag{7}$$

where $\delta = (p_1 - p_2)(q_1 - q_2)$.

To integrate the multi-population D^2 system after an admixture event, we require $\mathbb{E}[D^2_{\text{adm}}]$ and other second order terms in the basis (5) involving the admixed population. Using the same enumeration approach as for Equation 7, the expectation immediately following the admixture event is

$$\begin{aligned}\mathbb{E}[D^2_{\text{adm}}] =& f^2\mathbb{E}[D_1^2] + (1-f)^2\mathbb{E}[D_2^2] + 2f(1-f)\mathbb{E}[D_1D_2] \\ &+ 2f^2(1-f)\mathbb{E}[D_1\delta] + 2f(1-f)^2\mathbb{E}[D_2\delta] + f^2(1-f)^2\mathbb{E}[\delta^2].\end{aligned} \tag{8}$$

Each other required term can be found in a similar manner (Appendix A.2.3). In this way, the set of moments may be expanded to include the admixed population and integrated forward in time using Equation 6.

Numerical implementation

We rescale time by $2N_{\text{ref}}$ generations (N_{ref} is an arbitrary reference population size, often the ancestral population size), so that the recursion can be approximated as a differential equation

$$\dot{\mathbf{z}} = \left(\mathcal{D}_{\boldsymbol{\nu}(t)} + \mathcal{M}_{\tilde{\mathbf{m}}} + \mathcal{R}_{\rho/2} + \mathcal{U}_{\theta/2}\right)\mathbf{z}, \tag{9}$$

where $\boldsymbol{\nu}$ are the relative population sizes at time t ($\nu_i(t) = N_i(t)/N_{\text{ref}}$), $\tilde{\mathbf{m}}$ are the population size scaled migration rates $2N_{\text{ref}}m_{i,j}$, $\rho = 4N_{\text{ref}}r$, and $\theta = 4N_{\text{ref}}u$. Each matrix is sparse and has low order, so this equation can be solved efficiently using a standard Crank-Nicolson integration scheme. Our implementation allows users to define general models with standard demographic events (migrations, splits and mergers, size changes, etc.) similar to the `∂a∂i/moments` interface (Gutenkunst *et al.*, 2009; Jouganous *et al.*, 2017). A single evaluation of the four-population model shown in Figure A3 can be computed in roughly 0.1 second. We packaged our method with `moments` (Jouganous *et al.*, 2017) as `moments.LD`, a python module that computes expected statistics and performs likelihood-based inference from observed data (described below), available at `bitbucket.org/simongravel/moments`.

Validation

We validated our numerical implementation and estimation of statistics from simulated genomes using `msprime` (Kelleher *et al.*, 2016). Expectations for low-order statistics match closely with coalescent simulations. For example, Figure A3 shows the agreement for a four population model with non-constant demography, continuous migration, and an admixture event, for which we computed expectations using `moments.LD` that matched estimates from `msprime`. While approximating expectations from `msprime` required the time-consuming running and parsing of many simulations, expectations from `moments.LD` were computed in seconds on a personal computer.

Data and inference

Genotype data

Computing D using the standard definition requires phased haplotype data. Most currently available whole genome sequence data is unphased, so that we must rely on two-locus statistics based on genotype, instead

of haplotype, counts. One could estimate haplotype statistics using the Weir (1979) estimator

$$\hat{D} = \frac{1}{2n_d}\left(2n_{AABB} + n_{AABb} + n_{AaBB} + \frac{1}{2}n_{AaBb}\right) - \frac{n_A}{2n_d}\frac{n_B}{2n_d}, \tag{10}$$

where n_A is the count of A at the left locus, n_B the count of B at the right locus, n_d the number of diploid individuals in the sample, and $\{n_{AABB}, n_{AABb}, \ldots\}$ the counts of each observed genotype. However, the Weir estimator for D is biased. Fortunately, we can simply treat the Weir estimator $\hat{D}$ as a statistic and obtain an unbiased prediction for its expectation (Appendix A.7.3). Even though $\mathbb{E}[D^n]$ can be estimated from $2n$ phased haplotypes, more samples are required for accurate estimation with unphased data. Weir (1979) suggests sample sizes of $n_d = 20$ are sufficient for $\hat{D}$.

1000 Genome Project data

We computed statistics from intergenic data in the Phase 3 1000 Genomes Project data (The 1000 Genomes Project Consortium, 2015). The non-coding regions of the 1000 Genomes data is low coverage, which can lead to significant underestimation of low frequency variant counts, which distorts the frequency spectrum and can lead to biases in AFS-based demographic inference (e.g., Gravel *et al.* (2011)). However, low-order statistics in the Hill-Robertson system are robust to low coverage data in a large enough sample size (Figure A4), so that low coverage data are well suited for inference from LD statistics (see also Rogers (2014)).

To avoid possible confounding due to variable mutation rate across the genome, we calculated and compared statistics normalized by π_2, the joint heterozygosity: $\sigma_d^2 = \mathbb{E}[D^2]/\mathbb{E}[\pi_2]$, as in Rogers (2014). All figures showing σ_d^2-type statistics are normalized using π_2(YRI), the joint heterozygosity in the Yoruba population (YRI) from Nigeria.

This normalization removes all dependence of the statistics on the overall mutation rate, so that estimates of split times and population sizes are calibrated by the recombination rate per generation instead of the mutation rate. This is convenient given that genome-wide estimates of the recombination rate tend to be more consistent across experimental approaches than estimates of the mutation rate.

We considered all pairs of intergenic SNPs with $10^{-5} \leq r \leq 2 \times 10^{-2}$ using the African-American recombination map estimated by Hinch *et al.* (2011). The lower bound was chosen to further reduce the potential effect of short-range correlations of mutation rates and experimental error.

Likelihood-based inference on LD-curves

To compare observed LD statistics in the data to model predictions, and thus to evaluate the fit of the model to data, we used a likelihood approach. We binned pairs of SNPs based on the recombination distance separating them (Appendix A.7.2). Bins were defined by bin edges $\{r_0, r_1, \ldots, r_n\}$, roughly logarithmically spaced. The model is defined by the set of demographic parameters Θ. For a given recombination bin $(\rho_i, \rho_{i+1}]$, we computed statistics from the data $\hat{\mathbf{z}}_i$, and expected statistics from the model $\mathbf{M}_i$. We then estimated the likelihood as

$$\mathcal{L}(\Theta|\hat{\mathbf{z}}_i) = \mathcal{N}\left(\hat{\mathbf{z}}_i; \mathbf{M}_i, \Sigma_i\right),$$

taking the probability of observing data $\hat{\mathbf{z}}$ to be normally distributed with mean $\mathbf{M}$ and covariance matrix Σ.

We estimated Σ directly from the data by constructing bootstrap replicates from sampled subregions of the genome with replacement. This has the advantage of accounting for the covariance of statistics in our basis, as well as non-independence between distinct neighboring or overlapping pairs of SNPs. To compute the composite likelihood across ρ bins, we simply took the product of likelihoods over values of recombination bins indexed by i, so that

$$\mathcal{L}(\Theta) = \prod_i \mathcal{L}(\Theta|\hat{\mathbf{z}}_i).$$

To compute confidence intervals on parameters, we used the approach proposed by Coffman *et al.* (2016), which adjusts uncertainty estimates to account for non-independence between recombination bins and neighboring pairs of SNPs.

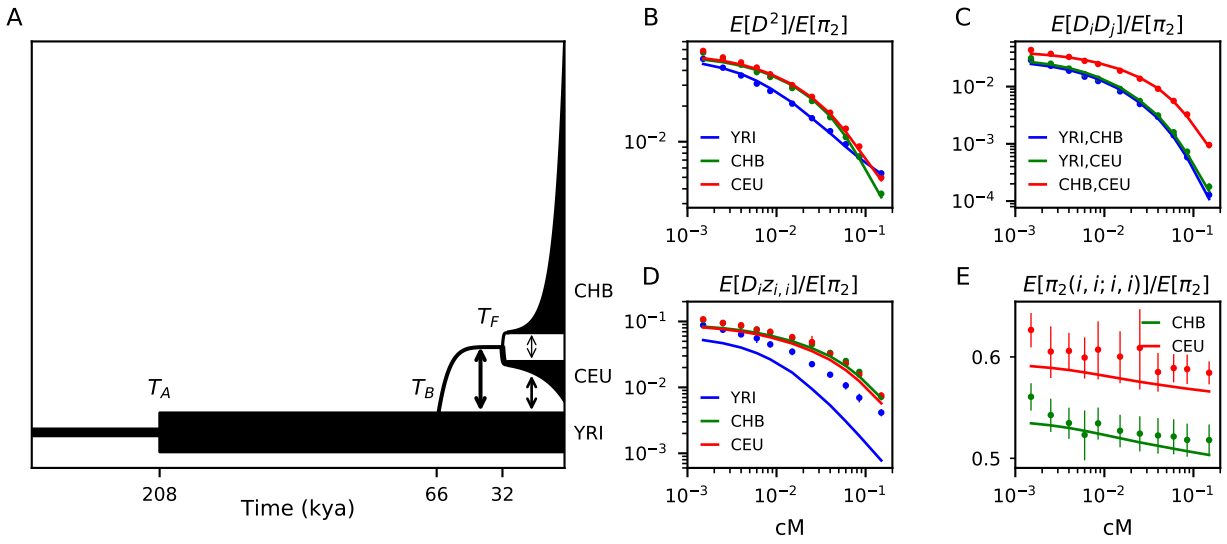

Figure 3: **Standard out-of-Africa model underestimates LD among rare variants** (A) We fit the 13-parameter model proposed by Gutenkunst *et al.* (2009) to statistics in the two-locus, multi-population Hill-Robertson system. The remaining 35 statistics from the Hill-Robertson basis used in the fit are shown in Figure A5. Best fit values for labeled parameters are given in Table A2. Most statistics were accurately predicted by this model, including (B) the decays of $\mathbb{E}[D^2]$ in each population, (C) the decay of the covariance of D between populations, and (E) the joint heterozygosity $\mathbb{E}[\pi_2(i)]$. (D) However, $\mathbb{E}[D_i(1-2p_i)(1-2q_i)]$ was fit poorly by this model, and we were unable to find a three-population model that recovered these observed statistics, including with additional periods of growth or recent admixture with other modern human populations. Error bars represent bootstrapped 95% confidence intervals on the statistic estimate.

Application to human data

Human expansion models underestimate LD between rare variants

The demographic model for human out-of-Africa (OOA) expansion proposed and inferred by Gutenkunst *et al.* (2009) has been widely used for subsequent simulation studies, and parameter estimates have been refined as more data became available (Gravel *et al.*, 2011; Tennessen *et al.*, 2012; Jouganous *et al.*, 2017). These models have typically been fit to the single-locus joint AFS, with Yoruba of Ibidan, Nigeria (YRI), Utah residents of Western European ancestry (CEU), and Han Chinese from Beijing (CHB) as representative panels. Gutenkunst *et al.* (2009) verified that the observed decay of r^2 was consistent with simulations under their inferred model.

We first asked if the OOA model (Figure 3(A)) is able to capture observed patterns of LD within and between these three populations. When fitting to all statistics in the multi-population basis, parameters diverged to infinite values, suggesting that the model is mis-specified. In particular, this model was unable to describe observed Dz-type statistics ($\mathbb{E}[D(1-2p)(1-2q)]$, D weighted by the joint rarity of alleles), with Dz-curves from the model drastically underestimating observations. We refit the OOA model without including Dz statistics, and we inferred best-fit parameters that generally align with estimates using the joint AFS (Table A2, left, and Figure 3). This model underestimated observed Dz in each population, especially in the YRI population (Figure 3(D)). Using AFS-inferred parameters from previous studies led to qualitatively similar results.

The Gutenkunst model is a vast oversimplification of human evolutionary history, so its failure to account for Dz is not all that surprising. However, given the good agreement of the model to both allele frequencies and D^2 decay, we did not expect such a large discrepancy. Having ruled out low coverage and spatial correlations in the mutation rate as explaining factors, our next hypothesis was a more complex demographic history. We generalized the Gutenkunst model with a number of additional parameters, such as recent growth

in the YRI population and recent mixing between populations, but none of these modifications provided satisfactory fit to the data.

Inference of archaic admixture

$\mathbb{E}[Dz]$ is a measurement of D weighted by the joint rarity of SNPs at the two loci, and pairs of rare alleles in high LD will have a substantial positive contribution to $\mathbb{E}[Dz]$. We therefore expect this statistic to be sensitive to the presence of rare, deep-coalescing lineages within the population, as those lineages will contribute haplotypes with a large number of tightly linked low frequency variants (see Discussion below).

Given prior genetic evidence for archaic introgression in Eurasia and Africa (Wall and Brandt, 2016), we proposed a model that includes two archaic branches, with one branch mixing with Eurasian ancestors beginning at the OOA event, and the second one mixing with the ancestors of the Yoruba population over a time period that could include the OOA event. In this scenario, this second branch could also contribute to Eurasians through admixture prior to the OOA event (Figure 4(A)). Modern and archaic humans coexisted on the African continent until quite recently (Berger *et al.*, 2017), and genetic evidence points to a history of archaic introgression across many modern African populations (Hammer *et al.*, 2011; Lachance *et al.*, 2012; Hsieh *et al.*, 2016; Durvasula and Sankararaman, 2018). It is likely that modern humans have met and mixed with archaic lineages many times through history, rather than receiving a single pulse of migrants (Browning *et al.*, 2018; Villanea and Schraiber, 2018). We chose to model the mixing of archaic and modern human branches as continuous and symmetric (Kuhlwilm *et al.*, 2016), parameterizing the migration rate between these branches and the times that migration began and ended.

We considered two topologies for the archaic branches: 1) both archaic branches split independently from that leading to modern humans (Figure 4(A)), and 2) one archaic branch split from the modern human branch, which some time later split into the two archaic populations (Figure A6(A)). Both models fit the data well with little statistical evidence to discriminate between these two models ($\Delta LL < 1$, as opposed to $\Delta LL = 1,730$ between models with and without archaic admixture). Consistent among the inferred models was the age of the split between archaic African and modern human branches at ~ 500 kya, though uncertainty remains with regard to the relationship between archaic humans in Africa and Eurasia. The sequencing of archaic genomes within Africa would clearly be helpful in resolving these topologies.

We inferred an archaic population to have contributed measurably to Eurasian populations. This branch (putatively Eurasian Neanderthal) split from the branch leading to modern humans between $\sim 470 - 650$ thousand years ago, and $\sim 1\%$ of lineages in modern CEU and CHB populations were contributed by this archaic population after the out-of-Africa split. This range of divergence dates compares to previous estimates of the time of divergence between Neanderthals and human populations, estimated at ∼650 kya (Prüfer *et al.*, 2014). The "archaic African" branch split from the modern human branch roughly $460 - 540$ kya and contributed $\sim 7.5\%$ to modern YRI in the model (Table A2).

We chose a separate population trio to validate our inference and compare levels of archaic admixture with different representative populations. This second trio consisted of the Luhya in Webuye, Kenya (LWK), Kinh in Ho Chi Minh City, Vietnam (KHV), and British in England and Scotland (GBR). We inferred the KHV and GBR populations to have experienced comparable levels of migration from the putatively Neanderthal branch. However, the LWK population exhibited lower levels of archaic admixture ($\sim 6\%$) in comparison to YRI, suggesting population differences in archaic introgression events within the African continent (Table A3).

Discussion

Multi-population two-locus diversity statistics

The application presented here relied on the four-haplotype statistics (D^2, Dz, π_2). Studying these low-order multi-population statistics in a likelihood framework allowed us to infer a demographic model with archaic admixture, even without archaic reference genomes. We have also shown that higher order statistics may be computed through this same framework. Extending higher order two-locus moment systems to multiple populations would potentially provide further information about demography, particularly for past encounters with archaic branches.

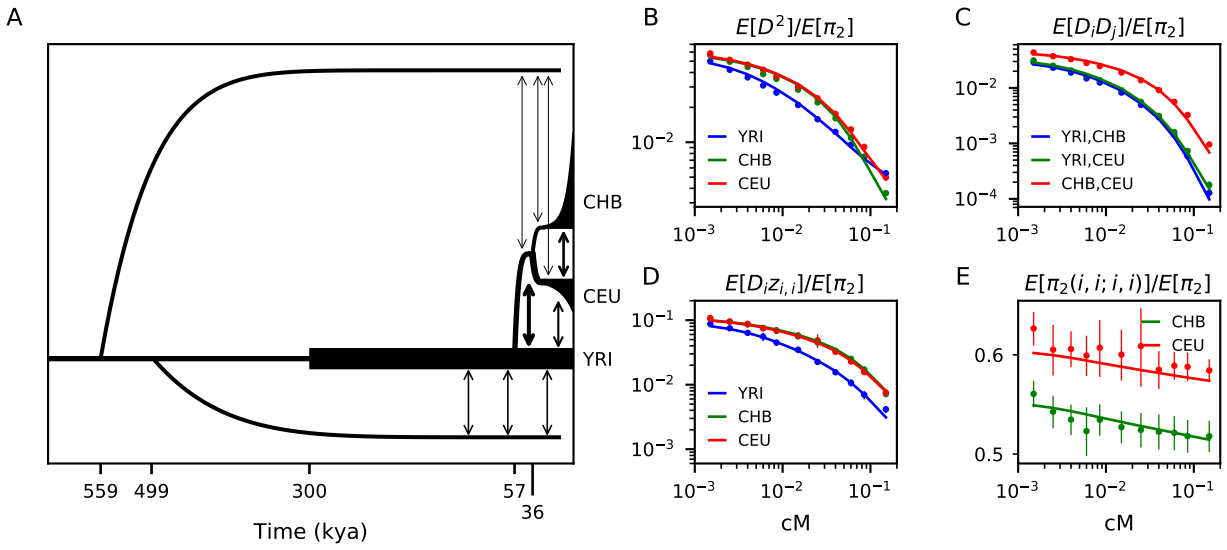

Figure 4: **Inferred OOA model with archaic admixture.** (A) We fit a model for out-of-Africa expansion related to the standard model in Figure 3(A). Demographic events for the three modern human populations are parameterized as above, but we also include two archaic branches with deep split from the ancestral population to modern humans. A putatively Neanderthal branch that remains isolated until the Eurasian split from YRI, and a deep archaic branch within Africa that is allowed to be isolated for some time before continuously exchanging migrants with the common ancestral branch and the YRI branch. (B-E) This model fits the data much better than the model without archaic admixture, and especially for the Dz statistics (D). Fits to 35 more curves and statistics are shown in Figure A5. The migration rates inferred between the archaic branch and YRI provides an estimate of $\sim 7.5\%$ archaic contribution.

