NON-EQUILIBRIUM ALLELE FREQUENCY SPECTRA VIA SPECTRAL METHODS

SERGIO LUKIĆ [1], JODY HEY, AND KEVIN CHEN

ABSTRACT. A major challenge in the analysis of population genomics data consists of isolating signatures of natural selection from background noise caused by random drift and gene flow. Analyses of massive amounts of data from many related populations require high-performance algorithms to determine the likelihood of different demographic scenarios that could have shaped the observed neutral single nucleotide polymorphism (SNP) allele frequency spectrum. In many areas of applied mathematics, Fourier Transforms and Spectral Methods are firmly established tools to analyze spectra of signals and model their dynamics as solutions of certain Partial Differential Equations (PDEs). When spectral methods are applicable, they have excellent error properties and are the fastest possible in high dimension. In this paper we present an explicit numerical solution, using spectral methods, to the forward Kolmogorov equations for a Wright-Fisher process with migration of K populations, influx of mutations, and multiple population splitting events.

1. INTRODUCTION

Natural selection is the force that drives the fixation of advantageous phenotypic traits, and represses the increase in frequency of deleterious ones. The growing amount of genome-wide sequence and polymorphism data motivates the development of new tools for the study of natural selection. Distinguishing between genuine selection and the effect of demographic history, such as gene-flow and population bottlenecks, on genetic variation presents a major technical challenge. A traditional population genetics approach to the problem, focuses on computing neutral allele frequency spectra to infer signatures of natural selection as deviations from neutrality. Diffusion theory provides a set of classical techniques to predict such frequency spectra, [8, 21, 4]; while the connection between diffusion and the theory of Partial Differential Equations (PDEs) allows to borrow well established high-perfomance algorithms from applied mathematics.

The theory of predicting the frequency spectrum under irreversible mutation was developed by Fisher, Wright and Kimura [6, 22, 13]. In particular Kimura [14] noted that this theory was applicable to many nucleotide positions and introduced the *infinite sites model*. The joint frequency spectra of neutral alleles can be obtained from the coalescent model [20] or by Monte-Carlo simulations [11]. The analysis in terms of diffusion theory is mathematically simpler and can incorporate natural selection easily, [8, 21, 4]. In this paper, we model the demographic history of K different populations that are descended by $K-1$ population splitting events from a common ancestral population. The populations evolve with gene exchange under an infinite sites mutation model. We introduce a powerful numerical scheme to solve the associated forward diffusion equations. After introducing a regularized discretization of the problem, we show how spectral methods are applied to compute theoretical Non-Equilibrium Frequency Spectra.

Date: November 2, 2010.
[1] Corresponding author.

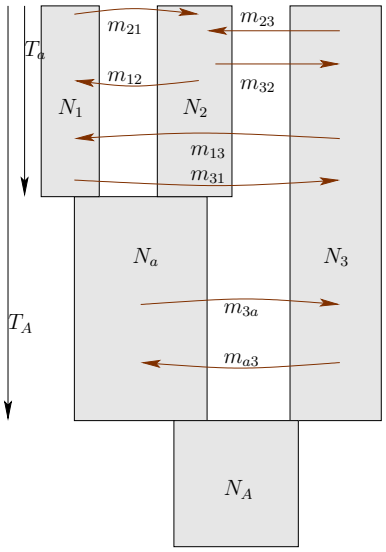

FIGURE 1. A graphical representation of a model for the demographic history of three populations.

The introduction of spectral methods is usually atributed to [16], although they are based on older precursors, such as finite element methods, and Ritz methods in quantum mechanics [17]. The basic idea consists of using finite truncations of expansions by complete bases of functions, to approximate the solutions of a PDE. This truncation allows to transform a diffusion PDE into a finite system of Ordinary Differential Equations (ODEs). The motivation to use these methods relies on their excellent error properties, and their high speed. In general, they are the preferred methods when the dimension of the domain is high [15], and the solutions to the PDE are smooth. This is because the number of basis functions that one needs to approximate the solutions of a PDE, is much lower than the number of grid-points that one needs in a finite difference scheme, working at the same level of accuracy [7].

As we show in this paper, the convergence of spectral methods depends on the smoothness of the solutions to be approximated. In many situations, solutions to diffusion equations have good analytical properties, and spectral methods can be applied. However, the application of these methods to the problem that interests us here, requires a proper discretization of the problem. Influx of mutations, population splitting events and boundary conditions have to be properly regularized before one applies these methods and exploits their high-perfomance properties.

1.1. **Non-Equilibrium Frequency Spectra.** The K-dimensional Allele Frequency Spectrum (AFS) summarizes the joint allele frequencies in K populations. We distinguish between the AFS, which consists of the unknown distribution of allele frequencies in K natural populations, and *observations* of the AFS. Given DNA sequence data from multiple individuals in K populations, the resulting *observation* of the AFS is a K-dimensional matrix with the allele counts (for a complete discussion on this see [20]). Each entry of the matrix consists of the number of diallelic polymorphisms in which the derived allele

was found. Or using other words, each entry of the AFS matrix is the observed number of derived alleles, j_a, found in the corresponding number of samples, n_a, from population a ($1 \leq a \leq K$).

The full AFS, is the real distribution of joint allele frequencies at the time when the samples were collected. If the total number of diploid individuals in the a^{th}-population is $N_a \gg n_a$, the natural allele frequencies $x_1 = i_1/2N_1$, $x_2 = i_2/2N_2$, $\ldots x_k = i_K/2N_K$ (with i_a the total number of derived polymorphisms in the a^{th} population) can be seen as points in the K-dimensional cube $[0,\,1]^K$. Thus, given the frequencies of every diallelic polymorphism (which we indexed by r) x_1^r, x_2^r, $\ldots$, x_K^r, the AFS can be expressed as the probability density function

$$\phi(x) = \frac{1}{S}\sum_{r=1}^{S}\delta(x - x_1^r)\delta(x - x_2^r)\cdots\delta(x - x_K^r). \tag{1.1}$$

Here, S is the total number of diallelic polymorphisms segregating in the K populations, and $\delta(\,)$ is the Dirac delta function.

Our goal is to determine such AFS under the infinite sites model. Any demographic scenario in the model is defined through a population tree topology T, such as in Fig. 1, and a set of paramaters that specify the effective population sizes $N_{e,a}$, splitting times t_A, and migration rates m_{ab} at different times. Hence, $2N_{e,a}m_{ab}$ is defined as the number of haploid genomes that the population a receives from b per generation. For simplicity, we refer to the set of parameters that specify a unique demographic scenario as Θ.

Thus, given a population tree topology and a choice of parameters, we will compute theoretical densities of derived joint allele frequencies as functions on $[0,1]^K$ of the type

$$\phi(x|\Theta,T) = \sum_{i_1=0}^{\Lambda-1}\sum_{i_2=0}^{\Lambda-1}\cdots\sum_{i_K=0}^{\Lambda-1}\alpha_{i_1,i_2,\ldots,i_K}(\Theta,T)R_{i_1}(x_1)R_{i_2}(x_2)\cdots R_{i_K}(x_K), \tag{1.2}$$

with Λ a truncation parameter, $\{R_i(x)\}_{i=0}^{\infty}$ a complete basis of functions on the Hilbert space $L^2[0,1]$ to be defined below, and $\alpha_{i_1\cdots i_K}$ the coefficients associated to the projection of $\phi(x|\Theta,T)$ onto the basis spanned by $\{R_{i_1}(x)R_{i_2}(x)\cdots R_{i_K}(x)\}$. These continuous densities relate to the expectation of an *observation* of the AFS via standard binomial sampling formulae

$$p(j_1|n_1;\ldots;j_K|n_K) = \int_{[0,1]^K}\phi(x|\Theta,T)\prod_{a=1}^{K}\frac{n_a!}{(n_a-j_a)!j_a!}x_a^{j_a}(1-x_a)^{n_a-j_a}dx_a. \tag{1.3}$$

Using Eq. (1.3) we can compare model and data, for instance, by means of maximum likelihood estimates.

The major goals of this paper are twofold. First, we present the finite Markov chain and diffusion approximation, associated to the infinite sites model used to compute neutral allele spectra. A special emphasis is made on the boundary conditions and the influx of mutations, because of their potential singular behavior. Second, we introduce spectral methods and show how to transform the diffusion equations into coupled systems of Ordinary Differential Equations (ODEs) that can be integrated numerically. In particular, we introduce a set of techniques to handle population splitting events, mutations and boundary interactions, that protect the numerical setup against Gibbs phenomena. A detailed analyses of the stability of the methods as function of the model parameters, and the control of the numerical error, is included at the end of the paper.

2. FINITE MARKOV CHAIN MODEL

The evolutionary dynamics of diallelic SNP frequencies in a randomly-mating diploid population, can be modeled using a finite Markov chain, with discrete time t representing non-overlapping generations. For simplicity, we consider first one population with N diploid individuals, and later will extend the results to more than one population.

The state of the Markov chain at time t, is described by the vector $f_j(t)$, with $1 \leq j \leq 2N$. Each entry, $f_j(t)$, consists of the expected number of loci at which the derived state is found on j chromosomes. Therefore, $\sum_{j=1}^{2N-1} f_j(t)$ is the expected number of polymorphic loci segregating in the population at time t, and $f_{2N}(t)$ is the expected number of loci fixed for the derived state. The model assumes that the total number of sites per individual is so large, and the mutation rate per site so low, that whenever a mutation appears, always does it on a previously homoallelic site, [14].

The vector $f_j(t)$, is also called the density of states. Its time evolution under random drift and mutation influx is described by the difference equations

$$f_j(t+1) = \sum_{i=1}^{2N} P(j|i) f_i(t) + \mu_j(t), \quad 1 \leq j \leq 2N. \tag{2.1}$$

In its simplest form, one assumes that the alleles in generation $t+1$ are derived by sampling with replacement from the alleles in generation t. Therefore, the transition coefficients in the chain Eq. (2.1) are

$$P(j|i) = \binom{2N}{j} (i/2N)^j \{1-(i/2N)\}^{2N-j}. \tag{2.2}$$

This describes stochastic changes in the state after a discrete generation, Fig. 2. The second term in Eq. (2.1) represents the influx of polymorphisms. Mutations are responsible for the creation of new polymorphisms in the population. The influx of mutations depends on the expected number of sites $2N\nu$, in which new mutations appear in the population each generation[1]. If we assume that each generation, every new mutation is found in just one chromosome, then

$$\mu_j(t) = 2N\nu\delta_{1,j}, \tag{2.3}$$

for the mutation alone, [4]. The term $\delta_{i,j}$ in Eq. (2.3) is the Kronecker symbol, with $\delta_{1,j} = 1$ if $j = 1$ and $\delta_{1,j} = 0$ otherwise.

2.1. **Effective Mutation Densities.**

In applications of the infinite sites model, one usually finds that the census population size and the effective population size that drives random drift in Eq. (2.2), are not the same, [14]. For this reason, we distinguish between N_e, the effective population size that defines the variance of the Wright-Fisher process in Eq. (2.2), from the census population size N that can be used to define the allele frequencies $x = i/2N$. Therefore, the smallest frequency, $x = 1/2N$, with which new mutations enter populations will be sensitive to small stochastic fluctuations in the census population size, even if the effective population size remains constant. This is important when we take the diffusion limit of Eq. (2.1), and the stochastic process is described by the continuous

[1]The expected number of sites $2N\nu$, relates to the expected number of mutations per base $2Nu$, by the total length L of the genomic sequence under study in units of base pairs, $\nu = u \times L$. Sometimes in this paper, in an abuse of notation we do not distinguish between ν and u, and they are seen as the same quantity expressed with different units.

variable $x = j/2N$, rather than the integer j. If we consider a constant census population size, the term Eq. (2.3) in the Markov chain is substituted by

$$\delta_{1,j} \mapsto \delta(x - 1/2N), \tag{2.4}$$

in the diffusion limit. However, if the census population size per generation is a stochastic variable distributed as $F(N)dN$, the diffusion limit of the mutation term will be

$$\delta_{1,j} \mapsto \mu(x) = \int_0^\infty \delta(x - 1/2N)F(N)dN. \tag{2.5}$$

We expect that $\mu(x)$ will have some general properties, independent of the particular characteristics of $F(N)dN$. For instance, in many realistic scenarios $\mu(x)$ will be a function that is nearly zero for frequencies $x > x_*$, with $x_* = 1/2N_{min}$ a very small characteristic frequency associated to the inverse of the minimum census population size.

Other phenomena that might not be properly captured by the simple mutational model in Eqs. (2.3) and (2.4), consist of organisms with partially overlapping generations, and organisms where mutations in gametes arise from somatic mutations. When an organism has a mating pattern that violates the assumption of non-overlapping generations (e.g. humans), the generation time in the model Eq. (2.2) is interpreted as an *average generation time*. Hence, during a generation unit, there is time enough for some individuals to be born with new mutations at the beginning of the generation time, and to reproduce themselves by the end of a generation unit. This implies that after one average generation, there can exist new identical mutations in more than one chromosome. Similarly, when the gametes of an organism originate from somatic tissue, they inherit de novo mutations that arised in the soma after multiple cell divisions. If the individuals of such organism can have more than one offspring per generation, one expects to find the same new mutation, in the same site, in more than one chromosome per generation.

Because of these different biological phenomena, we believe that the notion of *effective mutation density*, $\mu(x)$, is a more accurate way to describe mutations in natural populations. The *effective mutation density* describes the average frequency distribution of new mutations per generation, in one population, after taking into account the effects due to stochastic changes in census population size, non-overlapping generations and/or mutations of somatic origin. The major motivation here to use effective mutation densities rather than Dirac deltas, is because of their good behavior under spectral methods. As we show later when we discuss the continuous limit of the infinite sites model, different effective mutation densities can yield predictions which are identical to predictions of models based on Eq. (2.4).

