The expected sample allele frequencies from populations of changing size via orthogonal polynomials

Lynette Caitlin Mikula [a,*], Claus Vogl [b,c]

[a] *Centre for Biological Diversity, School of Biology, University of St. Andrews, St, Andrews KY16 9TH, UK*
[b] *Department of Biomedical Sciences and Pathobiology, Vetmeduni Vienna, Veterinärplatz 1, A-1210 Wien, Austria*
[c] *Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, A-1210 Wien, Austria*

ARTICLE INFO

ABSTRACT

In this article, discrete and stochastic changes in (effective) population size are incorporated into the spectral representation of a biallelic diffusion process for drift and small mutation rates. A forward algorithm inspired by Hidden-Markov-Model (HMM) literature is used to compute exact sample allele frequency spectra for three demographic scenarios: single changes in (effective) population size, boom-bust dynamics, and stochastic fluctuations in (effective) population size. An approach for fully agnostic demographic inference from these sample allele spectra is explored, and sufficient statistics for stepwise changes in population size are found. Further, convergence behaviours of the polymorphic sample spectra for population size changes on different time scales are examined and discussed within the context of inference of the effective population size. Joint visual assessment of the sample spectra and the temporal coefficients of the spectral decomposition of the forward diffusion process is found to be important in determining departure from equilibrium. Stochastic changes in (effective) population size are shown to shape sample spectra particularly strongly.

1. Introduction

Population demography is difficult to infer reliably, even without considering its interplay with different modes of selection (Johri et al., 2022). Here, we consider an analytically tractable framework for modelling and inferring demography that is efficient in terms of its use of the often limited information within site frequency data. Specifically, we utilise a previously proposed hidden Markov method (HMM) inspired approach (Bergman et al., 2018b) that allows for exact computation of the expected sample allele frequency spectrum of a single population across time, assuming this population evolves according to a diffusion process with a generator that permits a spectral decomposition. We will assume a Moran diffusion process for drift and small mutation rates (Vogl and Bergman, 2016). The HMM-inspired method, especially when applied to a diffusion with small mutation rates, is particularly efficient for modelling non-constant population size dynamics through approximation by sequences of epochs of piecewise constant population sizes, and this will form the backbone of our article. We develop, analyse, and discuss three specific demographic models: (i) a single deterministic shift in population size, to model simple population growth or shrinkage, (ii) series of alternating fixed population sizes, which mimic population boom-bust cycles, and (iii) stochastic changes in population size that occur according to an autoregression model. The deterministic demographic models (i) and (ii) are hardly new within this modelling context (Song and Steinrücken, 2012; Lukic et al., 2011; Zivkovic and Stephan, 2011; Zivkovic et al., 2015), however, our assumption of small mutation rates further simplifies the exact calculation of expected sample allele frequencies in these cases. To the best of our knowledge, the stochastic demographic models (iii) are novel. They utilise the theory of autoregression processes commonly used in econometrics (see Johnston, 1960, Chapter 7). Further, we find sufficient statistics for inferring changing population sizes (under simplifying assumptions). This enables us to propose a small-scale but reliable inference framework and contribute to the discussion around the limitations of inference of demography from site frequency data, particularly using spectral methods (Myers et al., 2008; Bhaskar and Song, 2014).

Both Kingman's *n*-coalescent (Wakeley, 2009, Chapter 3) and diffusion approaches have been used to generate the expected distribution of allele frequencies (*i.e.,* the site frequency spectrum) of population samples (*e.g.,* Sawyer and Hartl, 1992; Griffiths and Tavaré, 1998; Fu, 1995; Griffiths, 2003; Evans et al., 2007). Coalescent approaches yield accurate closed form solutions for small to moderate sample sizes, but for large sample sizes numerical problems generally arise. Coalescent approaches involve multiple steps: Starting with all observed lineages

---

* Corresponding author.
E-mail addresses: lcm29@st-andrews.ac.uk (L.C. Mikula), claus.vogl@vetmeduni.ac.at (C. Vogl).

at the present time and looking back, each expected coalescent time, *i.e.,* the duration for which the number of lineages remains constant, is determined. Conditional on the coalescent time and number of lineages present, the expected number of mutations on each thereby constructed branch of the genealogy is calculated. The expected sample allele frequencies follow from summing mutations along the bifurcating lineages. This process is cumbersome particularly with changing population sizes: To circumvent numerical pitfalls, computationally intense simulations are often employed (*e.g.,* Hudson, 2002; Excoffier et al., 2013). Even computationally more efficient analytical solutions for simplified scenarios can require slow high-precision numerics (Marth et al., 2004). Coalescent methods based on utilising the spectral decomposition of the transition matrix describing the decrease of distinct lineages across generations from the sample to the most recent common ancestor (*i.e.,* the "matrix coalescent", (Wooding and Rogers, 2002)) for determining the expected coalescent times, do not require such libraries for stepwise constant population sizes (Wooding and Rogers, 2002; Polanski and Kimmel, 2003) and exponential growth scenarios (Polanski and Kimmel, 2003; Bhaskar et al., 2015). Other growth patterns have also been well-approximated analytically (Chen, 2019). Aside from the incorporation of different growth patterns, the "matrix coalescent" approach has also been extended to the calculation of the expected site frequency spectra of general coalescent processes, which include coalescent models that explicitly include multiple and simultaneous coalescent events (Spence et al., 2016). A more recent, compellingly straightforward but not yet popularised matrix method for determining expected sample spectra assuming either the classic or a more general coalescent model utilises phase-type theory (Hobolth et al., 2019): So-called phase-type distributions are constructed via convolutions or mixtures of individual exponential distributions; since the individual branch lengths in coalescent trees are exponentially distributed, many measures of interest to population geneticists can easily be determined by application of results from phase-type theory.