Relation to other statistics

There are many approaches for computing expected statistics for diversity under a wide range of scenarios. Single-site statistics, which include expected heterozygosity and the AFS, may be computed efficiently using forward- or reverse-time approaches. Beyond the classical recursions for $\mathbb{E}[D]$ and $\mathbb{E}[D^2]$ (Hill and Robertson, 1968; Rogers, 2014), two-locus statistics are difficult to compute for non-equilibrium, multi-population demographic models. Sved (2009) proposed an IBD based recursion to compute $\mathbb{E}[r^2]$ across subdivided populations, but its accuracy and interpretation remain debated (Rogers, 2014).

The moments-based approach presented here generalizes the recursion for the single-site AFS presented in (Jouganous *et al.*, 2017). The moments system includes all heterozygosity statistics, so we recover expected F-statistics under arbitrary demography, which are commonly used to test for admixture (Reich *et al.*, 2009; Patterson *et al.*, 2012; Peter, 2016). Long-range patterns of elevated LD in putatively admixed populations are used to infer the timing of admixture events and relative contributions of parental populations (Moorjani *et al.*, 2011; Loh *et al.*, 2013). These approaches rely on the recursion for $\mathbb{E}[D]$ after admixture events that is used here (Equations 2 and 7). Thus the generalized Hill-Robertson system is sensitive to ancient admixture, but also captures statistics used to identify recent admixture history, with fewer assumptions about early history.

Plagnol and Wall (2006); Wall *et al.* (2009) introduced a statistic, S^*, specifically designed to scan for introgressed haplotypes without having sequence data from the putatively archaic population. S^* uses an ad-hoc score to identify SNPs that likely arose on haplotypes contributed from a deeply diverged population, and is estimated through simulation. These SNPs will tend to be rare and in high LD, and therefore also contribute to Dz. Thus even a small amount of archaic introgression will significantly elevate $\mathbb{E}[Dz]$ compared to that in an unadmixed population. Given its conceptual relationship to S^*, it may not be so surprising that this previously overlooked statistic is particularly well suited for model-based inference of archaic admixture.

Model	OOA (fit w/o Dz)		Archaic Introgression	
Parameter	Estimates	95% CI	Estimates	95% CI
N_0	2360	2190 – 2530	3600	2380 – 3920
N_{YRI}	13030	12200 – 13900	13900	12800 – 15000
N_{B}	1080	810 – 1350	880	670 – 1090
N_{CEU0}	1450	980 – 1920	2300	1810 – 2790
$r_{\text{CEU}}(\%)$	0.202	0.184 – 0.217	0.125	0.113 – 0.135
N_{CHB0}	410	340 – 480	650	540 – 750
$r_{\text{CHB}}(\%)$	0.498	0.445 – 0.531	0.372	0.333 – 0.398
$m_{\text{AF - B}}(\times 10^{-5})$	51.5	41.1 – 61.8	52.2	41.7 – 62.6
$m_{\text{YRI - CEU}}(\times 10^{-5})$	1.72	0.54 – 2.9	2.48	1.84 – 3.13
$m_{\text{YRI - CHB}}(\times 10^{-5})$	0	—	0	—
$m_{\text{CEU - CHB}}(\times 10^{-5})$	15.3	11.5 – 19.1	11.3	8.70 – 13.8
T_{AF} (kya)	208	196 – 220	300	277 – 323
T_{OOA} (kya)	65.7	52.4 – 79.0	60.7	50.3 – 64.2
$T_{\text{CEU - CHB}}$ (kya)	31.9	28.8 – 35.0	36.0	32.3 – 39.6
$T_{\text{Arch. Af. split}}$ (kya)			499	460 – 538
$T_{\text{Arch. Af. mig.}}$ (kya)			125	89.2 – 160
$m_{\text{AF - Arch. Af.}}(\times 10^{-5})$			1.98	1.15 – 2.82
$T_{\text{Nean. split}}$ (kya)			559	470 – 648
$m_{\text{OOA - Nean}}(\times 10^{-5})$			0.825	0.379 – 1.27
$T_{\text{Arch. adm. end}}$ (kya)			18.7	15.1 – 22.4

Table 1: **Inferred parameters for OOA models.** Two models for the out-of-Africa expansion. We fit the commonly used 13-parameter model to the multi-population Hill-Robertson statistics (left). The best fit parameters shown here were fit to the set of statistics without the $\mathbb{E}[Dz]$ terms, because the inclusion of those terms led to runaway parameter behavior in the optimization. This is often a sign of model mis-specification. On the right, the same 13-parameter model is augmented by the inclusion of two archaic branches, putatively Neanderthal and an unknown archaic African branch. We inferred that these branches split from the branch leading to modern humans roughly $500 - 700$ kya, and contributed migrants until quite recently ($\sim$14 kya). Times reported here assume a generation time of 29 years and are calibrated by the recombination (not mutation) rate.

Conclusion

We described an infinite hierarchy of multi-locus summaries of genomic diversity that are easy to compute under arbitrary, multi-population demographies. Some of these statistics are familiar, including expected heterozygosity, F-statistics, and LD decay, while others have been largely unexplored in multi-population models, such as the degree of LD between rare alleles (Dz) and the joint heterozygosity across sites and populations (π_2). The one-population Dz statistic, in particular, has an interesting history, as it has come up in early work as a mathematical stepping-stone on the way to computing D^2 (Hill and Robertson, 1968), but was, to our knowledge, never used in data analysis. As it happens, this 'ghost' statistic provides a unique window into human history.

Using this set of summary statistics, we explored a commonly used model of human demographic history derived from single-site AFS and validated using LD decay curves. While many statistics under this model fit the data well, the model dramatically underestimates levels of LD among rare alleles. Modeling archaic introgression worldwide resolved this discrepancy. We recovered the signal of Neanderthal introgression in Eurasian populations, and found evidence for substantial and long-lasting archaic admixture in two African populations.

This model deserve a more thorough investigation, including data from archaic humans and additional contemporary African populations. We leave this to future work for three reasons. First, proposing a detailed multi-population model of evolution in Africa will require carefully incorporating anthropological and archaeological evidence, which is a substantial endeavor. Second, the inclusion of two-locus statistics from ancient genomes will require vetting possible biases associated with ancient DNA sequencing, although

we see no problem with using two-locus statistics in modern populations jointly with one-locus statistics in ancient DNA.

Third, and more importantly, archaic admixture can hide in the blind spot of classical statistics, and widely used demographic models for simulating genomes underestimate LD between rare variants in populations around the globe, especially in Africa. This large bias affects neither the distribution of allele frequencies nor the amount of correlation measured by D^2, but it may impact analyses aiming to identify disease variants based on overrepresentation of rare variants in specific genes or pathways. Thus both statistical and population geneticists would benefit from including archaic admixture into baseline models of human genomic diversity.

Acknowledgements

The authors thank Brenna Henn, Mathias Steinrücken, and Chris Gignoux for useful discussions. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program, the NSERC discovery grant, and CIHR MOP-136855.

References

Auton, A. and G. McVean, 2007 Recombination rate estimation in the presence of hotspots. Genome Research **17**: 1219–1227.

Berger, L. R., J. Hawks, P. H. Dirks, M. Elliott, and E. M. Roberts, 2017 Homo naledi and Pleistocene hominin evolution in subequatorial Africa. eLife **6**: 1–19.

Browning, S. R., B. L. Browning, Y. Zhou, S. Tucci, and J. M. Akey, 2018 Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell **173**: 53–61.e9.

Cavalli-Sforza, L. L. and Bodmer, 1971 *The genetics of human populations*. W. H. Freeman and Company.

Chan, A. H., P. A. Jenkins, and Y. S. Song, 2012 Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. PLoS Genetics **8**.

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.

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

Donnelly, P. and T. G. Kurtz, 1999b Particle Representations for Measure-Valued Population Models. The Annals of Probability **27**: 166–205.

Durvasula, A. and S. Sankararaman, 2018 Recovering signals of ghost archaic admixture in the genomes of present-day Africans. bioRxiv p. 285734.

Ethier, S. N. and R. C. Griffiths, 1990 On the two-locus sampling distribution. Journal of Mathematical Biology **29**: 131–159.

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

Golding, G. B., 1984 The sampling distribution of linkage disequilibrium. Genetics **108**: 257–274.

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.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genetics **5**: e1000695.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, and J. D. Wall, 2011 Genetic evidence for archaic admixture in Africa. Proceedings of the National Academy of Sciences **108**: 15123–15128.

Hill, W. G. and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genetical Research **8**: 269.

Hill, W. G. and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theoretical and applied genetics **38**: 226–31.

Hinch, A. G., A. Tandon, N. Patterson, Y. Song, N. Rohland, *et al.*, 2011 The landscape of recombination in African Americans. Nature **476**: 170–175.

Hsieh, P. H., A. E. Woerner, J. D. Wall, J. Lachance, S. A. Tishkoff, *et al.*, 2016 Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. Genome Research **26**: 291–300.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159**: 1805–1817.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel, 2017 Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. Genetics **206**: 1549–1567.

Kamm, J. A., J. P. Spence, J. Chan, and Y. S. Song, 2016 Two-locus likelihoods under variable population size and fine-scale recombination rate estimation. Genetics **203**: 1381–1399.

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.

Karlin, S. and J. McGregor, 1968 Rates and probabilities of fixation for two locus random mating finite populations without selection. Genetics **58**: 141–159.

Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLoS Computational Biology **12**: 1–22.

Kimura, M., 1955 Random Genetic Drift in Multi-Allelic Locus. Evolution **9**: 419–435.

Kuhlwilm, M., I. Gronau, M. J. Hubisz, C. de Filippo, J. Prado-Martinez, *et al.*, 2016 Ancient gene flow from early modern humans into eastern neanderthals. Nature **530**: 429–433.

Lachance, J., B. Vernot, C. C. Elbers, B. Ferwerda, A. Froment, *et al.*, 2012 Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell **150**: 457–469.

Lan, T., H. Lin, W. Zhu, T. C. A. M. Laurent, M. Yang, *et al.*, 2017 Deep whole-genome sequencing of 90 Han Chinese genomes. GigaScience **6**: 1–7.

Li, H. and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. Nature **475**: 493–496.

Loh, P. R., M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, *et al.*, 2013 Inferring admixture histories of human populations using linkage disequilibrium. Genetics **193**: 1233–1254.

Lukic, 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.

Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry, 2004 The Allele Frequency Spectrum in Genome-Wide Human Variation Three Large World Populations. Genetics **372**: 351–372.

McVean, G. A. T., 2002 A genealogical interpretation of linkage disequilibrium. Genetics **162**: 987–991.

McVean, G. A. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley, *et al.*, 2004 The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. Science **304**: 581–584.

Moorjani, P., N. Patterson, J. N. Hirschhorn, A. Keinan, L. Hao, *et al.*, 2011 The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLoS Genetics **7**: e1001373.

Ohta, T. and M. Kimura, 1969a Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics **63**: 229–238.

Ohta, T. and M. Kimura, 1969b Linkage disequilibrium due to random genetic drift. Genetical Research **13**: 47.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, *et al.*, 2012 Ancient admixture in human history. Genetics **192**: 1065–1093.

Peter, B. M., 2016 Admixture, population structure, and f-statistics. Genetics **202**: 1485–1501.

Plagnol, V. and J. D. Wall, 2006 Possible ancestral structure in human populations. PLoS Genetics **2**: e105.

Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, *et al.*, 2014 The complete genome sequence of a Neanderthal from the Altai Mountains. Nature **505**: 43–49.

Ragsdale, A. P. and R. N. Gutenkunst, 2017 Inferring Demographic History Using Two-Locus Statistics. Genetics **206**: 1037–1048.

Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh, 2009 Reconstructing Indian population history. Nature **461**: 489–494.

Rogers, A. R., 2014 How population growth affects linkage disequilibrium. Genetics **197**: 1329–1341.

Schiffels, S. and R. Durbin, 2014 Inferring human population size and separation history from multiple genome sequences. Nature Genetics **46**: 919–925.

Song, Y. S. and J. S. Song, 2007 Analytic computation of the expectation of the linkage disequilibrium coefficient r2. Theoretical Population Biology **71**: 49–60.

Sved, J. A., 2009 Correlation measures for linkage disequilibrium within and between populations. Genetics Research **91**: 183–192.

Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, *et al.*, 2012 Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science **337**: 64–69.

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

Villanea, F. A. and J. G. Schraiber, 2018 Multiple episodes of interbreeding between Neanderthal and modern humans. Nature Ecology & Evolution p. 1.

Wall, J. D. and D. Y. C. Brandt, 2016 Archaic admixture in human history. Current Opinion in Genetics and Development **41**: 93–97.

Wall, J. D., K. E. Lohmueller, and V. Plagnol, 2009 Detecting ancient admixture and estimating demographic parameters in multiple human populations. Molecular Biology and Evolution **26**: 1823–1827.

Weir, B. S., 1979 Inferences about Linkage Disequilibrium. Biometrics **35**: 235–254.

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 **200**: 601–617.

A Appendix

A.1 Computing moment equations

Here we describe a systematic approach for computing transition functions for a system of statistics that include $\mathbb{E}[D^m]$. For any term $\mathbb{E}[f(D,p,q)]$, we change variables to the space of haplotype frequencies. c_1 is the frequency or count of type AB haplotype, c_2 for Ab, c_3 for aB, and c_4 for ab, so $D = c_1c_4 - c_2c_3$, $p = c_1 + c_2$, $q = c_1 + c_3$. We then compute transition probabilities on the monomial expansion of this transformation in 'c'-space, and then change variables back to (p,q,D)-space and simplify.

For example, for $\mathbb{E}[D]$, we transform the expectation to

$$\mathbb{E}[D] = \mathbb{E}[c_1c_4 - c_2c_3] = \mathbb{E}[c_1c_4] - \mathbb{E}[c_2c_3].$$

Then for each of the expectations, we calculate its change over one generation due to drift, recombination, mutation, or migration. For example, in the case of drift we find expectations after one generation by considering copying probabilities (described in section A.1.1). Then

$$\mathbb{E}[c_1c_4]_{t+1} = \left(1 - \frac{1}{2N(t)}\right)\mathbb{E}[c_1c_4]_t \tag{A1}$$

and

$$\mathbb{E}[c_2c_3]_{t+1} = \left(1 - \frac{1}{2N(t)}\right)\mathbb{E}[c_2c_3]_t. \tag{A2}$$

We then convert back to (p,q,D)-space and simplify, obtaining

$$\mathbb{E}[D]_{t+1} = \mathbb{E}[c_1c_4]_{t+1} - \mathbb{E}[c_2c_3]_{t+1} = \left(1 - \frac{1}{2N(t)}\right)\mathbb{E}[D]_t. \tag{A3}$$

Throughout, we assume populations are in Hardy-Weinberg equilibrium and randomly mating. In addition, we assume that the recombination rate r, migration rates m_{ij}, and mutation rate u are small enough, and population sizes N_i large enough, so that the product of r, m_{ij}, u, and $\frac{1}{N_i}$ may be ignored. In other words, we assume copying, recombination, mutation, and migration rates are small enough so that at most a single event occurs within the $n \ll N$ tracked lineages in a given generation.

A.1.1 Drift

We compute how $\mathbb{E}[c_1^{k_1}c_2^{k_2}c_3^{k_3}c_4^{k_4}]$ is expected to change in a given generation under the action of drift. We let $k = \sum_{i=1}^4 k_i$. We imagine tracking k lineages in the population, each of which is one of the four haplotypes. The expectation is proportional to the probability that k_1 of the lineages are in state c_1, k_2 are of type c_2, etc. We assume $k \ll N(t)$ so that the probability of a drift event (where one lineage copies itself over another within the k tracked lineages) is small, and we may assume that at most one such event occurs in any given generation.

The probability of a copying event among k lineages in generation t is

$$P(\text{copying event}) = \frac{1}{2N(t)}\binom{k}{2}. \tag{A4}$$

Given that a drift even occurs, we can compute the probability of any possible transition. For example,

$$\begin{aligned} P((k_1,k_2,k_3,k_4) \to (k_1+1,k_2-1,k_3,k_4)) &= P(\text{choose one of each } c_1 \text{ and } c_2) \cdot P(c_1 \text{ copies over } c_2) \\ &= \frac{1}{2}\frac{\binom{k_1}{1}\binom{k_2}{1}}{\binom{k}{2}} \\ &= \frac{k_1k_2}{k(k-1)}. \end{aligned} \tag{A5}$$

We combine all possible copying events and find $\mathbb{E}[c_1^{k_1}c_2^{k_2}c_3^{k_3}c_4^{k_4}]_{t+1}$ under drift

$$\mathbb{E}\left[c_1^{k_1}c_2^{k_2}c_3^{k_3}c_4^{k_4}\right]_{t+1} = (1 - P(\text{copying event}))\mathbb{E}\left[\cdot c_1^{k_1}c_2^{k_2}c_3^{k_3}c_4^{k_4}\right] + P(\text{copying event})\mathbb{E}\left[\cdot c_1^{k_1}c_2^{k_2}c_3^{k_3}c_4^{k_4} \cdot \sum_{1\le i,j\le 4}\left(\frac{k_i k_j}{k(k-1)}\frac{c_i}{c_j}\delta_{i\neq j} + \frac{k_i(k_i-1)}{k(k-1)}\delta_{i=j}\right)\right]. \tag{A6}$$

We wrote the D^2 drift matrix in the main text. The system of statistics for D^3 is

$$\mathbf{y} = \begin{pmatrix} \mathbb{E}[D^3] \\ \mathbb{E}[D^2 z] \\ \mathbb{E}[D\pi_2] \\ \mathbb{E}[\pi z] \\ \mathbb{E}[D\sigma_1] \\ \mathbb{E}[z\sigma_1] \\ \mathbb{E}[D[\\ \mathbb{E}[z] \end{pmatrix},$$

where $\sigma_1 = p(1-p) + q(1-q)$, and has transition matrix

$$\mathcal{D} = \frac{1}{2N}\begin{pmatrix} -6 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & -11 & 16 & 1 & -4 & 0 & 1 & 0 \\ 0 & 1 & -11 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 36 & -6 & -6 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -6 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 12 & -3 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{pmatrix}.$$

For the D^4 system, with terms

$$\mathbf{y} = \begin{pmatrix} \mathbb{E}[D^4] \\ \mathbb{E}[D^3 z] \\ \mathbb{E}[D^2\pi_2] \\ \mathbb{E}[D\pi z] \\ \mathbb{E}[\pi_2^2] \\ \mathbb{E}[D^2\sigma_1] \\ \mathbb{E}[Dz\sigma_1] \\ \mathbb{E}[\pi_2\sigma_1] \\ \mathbb{E}[\sigma_2] \\ \mathbb{E}[D^2] \\ \mathbb{E}[Dz] \\ \mathbb{E}[\pi_2] \\ \mathbb{E}[\sigma_1] \end{pmatrix},$$

the transition matrix is

$$\mathcal{D} = \frac{1}{2N}\begin{pmatrix}
-10 & 6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & -18 & 48 & 3 & 0 & -12 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\
0 & 1 & -21 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 36 & -19 & 0 & -6 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 4 & -12 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -12 & 1 & 1 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 12 & -12 & 0 & 0 & -4 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & -7 & 0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -6 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & -5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix}.$$

In general, the block upper diagonal structure reflects the hierarchy of the moments system (Figure A1), and the sparseness allows for rapid integration.