2.2. **More than one population.** Here, we show how to incorporate arbitrarily more populations, and migration flow between them. Generally, for the state in the chain we consider a discrete random variable $\vec{X}$ which takes values in the K-dimensional lattice of derived allele frequencies:

$$\vec{X} = \begin{pmatrix} i_1/2N_1 \\ i_2/2N_2 \\ \vdots \\ i_K/2N_K \end{pmatrix}, \tag{2.6}$$

with K the number of populations, and $0 \leq i_a \leq 2N_a$. For simplicity, we use a single index notation, $0 \leq I \leq \prod_a N_a$, to label the states where the random variable $\vec{X}$ takes values. The random variable $\vec{X} = I$ jumps to the state $\vec{X} = J$ at a discrete generation

unit, with prescribed probability $P(J|I)$. The density of states in this multi-population setup is $f_I(t)$, and the difference equations that describe its dynamics are equivalent to Eq. (2.1). The transition matrix $\hat{P} = P(J|I)$, incorporates random drift and migration events

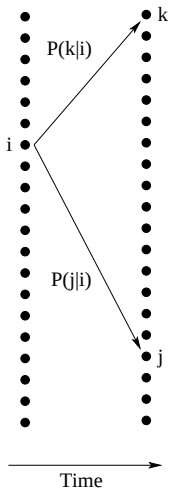

FIGURE 2. One unit of time transition in a finite Markov chain.

between populations. New mutations enter each population with an effective mutation vector $\vec{\mu}$

$$\vec{\mu} = \begin{pmatrix} 0 \\ \vdots \\ \mu_j^a \\ \vdots \\ 0 \end{pmatrix}; \tag{2.7}$$

in the standard model, the mutation density is $\mu_j^a = 2N_a\nu\delta_{1,j}$.

The Markov chain for a Wright-Fisher process for two independent populations is defined by the transition matrix $\hat{P}_{j_1 j_2; i_1 i_2} =$

$$Bi(j_1; 2N_{e,1}, i_1/2N_{e,1})Bi(j_2; 2N_{e,2}, i_2/2N_{e,2}), \tag{2.8}$$

where $Bi(j; k, p)$ stands for the binomial distribution with index k and parameter p. Also, we can introduce migration between populations, by sampling a constant number of alleles n_{ab} in population a that become part of the allele pool in population b. Thus, in a model of two populations with migration, one considers the transition matrix

$$\begin{aligned} \hat{P}_{j_1 j_2; i_1 i_2} = \sum_{k_1,k_2=0}^{k_1=n_{21},k_2=n_{12}} & Bi(j_1 - k_1; 2N_{e,1} - n_{21}, i_1/2N_{e,1})Bi(k_1; n_{21}, i_2/2N_{e,2}) \times \\ & \times Bi(j_2 - k_2; 2N_{e,2} - n_{12}, i_2/2N_{e,2})Bi(k_2; n_{12}, i_1/2N_{e,1}). \end{aligned} \tag{2.9}$$

This process is generalizable to an arbitrary number of populations in a straightforward way.

3. DIFFUSION APPROXIMATION

Diffusion approximations to finite Markov chains have a distinguished history in population genetics, dating back to Wright and Fisher. This approach can be used to describe the time evolution of derived allele frequencies in several populations, with relatively large population sizes. This approximation applies when the population sizes N_a are large (if $N_e > 50$, the binomial distribution with index $2N_e$ can be accurately approximated by the Gaussian distribution used in the diffusion limit) and migration rates are of order $1/N_e$.

In the large population size limit, the state space spanned by vectors such as Eq. (2.6), converges to the continuous space $[0,\,1]^K$. The density of states $f_I(t)$ on the state space, will converge to a continuous density $\phi(x,t)$ on $[0,\,1]^K$. The time evolution of $\phi(x,t)$ depends on how the inifinitesimal change $\delta\vec{X}$,

$$\vec{X}(t+\delta t) = \vec{X}(t) + \delta\vec{X},$$

is distributed. If the mean of the $\delta\vec{X}$ distribution is $M(\vec{X},t)$ and the covariance matrix is $C(\vec{X},t)$, the time continuous limit $\delta t \to 0^+$ of the process $\vec{X}(t)$ is well defined. In the small, but finite, limit of δt the stochastic process obeys the equation

$$\vec{X}(t+\delta t) = \vec{X}(t) + M(\vec{X},t)\delta t + \sigma(\vec{X},t)\vec{\epsilon}\sqrt{\delta t}, \tag{3.1}$$

where $\vec{\epsilon}$ is a standard normal random variable (with zero mean and unit covariance matrix) in $\mathbb{R}^K$, $\sigma(\vec{X},t)$ is a root square of the covariance matrix $C(\vec{X},t) = \sigma\sigma^T(\vec{X},t)$, and δt is a finite, but very small, time step.

In the diffusion approximation to the discrete Markov chain, the process is described as a time continuous stochastic process governed by Gaussian jumps of prescribed variance and mean. These processes can be denoted using the notation of stochastic differential equations:

$$dX_t^a = M^a(X_t,t)dt + \sum_{b=1}^{K}\sigma^{ab}(X_t,t)dW_t^b, \tag{3.2}$$

where dW^b is the infinitesimal element of noise given by standard Brownian motion in K-dimensions, and σ is the root square matrix of the covariance matrix $C = \sigma\sigma^T$, [19]. The diffusion generator associated to Eq. (3.2) is

$$\mathcal{L} = \sum_{a=1}^{K} M^a(x,t)\frac{\partial}{\partial x^a} + \frac{1}{2}\sum_{b=1}^{K} C^{ab}(x,t)\frac{\partial^2}{\partial x^a \partial x^b}. \tag{3.3}$$

Thus, if $\phi(x,t=0)$ is the density of allele frequencies at time 0, the time evolution of $\phi(x,t)$ will be governed by the forward Kolmogorov equation

$$\frac{\partial\phi(x,t)}{\partial t} = \sum_{a,b=1}^{K}\frac{1}{2}\frac{\partial^2}{\partial x^a\partial x^b}\left[C^{ab}(x,t)\phi(x,t)\right] - \sum_{a=1}^{K}\frac{\partial}{\partial x^a}\left[M^a(x,t)\phi(x,t)\right] + \rho(x,t), \tag{3.4}$$

with $\rho(x,t)$ an inhomogeneous terms, continuous limit of μ_j in Eq. (2.1), that describes the net influx of polymorphisms in the population per generation.

3.0.1. *Modeling migration flow and random drift.* The continuous limit of the Markov chain defined in Eq. (2.9), in the case of K diploid populations, and in the weak migration

limit has as associated momenta

$$M^a(x,t) = \sum_b m_{ab}(x_b - x_a), \tag{3.5}$$

$$C^{ab}(x,t) = \delta^{ab}\frac{x_a(1-x_a)}{2N_{e,a}}, \tag{3.6}$$

with δ^{ab} the Kronecker delta ($\delta^{ab} = 1$ if $a = b$ and $\delta^{ab} = 0$ otherwise). The matrix element $m_{ab} = n_{ab}/(2N_a)$ defines the migration rate from the b^{th} population to the a^{th} population.

Thus, associated to such process one has the Kolmogorov forward equations

$$\frac{\partial}{\partial t}\phi(x,t) = \sum_{a,b}\frac{1}{2}\frac{\partial^2}{\partial x_a \partial x_b}\left(\delta^{ab}\frac{x_a(1-x_a)}{2N_{e,a}}\phi(x,t)\right)$$
$$-\frac{\partial}{\partial x_a}\left(m_{ab}(x_b - x_a)\phi(x,t)\right) + \rho(x,t). \tag{3.7}$$

Eq. (3.7) describes the time evolution of the frequency spectrum density under random drift and migration events between populations, given an initial density and absorbing boundary conditions (see below). The inhomogenous term $\rho(x,t)$, models the total incoming/outgoing flow of SNPs per generation into the K-cube which is not due to the diffusion flow, $j^a = -M^a\phi + \partial^b(C^{ab}\phi)$, at the boundary. This total flow depends on mutation events that generate *de novo* SNPs, coming inflow from higher dimensional components of the allele density (see below), coming inflow from migration events from lower dimensional components of the allele density, and the outflow of migration events towards higher dimensional components. If there is not a positive influx of SNPs, the density would converge to $\phi(x,t) \to 0$ as $t \to \infty$. In order to understand the probability flow between different components of the density of alleles, we will have to study how the boundary conditions are defined precisely.

3.1. **Boundary Conditions.** Understanding the boundary conditions in this problem is one of the most challenging tasks. In the famous solution of Kimura to the problem of pure random drift in one population, [12], Kimura required the solutions to the diffusion equation to be finite at the boundaries $x = 0$ and $x = 1$. This boundary condition is absorbing. The points $x = 0$ and $x = 1$ describe states where SNPs reach the fixation of their ancestral or derived states.

If we consider K populations, the natural generalization of Kimura's boundary conditions can be derived by studying the possible stochastic histories of single diallelic SNPs segregating in the K populations. A SNP which is initially polymorphic in all the K populations, can reach the fixation of its derived or ancestral state in one population while still being polymorphic in the remaining $K-1$ populations. More generally, a SNP can be polymorphic in $K-\alpha$ populations, while its state can be fixated in the remaining α populations. A convenient way of visualizing this, is by looking at the geometry of the K-cube of allele frequencies, and the different subdimensional components of its boundary (see examples Fig. 3 and Fig. 4 for the 2-cube and 3-cube). A K-cube's boundary can be decomposed as a set of cubes of lower dimensionality, from $(K-1)$-cubes up to 0-cubes or points. The number of boundary components of codimension α, i.e. the number of $(K-\alpha)$-cubes, contained in the boundary of the K-cube is

$$\#\,(K-\alpha)\,\text{cubes} = \frac{2^\alpha K!}{(K-\alpha)!\alpha!}. \tag{3.8}$$

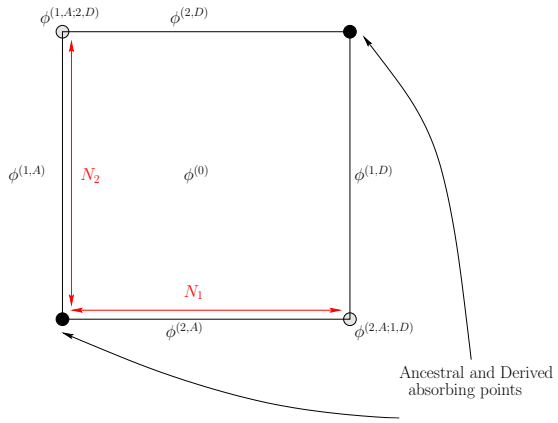

FIGURE 3. Decomposition of the singular probability density, for two populations, on the two dimensional bulk and the different subdimensional boundary components.

The most important set of boundary components are the $(K-1)$-cubes, because any other boundary component can be expressed as the intersection of a finite number of $(K-1)$-cubes at the boundary. We identify each $2K$ codimension-one boundary components, by the population where the SNPs are not polymorphic, and by the state that is fixated in such population (Derived or Ancestral). For example, the component (i, A) is defined as the set of points in the K-cube that obeys the equation $x_i = 0$, and the component (i, D) is defined by the equation $x_i = 1$. Therefore, any codimension α boundary component can be expressed as the intersection

(3.9)

$$(i_1\, S_1)\cap(i_2\, S_2)\cap\cdots(i_\alpha\, S_\alpha) = \{x \in [0,1]^K | x_{i_1} = \delta_{S_1\, D};\, x_{i_2} = \delta_{S_2\, D}; \ldots; x_{i_\alpha} = \delta_{S_\alpha\, D}\},$$

with $i_\alpha \neq i_\beta$ when $\alpha \neq \beta$, $\delta_{S,D} = 1$ for the derived state $S = D$, and $\delta_{S,D} = 0$ for the ancestral state $S = A$.

To each boundary component of codimension α we associate a $(K-\alpha)$-dimensional density of derived allele frequencies that are polymorphic just on the corresponding $K-\alpha$ populations, while are fixated in the other α populations. In this way, $\phi^{(0)}$ denotes the bulk probability density, $\{\phi^{(i,S)}\}_{i=1;S=A,D}^{i=K}$ are the $2K$ codimension one densities, $\{\phi^{(i,S_i;j,S_j)}\}_{i\neq j;S_i,S_j=A,D}$ the codimension 2 densities, etc. This decomposition is illustrated in the case of 2 and 3 populations in Fig. 4 and Fig. 3.

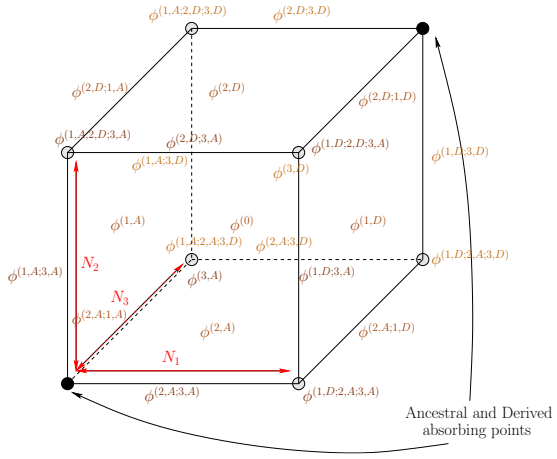

FIGURE 4. Decomposition of the singular probability density, for three populations, on the three dimensional bulk and the different subdimensional boundary components.

In this notation, we write the density of derived alleles segregating on K populations as the generalized probability function

$$\phi(x,t) = \phi^{(0)}(x,t) + \sum_{i=1}^{K}\Big(\phi^{(i,A)}(x,t)\delta(x_i) + \phi^{(i,D)}(x,t)\delta(x_i-1)\Big) + \sum_{i=1,j\neq i}^{K}\Big(\phi^{(i,A;j,A)}(x,t)\delta(x_i)\delta(x_j) + \phi^{(i,A;j,D)}(x,t)\delta(x_i)\delta(x_j-1) + \phi^{(i,D;j,D)}(x,t)\delta(x_i-1)\delta(x_j-1)\Big) + \sum_{i=1,j\neq i,k\neq i,k\neq j}^{K}\ldots \tag{3.10}$$

with $\delta(\cdot)$ the Dirac delta-function. The points $(1,\,A)\cap(2,\,A)\cap\cdots(K,\,A)$ and $(1,\,D)\cap(2,\,D)\cap\cdots(K,\,D)$ are the universal fixation boundaries, and they do not contribute to the total density of alleles in Eq. (3.10). It is useful to write the probability density $\phi(x,t)$ using such decomposition, because despite of being a singular generalized function, each boundary component $\phi^{(i,S_1;j,S_2,\ldots)}(x,t)$ will be, most of the times, a regular analytic function.

The dynamics of the boundary components $\phi^{(i_1,S_1;i_2,S_2,\ldots)}(x,t)$ are governed by diffusion equations, with an inhomogenous term, of the type

$$\frac{\partial}{\partial t}\phi^{(i_1,S_1;i_2,S_2,\ldots)}(x,t) = \sum_{a,b}\frac{1}{2}\frac{\partial^2}{\partial x_a\partial x_b}\left(\delta^{ab}\frac{x_a(1-x_a)}{2N_{e,a}}\phi^{(i_1,S_1;i_2,S_2,\ldots)}(x,t)\right) - \frac{\partial}{\partial x_a}\Big(m_{ab}(x_b-x_a)\phi^{(i_1,S_1;i_2,S_2,\ldots)}(x,t)\Big) + \rho^{(i_1,S_1;i_2,S_2,\ldots)}(x,t), \tag{3.11}$$

with $\rho^{(i_1,S_1;i_2,S_2,\ldots)}(x,t)$ the net incoming/outgoing flow into the boundary component $(i_1,S_1)\cap(i_2,S_2)\cap\ldots$. The ρ term can be interpreted as an interaction term between different boundary components.

More precisely, $\rho^{(i_1,S_1;i_2,S_2;\ldots)}(x,t)$ consists of four terms

$$\rho^{(i_1,S_1;i_2,S_2;\ldots;i_\alpha,S_\alpha)}(x,t) = \rho_{mut}(x,t)+\rho_{drift}(x,t)+\rho_{in_m}(x,t)+\rho_{out_m}(x,t). \tag{3.12}$$

$\rho_{mut}(x,t)$ is the influx of spontaneous mutations (only present in codimension $K-1$), $\rho_{drift}(x,t)$ consists of the boundary inflow from codimension $\alpha-1$ components that have $(i_1,S_1)\cap(i_2,S_2)\cap\ldots$ as a boundary component, $\rho_{in_m}(x,t)$ represents the incoming flow due to migration events from lower dimensional boundary components, and $\rho_{out_m}(x,t)$ is the outflow due to migration events from $(i_1,S_1)\cap(i_2,S_2)\cap\ldots(i_\alpha,S_\alpha)$ towards higher dimensional components that have $(i_1,S_1)\cap(i_2,S_2)\cap\ldots(i_\alpha,S_\alpha)$ as a boundary component.

We can write in a more precise way each term in $\rho^{(i_1,S_1;i_2,S_2,\ldots)}(x,t)$, as follows:

- ρ_{mut}:

$$\rho_{mut}^{(i_1,S_1;i_2,S_2;\ldots;i_\alpha,S_\alpha)}(x,t) = \sum_a 2N_{e,a}u\delta_{\alpha,K-1}\mu(x_a)\prod_{j=1}^{\alpha}(1-\delta_{i_j,a}), \tag{3.13}$$

  with $\delta_{\alpha,K-1}=1$ if $\alpha=K-1$, $\delta_{\alpha,K-1}=0$ if $\alpha\neq K-1$, and $\mu(x_a)$ is the mutation density (in the classical theory, $\mu(x_a)=\delta(x_a-1/2N_{e,a})$).
- ρ_{drift}:

$$\rho_{drift}^{(i_1,S_1;i_2,S_2;\ldots;i_\alpha,S_\alpha)}(x,t) = \sum_{j_\alpha=i_1}^{i_\alpha}\sum_a\Big(\delta_{S_{j_\alpha},A}\Big(\frac{1}{4N_{e,a}}-\sum_b m_{ab}x_b\Big) \tag{3.14}$$

$$+\delta_{S_{j_\alpha},D}\Big(\frac{1}{4N_{e,a}}-\sum_b m_{ab}(1-x_b)\Big)\Big)\phi^{(j_1,S_1;j_2,S_2;\ldots;j_{\alpha-1},S_{\alpha-1})}(x_{j_\alpha}=\delta_{S_{j_\alpha},D},x,t),$$

  where the sum over j_α is over all components of codimension $\alpha-1$ that have $(i_1,S_1)\cap(i_2,S_2)\cap\ldots(i_\alpha,S_\alpha)$ as a boundary component. The sum over a and b is over all populations that are not $i_1,i_2,\ldots i_\alpha$.
- ρ_{in_m}: Here, and throughout, c_α is a shorthand for the boundary component $(i_1,S_1;i_2,S_2;\ldots;i_\alpha,S_\alpha)$. $\rho_{in_m}^{(c_\alpha)}(x,t)$ represents the total incoming flow due to migration events of SNPs that are contained in densities of SNPs located at boundary components of c_α. If $\mathcal{B}_d(c_\alpha)$ is the set of boundary components of c_α with fixed codimension d ($\alpha<d\leq K$),

  $$\mathcal{B}_d=\{(i_1,S_1;i_2,S_2;\ldots;i_\alpha,S_\alpha;j_{\alpha+1},S_{\alpha+1};\ldots j_d,S_d)\}_{j_{\alpha+1},\ldots,j_d},$$

  then $\rho_{in_m}^{(c_\alpha)}(x,t)$ can be written as the sum of contributions from all boundary components in $\mathcal{B}_d$, for all codimensions $d=\alpha+1,\alpha+2,\ldots,P$, and for all possible migration events from elements q in $\mathcal{B}_d(c_\alpha)$ to c_α:

$$\rho_{in_m}^{(bc_\alpha)}(x,t) = \sum_{d=\alpha+1}^{K}\sum_{q\in\mathcal{B}_d}\phi^{(q)}(x,t)\times\left(\sum_{e\in\Gamma(q\to c_\alpha)}p(e)\prod_{k=\alpha+1}^{d}\delta(x_{i_k}-f_{i_k}^e)\right). \tag{3.15}$$

  Here, $\Gamma(q\to c_\alpha)$ is the set of all possible migrations events from SNPs in $\phi^{(q)}$ to $\phi^{(c_\alpha)}$, $p(e)$ denotes the probability of occurence of the migration event e, and $0<f_i^e\in c_\alpha$ denotes the expected frequency, in the i^{th}-population, of a SNP that enters c_α after the event e. We provide below a more precise description of Eq. (3.15), such as a description of the event space $\Gamma(q\to c_\alpha)$, the corresponding probabilities of occurence and expected frequencies.

- ρ_{out_m}: Denotes the outflow of SNPs due to migration events to higher dimensional boundary components. In other words, $\rho_{out_m}^{(c_\alpha)}$ measures the rate of loss of SNPs in $\phi^{(c_\alpha)}$, because of migration flow towards boundary components of codimension $d < \alpha$, that have c_α as a boundary component. Let $\mathbf{I}_{\partial q, c_\alpha}$ be a discrete function that returns 1 when c_α is a boundary component of q, and zero when it is not. Thus,

$$\rho_{out_m}^{(c_\alpha)}(x,t) = -\phi^{(c_\alpha)}(x,t) \times \sum_{d=0}^{\alpha-1} \sum_{q \in \mathcal{B}_d} \mathbf{I}_{\partial q, c_\alpha} \left(\sum_{e \in \Gamma(c_\alpha \to q)} p(e) \right). \tag{3.16}$$

To compute Eq. (3.15) and Eq. (3.16), it is unavoidable to use approximations. In principle, one could use the transition probabilities of the finite Markov chain to estimate the probabilities of different migration events and their expected allele frequencies. However, there is a simpler approximation, which is consistent with the weak migration limit in which the diffusion equation is derived.

This approximation follows from the observation that at the boundary $x_a = \delta_{S,D}$, the strength of random drift along the population a vanishes ($x_a(1-x_a) = 0$), and hence, the infinitesimal change in x_a obeys a deterministic equation:

$$\frac{dx_a}{dt} = \sum_b \delta_{S,A} m_{ab} x_b - \delta_{S,D} m_{ab}(1 - x_b). \tag{3.17}$$

Eq. (3.17) implies that a migration event from several populations b, to a target population a, can push the frequency x_a of a SNP, out of the boundary where it was initially fixated ($x_a = \delta_{S,D}$).

Therefore, given a K-cube, a boundary component c_α (of codimension α), and a boundary component q (of codimension $\beta > \alpha$) of c_α, we say that there will exist migration flow from q to c_α, if and only if

$$\frac{dx_{a_t}}{dt} = \sum_{b=1}^{K} \delta_{S_{a_t},A} m_{a_t b} x_b - \delta_{S_{a_t},D} m_{a_t b}(1 - x_b) \neq 0, \tag{3.18}$$

$$\frac{dx_n}{dt} = \sum_{b=1}^{K} \delta_{S_n,A} m_{nb} x_b - \delta_{S_n,D} m_{nb}(1 - x_b) = 0, \tag{3.19}$$

where $\{x_n\}_{n=1}^{\alpha}$ denote the allele frequencies of SNPs which are fixated at the boundary components c_α and q, $\{a_t\}_{a_t=\alpha+1}^{K}$ are the populations at c_α whose allele frequencies are polymorphic, and the frequencies x_b are defined at the boundary component q (which means that x_b is polymorphic as long as $b > \beta$, and is $\delta_{S_b,D}$ otherwise). It is important to realize that x_b can be 0 or 1 at q, and a migration event to c_α can still bring alleles of the opposite state that is fixated in the target population.

In this approximation, $\Gamma(q \to c_\alpha)$ consists of a single element, and $p(e)$ can be zero or one. If Eq. (3.18) and Eq. (3.19) are satisfied, the migration event in $\Gamma(q \to c_\alpha)$ has probability $p(e) = 1$, and the expected frequencies are

$$f_{a_t} = \sum_{b=1}^{K} \delta_{S_{a_t},A} m_{a_t b} x_b - \delta_{S_{a_t},D} m_{a_t b}(1 - x_b). \tag{3.20}$$

If Eq. (3.18) and Eq. (3.19) are not satisfied, $p(e) = 0$, and we say that there is not migration flow from q to c_α.

3.2. **Effective Mutation Densities.** Given a constant spontaneous mutation rate in the species under study, of u "base substitutions per site and per generation," and expected number of sites $\nu = L \times u$ where new mutations appear in the population each generation, the total number of *de novo* mutant sites that appear in the population a, every generation, is $2N_{e,a}\nu$. We can model this constant influx of mutations by adding a Dirac delta term

$$2N_{e,a}\nu\delta(x_a - 1/2N_{e,a}), \tag{3.21}$$

to the K diffusion equations that govern the ancestral components of codimension $K-1$ $\phi^{(i_1\,A)\cap(i_2\,A)\cap\cdots(i_{K-1}\,A)}(x,t)$. However, as we discussed above, more generally we work with a effective mutation density

$$2N_{e,a}\nu\mu(x_a). \tag{3.22}$$

As a particular example of an effective mutation density, we consider a stochastic census population size, which is a random variable distributed as

$$F(N)dN = \frac{c}{2N^2}\exp(-\kappa/2N)dN. \tag{3.23}$$

This distribution avoids extremely small populations by an exponential tail, while large population sizes are distributed as $\sim N^{-2}$, as shown in Fig. 5. In this model, we keep constant the effective population size N_e that defines the variance of random drift in Eq. (3.6). Thus, the mutation density will be

$$\mu(x) = \int_0^\infty \delta(x - 1/2M)\frac{c}{2M^2}\exp(-\kappa/2M)dM. \tag{3.24}$$

We can integrate exactly Eq. (3.24), by making the change of variables $y = 1/2M$, $dM = -dy/2y^2$:

$$\mu(x) = c\exp(-\kappa x). \tag{3.25}$$

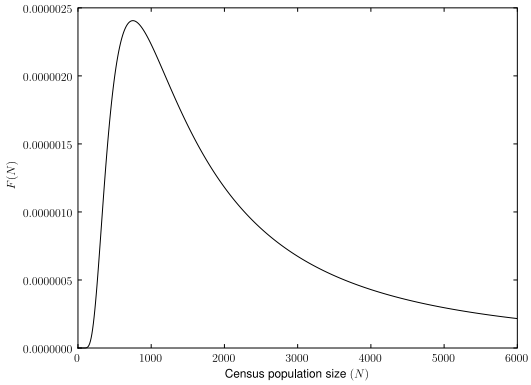

FIGURE 5. A model for a stochastic census population size, with exponential decay in the small population size limit, a quadratic decay $\sim N^{-2}$ in the large population size limit, and a population peak at $N = 1000$.