A.1.2 Recombination

Recombination changes D in one generation at rate directly proportional to the recombination distance between the two loci, and so the haplotype frequencies change over one generations according to

$$\begin{aligned}
c_1' &= c_1 - rD \\
c_2' &= c_2 + rD \\
c_3' &= c_3 + rD \\
c_4' &= c_4 - rD.
\end{aligned}$$

Then $\mathbb{E}[\prod c_i^{k_i}]_{t+1} = \mathbb{E}[\prod c_i'^{k_i}]$. For any moments $\mathbb{E}[D^\alpha f(p,q)]$, this simplifies to the recursion

$$\mathbb{E}[D^\alpha f(p,q)]_{t+1} = (1-\alpha r)\mathbb{E}[D^\alpha f(p,q)]_t. \tag{A7}$$

A.1.3 Mutation

In the infinite-sites model, mutations are assumed to occur once at any given locus (i.e., no recurrent or back mutation). Recurrent and reversible mutations are described in the next section. A two-locus pair that is observed to be polymorphic at both loci must have experienced a mutation first at one locus, and then a second to occur at the paired locus. Thus, while $\mathbb{E}[H] = \mathbb{E}[2p(1-p)] \propto \theta$, the joint heterozygosity $\mathbb{E}[p(1-p)q(1-q)] \propto \theta^2$.

In the Hill-Robertson system, a mutation can "create" one-locus diversity from an invariant sites, and create two-locus diversity from one-locus diversity:

$$\begin{aligned}
\Delta_{\mathcal{U}}\mathbb{E}[p(1-p)] &= \frac{\theta}{2}, \\
\Delta_{\mathcal{U}}\mathbb{E}[p(1-p)q(1-q)] &= \frac{\theta}{2}\mathbb{E}[p(1-p)] + \frac{\theta}{2}\mathbb{E}[q(1-q)],
\end{aligned}$$

where, $\Delta_{\mathcal{U}}$ is used to denote $\mathbb{E}[\cdot]_{t+1} - \mathbb{E}[\cdot]_t$ due to mutation. Other terms in the Hill-Robertson system are unchanged by mutation.

For higher order systems, we have terms of the form $\mathbb{E}[\pi_2\sigma_i] = \mathbb{E}[p(1-p)q(1-q)(p^i(1-p)^i + q^i(1-q)^i)]$, which change due to mutation as

$$\Delta_{\mathcal{U}}\mathbb{E}[p(1-p)q(1-q)(p^i(1-p)^i + q^i(1-q)^i)] = \frac{\theta}{2}\mathbb{E}[p^i(1-p)^i + q^i(1-q)^i].$$

Here, we have assumed that the mutation rates are equal at the left and right locus, although this approach allows for differing mutation rates between the two loci, as in Ohta and Kimura (1969a).

Reversible mutations are also handled in a similarly straightforward manner. Under reversible mutations, we have the same terms as above, and each moment in the system also decays as

$$\Delta_{\mathcal{U}}\mathbb{E}[f(p,q,D)] = -k\frac{\theta}{2}\mathbb{E}[f(p,q,D)],$$

where k is the sample-size order of the term (that is, the number of sampled haplotypes required to estimate that term, as described in the Drift section above) (Ohta and Kimura, 1969a).

A.1.4 Selection in the Hill-Robertson system

While the Hill-Robertson system closes under drift and recombination, the D^2 system (and all other orders) does not close under selection. Here, we consider a simple selection model, with additive selection acting on the A allele at the left locus (with selection strength s, $|s| \ll 1$), while the right locus remains neutral. Thus, selection acts for or against AB and Ab haplotypes, with strength $1+s$ relative to aB and ab haplotypes. This selection model is relevant to computing the expect LD between a selected site and a neutral marker separated by recombination distance r.

In this setting, we compute how selection is expected to change D^2, $D(1-2p)(1-2q)$, and $p(1-p)q(1-q)$, in expectation. Again we change variables to c-space and compute how selection is expected to change terms in the monomial expansion. We denote c_i' to be the expected frequency of type i after one generation. For example,

$$\begin{aligned} c_1' &= \frac{c_1(1+s)}{(c_1+c_2)(1+s)+c_3+c_4} \\ &= \frac{c_1(1+s)}{1+(c_1+c_2)s} \\ &\approx c_1(1+s)(1-(c_1+c_2)s) \approx c_1(1+(1-p)s), \end{aligned} \tag{A8}$$

to first order in s. Similarly,

$$\begin{aligned} c_2' &\approx c_2(1+(1-p)s), \\ c_3' &\approx c_3(1-ps), \\ c_4' &\approx c_4(1-ps). \end{aligned}$$

Then in one generation,

$$\begin{aligned} \mathbb{E}\left[{c_1'}^{k_1}{c_2'}^{k_2}{c_3'}^{k_3}{c_4'}^{k_4}\right] &\approx {c_1}^{k_1}{c_2}^{k_2}{c_3}^{k_3}{c_4}^{k_4}\left(1+k_1(1-p)s+k_2(1-p)s-k_3ps-k_4ps\right) \\ &\approx {c_1}^{k_1}{c_2}^{k_2}{c_3}^{k_3}{c_4}^{k_4}\left(1+(k_1+k_2-kp)s\right), \end{aligned}$$

where $k = \sum_i k_i$.

In one generation, the change in moments $\Delta_{\mathcal{S}}\mathbb{E}[\cdot] = \mathbb{E}[\cdot]_{t+1} - \mathbb{E}[\cdot]_t$ due to selection is

$$\begin{aligned} \Delta_{\mathcal{S}}\mathbb{E}[D^2] &= 2s\mathbb{E}[D^2(1-2p)] \\ \Delta_{\mathcal{S}}\mathbb{E}[D(1-2p)(1-2q)] &= -2s\mathbb{E}[D^2(1-2p)] + \frac{3}{2}s\mathbb{E}[D(1-2p)^2(1-2q)] - \frac{s}{2}\mathbb{E}[D(1-2p)] \\ \Delta_{\mathcal{S}}\mathbb{E}[p(1-p)q(1-q)] &= s\mathbb{E}[p(1-p)q(1-q)] + s\mathbb{E}[Dp(1-p)(1-2q)], \end{aligned}$$

Thus we need to include additional terms: $\mathbb{E}[D^2(1-2p)]$, $\mathbb{E}[D(1-2p)^2(1-2q)]$, etc. These additional terms require their own additional terms, so that this system grows and includes all terms (in expectation)

$$\begin{pmatrix} D^2p^i(1-p)^i(1-2p)^j \\ Dp^i(1-p)^i(1-2p)^j(1-2q) \\ p^i(1-p)^i(1-2p)^jq(1-q) \end{pmatrix}, \forall i,j \geq 0.$$

To solve this system, we will require a moments closure approximation. This could be achieved, for example, by approximating $D^2p^i(1-p)^i(1-2p)^j$ as a linear combination of terms of lower order, $\{D^2p^{i-k}(1-p)^{i-k}(1-2p)^{j-l}\}$, through a jackknife approximation. Alternatively, we could choose some values i and j to truncate the system and approximate the necessary higher order terms as $\mathbb{E}[f(p,q,D)]$ as $\mathbb{E}[g(p,q,D)]\mathbb{E}[h(p,q,D)]$ where $gh=f$, which leads to a nonlinear system of ODEs. While beyond the scope of this paper, work is ongoing to assess accuracy of closure approximations and incorporate models of selection into multi-population LD models.

A.2 Multiple populations

In this section, we describe the multi-population basis analogous to the Hill-Robertson system for a single populations. We derive recursion equations for the multi-population basis under migration and admixture events.

A.2.1 Population splits

Consider a single populations (denoted 0) that splits into two populations (denoted 1 and 2). In the time of the population split (t_0), expected two-locus statistics in populations 1 and 2 are equal to those in population 0. This can be seen by considering the probability of sampling haplotypes in the two split populations. We compute terms of the form $\prod_{j=1}^{2}\prod_{i=1}^{4} {c_{j,i}}^{k_{j,i}}$, where $c_{j,i}$ denotes the probability of sampling haplotype i in population j. We observe that at the time of the split, $c_{j,i}=c_{0,i}$, since expected haplotype frequencies in the split populations are equal to expected haplotype frequencies in the parental population. Then,

$$\prod_{j=1}^{2}\prod_{i=1}^{4} {c_{j,i}}^{k_{j,i}} = \prod_{j=1}^{2}\prod_{i=1}^{4} {c_{0,i}}^{k_{j,i}} = \prod_{i=1}^{4} {c_{0,i}}^{k_{1,i}+k_{2,i}}. \tag{A9}$$

Thus

$$\mathbb{E}[D_1^2]_{t_0} = \mathbb{E}[D_2^2]_{t_0} = \mathbb{E}[D_0^2]_{t_0},$$
$$\mathbb{E}[D_1(1-2p_1)(1-2q_1)]_{t_0} = \mathbb{E}[D_2(1-2p_2)(1-2q_2)]_{t_0} = \mathbb{E}[D_0(1-2p_0)(1-2q_0)]_{t_0},$$

and so on. Additionally, we consider $\mathbb{E}[D_1D_2]$, the covariance of D across populations 1 and 2, which initially is

$$\begin{aligned}
\mathbb{E}[D_1D_2]_{t_0} &= \mathbb{E}[(c_{1,1}c_{1,4}-c_{1,2}c_{1,3})(c_{2,1}c_{2,4}-c_{2,2}c_{2,3})]_{t_0} \\
&= \mathbb{E}[c_{1,1}c_{1,4}c_{2,1}c_{2,4}]_{t_0} - \mathbb{E}[c_{1,1}c_{1,4}c_{2,2}c_{2,3}]_{t_0} - \mathbb{E}[c_{1,2}c_{1,3}c_{2,1}c_{2,4}]_{t_0} + \mathbb{E}[c_{1,2}c_{1,3}c_{2,2}c_{2,3}]_{t_0} \\
&= \mathbb{E}[c_{0,1}^2c_{0,4}^2]_{t_0} - 2\mathbb{E}[c_{0,1}c_{0,2}c_{0,3}c_{0,4}]_{t_0} + \mathbb{E}[c_{0,2}^2c_{0,3}^2]_{t_0} \\
&= \mathbb{E}[D_0^2]_{t_0}.
\end{aligned}$$

In the absence of migration, the D^2-systems in each of the populations evolves according to the Hill-Robertson equations, while

$$\mathbb{E}[D_1D_2]_{t+1} = \left(1-\frac{1}{2N_1}-\frac{1}{2N_2}-2r\right)\mathbb{E}[D_1D_2]_t. \tag{A10}$$

A.2.2 Migration

With the inclusion of migration, additional moments are needed to obtain a closed system. We write the full basis with migration as

$$\mathbf{z} = \begin{pmatrix} \mathbb{E}[D_iD_j] \\ \mathbb{E}[D_iz_{j,k}] \\ \mathbb{E}[\pi_2(i,j;k,l)] \\ \mathbb{E}[H_{i,j}] \end{pmatrix},$$

where i,j,k,l index populations, and

$$D_iz_{j,k} = D_i(1-2p_j)(1-2q_k),$$

$$\pi_2(i,j;k,l) = \begin{cases} p_i(1-p_i)q_k(1-q_k), & i=j, k=l \\ \frac{1}{2}p_i(1-p_j)q_k(1-q_k) + \frac{1}{2}p_j(1-p_i)q_k(1-q_k), & i\neq j, k=l \\ \frac{1}{2}p_i(1-p_i)q_k(1-q_l) + \frac{1}{2}p_i(1-p_i)q_l(1-q_k), & i=j, k\neq l \\ \frac{1}{4}p_i(1-p_j)q_k(1-q_l) + \frac{1}{4}p_i(1-p_j)q_l(1-q_k) & \\ \quad + \frac{1}{4}p_j(1-p_i)q_k(1-q_l) + \frac{1}{4}p_j(1-p_i)q_l(1-q_k), & i\neq j, k\neq l \end{cases},$$

$$H_{i,j} = \begin{cases} p_i(1-p_i), & i=j \\ \frac{1}{2}p_i(1-p_j) + \frac{1}{2}p_j(1-p_i), & i\neq j \end{cases}.$$

Initial values for each cross term are found using Equation A9 in the above section Population splits.

The number of terms in the system grows in the number P of populations, as $P^3 + \left(\frac{P(P+1)}{2}\right)^2 + 3\cdot\frac{P(P+1)}{2}$. While the size of the full joint AFS grows exponentially in P, the multi-population Hill-Robertson system remains manageable for even large P: for example, the 10-population system has 4,190 terms, which is not at all computationally burdensome for a sparse, linear system. When the left and right loci have equal mutation rate, as we assume in most of our models, redundant terms exist in this basis so the system may be further reduced in size.

We want to know how migration changes expected statistics in our system. We again work in c-space, where c_{i1} represents the haplotype AB in population i, c_{i2} represents Ab in population i, c_{i3} represents aB in i, and c_{i4} represents ab in i. We consider terms of the general form

$$\mathbb{E}\left[\prod_{i=1}^{P}\prod_{l=1}^{4} c_{il}^{k_{il}}\right],$$

and set $k = \sum_{i,l} k_{il}$ as the total number of tracked lineages in our subsample across all populations, corresponding to the sample size (order) of the moment.

We denote migration rates m_{ij} as the probability that a lineage in population j is replaced by a migrant lineage from population i. We assume $m_{ij} \ll 1$ so that at most a single migrant replaces a lineage among the sample of size $k \ll N_j$ in population j in any generation. Then in one generation, the change due to migration is

$$\Delta_{\mathcal{M}}\mathbb{E}\left[\prod_{i=1}^{P}\prod_{l=1}^{4} c_{il}^{k_{il}}\right]_t = \sum_{i=1}^{P}\sum_{j=1}^{P}\delta_{i\neq j}m_{ij}\left(\sum_{l=1}^{4}\frac{k_{il}(c_{il}-c_{jl})}{c_{jl}}\right). \tag{A11}$$

In words, we sum over each pair of populations i, j and consider the probability that lineages of each type migrate from population i and copies over a lineage within our sample in population j.

As a simple example, consider the expected frequency at a single locus p_1 in population 1 (among P total populations). Using the above formula, after one generation we find

$$\Delta_{\mathcal{M}}\mathbb{E}[p_1] = \sum_{i=2}^{P} m_{i1}(\mathbb{E}[p_i] - \mathbb{E}[p_1]).$$

Expected changes due to migration may be found for each term in the system.

For two populations, with migration rates m_{12} from population 1 to 2, and m_{21} from population 2 to 1,

on the terms (in expectation) in the system

$$\mathbf{z} = \begin{pmatrix} D_1^2 \\ D_1 D_2 \\ D_2^2 \\ D_1 z_{1,1} \\ D_1 z_{1,2} \\ D_1 z_{2,1} \\ D_1 z_{2,2} \\ D_2 z_{1,1} \\ D_2 z_{1,2} \\ D_2 z_{2,1} \\ D_2 z_{2,2} \\ \pi_2(1,1;1,1) \\ \pi_2(1,1;1,2) \\ \pi_2(1,1;2,2) \\ \pi_2(1,2;1,1) \\ \pi_2(1,2;1,2) \\ \pi_2(1,2;2,2) \\ \pi_2(2,2;1,1) \\ \pi_2(2,2;1,2) \\ \pi_2(2,2;2,2) \\ H_{1,1} \\ H_{1,2} \\ H_{2,2} \end{pmatrix},$$

we obtain the migration operator

$$\mathcal{M} = m_{12} \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -2 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -2 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & -3 & 0 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & -4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 & 0
\end{pmatrix}$$

$$+ m_{21} \begin{pmatrix}
-2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -3 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 4 & -4 & 0 & -4 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

The two-population drift operator in the migration basis is

$$
\mathcal{D} = \frac{1}{2N_1}\begin{pmatrix}
-3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & -5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
+ \frac{1}{2N_2}\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{pmatrix}.
$$

For any number of populations, we have the closed recursion

$$
\mathbf{z}_{t+1} - \mathbf{z}_t = \left(\mathcal{D}_{\mathbf{N}} + \mathcal{M}_{\mathbf{m}} + \mathcal{R}_r + \mathcal{U}_u\right)\mathbf{z}_t, \tag{A12}
$$

where $\mathcal{D}, \mathcal{R}, \mathcal{U}$, and $\mathcal{M}$ are sparse matrices for drift, recombination, mutation, and migration, with population sizes $\mathbf{N}$ and migration rates $\mathbf{m}$.

A.2.3 Admixture

We find the expected value for terms in basis (5) that include a new admixed population. We assume the admixed population arises from two parental populations, with lineages arriving from the two populations with probability f and $1-f$. Computing terms involving the admixed population simply requires enumerating over the haplotype probabilities of lineages arising from each parental population, which have been computed up to the time of admixture, and then simplifying.

Suppose an admixture event between populations 1 (with proportion f) and 2 (with proportion $1-f$), with P total populations, forms a new population indexed $P+1$. Expected haplotype frequencies in the admixed population are given by f and the expected haplotype frequencies in the source populations. For example, $c_{P+1,1}$ (c_1 in the admixed population) equal to $fc_{1,1} + (1-f)c_{2,1}$, $c_{P+1,2} = fc_{1,2} + (1-f)c_{2,2}$, and so on. Cavalli-Sforza and Bodmer (1971) (page 69) used this approach to compute $\mathbb{E}[D]$ in an admixed population (Equation 7).