3.3. **Population splitting events.** So far we have studied how the allele frequency density changes as function of time while the number of populations K remains constant. When two populations split, the diffusion jumps to dimension $K+1$, and the corresponding density will obey the time evolution defined by Eq. (3.7) for $K+1$ populations, with different population sizes and migration parameters. The initial density $\phi_{K+1}(x, x_{K+1}, t)$ in the $K+1$ diffusion problem is determined from the density $\phi_K(x,t)$, before the populations split. Therefore, if population $K+1$ was formed by $N_{f,a}$ migrant founders from the a^{th} population, then

$$\phi_{K+1}(x, x_{K+1}, t) = \phi_K(x,t)\sqrt{\frac{N_{f,a}}{\pi x_a(1-x_a)}}e^{-N_{f,a}(x_a - x_{K+1})^2/(x_a(1-x_a))}. \tag{3.26}$$

In the limit $N_{f,a} \to \infty$, Eq. (3.26), simplifies to

$$\phi_{K+1}(x, x_{K+1}, t) = \phi_K(x,t)\delta(x_a - x_{K+1}), \tag{3.27}$$

with $\delta(x)$ the Dirac delta. Additionally, if the new population is formed by migrants from two populations merging, with a proportion f from population i and $1-f$ from population b, then

$$\phi_{K+1}(x, x_{K+1}, t) = \phi_K(x,t)\delta(fx_a + (1-f)x_b - x_{K+1}). \tag{3.28}$$

In the diffusion framework, one can also deal with populations that go extinct or become completely isolated. More precisely, if we remove the a^{th} population, the initial density in the $K-1$ dimensional problem will be

$$\phi_{K-1}(\tilde{x}, t) = \int_{[0,1]} \phi_K(x,t)dx_a. \tag{3.29}$$

4. Solution to the diffusion equations using spectral methods

The idea behind spectral methods consists in borrowing analytical methods from the theory of Hilbert spaces to transform a partial differential equation, such as Eq. (3.7), into an ordinary differential equation that can be integrated numerically using, for instance, a Runge-Kutta method.

In general, the problems that we are interested in are mixed initial-boundary value problems of the form

$$\frac{\partial\phi(x,t)}{\partial t} = L_{FP}(x,t)\phi(x,t) + \rho(x,t), \quad x \in D = [0,\,1]^K, \tag{4.1}$$

$$B(x)\phi(x,t) = 0, \quad x \in \partial D, t > 0, \tag{4.2}$$

$$\phi(x,0) = g(x), \quad x \in D, \tag{4.3}$$

where $D = [0,\,1]^K$ is the frequency spectrum domain with boundary ∂D, $L_{FP}(x,t)$ is a linear differential operator also known as the Fokker-Planck operator, $\rho(x,t)$ is a function, and $B(x)$ is the linear boundary operator that defines the boundary condition. In this paper, we are interested in the particular set of PDEs defined in Eq. (3.7), although we sometimes keep the notation of Eq. (4.1) as a shorthand.

We assume that $\phi(x,t)$ is, for all t, an element of a Hilbert space $\mathcal{H}$ of square integrable functions, and associated L^2-product $\langle\,,\,\rangle_{L^2}$. Furthermore, we assume that all functions in $\mathcal{H}$ satisfy the boundary conditions imposed by Eq. (4.2). In spectral methods we consider a complete orthogonal basis of functions for $\mathcal{H}$, that we denote by $\{\psi_i(x)\}_{i=0}^{\infty}$, which obeys

$$\langle\psi_i,\,\psi_j\rangle_{L^2} = h_i\delta_{ij}, \tag{4.4}$$

with h_i a function of i that depends on the particular choice of basis functions. One then approximates $\phi(x,t)$ as the truncated expansion

$$\mathcal{P}_\Lambda \phi(x,t) = \sum_{i=0}^{\Lambda-1} \alpha_i(t)\psi_i(x). \tag{4.5}$$

Similarly, one approximates the PDE in Eq. (4.1) by projecting it onto the finite dimensional basis $\{\psi_i(x)\}_{i=0}^{\Lambda-1}$, as

$$\frac{\partial}{\partial t}\mathcal{P}_\Lambda \phi(x,t) = \mathcal{P}_\Lambda L_{FP}(x,t)\mathcal{P}_\Lambda \phi(x,t) + \mathcal{P}_\Lambda \rho(x,t). \tag{4.6}$$

By $\mathcal{H}_\Lambda$ we denote the finite dimensional space spanned by $\{\psi_i(x)\}_{i=0}^{\Lambda-1}$, and by $\mathcal{P}_\Lambda$ the corresponding projector $\mathcal{H} \to \mathcal{H}_\Lambda$. If

$$\mathcal{P}_\Lambda L_{FP}(x,t)\mathcal{P}_\Lambda \phi(x,t) = \sum_{i,j=0}^{\Lambda-1} \omega_{ij}(t)\psi_i(x)\alpha_j(t)$$

$$\mathcal{P}_\Lambda \rho(x,t) = \sum_{i=0}^{\Lambda-1} \beta_i \psi_i(x),$$

we can re-write the ODE in Eq. (4.6) using just *modal* variables as

$$\frac{\partial \alpha_i(t)}{\partial t} = \sum_{j=0}^{\Lambda-1} \omega_{ij}(t)\alpha_j(t) + \beta_i. \tag{4.7}$$

One can solve Eq. (4.7) by discretizing the time variable t, and using a standard numerical method to integrate ODEs. Therefore, the spectral solution to the diffusion PDE is expressed in the form of a truncated expansion, like Eq. (4.5), and has coefficients determined by the integral of Eq. (4.7).

There are many different ways to construct sequences of approximating spaces $\mathcal{H}_\Lambda$ that converge to $\mathcal{H}$ in the limit $\Lambda \to \infty$, when the domain is the K-cube. Here, we follow other authors' preferred choice [7], and choose the basis of Chebyshev polynomials of the first kind. In the following section we introduce Chebyshev expansions and show why they are a preferred choice.

4.1. **Approximation of functions by Chebyshev expansions.** Let $\{T_i(x)\}_{i=0}^{\infty}$ be the basis of Chebyshev polynomials of the first kind. They are the set of eigenfunctions that solve the singular Sturm-Liouville problem

$$\frac{d}{dx}\left(\sqrt{1-x^2}\frac{dT_i(x)}{dx}\right) + \frac{i^2}{\sqrt{1-x^2}}T_i(x) = 0, \tag{4.8}$$

with $i = 0, 1, \ldots, \infty$, and $-1 \le x \le 1$. $\{T_i(x)\}_{n=0}^{\infty}$ are orthogonal under the L^2-product with weight function $w(x) = 1/\sqrt{1-x^2}$:

$$\int_{[-1,1]} T_i(x)T_j(x)\frac{dx}{\sqrt{1-x^2}} = \frac{\pi c_i}{2}\delta_{ij}, \tag{4.9}$$

where $c_0 = 2$ and $c_{i>0} = 1$. This basis of polynomials is a natural basis for the approximation of functions on a finite interval because the associated Gauss-Chebyshev quadrature formulae have an exact closed form, the evaluation of the polynomials is very efficient, and the convergence properties of the Chebyshev expansions are excellent [7].

The Chebyshev polynomials of the first kind can be evaluated by using trigonometric functions, because of the identity $T_i(x) = \cos(i \arccos(x))$. The derivatives of the basis functions can computed by utilizing the recursion

$$T_i(x) = -\frac{1}{2(i-1)} T_i'(x) + \frac{1}{2(i+1)} T_i'(x), \tag{4.10}$$

to express the derivate as

$$T_i'(x) = \sum_{j=0\,|j+i\ odd}^{i-1} \frac{1}{c_j} T_j(x). \tag{4.11}$$

Similar formulae can be found for higher derivatives. The coefficients in the expansion

$$\mathcal{P}_\Lambda f(x) = \sum_{i=0}^{\Lambda-1} a_i T_i(x), \tag{4.12}$$

can be calculated by using the orthogonality relations of the basis functions

$$a_i = \frac{2}{\pi c_i} \int_{-1}^{1} f(x) T_i(x) \frac{dx}{\sqrt{1-x^2}}. \tag{4.13}$$

However, a direct evaluation of the continuous inner product, Eq. (4.13), can be a source of considerable problems as it is in the case of the Fourier series. The classical solution lies in the introduction of a Gauss quadrature of the form

$$\frac{2}{\pi c_i} \int_{-1}^{1} f(x) T_i(x) \frac{dx}{\sqrt{1-x^2}} \simeq \frac{2}{c_i Q} \sum_{k=1}^{Q} f(x_k) T_i(x_k), \qquad x_k = \cos\left(\frac{2k-1}{2Q}\pi\right). \tag{4.14}$$

If $f(x)$ is smooth enough, the finite sum over Q grid points in Eq. (4.14) will converge quicker than $O(Q^{-1})$ to the exact formula, [15]. As Eq. (4.14) is equivalent to a discrete Fourier cosine transform, general results on the convergence of cosine transforms apply also to this problem. One can see this relationship by considering the change of variables $x = \cos y$:

$$\frac{2}{\pi c_i} \int_{-1}^{1} f(x) T_i(x) \frac{dx}{\sqrt{1-x^2}} = \frac{2}{\pi c_i} \int_{0}^{\pi} f(\cos y) \cos(iy) dy, \tag{4.15}$$

and choosing Q equally spaced abscissas in the interval $0 \le y \le \pi$,

$$\frac{2}{\pi c_i} \int_{0}^{\pi} f(\cos y) \cos(iy) dy \simeq \frac{2}{\pi c_i} \frac{\pi}{Q} \sum_{k=1}^{Q} f\left[\cos\left(\frac{2k-1}{2Q}\pi\right)\right] \cos\left(i\frac{2k-1}{2Q}\pi\right). \tag{4.16}$$

In order to study the convergence properties of the Chebyshev expansions Eq. (4.12), we exploit the rich analytical structure of the Chebyshev polynomials. By using the identity Eq. (4.8), one can re-write Eq. (4.13) as

$$a_i = -\frac{2}{\pi c_i i^2} \int_{-1}^{1} f(x) \frac{d}{dx}\left[\sqrt{1-x^2} \frac{dT_i(x)}{dx}\right] dx. \tag{4.17}$$

If $f(x)$ is $C^1([-1,1])$ (i.e., if its first derivative is continuous), we can integrate by parts twice Eq. (4.17) to obtain

$$a_i = -\frac{2}{\pi c_i i^2} \int_{-1}^{1} \sqrt{1-x^2} \frac{d}{dx}\left[\sqrt{1-x^2} \frac{df(x)}{dx}\right] \frac{T_i(x)}{\sqrt{1-x^2}} dx. \tag{4.18}$$

We can repeat this process as many times as $f(x)$ is differentiable; thus, if $f(x) \in C^{2q-1}([-1,1])$ then

$$a_i = -\frac{2}{\pi c_i i^{2q}} \int_{-1}^{1} \left[\left(\sqrt{1-x^2}\frac{d}{dx}\sqrt{1-x^2}\frac{d}{dx} \right)^q f(x) \right] T_i(x) \frac{dx}{\sqrt{1-x^2}}. \tag{4.19}$$

If we use the truncation error

$$\|f(x) - \mathcal{P}_\Lambda f(x)\|_{L^2} = \left[\int_{-1}^{1} \left| f(x) - \sum_{i=0}^{\Lambda-1} a_i T_i(x) \right|^2 \frac{dx}{\sqrt{1-x^2}} \right]^{1/2} = \left(\sum_{i=\Lambda}^{\infty} |a_i|^2 \right)^{1/2}. \tag{4.20}$$

as a measure of convergence of the Chebyshev expansion, we may estimate its asymptotic expansion by calculating the rate of decrease of a_i. But, as we showed in Eq. (4.19), $|a_i| = \frac{c(q)}{i^{2q}}$, for some constant $c(q)$, if $f(x) \in C^{2q-1}([-1,1])$. Therefore, for large Λ the error decreases as a power law

$$\|f(x) - \mathcal{P}_\Lambda f(x)\|_{L^2} = \left(\sum_{i=\Lambda}^{\infty} |a_i|^2 \right)^{1/2} \leq \frac{c}{\Lambda^{2q-1}}, \tag{4.21}$$

and if the function is infinitely differentiable ($q = \infty$), the corresponding Chebyshev series expansion will converge faster than any power of $1/\Lambda$.

In the applications of this paper we will work with re-scaled Chebyshev polynomials. As the Allele Frequency Spectrum is defined on the interval $[0,1]$, or direct products of it, we re-scale the Chebyshev polynomials to obtain an orthonormal basis on $[0,1]$. More precisely, the basis that we use is $\{R_i(x) = \alpha_i T_i((1-x)/2)\}_{i=0}^{\infty}$ with $x \in [0,1]$, $\alpha_0 = 1/\sqrt{\pi}$, $\alpha_{i>0} = \sqrt{2}/\sqrt{\pi}$, L^2-product:

$$\langle f,\, g\rangle = \int_{[0,1]} f(x) g(x) \frac{dx}{\sqrt{x(1-x)}}, \tag{4.22}$$

and orthonormality relations,

$$\int_{[0,1]} R_i(x) R_j(x) \frac{dx}{\sqrt{x(1-x)}} = \delta_{ij}. \tag{4.23}$$

4.1.1. *High-dimensional domains and spectral approximations of functional spaces.* The joint site frequency spectrum of K populations can be defined as a density on $[0,1]^K$. A natural basis of functions on the Hilbert space $L^2_w([0,1]^K)$, comes from the tensor product of one dimensional functions. More particularly, we consider the tensor product of Chebyshev polynomials

$$\psi_{i_1,i_2,\ldots i_K}(x) = R_{i_1}(x_1) R_{i_2}(x_2) \cdots R_{i_K}(x_K), \tag{4.24}$$

because $L^2_w([0,1]^K) = L^2_w([0,1]) \otimes \cdots \otimes L^2_w([0,1])$. Therefore, any square integrable function $F(x)$ under the L^2-product

$$\langle F(x),\, G(x)\rangle_w = \int_{[0,1]^K} F(x) G(x) \prod_{a=1}^{K} \frac{dx_a}{\sqrt{x_a(1-x_a)}}, \tag{4.25}$$

can be approximated as multi-dimensional Chebyshev expansion

$$F(x) = \sum_{i_1=0}^{\Lambda_1-1} \sum_{i_2=0}^{\Lambda_2-1} \cdots \sum_{i_K=0}^{\Lambda_K-1} \alpha_{i_1,i_2,\ldots i_K} R_{i_1}(x_1) R_{i_2}(x_2) \cdots R_{i_K}(x_K). \tag{4.26}$$

The truncation parameters Λ_1, $\Lambda_2 \ldots$ can be fixed depending on the analytical properties of the set of functions that one wants to approximate, and their corresponding truncation errors. There always exist a trade off between the accuracy of the approximation (the higher the Λ the more accurate the approximation) and the speed of the implementation of the algorithm (the lower the Λ, the faster the algorithm); therefore, choosing different values of Λ_i will yield more optimal implementations of the algorithm. Here, for simplicity in the notation, we use a unique truncation parameter $\Lambda = \Lambda_1 = \cdots = \Lambda_K$.

4.2. **Diffusion Operators in Modal Variables.** In order to approximate the PDEs defined in Eq. (3.7) by a system of ODEs in the modal Chebyshev variables such as Eq. (4.7), we need to project the Fokker-Planck operator in the Chebyshev basis spanned by Eq. (4.24). Later on we will show how to deal with the influx of mutations specified by the Dirac deltas.

A direct projection of the Fokker-Planck operator onto the Chebyshev basis spanned by Eq. (4.24), would require to store the coefficients in a huge matrix with Λ^{2K} matrix elements. Fortunately, the Fokker-Planck operator in our problem is very simple, and its non-trivial information can be stored in just four sparse $\Lambda \times \Lambda$ matrices. First, we need the random drift matrix

$$D_{ij} = \frac{1}{2}\int_0^1 R_i(x)\frac{d^2}{dx^2}\left(x(1-x)R_j(x)\right)\frac{dx}{\sqrt{x(1-x)}}, \tag{4.27}$$

and then, the three matrices needed to reconstruct the migration term

$$G_{ij} = \int_0^1 R_i(x)xR_j(x)\frac{dx}{\sqrt{x(1-x)}}, \tag{4.28}$$

$$H_{ij} = \int_0^1 R_i(x)\frac{dR_j(x)}{dx}\frac{dx}{\sqrt{x(1-x)}}, \tag{4.29}$$

$$J_{ij} = \int_0^1 R_i(x)x\frac{dR_j(x)}{dx}\frac{dx}{\sqrt{x(1-x)}}. \tag{4.30}$$

The matrix elements in Eqs. (4.27), (4.28), (4.29) and (4.30), can be quickly determined by means of the Gauss-Chebyshev quadrature defined in Eq. (4.14). Due to the properties of the Chebyshev polynomials, there exist important selection rules that make many matrix elements vanish. More particularly, D_{ij} and J_{ij} are upper triangular matrices (i.e., $D_{ij} = J_{ij} = 0$ if $i > j$), $H_{ij} = 0$ if $i \geq j$, and $G_{ij} = 0$ if $i > j+1$ or $i < j-1$. Thus, the total number of non-trivial matrix elements that we need to compute, for a given Λ, is just $\frac{3}{2}\Lambda^2 + \frac{7}{2}\Lambda - 2$. This is much smaller than the default number of matrix elements (i.e., Λ^{2K}).

Finally, the $\Lambda^K \times \Lambda^K$ matrix elements of the corresponding ω matrix in Eq. (4.7), can be easily recovered from the tensor product structure of the Λ^K-dimensional vector space that defines the Chebyshev expansion (as in Eq. (4.26)). Thus, $\omega_{i_1 \ldots i_P, j_1 \ldots j_K} =$

$$\sum_a \frac{1}{2N_{e,a}} D_{i_a,j_a} - \sum_{a,b} m_{ab}\left(H_{i_a,j_a}G_{i_b,j_b} - \delta_{i_a,j_a}\delta_{i_b,j_b} - J_{i_a,j_a}\delta_{i_b,j_b}\right), \tag{4.31}$$

with $\delta_{ij} = 1$, if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$.

4.3. **Influx of Mutations.** The inhomogeneous terms in Eq. (3.7) that model the influx of mutations are given by effective mutation densities. As we show in the appendices, a model of mutations given by an exponential distribution will be give the same results, up

to an exponentially small deviation, as a standard model with a Dirac delta. The motivation of using smooth effective mutation densities, is because they are better approximated by truncated Chebyshev expansions. As we briefly explained in the review on Chebyshev polynomials and its truncated expansions, the convergence of a truncated expansion depends strongly on the analytical properties of the function to be approximated. As Dirac deltas are not smooth functions, their truncated Chebyshev expansions present bad convergence properties. This is related to the problem of Gibbs phenomena, and we will give a more detailed account of its origin below (see *Sources of error and limits of numerical methods*).

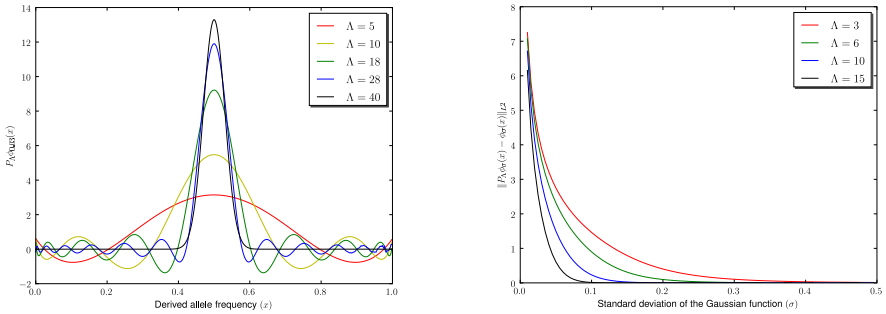

FIGURE 6. On the left, we show the plot of 5 different truncated Chebyshev expansions for a Gaussian peaked at $x = 0.5$ and $\sigma = 0.03$. On the right, we show the truncation error of different Chebyshev expansions (with $\Lambda = 3,\ 6,\ 10$ and 15) of a family of Gaussian functions peaked at $x = 0.5$ and parametrized by the standard deviation $0.01 \leq \sigma \leq 0.5$.

In this paper, we just consider a positive influx of mutations in boundary components of dimension one. In order to approximate the behavior under a mutation term given by a Dirac delta, an effective mutation density $\mu(x)$ has to satisfy the following

- The truncation error is bounded below the established parameter, ϵ, for the control of error. I.e. , $\|\mu(x) - \mathcal{P}_\Lambda \mu(x)\|_{L^2} < \epsilon$.
- The expected frequency of the mass mutation-function, is $\mathbb{E}_\mu(x) = \frac{1}{2N_e}$.
- The mutation-function is nearly zero for relatively large frequencies ($x > 0.05$), and it is as peaked as possible near $x = 1/2N_e$.

While the first and third qualitative requirements are straightforward, the second numerical condition is not. One can interpret such requirement, as equivalent to fix the neutral fixation rate to be u, because the probability that an allele at frequency $x = p$, reaches fixation at $x = 1$, is p. Thus, the average number of new mutants that reach fixation per generation is $2N_e u \times \mathbb{E}_\mu(x) = u$. This constraint also can be derived by studying the properties of the equilibrium density associated to this stochastic process. At equilibrium, the density $\phi_e(x)$ of derived alleles obeys

$$\frac{1}{4N_e}\frac{d^2\psi}{dx^2} = -2N_e u\mu(x), \tag{4.32}$$

with $\psi(x) = x(1-x)\phi_e(x)$. Therefore, the expected frequency of the mass mutation-function can be computed as

$$\int_0^1 x\mu(x)dx = -\frac{1}{8N_e^2 u}\int_0^1 x\frac{d^2\psi}{dx^2}dx. \tag{4.33}$$

Using the identity $x\frac{d^2\psi}{dx^2} = \frac{d}{dx}\left[x\frac{d\psi}{dx}\right] - \frac{d\psi}{dx}$, one can rewrite Eq. (4.33) as

$$\int_0^1 x\mu(x)dx = -\frac{1}{8N_e^2 u}\frac{d\psi}{dx}(x=0). \tag{4.34}$$

On the other hand, the probability flux associated to the equilibrium density of alleles at the boundary $x = 0$ is $j(0) = \frac{1}{4N_e}\psi'(0)$. We use this to write the expected frequency of the mutation density as:

$$\int_0^1 x\mu(x)dx = -\frac{1}{2N_e u}j(0). \tag{4.35}$$

In neutral evolution, the probability flux at the boundary $x = 1$ equals the fixation rate, which satisfies $j(1) = -u$. Every generation a total of $2N_e u$ *de novo* SNPs enter the population, and $2N_e u - u$ SNPs fixate at the boundary $x < 1/2N$. These known facts on the fixation rates of neutrally evolving alleles, require that the total probability flux at the boundary $x < 1/2N$ must be $j(0) = -u$. Therefore, Eq. (4.35) has to satisfy

$$\int_0^1 x\mu(x)dx = -\frac{1}{2N_e u}j(0) = -\frac{1}{2N_e u}(-u) = \frac{1}{2N_e}, \tag{4.36}$$

which is what we wanted to show.

Numerical experiments show that for a large class of functions $\mu(x)$, and in the frequency range $x > x_*$, the associated solutions to the different diffusion problems are identical (up to a very small deviation) to the standard model with a Dirac delta. x_* is a very small frequency, that depends on the choice of $\mu(x)$, and generally can be made arbitrarily small. It is in the region of the frequency space with $0 \leq x \leq x_*$, where the behavior of the different diffusion problems can deviate most.

The truncation error in the Chebyshev expansion depends on the smoothness of the function, and the choice of truncation paramater (see Fig. 6 for an example). For the effective mutation density $\mu(x)$, we use the exponential function

$$\mu(x) = \frac{1}{2N_e} \times \frac{\kappa^2}{1-\exp(-\kappa)-\kappa\exp(-\kappa)}\exp(-\kappa x), \tag{4.37}$$

where the values for $\kappa(\Lambda,\epsilon) \gg 1$, are determined by saturating the bound on error: $\|\mu(x) - \mathcal{P}_\Lambda\mu(x)\|_{L^2} < \epsilon$.

4.3.1. *Comparison of different mutation models at equilibrium.* We derive in the Appendix A the associated equilibrium distributions of derived alleles. For a model with a mutation density given by a Dirac delta, one finds the equilibrium density

$$\phi_e(x) = \frac{4N_e(2N_e-1)ux - 8N_e^2 u(x-1/2N_e)\theta(x-1/2N_e)}{x(1-x)}, \tag{4.38}$$

with $\theta(y)$ the Heaviside step function ($\theta(y) = 0$ for $y < 0$, $\theta(y) = 1/2$ for $y = 0$, and $\theta(y) = 1$ for $y > 0$). Which in the region $x > 1/2N_e$ simplifies to

$$\phi_e(x) = \frac{4N_e u}{x}. \tag{4.39}$$

In the case of $\mu_2(x) = c\exp(-\kappa x)$, the corresponding equilibrium density, is

(4.40)
$$\phi_e(x) = \frac{4N_e u(1-x) + \frac{8N_e^2 uc}{\kappa}\left(\exp(-\kappa)(1+1/\kappa) - \exp(-\kappa)x - \frac{1}{\kappa}\exp(-\kappa x)\right)}{x(1-x)}.$$

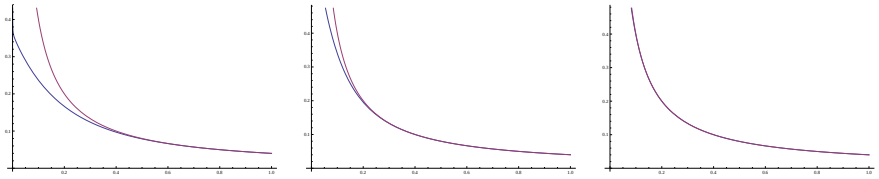

FIGURE 7. Three comparisons of the equilibrium densities associated to the exponential mutation density (blue) for several values of κ vs. the equilibrium density associated to the Dirac delta mutational model (red). For illustrative purposes, the population size used was $N = 10,000$ and the spontaneous mutation rate is $u = 10^{-6}$. On the left the equilibrium density associated to the exponential distribution with $\kappa = 10$ is shown, in the middle $\kappa = 20$, and on the right $\kappa = 40$. For a truncation parameter $\Lambda = 20$, one can choose mutation densities with κ up to 43, while keeping the truncation error below sensible limits.

Therefore, a pairwise comparison of both equilibrium densities, shows that the deviation from both models when $x > x_* = \kappa^{-1}$, is exponentially small when equilibrium is reached (see Fig. 7). We can show that the same is true in non-equilibrium.

4.3.2. *Non-equilibrium dynamics with effective mutation densities.* Here, we show how the non-equilibrium dynamics of a diffusion system under an exponential distribution mutation influx, is the same (up to an exponentially small deviation) to a system where mutations enter the population through the standard Dirac delta $\delta(1 - /2N_e)$, as long as the allele frequencies are bigger than certain minimum frequency x_*. Below x_* the dynamics will be sensitive to differences in the mutation densities.

Let $\varphi(x)$ be an arbitrary initial density of alleles. Let $\phi_1(x,t)$ be the solution to the diffusion equations under pure random drift and a mutation influx given by $\delta(x - 1/2N_e)$. $\phi_2(x,t)$ is the solution of the diffusion equations under pure random drift and mutation influx given by the exponential effective mutation density Eq. (4.37). In the Appendix B, we prove the following identity in the large t limit

$$\int_0^1 |\phi_1(x,t) - \phi_2(x,t)|x(1-x)dx = \frac{4N_e u}{\kappa}(1-\exp(-t/2N_e)) + O(\exp(-\kappa)), \tag{4.41}$$

with $\phi_1(x,0) = \phi_2(x,0) = \varphi$, and $\kappa \leq 2N_e$. If we normalize Eq. (4.41) by

$$\lim_{t\to\infty}\int_0^1 |\phi_1(x,t)|x(1-x)dx,$$

the normalized deviation of $\phi_2(x,t)$ from $\phi_1(x,t)$ is, for t large,

$$\frac{\int_0^1 |\phi_1(x,t) - \phi_2(x,t)|x(1-x)dx}{\lim_{t\to\infty}\int_0^1 |\phi_1(x,t)|x(1-x)dx} = \frac{2}{\kappa}(1-\exp(-t/2N_e)) + O(e^{-\kappa}, N_e^{-1}). \tag{4.42}$$

We can also show, by applying the Minkowski identity to Eq. (B.10) in Appendix B, that

$$\frac{\int_0^1 |\phi_1(x,t)-\phi_2(x,t)|x(1-x)dx}{\lim_{t\to\infty}\int_0^1 |\phi_1(x,t)|x(1-x)dx} \leq \frac{2}{\kappa}(1+\exp(-t/2N_e)) + O(e^{-\kappa}, N_e^{-1}), \tag{4.43}$$

for all $t > 0$. What means, that the deviation is bounded by $O(\kappa^{-1})$ for all t, and therefore the non-equilibrium dynamics of $\phi_1(x,t)$ and $\phi_2(x,t)$ are identical in the large κ limit.