For any statistic $f(\mathbf{p}, \mathbf{q}, \mathbf{D})$, we first convert to c-space variables. Then for each term in the c-space expansion $\prod_{i=1}^{P+1}\prod_{l=1}^{4} {c_{i,l}}^{k_{i,l}}$, where $c_{i,l}$ represents haplotype l in population i, we compute

$$\prod_{i=1}^{P+1}\prod_{l=1}^{4} {c_{i,l}}^{k_{i,l}} = \left(\prod_{i=1}^{P}\prod_{l=1}^{4} {c_{i,l}}^{k_{i,l}}\right) \left(fc_{1,1} + (1-f)c_{2,1}\right)^{k_{P+1,1}} \left(fc_{1,2} + (1-f)c_{2,2}\right)^{k_{P+1,2}} \cdot \left(fc_{1,3} + (1-f)c_{2,3}\right)^{k_{P+1,3}} \left(fc_{1,4} + (1-f)c_{2,4}\right)^{k_{P+1,4}}.$$

We then convert back to the $(\mathbf{p}, \mathbf{q}, \mathbf{D})$ variables and simplify. If a term does not include any term with p_{P+1}, q_{P+1} or D_{P+1}, it remains remains unchanged. Otherwise, the term can be written as a linear combination of terms in the P-population basis.

For example, with two populations, with contribution f from population 1 and $1-f$ from population 2,

$$\begin{aligned}\mathbb{E}[D^2_{\text{adm}}] =& f^2\mathbb{E}[D_1^2] + (1-f)^2\mathbb{E}[D_2^2] + 2f(1-f)\mathbb{E}[D_1 D_2] \\ &+ 2f^2(1-f)\mathbb{E}[D_1\delta] + 2f(1-f)^2\mathbb{E}[D_2\delta] + f^2(1-f)^2\mathbb{E}[\delta^2].\end{aligned} \tag{A13}$$

$\mathbb{E}[D_i\delta]$ and $\mathbb{E}[\delta^2]$ can be written as linear combinations of terms in the multi-population basis (5). Similar equations exist for each new term in the basis with the additional admixed population, and this system can then be integrated forward in time using Equation 6.

A.3 Haplotype frequency spectrum

The allele frequency spectrum (AFS) is the distribution of allele counts in a sample of size n, denoted Φ_n. Because Φ_n is sensitive to demographic and evolutionary processes, it is widely used to infer demographic history, patterns of selection, and mutation rates. While the coalescent computes expectations of the entries of Φ_n looking backward in time through the branch lengths of a genealogy, forward in time approaches solve the underlying diffusion equation, which describes the time-evolution of the distribution of allele frequencies in one or more populations. In one populations, this takes the form

$$\frac{\partial\phi}{\partial t} = \frac{1}{2N}\frac{\partial^2}{\partial x^2}x(1-x)\phi - s\frac{\partial}{\partial x}(h + (1-2h)x)x(1-x)\phi, \tag{A14}$$

where N is the effective population size and can change over time. Φ_n can then be found by integrating ϕ against the binomial sampling distribution.

Analytic solutions to Equation A14 have only been found for simple scenarios, such as steady-state solutions. To compute Φ_n for non-equilibrium demography, we turn to numerical solutions, e.g. (Evans *et al.*, 2007; Gutenkunst *et al.*, 2009; Lukic and Hey, 2012). Recently, Jouganous *et al.* (2017) recognized that the entries of Φ_n themselves comprise a moments system that allows for direct integration of Φ_n without having to numerically solve Equation A14. This system closes under drift, while selection requires a moment-closure approximation.

The two-locus frequency spectrum (Ψ_n) is defined similarly to the single locus AFS, but instead tracks the haplotype frequencies of two-locus pairs. We consider a model that permits two alleles at each of the two loci: A/a at the left locus, and B/b at the right. Four haplotypes are possible in the two-locus model (AB, Ab, aB, ab) whose frequencies in the population sum to one. Then $\Psi_n(i,j,k)$ is the expected number of two-locus pairs in a sample of size n in which we observe i copies of type AB, j of type Ab, k of type aB, and $n-i-j-k$ of type ab.

The two-locus frequency spectrum can be found at steady-state using a recursion due to Golding (1984). For non-equilibrium demography, Kamm *et al.* (2016) presented a coalescent approach for Ψ_n via a Moran model closely related to the look-down approach of Donnelly and Kurtz (1999a). Alternatively, Ψ_n can be found by first solving the associated two-locus diffusion equation ψ (Kimura, 1955; Hill and Robertson, 1966), the density of two-locus haplotype frequencies in the population,

$$\frac{\partial\psi}{\partial t} = \frac{1}{2}\sum_{1\le i,j\le 3}\sum\frac{\partial^2}{\partial x_i\partial x_j}\frac{x_i(\delta_{i,j}-x_j)\psi}{N(t)} + \frac{\rho}{2}\left(\frac{\partial}{\partial x_1}D\psi - \frac{\partial}{\partial x_2}D\psi - \frac{\partial}{\partial x_3}D\psi\right), \tag{A15}$$

shown here without terms for selection and $D = x_1(1 - x_1 - x_2 - x_3) - x_2x_3$, and then integrating ψ against the multinomial sampling distribution. Ragsdale and Gutenkunst (2017) solved Equation A15 using finite differences, which they used in single-population demographic inference. The advantage to the diffusion approach is that selection is easily incorporated at one or both loci, which allows us to directly model the effect of linked selection with any recombination recombination rate and non-equilibrium demography.

Here, instead of finding a numerical solution to Equation A15, we show that we can directly solve for Ψ_n through a recursion on its entries just as Jouganous *et al.* (2017) proposed for the single-locus AFS. By considering how haplotype frequencies within n tracked lineages are expected to change due to drift, recombination, selection, and mutation, we obtain the recursion

$$\Psi_n^{t+1}(i,j,k) - \Psi_n^t(i,j,k) = \mathcal{D}_{2N(t),n;i,j,k}\Psi_n^t + \mathcal{U}_{\mathbf{u}}\Psi_n^t + \mathcal{R}_r\Psi_{n+1}^t + \mathcal{S}_{\mathbf{s},h}\Psi_{n+2}^t. \tag{A16}$$

Here, $\mathcal{D}$ is a sparse matrix to account for drift, $\mathcal{R}$ accounts for recombination with rate r, $\mathcal{S}$ for selection with arbitrary selection and dominance coefficients for each haplotype, and $\mathcal{U}$ is a mutation operator for either an infinite sites or reversible mutation model.

Under drift and mutation, in the neutral case and with no recombination, Equation A16 is closed and can be solved exactly. With selection and recombination Ψ_n relies on the slightly larger frequency spectra Ψ_{n+1} and Ψ_{n+2}, and so the system does not close. Intuitively, drift closes because we are just concerned with lineages copying over each other within the subsample of n tracked lineages. However, we require additional lineages in the case of non-zero recombination and selection. If a recombination event occurs within our n tracked lineages, we require an additional lineage to be drawn from the full population for the recombining lineage to be paired with. And in the case of selection, a replacement lineage must be drawn from the entire population, so we need to know the expected distribution of a larger size sample. We close the system using a jackknife extrapolation, which estimates Ψ_{n+l} using the known Ψ_n, as was done for the single locus frequency spectrum in Jouganous *et al.* (2017). In practice, we find the jackknife to be reasonably accurate for moderate sample sizes ($n \gtrsim 20$), with accuracy increasing in n (Table A1).

Below, we derive each operator in turn. We consider tracking a subsample of n lineages in the population and find how drift, mutation, recombination, and selection are each expected to change probabilities of two-locus haplotype frequencies in a given generation.

A.3.1 Drift

Just as in the Hill-Robertson section above (A.1.1), transition probabilities for Ψ_n are derived by considering haplotype frequencies within a sample of size n. Allele frequencies within the subsample of n lineages change due to drift in one generation if one lineage within the n subsampled lineages copies itself onto another lineage within our subsample. For simplicity, we assume $n \ll N$, so that at most a single copying event occurs between two of the n lineages in any given generation. Generalization to multiple coalescences per generation could be performed as in Jouganous *et al.* (2017).

The probability of a single copying event within the sample in a single generation at time t (to leading order in $1/N$) is

$$P_{n,N}(1 \to 2) = \frac{1}{2N(t)}\binom{n}{2}, \tag{A17}$$

the classical large-population limit from coalescence theory. In the case of a copying event, one lineage is drawn twice and another is not drawn. Haplotype frequencies will only change if the two drawn lineages differ in state.

We compute the probability of all possible haplotype frequency changes ($(i,j,k) \to (i',j',k')$) in one generation, given a coalescent event occurs. For example, frequencies change from (i,j,k) to $(i+1,j-1,k)$ if an AB lineage is chosen to copy over an Ab lineage. For a given frequency bin (i,j,k), this occurs if we first choose an AB lineage that copies itself twice (with probability i/n), then choose an Ab lineage to be copied over (with probability $j/(n-1)$), scaled by the density of two-locus haploypes with those frequencies ($\Psi_n(i,j,k)$). Together we get

$$P_{1\to2}((i,j,k) \to (i+1,j-1,k)) = \frac{i}{n}\frac{j}{n-1}\Psi_n(i,j,k). \tag{A18}$$

We account for all possible changes in haplotype frequencies due to copying events in this way:

$$
\begin{aligned}
P_{1\to 2}((i,j,k)\to(i+1,j,k)) &= \frac{i}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i-1,j,k)) &= \frac{i}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j+1,k)) &= \frac{j}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j-1,k)) &= \frac{j}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j,k+1)) &= \frac{k}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j,k-1)) &= \frac{k}{n}\frac{n-i-j-k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i+1,j-1,k)) &= \frac{i}{n}\frac{j}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i+1,j,k-1)) &= \frac{i}{n}\frac{k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i-1,j+1,k)) &= \frac{i}{n}\frac{j}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i-1,j,k+1)) &= \frac{i}{n}\frac{k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j+1,k-1)) &= \frac{j}{n}\frac{k}{n-1}\Psi_n(i,j,k)\\
P_{1\to 2}((i,j,k)\to(i,j-1,k+1)) &= \frac{j}{n}\frac{k}{n-1}\Psi_n(i,j,k)
\end{aligned}
$$

$$
\begin{aligned}
P_{1\to 2}((i-1,j,k)\to(i,j,k)) &= \frac{i-1}{n}\frac{n-i-j-k+1}{n-1}\Psi_n(i-1,j,k)\delta_{i>0}\\
P_{1\to 2}((i+1,j,k)\to(i,j,k)) &= \frac{i+1}{n}\frac{n-i-j-k-1}{n-1}\Psi_n(i+1,j,k)\delta_{i<n}\\
P_{1\to 2}((i,j-1,k)\to(i,j,k)) &= \frac{j-1}{n}\frac{n-i-j-k+1}{n-1}\Psi_n(i,j-1,k)\delta_{j>0}\\
P_{1\to 2}((i,j+1,k)\to(i,j,k)) &= \frac{j+1}{n}\frac{n-i-j-k-1}{n-1}\Psi_n(i,j+1,k)\delta_{j<n}\\
P_{1\to 2}((i,j,k-1)\to(i,j,k)) &= \frac{k-1}{n}\frac{n-i-j-k+1}{n-1}\Psi_n(i,j,k-1)\delta_{k>0}\\
P_{1\to 2}((i,j,k+1)\to(i,j,k)) &= \frac{k+1}{n}\frac{n-i-j-k-1}{n-1}\Psi_n(i,j,k+1)\delta_{k<n}\\
P_{1\to 2}((i-1,j+1,k)\to(i,j,k)) &= \frac{i-1}{n}\frac{j+1}{n-1}\Psi_n(i-1,j+1,k)\delta_{i>0}\\
P_{1\to 2}((i-1,j,k+1)\to(i,j,k)) &= \frac{i-1}{n}\frac{k+1}{n-1}\Psi_n(i-1,j,k+1)\delta_{i>0}\\
P_{1\to 2}((i+1,j-1,k)\to(i,j,k)) &= \frac{i+1}{n}\frac{j-1}{n-1}\Psi_n(i+1,j-1,k)\delta_{j>0}\\
P_{1\to 2}((i+1,j,k-1)\to(i,j,k)) &= \frac{i+1}{n}\frac{k-1}{n-1}\Psi_n(i+1,j,k-1)\delta_{k>0}\\
P_{1\to 2}((i,j-1,k+1)\to(i,j,k)) &= \frac{j-1}{n}\frac{k+1}{n-1}\Psi_n(i,j-1,k+1)\delta_{j>0}\\
P_{1\to 2}((i,j+1,k-1)\to(i,j,k)) &= \frac{j+1}{n}\frac{k-1}{n-1}\Psi_n(i,j+1,k-1)\delta_{k>0}
\end{aligned}
$$

Taken together, we obtain $\mathcal{D}_n(i,j,k)$.

A.3.2 Mutation

The moment system for Ψ_n allows for flexible mutation operators. Both the infinite sites model (ISM) and a reversible mutation model are straightforward to derive and implement.

For the ISM model, mutations at each locus occur only once from ancestral to derived state. We suppose the mutation rate at the left locus is u_1 (for $a \to A$) and at the right locus is u_2 ($b \to B$). For a two-locus pair to segregate at both loci, a mutation must first occur at one of the loci, and then a mutation must occur at the second locus while the first locus is still segregating. For the first mutation, we introduce density in the singleton bins (one copy of either Ab or aB) as

$$\mathcal{U}\Psi_n(0,1,0) = n\,u_1, \tag{A19}$$
$$\mathcal{U}\Psi_n(0,0,1) = n\,u_2. \tag{A20}$$

The second mutation, which occurs while the first locus is already segregating, introduces pairs of segregating loci proportional to the marginal single locus AFS at each locus. We account for whether, for example, a mutation $a \to A$ occurs on a B or b background:

$$\mathcal{U}\Psi_n(1,j,0) = u_2(j+1)\Psi_n(0,j+1,0) \qquad \forall j \in \{0,1,\ldots,n-1\} \tag{A21}$$
$$\mathcal{U}\Psi_n(0,j,1) = u_2(n-j)\Psi_n(0,j,0) \qquad \forall j \in \{1,2,\ldots,n\} \tag{A22}$$
$$\mathcal{U}\Psi_n(1,0,k) = u_1(k+1)\Psi_n(0,0,k+1) \qquad \forall k \in \{0,1,\ldots,n-1\} \tag{A23}$$
$$\mathcal{U}\Psi_n(0,1,k) = u_1(n-k)\Psi_n(0,0,k) \qquad \forall k \in \{1,2,\ldots,n\}. \tag{A24}$$

We can allow recurrent, reversible mutations, with rates

$$a \underset{v_1}{\overset{u_1}{\rightleftharpoons}} A \text{ and } b \underset{v_2}{\overset{u_2}{\rightleftharpoons}} B. \tag{A25}$$

In this case, there are no absorbing states. For example, the probability that a mutation event $a \to A$ occurs (u_1) that changes an aB haplotype to an AB haplotype depends on the number of aB haplotypes present in the sample (k) and the probability of observing the required haplotype frequencies $\Psi_n(i,j,k)$. Then $P_{\text{mut}}((i,j,k) \to (i+1,j,k-1)) = u_1(k)\Psi_n(i,j,k)$. All together, the mutation operator is

$$\begin{aligned}
\mathcal{U}_{\text{rev}}\Psi_n(i,j,k) = {} & u_1(k+1)\Psi_n(i-1,j,k+1) - u_1(k)\Psi_n(i,j,k) \\
& + u_1(n-i-j-k+1)\Psi_n(i,j-1,k) - u_1(n-i-j-k)\Psi_n(i,j,k) \\
& + v_1(i+1)\Psi_n(i+1,j,k-1) - v_1(i)\Psi_n(i,j,k) \\
& + v_1(j+1)\Psi_n(i,j+1,k) - v_1(j)\Psi_n(i,j,k) \\
& + u_2(j+1)\Psi_n(i-1,j+1,k) - u_2(j)\Psi_n(i,j,k) \\
& + u_2(n-i-j-k+1)\Psi_n(i,j,k-1) - u_2(n-i-j-k)\Psi_n(i,j,k) \\
& + v_2(i+1)\Psi_n(i+1,j-1,k) - v_2(i)\Psi_n(i,j,k) \\
& + v_2(k+1)\Psi_n(i,j,k+1) - v_2(k)\Psi_n(i,j,k).
\end{aligned}$$