As $|\phi_1(x,0) - \phi_2(x,0)| = 0$ at time zero, and the deviation of $\phi_2(x,t)$ from $\phi_1(x,t)$ attains equilibrium in the large t limit, we can study the frequency dependence of such deviation by looking at the equilibrium

$$\lim_{t\to\infty} \phi_1(x,t) - \phi_2(x,t) = \frac{4N_e u e^{-\kappa x}}{x(1-x)} + O(e^{-\kappa}). \tag{4.44}$$

Here, the $O(e^{-\kappa})$ term exactly cancel the singularity at $x = 1$, and the deviation decays exponentially as function of the frequency. This shows that for frequencies $x > x_* = \kappa^{-1} \geq 1/2N_e$, the dynamics of a model with mutation influx given by a Dirac delta, is the same, up to an exponentially small deviation, to the non-equilibrium dynamics of a model with exponential mutation density.

4.4. **Branching-off of populations.** Modeling a population splitting event also involves the use of Dirac deltas, as in Eq. (3.27), or peaked functions such as Eq. (3.26), whose truncated Chebyshev expansions may present bad convergence properties. These Gibbs-like phenomena can be dealt in a similar way as we did with the mutation term of Eq. (3.7).

We implemented two different solutions to this problem, and both solutions yielded similar results. First, we constructed a smoothed approximation to the Dirac delta by using Gaussian functions:

$$\tilde{\delta}(x_a - x_{K+1}) = \frac{1}{w(x_a)} \exp\left(-\frac{(x_a - x_{K+1})^2}{2\sigma(x_a)^2}\right), \tag{4.45}$$

with

$$w(x_a) = \int_0^1 \exp\left(-\frac{(x_a - x_{K+1})^2}{2\sigma(x_a)^2}\right) dx_{K+1}, \tag{4.46}$$

and $\sigma(x_a)$ is a standard deviation which is chosen as small as possible while preserving the bound on error, $\|\tilde{\delta}(x_a - x) - \mathcal{P}_\Lambda \tilde{\delta}(x_a - x)\|_{L^2[0\leq x\leq 1]} < \epsilon$, for any value of $x_a \in [0,1]$. In order to map the $\tilde{\delta}$-function in Eq. (4.45) to a truncated Chebyshev expansion,

$$\mathcal{P}_\Lambda \tilde{\delta}(x_a - x_{K+1}) = \sum_{i=0}^{\Lambda-1}\sum_{j=0}^{\Lambda-1} \Delta_{ij} R_i(x_a) R_j(x_{K+1}), \tag{4.47}$$

one has to perform a Gauss-Chebyshev quadrature in 2 dimensions, $0 \leq x_a \leq 1$, $0 \leq x_{K+1} \leq 1$:

$$\Delta_{ij} = \int_0^1 \int_0^1 \tilde{\delta}(x_a - x) R_i(x_a) R_j(x) \frac{dx_a}{\sqrt{x_a(1-x_a)}} \frac{dx}{\sqrt{x(1-x)}}. \tag{4.48}$$

The second approach, exploits the analytical behavior of diffusion under pure random drift (i.e., with no migration). By Kimura's solution to the diffusion PDE in terms of the

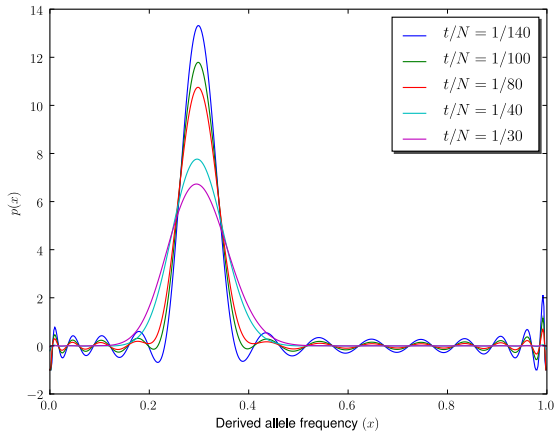

FIGURE 8. Diffusion under pure random drift acts by smoothing out the initial density at $t = 0$. Here we show numerical solutions to the diffusion equations with $\Lambda = 28$, at 5 different times, with initial condition $\phi(x, t = 0) = \delta(x - 0.3)$. As time passes, the numerical solution approaches faster to the exact solution, and the Gibbs phenomena disappear.

Gegenbauer polynomials $\{\mathcal{G}_i(z)\}$, see [12], we know that the time evolution of 1-d density is

$$\mathcal{P}_\Lambda \phi(x,t) = \sum_{i=0}^{\Lambda-1} \frac{2(i+1)+1}{(i+1)(i+2)} (1-r^2)\mathcal{G}_i(r)\mathcal{G}_i(z)\exp(-(i+1)(i+2)t/4N), \tag{4.49}$$

with $r = 1 - 2p$, $z = 1 - 2x$ and $\phi(x, 0) = \delta(x - p)$. Thus, as the degree of $\mathcal{G}_i(z)$ is i, and we can express any Gegenbauer polynomial as a linear combination of Chebyshev polynomials, we know that the exact solution to the diffusion equation acts on the high modes by the suppressing factor $\exp(-(i + 1)(i + 2)t/4N)$. This means that diffusion smooths out the Dirac delta at initial time. Fig. 8, represents the evolution of the density at different times.

Thus, we can use diffusion under pure random drift to smooth the density introduced after the population splitting event. Let $\phi_K(x, t)$ the density before the splitting, and let a be the population from which population $K + 1$ is founded, then we consider initially the density

$$\phi_{K+1}(x, x_{K+1}, t' = 0) = \phi_K(x, t)\delta(x_a - x_{K+1}). \tag{4.50}$$

The associated Chebyshev expansion $\mathcal{P}_\Lambda \phi_{K+1}(x, x_{K+1}, t)$ will present Gibbs-phenomena. However, by diffusing for a short time τ under pure random drift

$$\frac{\partial}{\partial t}\phi_{K+1}(x, x_{K+1}, t) = \sum_{b=1}^{K+1} \frac{1}{2}\frac{\partial^2}{\partial x_b^2}\left(\frac{x_b(1-x_b)}{2S_b}\phi_{K+1}(x, x_{K+1}, t)\right), \tag{4.51}$$

(with $S_a = S_{K+1} = W$, $S_b = V$ for $K + 1 \neq b \neq a$, and $V \gg W$), $\phi_{K+1}(x, x_{K+1}, \tau)$ becomes tractable under Chebyshev expansions. In other words, by choosing τ such that

the error bound is satisfied

$$\|\phi_{K+1}(x, x_{K+1}, \tau) - \mathcal{P}_\Lambda \phi_{K+1}(x, x_{K+1}, \tau)\|_{L^2} < \epsilon,$$

we obtain a smooth density, after population splitting, which can follow the regular diffusion with migration prescribed in the problem, and approximate accurately the branching-off event. In some limits, this approximation can fail, though we leave the corresponding analysis for the next section.

Here, we do not consider the numerical solution to the problem of splitting with admixture, although we are confident that it should be possible to solve it along similar arguments.

4.5. **Sources of error and limits of numerical methods.** There are two major sources of error in these numerical methods. First, the solution of the diffusion equation is itself a time-continuous approximation to the time evolution of a probability density evolving under a discrete Markov chain. Hence, whenever the diffusion approximation fails, its numerical implementation will also fail. Secondly, a numerical solution by means of spectral expansions, involves an approximation of the infinite dimensional space of functions on a domain by a finite dimensional space generated by bases of orthonormal functions under certain L^2-product. As we show below, under a broad set of conditions the numerical solution will converge accurately to the exact solution; otherwise, the numerical solution can fail to approximate the exact solution. A third source of error appears because one has to discretizate time in order to integrate the high-dimensional ODE that approximates the PDE. Fortunately, this source of error is neglectible because the diffusion generators yield a stable time evolution.

We summarize below the main conditions that have to be satisfied in order to obtain high-quality numerical solutions to the PDEs studied in this work.

4.5.1. *Limits of diffusion equations.* In the diffusion approximation to a Markov chain, the transition probability is approximated by a Gaussian distribution, [5]. Here, we review the derivation of the diffusion equation as the continuous limit of a Markov chain, in order to emphasize the assumptions made and determine the limits of such approximation.

Given a Markov process defined by a discrete state space $\mathcal{S}$, transition matrices $p(I|J)$, initial value $K \in \mathcal{S}$ and discrete time $\tau = 0, 1, \ldots$, the probability that the state will be at I at time τ is $f(I|K,\, \tau)$, where $f(I|K,\, \tau)$ obeys the recurrence relation

$$f(I|K,\, \tau) = \sum_{J \in \mathcal{S}} p(I|J) f(J|K,\, \tau - 1). \tag{4.52}$$

In the diffusion approximation one considers a sequence of discrete state spaces $\{\mathcal{S}_\lambda\}_{\lambda \in \mathbb{Z}_+}$, such that in the limit $\lambda \to \infty$ the state space $\mathcal{S}_\infty$ converges to a smooth manifold (in practical applications, a compact domain $D \subset \mathbb{R}^K$).

In this paper, we take $\mathcal{S}_\lambda$ to be $[0,\, \lambda]^K$, and $\mathcal{S}_\infty \sim [0, 1]^K$. Therefore, the state variables can be re-scaled as $K_a/\lambda = x_a$, with $a = 1, \ldots, K$ and $K_a \in [0,\, \lambda]^K$. Similarly, we introduce the time variable $t = \tau/\lambda$. In the large λ limit, the transition probability for the change of the state from time τ/λ to time $(\tau + 1)/\lambda$, is governed by a distribution with momenta

$$\mathbb{E}(\delta x_a | x) = M_a(x)/\lambda + O(1/\lambda^2), \tag{4.53}$$

$$\mathbb{E}(\delta x_a \delta x_b | x) - \mathbb{E}(\delta x_a | x)\mathbb{E}(\delta x_b | x) = C_{ab}(x)/\lambda + O(1/\lambda^2), \tag{4.54}$$

$$\mathbb{E}(\delta x_a^3 | x) = O(1/\lambda^2). \tag{4.55}$$

In this continuous limit, the equation that describes the time evolution of the Markov chain in Eq. (4.52), can be writen as a forward Kolmogorov equation if we neglect terms of order $O(1/\lambda^2)$. However, if $M_a(x)$ is proportional to λ, the $O(1/\lambda^2)$ terms in Eq. (4.53) will not be neglectible and the diffusion approximation will not be valid. As in this paper we take $\lambda = 2N_e$, and $M_a(x)$ proportional to the migration rates m_{ab}, if the migration rates obey $2N_{e,a}m_{ab} \leq O(1)$ the diffusion approximation will be valid. Indeed, computer experiments show that the numerical solutions become unstable and yield wrong results if this bound is violated. This limit precisely defines when two populations can be considered the same, [9]. Therefore, in cases when $2N_{e,a}m_{ab} \gg O(1)$, we can consider populations a and b as two parts of the same population. Another assumption in the diffusion approximation is that a binomial distribution with $2N_{e,a}$ degrees of freedom can be approximated by a Gaussian distribution. This will be a valid approximation as long as $N_{e,a}$ is large enough. Numerical experiments show that the approximation is accurate if $N_{e,a} > 100$; otherwise, effects due to the finiteness of the Markov chain cannot be neglected and the approximation will fail.

$$N_{e,a} > 100$$
$$2N_{e,a}m_{ab} \leq O(1)$$

4.5.2. *Limits of spectral expansions.* Spectral methods, as any numerical scheme to solve PDEs, require several assumptions about the behavior of the solution of the PDE. The most important one is that one can approximate the solution as a series of smooth basis functions,

$$\mathcal{P}_\Lambda \phi(x,t) = \sum_{i_1=0}^{\Lambda-1} \cdots \sum_{i_P=0}^{\Lambda-1} \alpha_{i_1,i_2,\ldots i_P}(t) T_{i_1}(x_1) T_{i_2}(x_2) \cdots T_{i_P}(x_P). \tag{4.56}$$

In other words, the projection of the solution $\mathcal{P}_\Lambda \phi(x,t)$ is assumed to approximate $\phi(x,t)$ well, in some appropriate norm, for Λ large enough. As one has to choose finite values for Λ, Eq.(4.56) will sometimes fail to approximate correctly the solution of the PDE.

In the applications of this paper, the basis of functions that we use consist of the Chebyshev polynomials of the first kind[2]. Below we provide bound estimates for the truncation error $\|\mathcal{P}_\Lambda \phi(x,t) - \phi(x,t)\|_{L^2[-1,1]^K}$, to understand the quality of the approximate solutions for different values of Λ, (see also [7, 1] for different choices of basis functions).

More precisely, as the L^2 inner product and norm in the Chebyshev problem are:

$$\langle f,\, g\rangle_{L^2[-1,1]^K} = \int_{[-1,1]^K} f(x) g(x) \prod_{i=1}^{K} \frac{dx_i}{\sqrt{1-x_i^2}}, \tag{4.57}$$

and

$$\|f\|^2_{L^2[-1,1]^K} = \int_{[-1,1]^K} |f(x)|^2 \prod_{i=1}^{K} \frac{dx_i}{\sqrt{1-x_i^2}}, \tag{4.58}$$

[2] One can work either with the basis of functions $\{T_i(x)\}$ on $x \in [-1,1]$, or with the re-scaled basis $\{R_i(x)\}$ defined on $x \in [0,1]$, by performing a simple scale transformation.

the terms in the expansion Eq. (4.56) can be computed by performing inner products

(4.59)
$$\alpha_{i_1,i_2,\ldots i_K}(t) = \left(\frac{2}{\pi}\right)^K \int_{[-1,1]^K} \phi(x,t) T_{i_1}(x_1) T_{i_2}(x_2) \cdots T_{i_K}(x_K) \prod_{j=1}^{K} \frac{dx_j}{c_{i_j}\sqrt{1-x_j^2}},$$

with $c_0 = 2$ and $c_j = 1$ ($j > 0$). A consequence of the orthogonality of the basis functions, is that the squared truncation error admits a simple formulation in terms of the coefficients in the expansion:

$$\|\mathcal{P}_\Lambda \phi(x,t) - \phi(x,t)\|^2_{L^2[-1,1]^K} = \left(\frac{2}{\pi}\right)^K \sum_{i_1 \geq \Lambda} \cdots \sum_{i_K \geq \Lambda} |\alpha_{i_1,i_2,\ldots i_K}(t)|^2. \tag{4.60}$$

Thus, the truncation error depends only on the decay of the higher modes $|\alpha_{i_1,i_2,\ldots i_K}|$ in the expansion. On the other hand, the decay of these higher modes depend on the analytical properties of $\phi(x,t)$ itself. For instance, if $\phi(x,t) \in C^{2q_1-1,2q_2-1,\ldots,2q_K-1}\left([-1,1]^K\right)$, i.e. if

$$\left\| \left(\frac{\partial}{\partial x_1}\right)^{2q_1-1} \left(\frac{\partial}{\partial x_2}\right)^{2q_2-1} \cdots \left(\frac{\partial}{\partial x_K}\right)^{2q_K-1} \phi(x,t) \right\|_{L^2[-1,1]^K} < \infty, \tag{4.61}$$

we can integrate by parts Eq. (4.59), as we did in Eq. (4.18), to write the decay of each mode as

(4.62)
$$|\alpha_{i_1,i_2,\ldots i_K}(t)| = \left(\frac{2}{\pi}\right)^K \left| \int_{[-1,1]^K} \left[\prod_{j=1}^{K} \left(\sqrt{1-x_j^2}\frac{\partial}{\partial x_j}\right)^{2q_j} \phi(x,t) \right] \times \frac{T_{i_1}(x_1)}{c_{i_1} i_1^{2q_1}\sqrt{1-x_1^2}} dx_1 \cdots \frac{T_{i_K}(x_K)}{c_{i_K} i_K^{2q_K}\sqrt{1-x_K^2}} dx_K \right|.$$

Eq. (4.62) implies that the truncation error is directly related to the smoothness of $\phi(x,t)$; it follows that we can bound the truncation error as a function of Λ:

(4.63)
$$\|\mathcal{P}_\Lambda \phi(x,t) - \phi(x,t)\|_{L^2[-1,1]^K} \leq C(q) \Lambda^{-\sum_j q_j} \left\| \prod_{j=1}^{K} \left[\sqrt{1-x_j^2}\frac{\partial}{\partial x_j}\right]^{q_j} \phi(x,t) \right\|_{L^2[-1,1]^K}.$$

Another convenient measure of smoothness is the Sobolev norm:

$$\|\Phi(x)\|^2_{W^{q_1,\ldots,q_K}[-1,1]^K} = \sum_{s_1=0}^{q_1} \cdots \sum_{s_P=0}^{q_K} \left\| \prod_{j=1}^{K} \left(\frac{\partial}{\partial x_j}\right)^{s_j} \Phi(x) \right\|_{L^2[-1,1]^K}; \tag{4.64}$$

in terms of the Sobolev norm, the truncation error is bounded as

$$\|\mathcal{P}_\Lambda \phi(x,t) - \phi(x,t)\|_{L^2[-1,1]^K} \leq C \Lambda^{-\sum_j q_j} \|\phi(x,t)\|_{W^{q_1,\ldots,q_K}[-1,1]^K}. \tag{4.65}$$

A corollary of Eq. (4.65) is that if $\phi(x,t)$ is smooth, $\mathcal{P}_\Lambda \phi(x,t)$ converges to $\phi(x,t)$ more rapidly than any finite power of Λ^{-1}. This is indeed the basic property that has given name to spectral methods.

In the case of the diffusion problem that we consider in this paper, the boundary conditions are absorbing and there is not flux of mutations when $K > 1$. It is easy to see how $|\alpha_{i_1,i_2,\ldots i_K}(t)| \to 0$ as $t \to \infty$ in the absence of influx of polymorphisms in the population.

This means that diffusion acts as a smoothing operator by decreasing the Sobolev norm of the initial density. Eq. (4.65) imposes important limits to the set of parameters of the model where spectral methods work; for instance, in the absence of influx of polymorphisms and for long diffusion times the truncation error will remain low.

After two populations split and the K-dimensional diffusion becomes a $K+1$ dimensional process, the $K+1$ dimensional density becomes a distribution concentrated in the linear subspace of $[-1,1]^{K+1}$ defined by $x_a = x_{a+1}$ (with a and $a+1$ labeling the two daughter populations that just splited). Such density has an infinite Sobolev norm and cannot be represented as a finite sum of polynomials. Fortunately, the diffusion generator acts on the density by smoothing it out, and by bringing the density to a density with finite Sobolev norm. The main variables involved in this process are: the time difference between the current splitting event and the next one, $T_{A+1} - T_A$, and the effective population sizes $N_{e,a}$ and $N_{e,a+1}$ of the daughters populations. Therefore, depending on the choice of the truncation parameter Λ, a minimun diffusion time $t_m(N_{e,a}, N_{e,a+1}, \Lambda)$ will be necessary to bring the truncation error within desired limits $\|\mathcal{P}_\Lambda \phi(x,t_m) - \phi(x,t_m)\|_{L^2} \leq \epsilon$. Here, ϵ, is the control parameter on numerical error. Therefore, the bigger is the largest effective population size of the two daughters populations, the bigger will be such minimum diffusion time. If the time difference between the current splitting event and the next one is bigger than

$$t_m(N_{e,a}, N_{e,a+1}, \Lambda) = C(\Lambda)\max(N_{e,a}, N_{e,a+1}), \tag{4.66}$$

(where $C(\Lambda)$ is a function that can be computed numerically), the resulting numerical error will stay below the desired limits. As our model aims to reproduce the real SNP Allele

$$|T_{A+1} - T_A| > C(\Lambda)\max(N_{e,a+1}, N_{e,a})$$
$$\|\mathcal{P}_\Lambda \hat{\gamma}(x) - \hat{\gamma}(x)\|_{L^2} < \epsilon$$

Frequency Spectrum density, there should exist low error approximations of such density, that we denote as $\hat{\gamma}(x)$, in terms of polynomial expansions. Otherwise, the methods here presented will fail to solve the problem. This can only happen if $\hat{\gamma}(x)$ is so rugged, i.e. the corresponding Sobolev norm is so high, that the largest finite choice for Λ that we can implement in our computer-code is not large enough to approximate accurately $\hat{\gamma}(x)$:

$$\|\mathcal{P}_{\Lambda_{max}}\hat{\gamma}(x) - \hat{\gamma}(x)\|_{L^2[-1,1]^K} \sim C\Lambda_{max}^{-\sum_j q_j}\|\hat{\gamma}(x)\|_{W^{q_1,\ldots,q_K}[-1,1]^K} \gg \epsilon. \tag{4.67}$$

In case that Eq. (4.67) is obeyed, it is likely that 2 or more populations are so closely related that we can treat them as if they were one identical population. If we reduce the dimensionality of the problem in this way, (by only incorporating differentiated populations), the correlations will disappear and the Sobolev norm of $\hat{\gamma}'(x)$, will be such, that we will be able to find a sensible parameter Λ to approximate $\hat{\gamma}'(x)$ as a truncated Chebyshev expansion.

5. CONCLUSION

In this paper we have introduced a forward diffusion model of the joint allele frequency spectra, and a numerical method to solve the associated PDEs. Our approach is inspired by recent work where similar models were proposed [21, 4, 8]. Analogously, our methods

are quite general and can accomodate selection coefficients and time dependent effective population sizes.

The major novelties of the model here presented with respect previous work are:

- The introduction of spectral methods/finite elements in the context of forward diffusion equations and infinite sites models. Traditionally, these techniques yield better results than finite differences schemes when the dimension of the domain is high (i.e., when the final number of populations is high), and the solutions are smooth. A comparison of our implementation using spectral methods, and previous implementations using finite differences [8], will be the matter of future work.
- A set of boundary conditions that deals with the possibility that some polymorphisms reach fixation in some populations while remain polymorphic in other populations. When the differences in effective population sizes between different populations are large, this phenomenon can become very important. Here, we have introduced a solution to address such possible scenario. Previous work imposed zero flux at the boundaries [8], and hence avoided the fixation of polymorphisms in some populations while remaining polymorphic in the rest.
- The introduction of effective mutation densities, which generalize previous models for the influx of mutations, [4]. We have emphasized how different ways to inject mutations at very low frequencies converge to the same solution for larger frequencies.

The non-equilibrium theory of Allele Frequency Spectra is of primary importance to analyze population genomics data. Although it does not make use of information about haplotype structure or linkage non-equilibrium, the analysis of AFS allows the study of demographic history and the inference of natural selection. In this work, we have extended the diffusion theory of the multi-population AFS, to accomodate spectral methods, a general framework for the influx of mutations, and non-trivial boundary interactions.

6. Acknowledgements

We thank Rasmus Nielsen, Anirvan Sengupta and Vitor Sousa for helpful conversations. This work was partially supported by the National Institutes of Health (grant number R00HG004515-02 to K. C.).

APPENDIX A. COMPARISON OF MUTATIONAL MODELS AT EQUILIBRIUM

In this appendix we compute the equilibrium densities associated to Wright-Fisher processes with mutation. Two types of mutation processes are considered, both modeled by a mutation density. The first mutation density is a Dirac delta, while the second one is an exponential distribution.

As the diffusion equation that describe the time evolution of the density of alleles for diallelic SNPs, is

$$\frac{\partial \phi(x,t)}{\partial t} = \frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi(x,t)\right] + 2N_e u\mu(x), \tag{A.1}$$

the equilibrium density $\phi_e(x)$ satisfies $\frac{\partial \phi_e(x)}{\partial t} = 0$. By using instead the function $\psi(x) = x(1-x)\phi_e(x)$, the associated second order ordinary differential equation becomes

$$\frac{1}{4N_e}\frac{\partial^2 \psi(x)}{\partial x^2} + 2N_e u\mu(x) = 0. \tag{A.2}$$

As Eq. (A.2) is just defined for $x > 0$, we can use Laplace transforms to solve the equation. Let

$$\tilde{\mu}(s) = \int_0^\infty \mu(x)\exp(-sx)dx, \tag{A.3}$$

be the Laplace transform associated to the mutation density $\mu(x)$, and

$$\int_0^\infty \frac{\partial^2 \psi(x)}{\partial x^2}\exp(-sx)dx = s^2\tilde{\psi}(s) - s\psi(0) - \psi'(0), \tag{A.4}$$

the Laplace transform associated to $\psi''(x)$, with $\psi(0)$ and $\psi'(0)$ integration constants. Therefore, in the s domain, $\tilde{\psi}(s)$ is

$$\tilde{\psi}(s) = \frac{s\psi(0) + \psi'(0) - 8N_e^2 u\tilde{\mu}(s)}{s^2}, \tag{A.5}$$

and by performing the inverse Laplace transform we obtain the solution to the equilibrium density

$$\phi_e(x) = \frac{1}{x(1-x)}\frac{1}{2\pi i}\lim_{T\to\infty}\int_{\epsilon-iT}^{\epsilon+iT}\frac{s\psi(0) + \psi'(0) - 8N_e^2 u\tilde{\mu}(s)}{s^2}\exp(sx)ds. \tag{A.6}$$

We fix the integration constants, $\psi(0)$ and $\psi'(0)$, by requiring $\phi_e(x)$ to be finite at $x = 1$, and the probability flow at $x = 1$ to be equal to u,

$$j(1) = -\frac{1}{4N_e}\psi'(0) = u. \tag{A.7}$$

As an example, we can evaluate exactly Eq. (A.6), for $\mu_1(x) = \delta(x - 1/2N_e)$ and $\mu_2(x) = c\exp(-\kappa x)$. For the Dirac delta, the Laplace transform is

$$\tilde{\mu}_1(s) = \exp(-s/2N_e). \tag{A.8}$$

If we compute the corresponding inverse Laplace transform in Eq. (A.6), and fix the integration constants as explained above, we find the equilibrium density

$$\phi_e(x) = \frac{4N_e(2N_e - 1)ux - 8N_e^2 u(x - 1/2N_e)\theta(x - 1/2N_e)}{x(1-x)}, \tag{A.9}$$

with $\theta(y)$ the Heaviside step function ($\theta(y) = 0$ for $y < 0$, $\theta(y) = 1/2$ for $y = 0$, and $\theta(y) = 1$ for $y > 0$). If $x > 1/2N$, Eq. (A.9) simplifies to

$$\phi_e(x) = \frac{4N_e u}{x}. \tag{A.10}$$

In the case of $\mu_2(x) = c\exp(-\kappa x)$, the Laplace transform is

$$\tilde{\mu}_2(s) = \frac{c}{s+\kappa}, \tag{A.11}$$

and the corresponding equilibrium density, after integrating Eq. (A.6), is

$$\phi_e(x) = \frac{4N_e u(1-x) + \frac{8N^2 uc}{\kappa}\left(\exp(-\kappa)(1+1/\kappa) - \exp(-\kappa)x - \frac{1}{\kappa}\exp(-\kappa x)\right)}{x(1-x)}, \tag{A.12}$$

which in the large κ limit, and for $x \gg 1/\kappa$, converges exponentially fast to

$$\phi_e(x) = \frac{4N_e u}{x}. \tag{A.13}$$

In the limit $x \to 0$, $\phi_e(x)$ is finite just iff $c = \frac{1}{2N_e} \times \frac{\kappa^2}{1-\exp(-\kappa)-\kappa\exp(-\kappa)}$, which is the normalization choice made in Eq. (4.37), and the only one satisfying

$$\int_0^1 x\mu_2(x) = \frac{1}{2N_e}. \tag{A.14}$$

This shows how a mutation model defined by a certain class of smooth mutation densities, reaches the same equilibrium density, up to a small deviation, as the standard model with a Dirac delta.