A.3.3 Recombination

Here we derive the probabilities for transitions in frequencies due to recombination events, where a lineage in the sample is chosen to recombine with a lineage drawn from the full population, and one of the two recombinant types replaces the chosen lineage. Because we need to choose an extra lineage from the full population, $\mathcal{R}\Psi_n$ depends on Ψ_{n+1}, leading to the system being unclosed. For example, we write the probability of recombination between an ab type lineage from within our n subsampled lineages and an AB haplotype from the full population, and subsequently changing frequencies from (i,j,k) to (i',j',k'), as $\mathcal{R}_{ab\times AB}((i,j,k) \to (i',j',k'))$. These probabilities are

$$\mathcal{R}_{ab\times AB}((i,j,k) \to (i,j+1,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{n-i-j-k}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{ab\times AB}((i,j,k) \to (i,j,k+1)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{n-i-j-k}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{ab\times Ab}((i,j,k) \to (i,j+1,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{n-i-j-k}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{ab\times aB}((i,j,k) \to (i,j,k+1)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{n-i-j-k}{n+1}\frac{k+1}{n}$$

$$\mathcal{R}_{aB\times AB}((i,j,k)\to(i+1,j,k-1)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{k}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{aB\times Ab}((i,j,k)\to(i+1,j,k-1)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{k}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{aB\times Ab}((i,j,k)\to(i,j,k-1)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{k}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{aB\times ab}((i,j,k)\to(i,j,k-1)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{k}{n+1}\frac{n-i-j-k+1}{n}$$
$$\mathcal{R}_{Ab\times AB}((i,j,k)\to(i+1,j-1,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{j}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{Ab\times aB}((i,j,k)\to(i+1,j-1,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{j}{n+1}\frac{k+1}{n}$$
$$\mathcal{R}_{Ab\times aB}((i,j,k)\to(i,j-1,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{j}{n+1}\frac{k+1}{n}$$
$$\mathcal{R}_{Ab\times ab}((i,j,k)\to(i,j-1,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{j}{n+1}\frac{n-i-j-k+1}{n}$$
$$\mathcal{R}_{AB\times Ab}((i,j,k)\to(i-1,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{i}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{AB\times aB}((i,j,k)\to(i-1,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{i}{n+1}\frac{k+1}{n}$$
$$\mathcal{R}_{AB\times ab}((i,j,k)\to(i-1,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{i}{n+1}\frac{n-i-j-k+1}{n}$$
$$\mathcal{R}_{AB\times ab}((i,j,k)\to(i-1,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{i}{n+1}\frac{n-i-j-k+1}{n}$$

and

$$\mathcal{R}_{ab\times AB}((i,j-1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j-1,k)\frac{n-i-j-k+1}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{ab\times AB}((i,j,k-1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k-1)\frac{n-i-j-k+1}{n+1}\frac{i+1}{n}$$
$$\mathcal{R}_{ab\times Ab}((i,j-1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{n-i-j-k+1}{n+1}\frac{j}{n}$$
$$\mathcal{R}_{ab\times aB}((i,j,k-1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k)\frac{n-i-j-k+1}{n+1}\frac{k}{n}$$
$$\mathcal{R}_{aB\times AB}((i-1,j,k+1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{k+1}{n+1}\frac{i}{n}$$
$$\mathcal{R}_{aB\times Ab}((i-1,j,k+1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i-1,j+1,k+1)\frac{k+1}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{aB\times Ab}((i-1,j,k+1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k+1)\frac{k+1}{n+1}\frac{j+1}{n}$$
$$\mathcal{R}_{aB\times ab}((i,j,k+1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j,k+1)\frac{k+1}{n+1}\frac{n-i-j-k}{n}$$
$$\mathcal{R}_{Ab\times AB}((i-1,j+1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{j+1}{n+1}\frac{i}{n}$$
$$\mathcal{R}_{Ab\times aB}((i-1,j+1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i-1,j+1,k+1)\frac{j+1}{n+1}\frac{k+1}{n}$$
$$\mathcal{R}_{Ab\times aB}((i,j+1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k+1)\frac{j+1}{n+1}\frac{k+1}{n}$$
$$\mathcal{R}_{Ab\times ab}((i,j+1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i,j+1,k)\frac{j+1}{n+1}\frac{n-i-j-k}{n}$$
$$\mathcal{R}_{AB\times Ab}((i+1,j-1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{i+1}{n+1}\frac{j}{n}$$
$$\mathcal{R}_{AB\times aB}((i+1,j,k-1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k)\frac{i+1}{n+1}\frac{k}{n}$$
$$\mathcal{R}_{AB\times ab}((i+1,j-1,k)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j-1,k)\frac{i+1}{n+1}\frac{n-i-j-k+1}{n}$$

$$\mathcal{R}_{AB\times ab}((i+1,j,k-1)\to(i,j,k)) = \frac{1}{2}\Psi_{n+1}(i+1,j,k-1)\frac{i+1}{n+1}\frac{n-i-j-k+1}{n}$$

Multiplying by the probability that a recombination event occurs on a lineage in our sample (nr, assuming $r \ll 1$), and then cancelling terms and simplifying, we find

$$\begin{aligned}
\mathcal{R}_{n,r}\Psi_n(i,j,k) = nr\times \Big[&\Psi_{n+1}(i+1,j-1,k)\frac{i+1}{n+1}\frac{n-i-j-k+1}{n}\\
&+\Psi_{n+1}(i+1,j,k-1)\frac{i+1}{n+1}\frac{n-i-j-k+1}{n}\\
&+\Psi_{n+1}(i-1,j+1,k+1)\frac{j+1}{n+1}\frac{k+1}{n}\\
&+\Psi_{n+1}(i-1,j+1,k+1)\frac{j+1}{n+1}\frac{k+1}{n}\\
&-\Psi_{n+1}(i+1,j,k)\frac{i+1}{n+1}\frac{n-i-j-k}{n}\\
&-\Psi_{n+1}(i,j+1,k)\frac{j+1}{n+1}\frac{k}{n}\\
&-\Psi_{n+1}(i,j,k+1)\frac{j}{n+1}\frac{k+1}{n}\\
&-\Psi_{n+1}(i,j,k)\frac{i}{n+1}\frac{n-i-j-k+1}{n}\Big].
\end{aligned}$$

A.3.4 Selection

Here, we consider a model that allows selection at the left locus (A/a). In this setting we model a neutral locus linked to a selected locus separated by arbitrary recombination rate. We suppose A is selected against or for with selection and dominance coefficients s (with $|s| \ll 1$) and h, so that a diploid carrying AA has relative fitness $(1+s)$ compared to a aa diploid, while the heterozygote Aa has relative fitness $(1+hs)$. If $h = 1/2$, this reduces to the simple haploid selection model with A having fitness $(1+s)$ relative to a.

We consider how selection is expected to change haplotype frequencies over a given generation. As was the case for recombination and drift events above, there are many possible haplotype frequency changes that may occur due to a selection event.

For example, suppose that selection acts against the A allele ($s < 0$). We want to estimate the probability of events where, for example, a selection event occurs in which an AB lineage fails to replicate to the subsequent generation and is replaced by an Ab haplotype, drawn from the full population. The AB lineage is eliminated with probability $-s$ if the parent is a homozygote at the A/a locus (i.e. its diploid pair is type AB or Ab), and it is eliminated with probability $-sh$ if a heterozygote (i.e. paired with aB or ab). To compute the probability of any such event, we must draw two additional lineages from the population (one for the diploid pairing and one for the replacement lineage) in addition to the n tracked lineages in Ψ_n. Thus, we require Ψ_{n+2} to compute the evolution of Ψ_n.

Here, we write the probability that an AB lineage paired with aB is replaced by an Ab lineage as $\mathcal{S}_{(aB,\mathbf{AB})\leftarrow Ab}$. This event has probability

$$\mathcal{S}_{(aB,\mathbf{AB})\leftarrow Ab} = -sh\cdot n\cdot\Psi_{n+2}(i,j+1,k+1)\frac{\binom{i}{1}\binom{j+1}{1}\binom{k+1}{1}}{\binom{n}{3}}\frac{1}{6}, \tag{A26}$$

where we account for choosing the three necessary lineages in the correct order. Probabilities for the other three possible diploid pairings take a similar form:

$$\mathcal{S}_{(AB,\mathbf{AB})\leftarrow Ab} = -s\cdot n\cdot\Psi_{n+2}(i+1,j+1,k+1)\frac{\binom{i+1}{2}\binom{j+1}{1}}{\binom{n}{3}}\frac{1}{3}$$

$$\mathcal{S}_{(Ab,\mathbf{AB})\leftarrow Ab} = -s\cdot n\cdot\Psi_{n+2}(i,j+2,k)\frac{\binom{i}{1}\binom{j+2}{2}}{\binom{n}{3}}\frac{1}{3}$$

$$\mathcal{S}_{(ab,\mathbf{AB})\leftarrow Ab} = -sh \cdot n \cdot \Psi_{n+2}(i, j+1, k) \frac{\binom{j+1}{1}\binom{k+1}{1}\binom{n-i-j-k+1}{1}}{\binom{n}{3}} \frac{1}{6}.$$

Accounting for all possible selection events (including the replaced lineage, its diploid pair, and the replacement lineage), we find the selection operator

$$\begin{aligned}
\mathcal{S}_{n,s,h} = s\Bigg\{ & \frac{h}{n+1}\Big[(i+j)(k+1)\Psi_{n+1}(i,j,k+1) + \\
& + (i+j)(n-i-j-k+1)\Psi_{n+1}(i,j,k) \\
& - (i+1)(n-i-j)\Psi_{n+1}(i+1,j,k) \\
& - (j+1)(n-i-j)\Psi_{n+1}(i,j+1,k)\Big] \\
& + \frac{1-2h}{(n+2)(n+1)}\Big[(i+1)(k+1)(i+j)\Psi_{n+2}(i+1,j,k+1) \\
& + (i+1)(n-i-j-k+1)(i+j)\Psi_{n+2}(i+1,j,k) \\
& + (j+1)(k+1)(i+j)\Psi_{n+2}(i,j+1,k+1) \\
& + (j+1)(n-i-j-k+1)(i+j)\Psi_{n+2}(i,j+1,k) \\
& - (i+2)(i+1)(n-i-j)\Psi_{n+2}(i+2,j,k) \\
& - 2(i+1)(j+1)(n-i-j)\Psi_{n+2}(i+1,j+1,k) \\
& - (j+2)(j+1)(n-i-j)\Psi_{n+2}(i,j+2,k)\Big]\Bigg\}.
\end{aligned} \tag{A27}$$

Here we've used the downsampling formula to simplify the additive terms: if $h = 1/2$, $\mathcal{S}$ only requires Ψ_{n+1} since selection is independent of the state of the paired lineage.

A.3.5 Moment closure approximation for Ψ_n

Because neither recombination nor selection close in the full two-locus haplotype frequency system, we require a moment closure approximation to integrate Equation A16 forward in time. Here we will use a jackknife extrapolation, similar to Jouganous *et al.* (2017) for the single locus AFS, to express Ψ_{n+2} and Ψ_{n+1} as linear combinations of Ψ_n, so that

$$\Psi_{n+l}(i,j,k) \approx \sum_{(i',j',k')\in I_{(i,j,k)}} w_{(i',j',k')}\Psi_n(i',j',k'), \tag{A28}$$

where $l \in 1, 2$ and the entries $(i', j', k') \in I_{(i,j,k)}$ are chosen so that $(i'/n, j'/n, k'/n)$ are close to $(i/(n+l), j/(n+l), k/(n+l))$. Each choice l will have its own set I and weights w.

We then find the appropriate set of entries I and weights w, for a given l and entry in that frequency spectrum. First, we note that for any continuous function ψ that solves Equation A15, we can find the entries of $\Psi_n(i,j,k)$ for any n, i, j, k, using the multinomial sampling formula

$$\Psi_n(i,j,k) = \iiint\limits_{\substack{x,y,z\geq 0,\\ x+y+z\leq 1}} \psi(x,y,z)\binom{n}{i,j,k} x^i y^j z^k (1-x-y-z)^{n-i-j-k}\, dx\, dy\, dz. \tag{A29}$$

We make the assumption here that ψ can be approximated *locally* as a quadratic, so that

$$\psi(x,y,z) \approx a_1 + a_2 x + a_3 y + a_4 z + a_5 x^2 + a_6 xy + a_7 xz + a_8 y^2 + a_9 yz + a_{10} z^2. \tag{A30}$$

Using this approximation of ψ, the multinomial sampling integral can be computed analytically, to get

$$\tilde{\Psi}_n(i,j,k) = \frac{1}{(n+5)(n+4)(n+3)(n+2)(n+1)}\Big[a_1(n+5)(n+4) + a_2(n+5)(i+1) \tag{A31}$$

$$
\begin{aligned}
&+ a_3(n+5)(j+1) + a_4(n+5)(k+1) \\
&+ a_5(i+2)(i+1) + a_6(i+1)(j+1) \\
&+ a_7(i+1)(k+1) + a_8(j+2)(j+1) \\
&+ a_9(j+1)(k+1) + a_{10}(k+2)(k+1)\Big]
\end{aligned}
$$

For a given $\Psi_n(i,j,k)$ we take the ten closest values $I = \{(i'_1,j'_1,k'_1),(i'_2,j'_2,k'_2),\ldots,(i'_{10},j'_{10},k'_{10})\}$ as described above, with the added condition that the sets of $\{i'\},\{j'\}$, and $\{k'\}$ each have size at least three. We then set

$$\tilde{\Psi}_{n+l}(i,j,k) = \sum w_{(i',j',k')}\tilde{\Psi}_n(i',j',k'). \tag{A32}$$

This holds for any ψ, so we set $(a_1,a_2,\ldots,a_{10}) = (1,0,\ldots,0)$, $(a_1,a_2,\ldots,a_{10}) = (0,1,\ldots,0)$, etc, in turn to get a system of equations for the weights $w_m = w_{(i'_m,j'_m,k'_m)}$, $m = 1,\ldots,10$:

$$
\begin{cases}
\frac{(n+3)!(n+l)!}{n!(n+3+l)!} = \sum_{m=1}^{10} w_m \\
\frac{(n+4)!(n+l)!}{n!(n+4+l)!} = \sum_{m=1}^{10} (i'_m+1) w_m \\
\frac{(n+4)!(n+l)!}{n!(n+4+l)!} = \sum_{m=1}^{10} (j'_m+1) w_m \\
\frac{(n+4)!(n+l)!}{n!(n+4+l)!} = \sum_{m=1}^{10} (k'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (i'_m+2)(i'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (i'_m+1)(j'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (i'_m+1)(k'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (j'_m+2)(j'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (j'_m+1)(k'_m+1) w_m \\
\frac{(n+5)!(n+l)!}{n!(n+5+l)!} = \sum_{m=1}^{10} (k'_m+2)(k'_m+1) w_m,
\end{cases}
$$

which can be solved either analytically or numerically.

A.3.6 Comparing methods to compute Ψ_n

We compared the accuracy and computational time needed for computing Ψ_n using the numerical PDE approach presented in Ragsdale and Gutenkunst (2017) and the moment approach presented here. We performed this comparison for $n \in \{30,50\}$ and $\rho \in \{0,10\}$. We considered two demographic models for each sample size and recombination rate: equilibrium demography with no size changes, and a bottleneck demography. In the bottleneck demographic model the population is initially at steady state, then instantaneously changes in size to 1/10 the original size for 0.05 time units (measured in $2N_e$), and then it recovers to the original size for 0.2 time units.

For the equilibrium Ψ_n, we compared to Hudson's Monte Carlo implementation (described in Hudson (2001) and available from the author's website) for $n = 100$ projected down to sample sizes $n = 30$ and $n = 50$. We projected from a larger sample size for improved accuracy in Hudson's estimate and considered this distribution the "true" distribution to compare against. For the bottleneck distribution, we computed a numerical approximation with a larger sample size ($n = 80$) and 100x smaller time step than the default using `moments.TwoLocus` and then projected to $n = 30$ and $n = 50$. To compute Ψ_n in ∂a∂i, we numerically solve for Ψ_n for three grid spacings and three time steps, and then perform Richardson extrapolation (detailed in Ragsdale and Gutenkunst (2017)). We set integration time steps to $[0.005, 0.0025, 0.001]$ and grid points $[40, 50, 60]$ for $n = 30$ and $[60, 70, 80]$ for $n = 50$. We measured accuracy as $\sum(\Psi_n(\text{model}) - \Psi_n(\text{True}))^2/\Psi_n(\text{True})$. In general, `moments` performs favorably compared to ∂a∂i, with orders of magnitude improved accuracy and faster integration and evaluation (Table A1).

A.4 Deriving moment equations from the PDE

So far we have derived recursion equations in both the Hill-Robertson basis and for the full haplotype frequency spectrum by tracking an appropriately sized subset of lineages within the full population and

considering the effects such as drift and recombination within these lineages. The statistics in these recursion equations are all non-canonical moments of the full two-locus distribution. Thus an alternative route to deriving all of the recursion equations presented in this paper is directly through the partial differential equation (PDE) describing the evolution of this full distribution.

Classically, two equivalent PDEs describing this distribution were studied, one in the variables of haplotype frequencies (x_1, x_2, x_3, x_4) and the other in the variables (p, q, D) (Hill and Robertson, 1966; Ohta and Kimura, 1969b). In this section, we outline the approach to obtain the Hill-Robertson D^2 system from the latter of these bases, do the same for the haplotype frequency spectrum Ψ_n, and then intuitively discuss why recombination closes in the Hill-Robertson basis but not the haplotype frequency basis from a PDE perspective.

Without selection, the neutral two-locus distribution $\psi(p, q, D)$ follows the forward Kolmogorov equation,

$$\begin{aligned}\frac{\partial \psi}{\partial \tau} =& \frac{1}{2}\frac{\partial^2}{\partial p^2}\left(p(1-p)\psi\right) + \frac{1}{2}\frac{\partial^2}{\partial q^2}\left(q(1-q)\psi\right) + \frac{\partial^2}{\partial p \partial q}\left(D\psi\right) + \frac{\partial^2}{\partial p \partial D}\left(D(1-2p)\psi\right) + \frac{\partial^2}{\partial q \partial D}\left(D(1-2q)\psi\right) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial D^2}\left(\left(p(1-p)q(1-q) + D(1-2p)(1-2q) - D^2\right)\psi\right) + \frac{\partial}{\partial D}\left(D\left(1+\frac{\rho}{2}\right)\psi\right), \end{aligned} \tag{A33}$$

where time τ is measured in $2N$ generations. To obtain the time evolution of any moment of this distribution, we take

$$\begin{aligned}\partial_\tau \mathbb{E}[f(p,q,D)] &= \frac{\partial}{\partial \tau}\int \psi f(p,q,D) \\ &= \int \frac{\partial \psi}{\partial \tau} f(p,q,D) \\ &= \int (\text{RHS}) f(p,q,D),\end{aligned}$$

where RHS are the terms in Equation A33. Here, we have abused the integral notation, and it should be understood to be the triple integral over the domain of the function ψ.

For the D^2 system, we first find $\partial_\tau \mathbb{E}[D^2]$:

$$\begin{aligned}\partial_\tau \mathbb{E}[D^2] =& \frac{1}{2}\int D^2 \frac{\partial^2}{\partial p^2} p(1-p)\psi + \frac{1}{2}\int D^2 \frac{\partial^2}{\partial q^2} q(1-q)\psi + \int D^2 \frac{\partial}{\partial p \partial q} pq\psi \\ &+ \int D^2 \frac{\partial^2}{\partial p \partial D} D(1-2p)\psi + \int D^2 \frac{\partial^2}{\partial p \partial D} D(1-2p)\psi \\ &+ \frac{1}{2}\int D^2 \frac{\partial^2}{\partial D^2}\left(p(1-p)q(1-q) + D(1-2p)(1-2q) - D^2\right)\psi \\ &+ \int D^2 \frac{\partial}{\partial D} D\left(1+\frac{\rho}{2}\right)\psi \\ \overset{\text{IBP}}{=}& \int \left(p(1-p)q(1-q) + D(1-2p)(1-2q) - D^2\right)\psi - 2\int D^2\left(1+\frac{\rho}{2}\right)\psi \\ =& -3\mathbb{E}[D^2] + \mathbb{E}[D(1-2p)(1-2q)] + \mathbb{E}[p(1-p)q(1-q)] - \rho\mathbb{E}[D^2],\end{aligned}$$

which recovers, to first order, the Hill-Robertson equation for $\mathbb{E}[D^2]$. Surface terms vanish since the functions decay to zero at the boundary. The other two terms in the system can by found by similarly integrating by parts:

$$\begin{aligned}\partial_\tau \mathbb{E}[D(1-2p)(1-2q)] &= 4\mathbb{E}[D^2] - 5\mathbb{E}[D(1-2p)(1-2q)] - \frac{\rho}{2}\mathbb{E}[D(1-2p)(1-2q)] \\ \partial_\tau \mathbb{E}[p(1-p)q(1-q)] &= \mathbb{E}[D(1-2p)(1-2q)] - 2\mathbb{E}[p(1-p)q(1-q)]\end{aligned}$$

An equivalent PDE for ψ is expressed in haplotype frequencies (x_1, x_2, x_3):

$$\frac{\partial \psi}{\partial \tau} = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\partial^2}{\partial x_i \partial x_j} x_i(\delta_{i=j} - x_j)\psi + \frac{\rho}{2}\left(\frac{\partial}{\partial x_1}D\psi - \frac{\partial}{\partial x_2}D\psi - \frac{\partial}{\partial x_3}D\psi\right), \tag{A34}$$

where $\delta_{i=j} = 1$ if $i = j$ and 0 otherwise, and formally $D = x_1(1 - x_1 - x_2 - x_3) - x_2x_3$. The first set of terms in the double sum account for drift, and the second set of terms account of recombination with scaled rate ρ.

We want to find evolution equations for $\Psi_n(i, j, k)$, and we take a similar approach as above. Given the continuous distribution ψ, we compute $\Psi_n(i, j, k)$ by integrating ψ agains the multinomial distribution:

$$\Psi_n(i,j,k) = \iiint \binom{n}{i,j,k,n-i-j-k} x_1^i x_2^j x_3^k (1 - x_1 - x_2 - x_3)^{n-i-j-k} \psi(x_1, x_2, x_3),$$

where $\binom{n}{i,j,k,n-i-j-k} = \frac{n!}{i!j!k!(n-i-j-k)!}$ is the multinomial coefficient. Then integrating both sizes of Equation A34 against this sampling function, we can find

$$\partial_\tau \Psi_n(i,j,k) = \iiint \binom{n}{i,j,k,n-i-j-k} x_1^i x_2^j x_3^k (1 - x_1 - x_2 - x_3)^{n-i-j-k} \left(\frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{\partial^2}{\partial x_i \partial x_j} x_i(\delta_{i=j} - x_j)\psi + \frac{\rho}{2} \left(\frac{\partial}{\partial x_1} D\psi - \frac{\partial}{\partial x_2} D\psi - \frac{\partial}{\partial x_3} D\psi \right) \right).$$

In brief, for the drift terms we integrate by parts twice, obtain a series of multinomial sampling functions against ψ, and then simplify. This results in the same set of equations as derived above by computing copying probabilities. This derivation follows closely the approach described in Jouganous *et al.* (2017) for the single site AFS.

To understand closure properties, we can consider whether the order of terms on the RHS is always equal or smaller to n. If so, the set of all moments of order n will only ever require moments of order less than or equal to n, ensuring closure of the moment equation. When counting the order of terms on the RHS, we must subtract one for each derivative: in the integration by part, each derivative reduces the degree of the polynomial coefficient by at least one. Thus the drift term on the RHS has order n and closes, whereas the recombination term has order $n + 1$, and as a result does not close.

A.4.1 Closure of Hill-Robertson moments

Any order moment system for $\mathbb{E}[D^m]$ closes under drift and recombination. First, we observe that the recombination transition matrix will be diagonal by considering the PDE without the drift terms: $\psi_\tau = \frac{\rho}{2}(D\psi)_D$. For any moment $\mathbb{E}[D^\alpha f(p, q)]$, we get

$$\begin{aligned} \partial_\tau \mathbb{E}[D^\alpha f(p,q)] &= \frac{\rho}{2} \int D^\alpha f(p,q) \frac{\partial}{\partial D} D\psi \\ &= -\alpha \frac{\rho}{2} \int D^\alpha f(p,q)\psi = -\alpha \frac{\rho}{2} \mathbb{E}[D^\alpha f(p,q)]. \end{aligned}$$

Thus any basis of moments expressed as a functions of (p, q, D) will close under recombination.

Intuitively, for any moment $f(p, q, D)$ we can expect to find a closed system for its evolution under drift. This can be seen from the PDE for ψ (Equation A33): the coefficient of each spatial derivative has order equal to the derivative. Thus integrating by parts does not result in any moments of ψ of higher order than the original moment f. Since there are a finite number of moments of any given order, this system must also be finite in size, and thus necessarily close.

Here we compute the recursion for the D^m system (for m even) and explicitly demonstrate that it closes by considering the new terms required to compute $\mathbb{E}[D^m]$. Again, we define $\pi_2 = p(1 - p)q(1 - q)$, $z = (1 - 2p)(1 - 2q)$, and $\sigma_j = p^j(1 - p)^j + q^j(1 - q)^j$.

First, suppose that the D^{m-2} system closes. The D^{m-2} system contains all terms

$$\left\{ \mathbb{E}[D^{m-2i-2j-k-2} \pi_2^i z^k \sigma_j] \right\}, \text{ for } k \in \{0, 1\}, j \in \left\{ 0, 1, \ldots, \frac{m-2k-2}{2} \right\}, i \in \left\{ 0, 1, \ldots, \frac{m-2k-2j-2}{2} \right\}.$$

It also contains all terms in the D^{m-4} system, which take the same form as above, and so on. By computing $\partial_\tau \mathbb{E}[D^m]$ and its dependent terms, we find that all terms are present in the D^{m-2} system or are in

$$\left\{ \mathbb{E}[D^{m-2i-2j-k} \pi_2^i z^k \sigma_j] \right\}, \text{ for } k \in \{0, 1\}, j \in \left\{ 0, 1, \ldots, \frac{m-2k}{2} \right\}, i \in \left\{ 0, 1, \ldots, \frac{m-2k-2j}{2} \right\}. \quad \text{(A35)}$$

For the m^{th} moment of D, we can use this same approach to obtain moment dependencies and compute evolution equations, just as we did in the previous section for the Hill-Robertson system. For $\mathbb{E}[D^m]$ (here showing just the terms for drift), by integrating by parts, we have

$$\begin{aligned}
\partial_\tau \mathbb{E}[D^m] &= \frac{1}{2}\int D^m \frac{\partial^2}{\partial D^2}\left(p(1-p)q(1-q) + D(1-2p)(1-2q) - D^2\right)\psi + \int D^m \frac{\partial}{\partial D} D\psi \\
&= \frac{m(m-1)}{2}\int D^{m-2}\left(p(1-p)q(1-q) + D(1-2p)(1-2q) - D^2\right)\psi - m\int D^m\psi \\
&= -\frac{m(m+1)}{2}\mathbb{E}[D^m] + \frac{m(m-1)}{2}\mathbb{E}[D^{m-1}(1-2p)(1-2q)] + \frac{m(m-1)}{2}\mathbb{E}[D^{m-2}p(1-p)q(1-q)].
\end{aligned}$$

In the same way, we compute time derivative for each dependent moment:

$$\begin{aligned}
\partial_\tau \mathbb{E}[D^{m-1}z] &= 4\mathbb{E}[D^m] - \frac{(m+8)(m-1)}{2}\mathbb{E}[D^{m-1}z] + 8(m-1)(m-2)\mathbb{E}[D^{m-2}\pi_2] \\
&\quad - 2(m-1)(m-2)\mathbb{E}[D^{m-2}\sigma_1] + \frac{(m-1)(m-2)}{2}\mathbb{E}[D^{m-2}] + \frac{(m-1)(m-2)}{2}\mathbb{E}[D^{m-3}\pi_2 z] \\
\partial_\tau \mathbb{E}[D^{m-2}\pi_2] &= \mathbb{E}[D^{m-1}z] - \frac{1}{2}(m^2+13m-26)\mathbb{E}[D^{m-2}\pi_2] + (m-2)\mathbb{E}[D^{m-2}\sigma_1] \\
&\quad + \frac{(m-2)(m-3)}{2}\mathbb{E}[D^{m-3}\pi_2 z] + \frac{(m-2)(m-3)}{2}\mathbb{E}[D^{m-4}\pi_2^2], \\
&\vdots \\
\partial_\tau \mathbb{E}[D^{m-2i}\pi_2^i] &= i^2\mathbb{E}[D^{m-2i+1}\pi_2^{i-1}z] - \frac{m^2+(1+12i)m-2i(3+10i)}{2}\mathbb{E}[D^{m-2i}\pi_2^i] \\
&\quad (im - i(1+3i)/2)\mathbb{E}[D^{m-2i}\pi_2^{i-1}\sigma_1] + \frac{(m-i+1)(m-i)}{2}\mathbb{E}[D^{m-2i-1}\pi_2^i z] \\
&\quad + \frac{(m-i+1)(m-i)}{2}\mathbb{E}[D^{m-2i-2}\pi_2^{i+1}] \\
\partial_\tau \mathbb{E}[D^{m-2i-1}\pi_2^i z] &= 4(1+2i)^2\mathbb{E}[D^{m-2i}\pi_2^i] - 2i(1+2i)\mathbb{E}[D^{m-2i}\pi_2^{i-1}z] + i^2\mathbb{E}[D^{m-2i}\pi_2^{i-1}] \\
&\quad - \frac{m^2+(7+12i)m-2(10i^2+13i+4)}{2}\mathbb{E}[D^{m-2i-1}\pi_2^i z] \\
&\quad + (im - 3i(i+1)/2)\mathbb{E}[D^{m-2i-1}\pi_2^{i-1}z\sigma_1] + 8(m-2i-2)(m-2i-1)\mathbb{E}[D^{m-2i-2}\pi_2^{i+1}] \\
&\quad - 2(m-2i-2)(m-2i-1)\mathbb{E}[D^{m-2i-2}\pi_2^i\sigma_1] + \frac{(m-2i-2)(m-2i-1)}{2}\mathbb{E}[D^{m-2i-2}\pi_2^i] \\
&\quad + \frac{(m-2i-2)(m-2i-1)}{2}\mathbb{E}[D^{m-2i-3}\pi_2^{i+1}z] \\
&\vdots \\
\partial_\tau \mathbb{E}[D\pi_2^{\frac{m-2}{2}}z] &= \frac{(m-1)^2}{4}\mathbb{E}[D^2\pi_2^{\frac{m-2}{2}}] - (m-2)(m-1)\mathbb{E}[D^2\pi_2^{\frac{m-4}{2}}\sigma_1] + \frac{(m-2)^2}{4}\mathbb{E}[D^2\pi_2^{\frac{m-4}{2}}] \\
&\quad - (m^2+m-1)\mathbb{E}[D\pi_2^{\frac{m-2}{2}}z] + \frac{m(m-2)}{2}\mathbb{E}[D\pi_2^{\frac{m-4}{2}}z\sigma_1] \\
\partial_\tau \mathbb{E}[\pi_2^{\frac{m}{2}}] &= \frac{m^2}{4}\mathbb{E}[D\pi_2^{\frac{m-2}{2}}z] - m(m-1)\mathbb{E}[\pi_2^{\frac{m}{2}}] + \frac{m(m-2)}{8}\mathbb{E}[\pi_2^{\frac{m-2}{2}}\sigma_1]
\end{aligned}$$

The evolution for the terms computed here require those already considered, terms in the D^{m-2} system, or additional terms of the form

$$\{D^{m-2}\sigma_1, D^{m-3}z\sigma_1, D^{m-4}\pi_2\sigma_1, \ldots, \pi_2^{\frac{m-2}{2}}\sigma_1\}.$$

In general, for $j > 0$, $k \in \{0,1\}$, $i \in \{0,\ldots,(m-2j-2k)/2\}$, we have terms of the form $\mathbb{E}[D^{m-2i-2j-k}\pi_2^i z^k \sigma_j]$. We compute transition probabilities for each of these terms:

$$\begin{aligned}
\partial_\tau \mathbb{E}[D^{m-2j}\sigma_j] &= -\frac{m^2+(4j+1)m-4i(1+2i)}{2}\mathbb{E}[D^{m-2j}\sigma_j] + (im + i(1+3i)/2)\mathbb{E}[D^{m-2j}\sigma_{j-1}] \\
&\quad + \frac{(m-2j-1)(m-2j)}{2}\mathbb{E}[D^{m-2j-1}z\sigma_j] + \frac{(m-2j-1)(m-2j)}{2}\mathbb{E}[D^{m-2j-2}\pi_2\sigma_j]
\end{aligned}$$

$$
\begin{aligned}
\partial_\tau \mathbb{E}[D^{m-2j-1} z\sigma_j] &= 4(1+2i)\mathbb{E}[D^{m-2j}\sigma_j] - 2i\mathbb{E}[D^{m-2j}\sigma_{j-1}] \\
&\quad - \frac{m^2+(7+4i)m-4(i+2(1+i)^2)}{2}\mathbb{E}[D^{m-2j-1}z\sigma_j] \\
&\quad + (4m-3i(i+1)/2)\mathbb{E}[D^{m-2j-1}z\sigma_{j-1}] + 8(m-2j-2)(m-2j-1)\mathbb{E}[D^{m-2j-2}\pi_2\sigma_j] \\
&\quad - 2(m-2j-2)(m-2j-1)\mathbb{E}[D^{m-2j-2}\pi_2\sigma_{j-1}] - 2(m-2j-2)(m-2j-1)\mathbb{E}[D^{m-2j-2}\sigma_{j+1}] \\
&\quad + \frac{(m-2j-2)(m-2j-1)}{2}\mathbb{E}[D^{m-2j-2}\sigma_j] + \frac{(m-2j-2)(m-2j-1)}{2}\mathbb{E}[D^{m-2j-3}\pi_2 z\sigma_j] \\
&\vdots \\
\partial_\tau \mathbb{E}[D\pi_2^{\frac{m-2-2j}{2}} z\sigma_j] &= 4(m-2j-1)(m-1)\mathbb{E}[D^2\pi_2^{\frac{m-2j-2}{2}}\sigma_j] - (m-2j-1)(m-2)\mathbb{E}[D^2\pi_2^{\frac{m-2j-2}{2}}\sigma_{j-1}] \\
&\quad - (m-2j-2)(m-1)\mathbb{E}[D^2\pi_2^{\frac{m-2j-4}{2}}\sigma_{j+1}] + \frac{(m-2j-2)(m-2)}{4}\mathbb{E}[D^2\pi_2^{\frac{m-2j-4}{2}}\sigma_{j-1}] \\
&\quad - (m^2-(2j-1)m+(2j+1)(j-1))\mathbb{E}[D\pi_2^{\frac{m-2j-2}{2}} z\sigma_j] + \frac{m(m-2)}{8}\mathbb{E}[D\pi_2^{\frac{m-2j-2}{2}} z\sigma_{j-1}] \\
&\quad + \frac{(m-2j-2)(m-2j)}{8}\mathbb{E}[D\pi_2^{\frac{m-2j-4}{2}} z\sigma_{j+1}] \\
\partial_\tau \mathbb{E}[\pi_2^{\frac{m-2j}{2}}\sigma_j] &= \frac{m(m-2j)}{4}\mathbb{E}[D\pi_2^{\frac{m-2j-2}{2}} z\sigma_j] - (m^2-(2j+1)m+j(2j+1))\mathbb{E}[\pi_2^{\frac{m-2j-2}{2}}\sigma_j] \\
&\quad + \frac{m(m-2)}{8}\mathbb{E}[\pi_2^{\frac{m-2j-2}{2}}\sigma_j] + \frac{(m-2j-2)(m-2j)}{8}\mathbb{E}[\pi_2^{\frac{m-2j-4}{2}}\sigma_{j+1}].
\end{aligned}
$$

Each term appearing here belongs to the D^{m-2} system or is found in the set of new moments enumerated in (A35).

A.5 Sampling bias and the relationship between Ψ_n and Hill-Robertson statistics

$\mathbb{E}[D]$ is a two-haplotype statistic, meaning we require one phased diploid genome to estimate D genome-wide. In practice, to estimate D we count the number of times we observe a $(AB|ab)$ pairing, subtract the counts of observed $(Ab|aB)$ pairings, and normalize by the total number of two-locus pairs considered. From the two-locus haplotype frequency spectrum, computed under the appropriate per-base mutation rate,

$$\mathbb{E}[D] = \frac{1}{2}\left(\Psi_2(1,0,0) - \Psi_2(0,1,1)\right).$$

A more accurate estimate may be obtained by considering haplotype frequencies in a larger sample size than $n=2$, although the estimate would then need to be corrected due to sampling bias. Alternatively, we can use hypergeometric projection to directly calculate an unbiased estimate of $\mathbb{E}[D]$ for $n>2$ samples. A two-locus pair with sample size n and observed haplotype counts $(n_{AB}, n_{Ab}, n_{aB}, n_{ab})$ contributes to $\mathbb{E}[D]$ by

$$\frac{1}{2}\frac{\binom{n_{AB}}{1}\binom{n_{ab}}{1}}{\binom{n}{2}} - \frac{1}{2}\frac{\binom{n_{Ab}}{1}\binom{n_{aB}}{1}}{\binom{n}{2}} = \frac{n_{AB}n_{ab}}{n(n-1)} - \frac{n_{Ab}n_{aB}}{n(n-1)}. \tag{A36}$$

This approach not only provides an unbiased estimate of $\mathbb{E}[D]$, but also allows us to compute $\mathbb{E}[D]$ over pairs of sites with different sample sizes (e.g. to deal with missing data).

We can express $\mathbb{E}[D^2]$ or any other term in the Hill-Robertson system as a linear combination of entries in Ψ_4. For example,

$$
\begin{aligned}
\mathbb{E}[D^2] &= \mathbb{E}[(f_{AB}f_{ab} - f_{Ab}f_{aB})^2] \\
&= 2\left(\frac{1}{\binom{4}{2,0,0,2}}\Psi_4(2,0,0) + \frac{1}{\binom{4}{0,2,2,0}}\Psi_4(0,2,2) - \frac{2}{\binom{4}{1,1,1,1}}\Psi_4(1,1,1)\right) \\
&= \frac{1}{3}\left(\Psi_4(2,0,0) + \Psi_4(0,2,2) - \frac{1}{2}\Psi_4(1,1,1)\right).
\end{aligned} \tag{A37}
$$

The multinomial factors arise because the entries of Ψ_n are unsorted configuration probabilities, while $\mathbb{E}[D^2]$ implies a particular order of drawn haplotypes. This implies that we may obtain an unbiased estimate for any quantity (such as $\mathbb{E}[D^2]$) from an arbitrary sample size n through hypergeometric projection to the appropriate sample size. For example, in a sample of size n a two-locus pair (with observed haplotype counts $(n_{AB}, n_{Ab}, n_{aB}, n_{ab})$) contributes to $\mathbb{E}[D^2]$

$$\frac{1}{3}\frac{\binom{n_{AB}}{2}\binom{n_{ab}}{2}}{\binom{n}{4}} + \frac{1}{3}\frac{\binom{n_{Ab}}{2}\binom{n_{aB}}{2}}{\binom{n}{4}} - \frac{1}{6}\frac{\binom{n_{AB}}{1}\binom{n_{Ab}}{1}\binom{n_{aB}}{1}\binom{n_{ab}}{1}}{\binom{n}{4}}. \tag{A38}$$

This allows for direct comparison between observed haplotype data and expectations from the model without having to correct for sample size bias. We discuss our approach for unphased data below in Data processing.