APPENDIX B. COMPARISON OF MUTATIONAL MODELS AT NON-EQUILIBRIUM

In this appendix we compare the non-equilibrium dynamics of models with a mutation influx described by exponential distributions, with models that consider a standard Dirac delta.

More particularly, we prove that if $\phi_1(x,t)$ is the solution to an infinite sites model PDE, with absorbing boundaries,

$$\frac{\partial \phi_1(x,t)}{\partial t} = \frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi_1(x,t)\right] + 2N_e u\delta(x-1/2N_e), \tag{B.1}$$

and $\phi_2(x,t)$ is the solution to the same model, but with an exponential mutation density

$$\frac{\partial \phi_2(x,t)}{\partial t} = \frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi_2(x,t)\right] + u\frac{\kappa^2\exp(-\kappa x)}{1-\exp(-\kappa)-\kappa\exp(-\kappa)}, \tag{B.2}$$

then, the *deviation* of $\phi_2(x,t)$ with respect to $\phi_1(x,t)$, as function of time and for any initial condition $\phi_2(x,t=0)=\phi_1(x,t=0)=\varphi(x)$, is, in the large t limit,

$$\int_0^1 |\phi_1(x,t)-\phi_2(x,t)|x(1-x)dx = \frac{4N_e u}{\kappa}(1-\exp(-t/2N_e)) + O(e^{-\kappa}). \tag{B.3}$$

Here, $|\cdot|$ is the absolute value, and $O(e^{-\kappa})$ are terms that decay exponentially as function of κ, which can be neglected in the large κ limit.

As the total number of SNPs that are polymorphic in one population depend on the population size and the mutation rate, it is convenient to normalize the deviation Eq. (B.3) by $\lim_{t\to\infty}\int_0^1 |\phi_1(x,t)|x(1-x)dx = (2N_e-1)u$. In this normalization we have

$$\frac{\int_0^1 |\phi_1(x,t)-\phi_2(x,t)|x(1-x)dx}{\lim_{t\to\infty}\int_0^1 |\phi_1(x,t)|x(1-x)dx} = \frac{2}{\kappa}(1-\exp(-t/2N_e)) + O(e^{-\kappa}, N_e^{-1}). \tag{B.4}$$

To prove Eq. (B.3), we first describe the solutions to Eq. (B.1) and Eq. (B.2). Both equations consist of a homogeneous term and an inhomogeneous contribution given by the mutation density. As they are linear equations, the solution to the PDE is the sum of a homogeneous and an inhomogeneous term

$$\phi_1(x,t) = \phi_1^h(x,t) + \phi_1^e(x), \tag{B.5}$$

satisfying

$$\begin{aligned}\frac{\partial \phi_1^h(x,t)}{\partial t} &= \frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi_1^h(x,t)\right] + \\ &\frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi_1^e(x)\right] + 2N_e u\delta(x-1/2N_e).\end{aligned} \tag{B.6}$$

Hence, the only time-independent term $\phi_1^e(x)$ that solves Eq. (B.6) is the equilibrium density Eq. (A.9), and $\phi_1^h(x,t)$ obeys a standard diffusion equation with no mutation density, and with initial condition $\phi_1^h(x,t=0)=\varphi(x)-\phi_1^e(x)$. If $\mathcal{L}\phi_1^h(x,t)$ denotes the Fokker-Planck operator acting on $\phi_1^h(x,t)$,

$$\mathcal{L}\phi_1^h(x,t) = \frac{1}{4N_e}\frac{\partial^2}{\partial x^2}\left[x(1-x)\phi_1^h(x,t)\right],$$

we can write the solution to Eq. (B.6) in the following compact form

$$\phi_1(x,t) = \exp\left(t\mathcal{L}\right)\left(\varphi(x)-\phi_1^e(x)\right) + \phi_1^e(x). \tag{B.7}$$

Here, $\exp(t\mathcal{L})$ is the time-dependent action of the diffusion operator on the initial density $\varphi(x) - \phi_1^e(x)$ while preserving the absorbing boundary conditions. Such operator can be diagonalized in the basis of Gegenbauer polynomials on $L^2([0,1])$, see [12]. The corresponding eigenvalues of $\exp(t\mathcal{L})$ are $\exp(-(i+1)(i+2)t/4N_e)$ with $i \in [0, \infty)$.

We can solve Eq. (B.2) in a similar way, by using the decomposition

$$\phi_2(x,t) = \phi_2^h(x,t) + \phi_2^e(x). \tag{B.8}$$

In this case, $\phi_2^e(x)$ is the equilibrium density associated to the exponential mutation density, as defined in Eq. (A.12). The term $\phi_2^h(x,t)$ evolves under pure random drift, with no mutation influx, and initial condition $\phi_2^h(x,t=0) = \varphi(x) - \phi_2^e(x)$:

$$\phi_2(x,t) = \exp(t\mathcal{L})\left(\varphi(x) - \phi_2^e(x)\right) + \phi_2^e(x). \tag{B.9}$$

By substracting Eq. (B.9) from Eq. (B.7), we can describe the time evolution of the deviation as

$$\phi_1(x,t) - \phi_2(x,t) = -\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right) + \phi_1^e(x) - \phi_2^e(x), \tag{B.10}$$

which is independent of the initial condition $\varphi(x)$.

One can show that $\phi_1^e(x) - \phi_2^e(x)$ is non negative on $[0,1]$, if $\kappa \leq 2N_e$. This can be seen more clearly by computing $\phi_1^e(x) - \phi_2^e(x)$ in the large κ limit

$$\phi_1^e(x) - \phi_2^e(x) = \frac{4N_e u}{1-x}(2N_e - \kappa), \qquad x \in [0, 1/2N_e), \tag{B.11}$$

$$\phi_1^e(x) - \phi_2^e(x) = \frac{4N_e u e^{-\kappa x}}{x(1-x)} + O(e^{-\kappa}/(1-x)), \qquad x \in (1/2N_e, 1]. \tag{B.12}$$

The terms of order $e^{-\kappa}$ in Eq. (B.12) exactly cancel the divergence at $x = 1$. Therefore, the action of the diffusion operator on $\phi_1^e(x) - \phi_2^e(x)$, will preserve the non-negativity of the density

$$\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right) \geq 0, \qquad \forall x \in [0,1], \quad \forall t > 0. \tag{B.13}$$

Because of this inequality, the absolute value $|\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right)|$, is the same as $\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right)$, and we can evaluate exactly the integral

$$\int_0^1 |\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right)| x(1-x)dx = \int_0^1 \exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right) x(1-x)dx, \tag{B.14}$$

by expanding $\exp(t\mathcal{L})\left(\phi_1^e(x) - \phi_2^e(x)\right)$ in the eigenbasis of $\exp(t\mathcal{L})$. This basis is orthogonal under the L^2-product defined by the weight $x(1-x)$, and the constant function on $[0,1]$ corresponds to the eigenfunction with smallest eigenvalue. In this way we can interpret the right hand side of Eq. (B.14) as a projection on such eigenfunction, and evaluate exactly the integral.

The eigenbasis of $\exp(t\mathcal{L})$ is defined by the Gegenbauer polynomials. As an example, the first three Gegenbauer polynomials on $[0,1]$, orthonormal under the L^2-product with weight $x(1-x)$, are

$$T_0(x) = \sqrt{6}, \tag{B.15}$$

$$T_1(x) = \sqrt{30}(1-2x), \tag{B.16}$$

$$T_2(x) = \sqrt{84}(1 - 5x + 5x^2). \tag{B.17}$$

The corresponding eigenvalues in $\exp(t\mathcal{L})$, are eigenvalues $\exp(-t/2N_e)$, $\exp(-3t/2N_e)$, and $\exp(-3t/N_e)$. Thus, Eq. (B.14) is the same as

(B.18)
$$\int_0^1 \exp(t\mathcal{L})\,(\phi_1^e(x)-\phi_2^e(x))x(1-x)dx = \int_0^1 \exp(t\mathcal{L})\,(\phi_1^e(x)-\phi_2^e(x))\frac{T_0(x)}{\sqrt{6}}x(1-x)dx,$$

and

(B.19)
$$\exp(-t/2N_e)\int_0^1 (\phi_1^e(x)-\phi_2^e(x))x(1-x)dx = \frac{4N_e u}{\kappa}\exp(-t/2N_e)+O(e^{-\kappa}).$$

As $0 \leq \exp(t\mathcal{L})\,(\phi_1^e(x)-\phi_2^e(x)) \leq \phi_1^e(x)-\phi_2^e(x)$ for all $x \in [0,1]$ and for $t \gg N_e$, we finally compute Eq. (B.3), as

$$\int_0^1 |\phi_1(x,t)-\phi_2(x,t)|x(1-x)dx = \int_0^1 (\phi_1^e(x)-\phi_2^e(x))x(1-x)dx$$

(B.20)
$$-\int_0^1 \exp(t\mathcal{L})\,(\phi_1^e(x)-\phi_2^e(x))x(1-x)dx,$$

which is

(B.21)
$$\int_0^1 |\phi_1(x,t)-\phi_2(x,t)|x(1-x)dx = \frac{4N_e u}{\kappa}\left(1-\exp(-t/2N_e)\right)+O(e^{-\kappa}),$$

as we wanted to show.

REFERENCES

[1] C. Canuto, M.Y. Hussaini, A. Quarteroni and T.A. Zang, *Spectral Methods in Fluid Dynamics*. Springer Series in Computational Physics, Springer, (1988).

[2] J. F. Crow and M. Kimura, *Some genetic problems in natural populations*. Proc. Third Berkeley Symp. on Math. Statist. and Prob. **4**, 1-22, (1956).

[3] J. F. Crow and M. Kimura, *An introduction to population genetics theory*. Harper & Row publishers (1970).

[4] S. N. Evans, Y. Shvets, and M. Slatkin, *Non-equilibrium theory of the allele frequency spectrum*. Theoretical Population Biology, Vol. 71, Issue 1, pp. 109-119 (2007).

[5] W. J. Ewens, *Mathematical Population Genetics. I. Theoretical Introduction*. Interdisciplinary Applied Mathematics, Vol. **27**, Springer (2004).

[6] R. A. Fisher, *The distribution of gene ratios for rare mutations*. Proc. Royal Soc. Edinburgh 50, 205-220, (1930).

[7] D. Gottlieb and S. A. Orszag, *Numerical analysis of spectral methods: Theory and Applications*. Society for Industrial and Applied Mathematics, (1977).

[8] R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, *Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data*. PLoS Genet, Vol. 5, No. 10. (2009).

[9] J. Hey, *Isolation with migration models for more than two populations*. Mol. Biol. Evol. **27**(4) 905-920, (2010).

[10] J. Hey and R. Nielsen, *Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis*. Genetics, Vol. 167, 747-760, (2004).

[11] R. R. Hudson, *Generating samples under a Wright-Fisher neutral model of genetic variation.* Bioinformatics **18**:337-338 (2002).

[12] M. Kimura, *Solution of a process of random genetic drift with a continuous model*. Proc. Nat. Acad. Sci. **41** 144-150, (1955).

[13] M. Kimura, *Diffusion models in population genetics*. Journal of Applied Probability 1, 177-232 (1964).

[14] M. Kimura, *The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations*. Genetics **61** 893-903, (1969).

[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition (2007).

[16] S. A. Orszag, *Numerical Methods for the Simulation of Turbulence*. Phys. Fluids Supp. II, 12, 250-257, (1969).

[17] W. Ritz, *Uber eine neue Methode zur Losung gewisser Variationsprobleme der mathematischen Physik.* Journal fr die Reine und Angewandte Mathematik, vol. 135, pp 161, (1909).

[18] S. A. Sawyer and D. L. Hartl, *Population genetics of polymorphisms and divergence.* Genetics 132 (4), 1151-1176, (1992).

[19] S. E. Shreve, *Stochastic Calculus II. Continous-Time Models*. Springer (2004).

[20] J. Wakeley and J. Hey, *Estimating ancestral population parameters.* Genetics 145:847-855. (1997).

[21] S. H. Williamson, R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen, et. al., *Simultaneous inference of selection and population growth from patterns of variation in the human genome*. Proc. Nat. Acad. Sci. USA **102** 7882-7887, (2005).

[22] S. Wright, *Evolution in Mendelian populations*. Genetics 16, 97-159 (1931).

SERGIO LUKIĆ, DEPARTMENT OF GENETICS AND BIOMAPS INSTITUTE, RUTGERS UNIVERSITY, PISCATAWAY NJ 08854, USA

E-mail address: lukic@biology.rutgers.edu

JODY HEY, DEPARTMENT OF GENETICS, RUTGERS UNIVERSITY, PISCATAWAY NJ 08854, USA

KEVIN CHEN, DEPARTMENT OF GENETICS AND BIOMAPS INSTITUTE, RUTGERS UNIVERSITY, PISCATAWAY NJ 08854, USA