A.6 Low coverage data

Low coverage sequencing data is known to miss a sizable proportion of low frequency variation, so that singleton and doubleton bins of the AFS may be significantly underestimated (Gravel *et al.*, 2011). This may bias demographic inference based on the AFS from low coverage data, particularly for recent population size or growth parameters. Low order LD statistics studied in this paper are less sensitive to low coverage data, because low frequency variants contribute relatively little to aggregate statistics in the Hill-Robertson system across the genome (Rogers, 2014). Rogers (2014) argued that σ_d^2 type statistics are insensitive to variants with low heterozygosity, and thus insensitive to low coverage data or sequencing error.

To confirm this claim with real data, we examined the effect of low coverage on 40 individuals from the CHB population in the 1000 Genomes data that were also sequenced at high (∼80x) coverage (in the 90 Han Chinese high coverage genomes dataset (Lan *et al.*, 2017)). By comparing statistics computed from intergenic regions in the same individuals between the two datasets, we can see if low coverage biases our estimates. In Figure A4, we compare $\sigma_d^2 = \mathbb{E}[D^2]/\mathbb{E}[\pi_2]$, $\mathbb{E}[Dz]/\mathbb{E}[\pi_2]$, and the folded single-site AFS. We find that the low order two-locus statistics are unaffected by low coverage data, but low frequency bins of the AFS are underestimated in the low coverage data (17.5% fewer singletons, 7% fewer doubletons).

A.7 Data processing

A.7.1 Intergenic data

We used data from all intergenic regions on autosomal chromosomes, as identified using the GRCh37 build from the Genome Reference Consortium. We chose intergenic regions to reduce possible biases in statistic estimation due to selection (Figure A7). We considered only keeping SNPs at least a given distance from the nearest gene, to further reduce selective effects, but that would come at the cost of more noise in the statistic estimates, as we would be left with fewer loci from which to estimate statistics. We chose to use all intergenic SNPs because there was little difference between statistics estimated using all intergenic loci and using all loci at least a given distance from genes (Figure A8).

A.7.2 Recombination map and binning pairs by recombination distance

We considered all pairs of biallelic SNPs in intergenic regions separated by recombination distances $0.00001 \leq r < 0.002$, and binned pairs of SNPs with bin edges (0.00001, 0.00002, 0.00003, 0.00005, 0.00007, 0.0001, 0.0002, 0.0003, 0.0005, 0.0007, 0.001, 0.002). Recombination distances for each pair of SNPs were computed using the African American recombination map from Hinch *et al.* (2011).

A.7.3 Computing LD statistics from unphased data and accounting for sampling bias

For a given pair of biallelic SNPs, we computed their contribution to the set of statistics in the multi-population Hill-Robertson system (for three populations: YRI, CHB, and CEU). We determined two-locus genotype counts within each population $(n_{AABB}, n_{AaBB}, n_{aaBB}, n_{AABb}, \ldots)$, as well as allele frequencies at the left and right loci within each population. Because we worked with genotype data instead of phased

haplotypes, we used the estimator $\hat{D}$ of Weir (1979),

$$\hat{D} = \frac{1}{2n_d}\left(2n_{AABB} + n_{AABb} + n_{AaBB} + \frac{1}{2}n_{AaBb}\right) - \frac{n_A}{2n_d}\frac{n_B}{2n_d}.$$

We then computed the values $\hat{D}_i\hat{D}_j$, $\hat{D}_i(1-2p_j)(1-2q_k)$, and so on from genotype data.

$\hat{D}$ is known to be a biased estimator. To compare model predictions to data, we treated $\hat{D}$ as a statistic and computed unbiased predictions for it. Without considering bias in estimates due to sample size, $\mathbb{E}[\hat{D}^\alpha f(p,q)] = \frac{1}{2^\alpha}\mathbb{E}[D^\alpha f(p,q)]$.

We also needed to account for biased estimates for a given sample sizes in each population $(n_1, n_2, \ldots)$. We adjusted expectations to match observed biased statistics by computing expected values for each statistic under the multinomial sampling process.

For a given statistic, we converted the $\hat{D}$ statistic to genotype frequency space (e.g., using $\hat{D} = (g_1 + g_2/2 + g_3/2 + g_5/4) - (g_1 + g_2 + g_3 + g_4/2 + g_5/2 + g_6/2)(g_1 + g_2/2 + g_4 + g_5/2 + g_7 + g_8/2)$, where $(g_1, g_2, \ldots, g_9) = (n_{AABB}/n_d,\ n_{AABb}/n_d,\ n_{AAbb}/n_d,\ n_{AaBB}/n_d,\ n_{AaBb}/n_d,\ n_{Aabb}/n_d,\ n_{aaBB}/n_d,\ n_{aaBb}/n_d,\ n_{aabb}/n_d)$).

Then for each term in the expansion in g-space, we computed the expected sampling probabilities through the multinomial moment generating function (Weir, 1979). We then converted those adjusted sampling probabilities to expected haplotype probabilities, and then to terms in the Hill-Robertson basis. For example, in the Hill-Robertson system, for a sample size of n diploid genomes, we have

$$\begin{pmatrix} \tilde{D^2} \\ \tilde{Dz} \\ \tilde{\pi_2} \\ \tilde{H} \end{pmatrix} = \begin{pmatrix} \frac{(n^2-n+1)(n-1)}{4n^3} & \frac{(n-1)^2}{8n^3} & \frac{n-1}{4n^2} & 0 \\ \frac{(n-1)^2}{n^3} & \frac{(n-1)^3}{2n^3} & 0 & 0 \\ \frac{(2n-1)}{4n^3} & \frac{(2n-1)^2}{8n^3} & \frac{(2n-1)^2}{4n^2} & 0 \\ 0 & 0 & 0 & \frac{2n-1}{2n} \end{pmatrix} \begin{pmatrix} \mathbb{E}[D^2] \\ \mathbb{E}[Dz] \\ \mathbb{E}[\pi_2] \\ \mathbb{E}[H] \end{pmatrix},$$

where tildes denote expected $\hat{D}$ statistics corrected for sample size bias.

In practice, we worked with σ_d^2-type statistics of the form $\mathbb{E}[\cdot]/\mathbb{E}[\pi_2(\text{YRI})]$ and compared to the same statistics computed from `moments.LD`. To compute $\mathbb{E}[D^2]$ or any other statistic for a given recombination bin, we sum all contributions of pairs of SNPs with recombination distance falling within that bin, and then divide by the total number of pairs of sites along the genome which could have contributed to that bin, whether they are variable or not. For σ_d^2-type statistics, we don't need to compute the total possible number of pairs per bin, and need only sum all contributions and divide by the total sum of contributions to $\mathbb{E}[\pi_2(\text{YRI})]$.

A.7.4 Bootstraps for likelihood computation

We used a multivariate normal function to estimate likelihood from expected values. For a given recombination bin, we compared the observed vector of LD statistics to expectations computed under the model with adjustments for phasing and sample size. To compute likelihoods, we needed to estimate the covariances of statistics in this vector. To compute the covariance matrix, we divided the genome into 500 regions, each with approximately the same length of intergenic regions. We computed statistics over each of the 500 regions, and then constructed 500 bootstrap replicate sets of statistics by sampling with replacement 500 times. We used these bootstrap replicates to estimate the covariance matrix Σ for each recombination bin. This same set of bootstrap replicates and covariance matrix was used to estimate confidence intervals, as proposed by Coffman *et al.* (2016), to account for non-independence of pairs of SNPs in the same region or pairs with overlapping SNPs.

A.8 Supplementary figures and tables

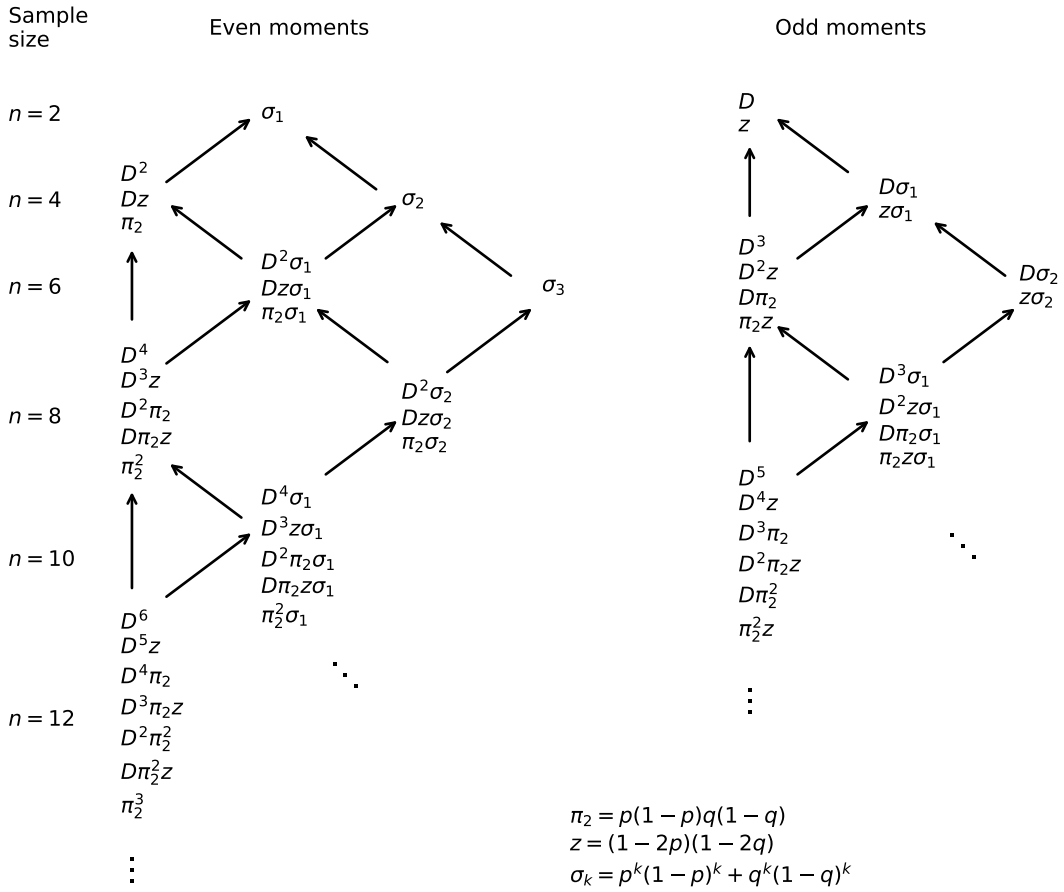

Figure A1: **Hierarchy of moments.** Even and odd moments separate into distinct hierarchical systems. Arrows indicate dependence of sets of moments.

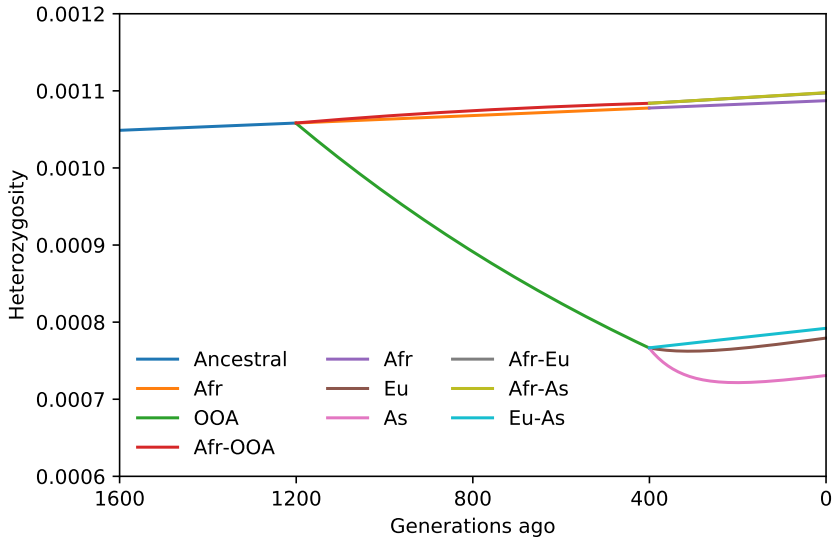

Figure A2: **Within and between population heterozygosity.** Toy model for out-of-Africa expansion, with subsequent migration between split populations. The OOA population experiences a steady decay of heterozygosity due to the prolonged bottleneck, and different bottleneck strengths and exponential growth rates between more recent Eu and As populations account for differences in observed heterozygosity in those populations. Drift does not directly affect cross-population heterozygosity, which increases linearly in the absence of migration, and more slowly with low levels of migration. Strong migration would lead to cross-population H intermediate between the two populations.

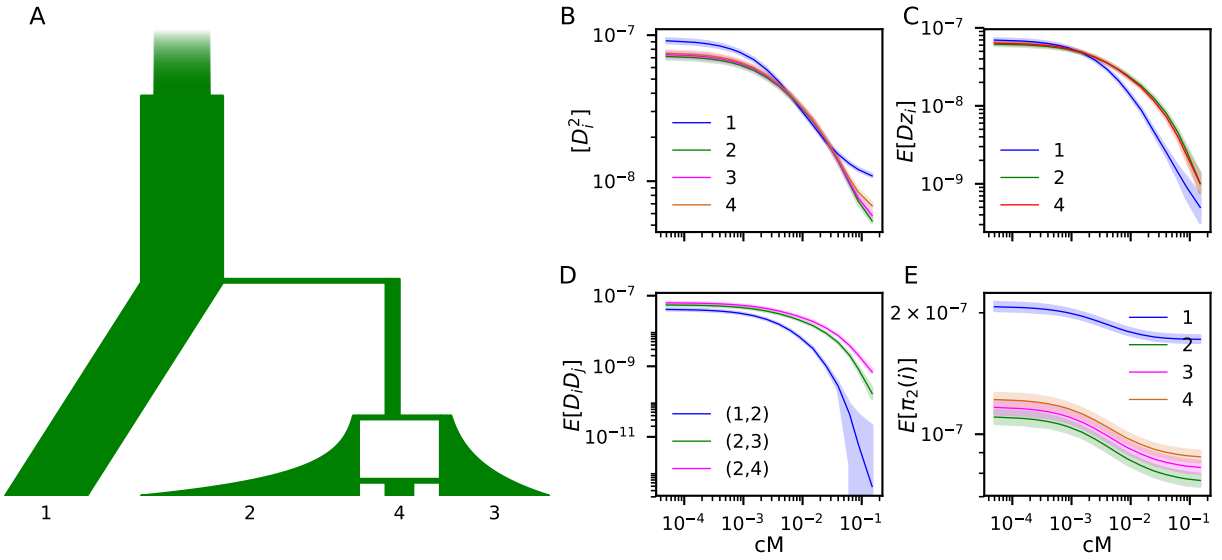

Figure A3: **Validation: computed LD curves match simulation.** (A) We simulated 200 replicates of 100Mb genome using `msprime` (Kelleher *et al.*, 2016) under the illustrated four population demography. Demographic parameters for this simulation were $\nu_A = 1.5$, $T_A = 0.3$, $\nu_B = 0.25$, $T_B = 0.16$, $\nu_{2,0} = 0.1$, $\nu_{2_f} = 4.0$, $\nu_{3,0} = 0.2$, $\nu_{3,f} = 2.0$, $T_3 = 0.06$, $\nu_4 = 0.5$, $T_4 = 0.01$, and $f = 0.5$, where ν_i is the relative size of population i compared to the reference ancestral size, $N_e = 10^4$. $u = 2 \times 10^{-8}$ and $r = 2 \times 10^{-8}$ per base pair. (B-E) Shaded regions indicate 95% confidence intervals of statistics from the 200 simulations, while solid curves are expectations from `moments.LD`.

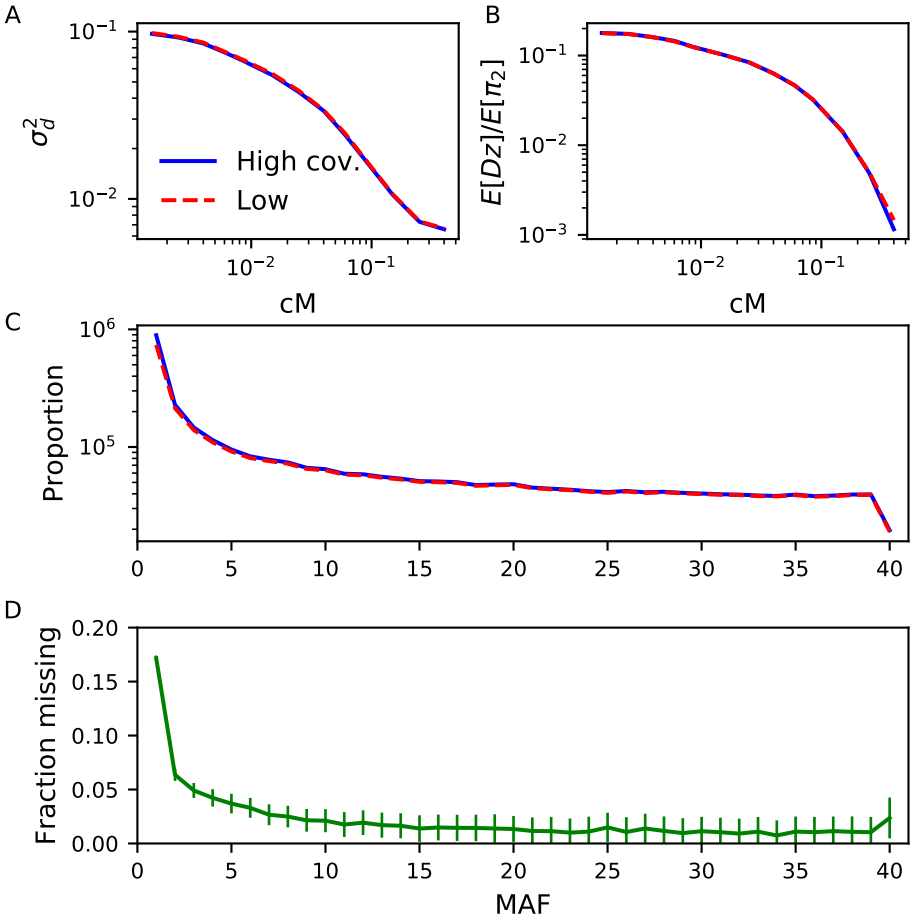

Figure A4: **Effect of low coverage on statistics in the CHB population.** We used 40 individuals that overlapped between the 1000 Genomes data and the 90 Han Chinese data to compute (A) σ_d^2, (B) $\mathbb{E}[Dz]/\mathbb{E}[\pi_2]$ and (C) the folded AFS across intergenic sites. The 90 Han Chinese data was high coverage, while 1000 Genomes data was low coverage. The LD curves are largely unaffected by low coverage, while the singleton bin of the AFS is significantly underestimated in the 1000 Genomes data (D).

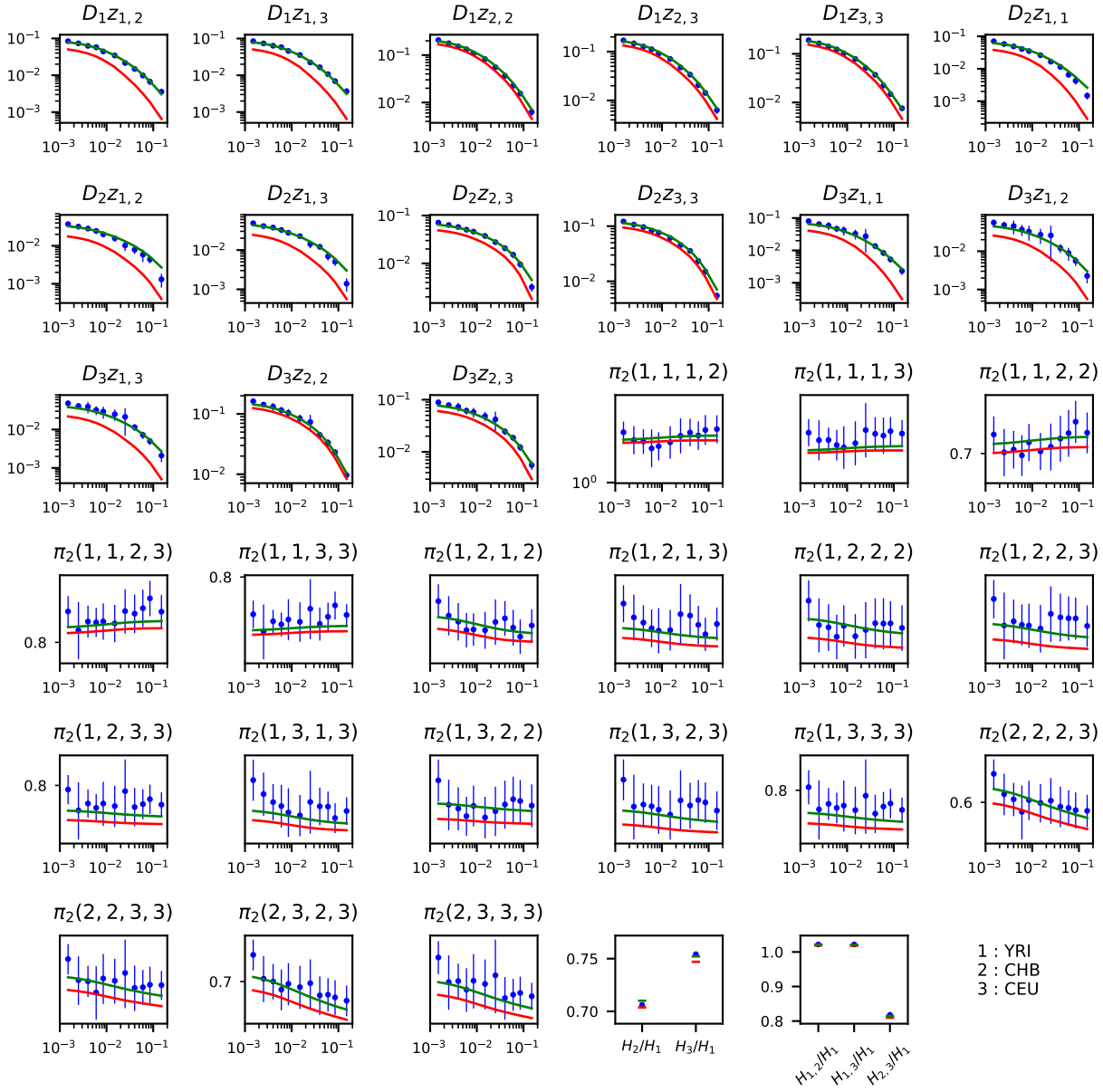

Figure A5: **Additional statistics from model fits.** Figures 3 and 4 show model fits to a handful of statistics in the multi-population basis. Here, we show fits for the remainder of the statistics, and each statistic is normalized by $\pi_2(1,1,1,1) = \pi_2(YRI)$. Indices in the titles indicate populations: YRI is population 1, CHB is population 2, and CEU is population 3. Red curves: standard OOA model. Green curves: OOA model with archaic branches. Error bars on the data indicate 95% confidence intervals of estimates. Best fit parameters and 95% confidence intervals for each are given in Table A2.

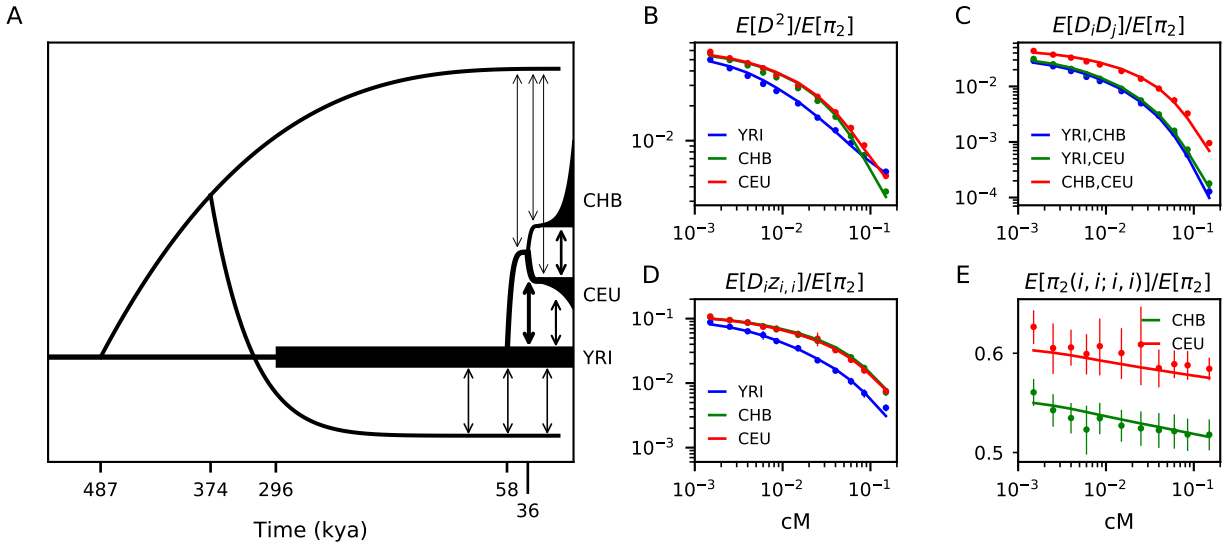

Figure A6: **Alternate topology of archaic branches.** (A) In addition to the scenario where each archaic branch splits independently from the modern human branch (Figure 4), we considered a model where a single archaic lineage splits from modern humans, and then some time later splits into the Eurasian and African archaic branches. Aside from the archaic split times and topology, parameterization was the same between the two models. (B-E) This archaic admixture model provided a good fit the LD data, roughly equal to the archaic admixture model shown in the main text.

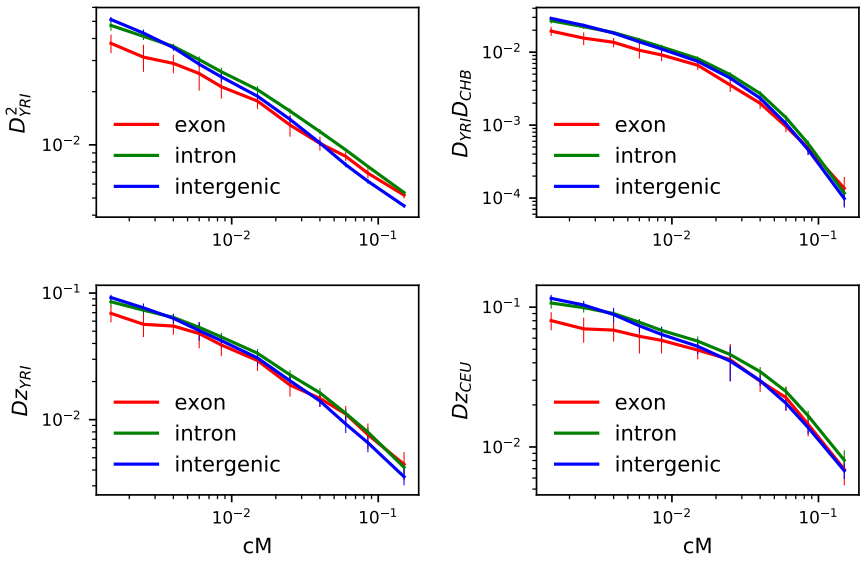

Figure A7: **LD statistics from genome regions.** We compared intergenic data, which we used in our analyses, to LD decay curves from intron and exon regions. Each statistic is normalized by $\pi_2(YRI)$. Exon regions have LD decay curves that differ significantly from intron and intergenic regions, and intron and intergenic regions also differ. Selection is known to affect the expected AFS and LD, so we excluded genic regions from our analyses.

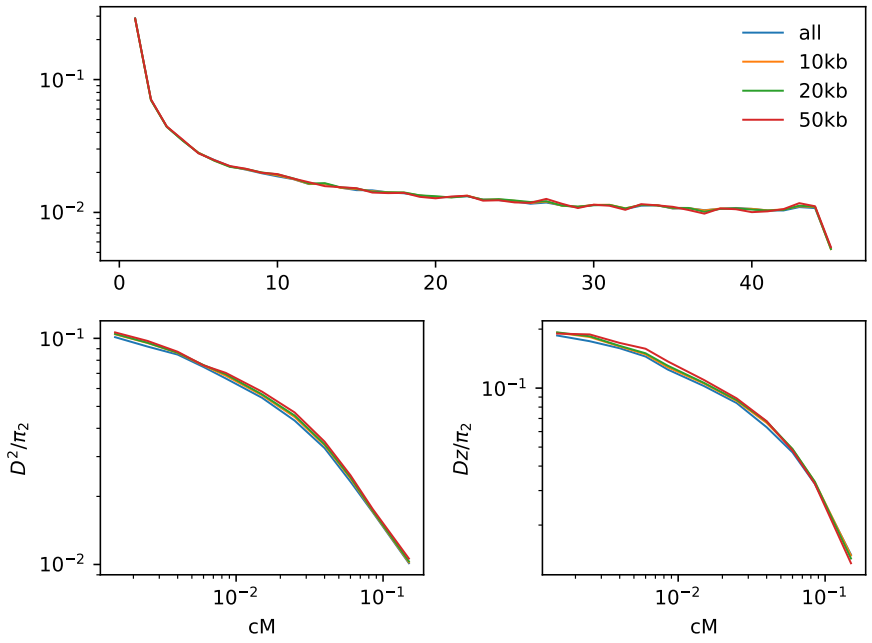

Figure A8: **LD statistics from intergenic regions.** We compared statistics for SNPs across all intergenic regions to SNPs from intergenic regions at least a given distance from the nearest gene. Overall, statistics were similar for each choice. We chose to include all intergenic SNPs in order to increase the number of observed pairs in our data.

	Relative error($\times 10^{-3}$)		Comp. time (minutes)	
Model	∂a∂i	moments	∂a∂i	moments
Equilibrium, $n = 30$, $\rho = 0$	1.8	0.0017	0.25	$< 10^{-3}$
Equilibrium, $n = 30$, $\rho = 10$	15.6	0.40	0.25	$< 10^{-3}$
Equilibrium, $n = 50$, $\rho = 0$	4.6	0.0033	3.5	$< 10^{-3}$
Equilibrium, $n = 50$, $\rho = 10$	34	0.11	3.5	$< 10^{-3}$
Bottleneck, $n = 30$, $\rho = 0$	0.41	0.031	7.8	0.037
Bottleneck, $n = 30$, $\rho = 10$	38.32	0.25	7.9	0.054
Bottleneck, $n = 50$, $\rho = 0$	3.97	0.036	21.8	0.60
Bottleneck, $n = 50$, $\rho = 10$	63.53	0.058	22.3	0.94

Table A1: **Comparison of `moments.TwoLocus` and ∂a∂i`.TwoLocus`.** For the equilibrium distribution, we compared to Hudson's 2001 implementation for $n = 100$ projected down to sample sizes $n = 30$ and $n = 50$. For the bottleneck distribution, we computed a numerical approximation with a larger sample size ($n = 80$) and shorter integration time step using `moments` and then projected to the required size. We measured relative error as $\sum(\Psi_n(\text{model}) - \Psi_n(\text{True}))^2/\Psi_n(\text{True})$. Equilibrium solutions were cached, as is default in both programs, but ∂a∂i still needs to integrate ψ against the multinomial distribution to obtain Ψ_n, accounting for the time differences in the Equilibrium case. With recombination, `moments` improves in accuracy as n increases, since the Jackknife approximation becomes more and more accurate with larger n.

Model	Archaic Introgression B	
Parameter	ML Estimates	95% CI
N_0	3700	$3020 - 4370$
N_{YRI}	14000	$11800 - 16100$
N_{B}	860	$110 - 1610$
N_{CEU0}	2300	$1430 - 3210$
$r_{\text{CEU}}(\%)$	0.122	$0.081 - 0.149$
N_{CHB0}	650	$340 - 960$
$r_{\text{CHB}}(\%)$	0.362	$0 - 0.435$
$m_{\text{AF - B}}(\times 10^{-5})$	53.4	$11.8 - 95.0$
$m_{\text{YRI - CEU}}(\times 10^{-5})$	2.43	$1.62 - 3.24$
$m_{\text{YRI - CHB}}(\times 10^{-5})$	0	—
$m_{\text{CEU - CHB}}(\times 10^{-5})$	11.6	$7.27 - 15.9$
T_{AF} (kya)	296	$244 - 347$
T_{OOA} (kya)	57.7	$35.4 - 80.0$
$T_{\text{CEU - CHB}}$ (kya)	36.3	$28.8 - 43.8$
$T_{\text{Arch. split}}$ (kya)	487	$167 - 807$
$T_{\text{Arch. Af. - Nean. split}}$ (kya)	374	$25.5 - 723$
$T_{\text{Arch. Af. mig.}}$ (kya)	110	$0 - 476$
$m_{\text{AF - Arch. Af.}}(\times 10^{-5})$ ...	2.43	$0 - 5.24$
$m_{\text{OOA - Nean}}(\times 10^{-5})$	1.51	$0 - 3.19$
$T_{\text{Arch. adm. end}}$ (kya)	20.7	$16.3 - 25.1$

Table A2: **Maximum likelihood parameters for alternate archaic topology.** For the most part, estimates were qualitatively similar to the archaic model presented in the main text. In each model, the split between modern humans and the branch leading to the archaic African population occurred about 500 kya. However, here the Neanderthal lineage split from this branch more recently than 500 kya, which is considerably more recent than most estimates or our estimate from the alternative model. This is expected, as the largest discrepancy between the non-archaic model and data occurred for Dz in African populations, so the inference of the split date in this model is primarily driven by the signal in YRI. Without including archaic genomes in this analysis, we did not have statistical power to discriminate between the two proposed topologies, although we speculate that the model independent splits (Figure 4(A)) will prove to be the more likely topology.

Model	OOA (fit w/o Dz)		Archaic Introgression	
Parameter	Estimates	95% CI	Estimates	95% CI
N_0	2360	$2140 - 2580$	2860	$2460 - 3260$
N_{LWK}	14600	$11500 - 17700$	15300	$10400 - 20300$
N_{B}	1130	$700 - 1570$	1020	$730 - 1310$
N_{GBR0}	1560	$760 - 2370$	2210	$1500 - 2920$
$r_{\text{GBR}}(\%)$	0.229	$0.011 - 0.293$	0.157	$0.135 - 0.175$
N_{KHV0}	390	$0 - 980$	630	$480 - 790$
$r_{\text{KHV}}(\%)$	0.680	$0 - 0.888$	0.471	$0 - 0.545$
$m_{\text{AF - B}}(\times 10^{-5})$	48.9	$29.2 - 68.6$	50.7	$41.2 - 60.2$
$m_{\text{LWK - GBR}}(\times 10^{-5})$	2.71	$0 - 8.92$	2.87	$1.45 - 4.29$
$m_{\text{LWK - KHV}}(\times 10^{-5})$	0	$0 - 1.32$	0	—
$m_{\text{GBR - KHV}}(\times 10^{-5})$	9.94	$4.41 - 15.5$	7.29	$0 - 16.8$
T_{AF} (kya)	215	$139 - 290$	249	$219 - 279$
T_{OOA} (kya)	68.2	$52.6 - 83.7$	61.5	$44.0 - 79.0$
$T_{\text{GBR - KHV}}$ (kya)	28.3	$21.9 - 34.8$	30.9	$26.2 - 35.5$
$T_{\text{Arch. Af. split}}$ (kya)	—		511	$456 - 566$
$T_{\text{Arch. Af. mig.}}$ (kya)	—		250	$160 - 341$
$m_{\text{AF - Arch. Af.}}(\times 10^{-5})$	—		0.752	$0.288 - 1.22$
$T_{\text{Nean.}}$ (kya)	—		540	$381 - 700$
$m_{\text{OOA - Nean}}(\times 10^{-5})$	—		0.414	$0 - 0.993$
$T_{\text{Arch. end}}$ (kya)	—		13.0	$4.3 - 21.7$

Table A3: **Models fits to alternate trio.** We fit the same out-of-Africa model with and without archaic branches to a separate trio in the 1000 Genomes data: (Luhya from Kenya (LWK), Kinh from Vietnam (KHV), and British from England and Scotland (GBR)). Best fit parameters compare qualitatively to those fit to the YRI-CHB-CEU data, although confidence intervals were wider for this trio.