Incorporation of non-constant population sizes into diffusion equations is considered comparatively conceptually straightforward to the classic coalescent approaches; however, closed form solutions for expected sample allele spectra are available only when considering concrete sampling schemes in combination with the backward diffusion process (Griffiths, 2003). Analytically tractable solutions to the diffusion density itself have long been studied in population genetics: Kimura (1955a) showed that the biallelic Wright–Fisher diffusion for population allele frequencies can be re-parameterised to the so-called oblate spheroidal differential equation. The general solution to the system is an infinite sum over: i) A spatial component constituted by the eigenvectors that depends solely on the allele frequencies. It is given by a class of weighted orthogonal polynomials with a normalisation constant that is determined by the initial population conditions. ii) A temporal component modulating the change of polymorphic allele frequencies over time by the corresponding eigenvalues. Kimura often assumed the infinite sites mutation model (Kimura, 1969), where each novel mutation hits a previously unmutated site. Formal treatment of neutral parent-independent mutation diffusions by others followed Shimakura (1977), Griffiths (1979), and a more direct mathematical approach was introduced by Song and Steinrücken (2012): If the generator associated with the biallelic diffusion process fulfils certain mathematical requirements, a spectral decomposition can be applied to find an analytical representation of the allele transition density (extension to the multiallelic case in Steinrücken et al. (2013)). Applications to diffusion processes with recurrent mutations and selection of arbitrary strength have been demonstrated (which goes beyond Kimura, who models only weak selection). Note that the spectral decomposition of the diffusion generator for populations with deterministic, piecewise constant population sizes under selection and with mutations segregating as per the infinite sites model has also been derived Zivkovic et al. (2015). To obtain explicit results for the expected sample allele frequencies from the transition densities for populations obtained using either of the above methods, binomial sampling at the extant time is required (Griffiths, 2003). This sampling, coupled with a system of ordinary differential equations on the moments of the sample spectrum derived from the time-reversed Wright–Fisher diffusion (Evans et al., 2007), also provides analytical equations for the sample site frequency spectrum while by-passing exact calculation of the transition density itself (Zivkovic and Stephan, 2011; Zivkovic et al., 2015).

In Bergman et al. (2018b), analytical formulae for the transition densities are used to obtain the sample allele spectra via a dynamic programming approach: The continuous allele proportions of biallelic Moran diffusion processes are treated as latent/hidden variables, and realisations of discrete sample allele frequencies at predetermined intervals across time are treated as emission/observed variables. The logic of the classic forward–backward algorithm from Hidden Markov Model (HMM) literature (Rabiner and Juang, 1986) can then be applied for inference: *e.g.,* the probability of observed sample allele frequencies at extant time given assumptions on the population parameters can be determined via the forward algorithm, and the combination of the forward and backward algorithms can be utilised to calculate the joint probabilities of population and sample allele proportions at any point in time. Recall that standard Moran/Wright–Fisher diffusions are considered dual processes to the coalescent in the sense that the expected allele proportions that these models yield are identical, once the state spaces are made comparable by an appropriate function. It can, *e.g.,* be shown that the neutral biallelic Moran diffusion model with recurrent mutations (which is equivalent to parent-independent mutations) in the first half of Bergman et al. (2018b) is dual to variants of Kingman's coalescent with respect to the binomial sampling scheme (Tavare, 1984; Etheridge and Griffiths, 2009a; Chaleyat-Maurel and Genon-Catalot, 2009; Papaspiliopoulos and Ruggiero, 2014). From the HMM theory it follows that the hidden probabilities of population allele proportions can be computed at a cost polynomial in the number of observations via finite mixtures of distributions involving the coalescing dual process (note that this is for constant population sizes, Papaspiliopoulos and Ruggiero, 2014). By comparison, the transition densities of the forward diffusion processes in Bergman et al. (2018b) are already made tractable by expansion into infinite spectral sums involving polynomials, and representation of the binomial sampling scheme as a finite sum of corresponding orthogonal polynomials then enables all HMM tasks to be performed using finite spectral sums of the order of the sample size. This ties to the long-established duality between the spectral representation and the diffusion (Griffiths and Li, 1983; Griffiths and Spanò, 2010). However, the finite expansion up to the haploid sample size K (rather than an infinite expansion) also reduces the computational burden of the system. Note that similar reasoning is implicit in the moment method of Zivkovic and Stephan (2011), Zivkovic et al. (2015), who forgo exact calculation of the transition densities.

In the second half of Bergman et al. (2018b), the same HMM approach is applied to the spectral representation of the diffusion limit of the so-called boundary mutation Moran model (Vogl and Clemente, 2012; Vogl and Bergman, 2016). It is assumed that the expected heterozygosity is less than about $2.5 \cdot 10^{-2}$ (Vogl and Clemente, 2012), which is true for most eukaryotes (Lynch et al., 2016). Through a first-order approximation in the overall scaled mutation rate to the general mutation Moran model, mutations segregating in a sample of moderate size can then be modelled to occur only at previously monomorphic sites and normalised so they enter the polymorphic region at a constant equilibrium rate in relation to drift. While "back-mutations" are generally allowed, they are excluded while the allele is polymorphic.

In this article, we extend the spectral representation of the boundary mutation Moran model for drift and small mutation rates (Vogl and Bergman, 2016) (recap in Section 2.4) to both deterministic and stochastic piecewise constant (effective) population sizes (see Sections 3.1, 3.3, 3.4, 3.5). At this point we would like to note that any appropriate measure of the census or effective population size could

be assumed to vary within our models (hence our use of (effective) population sizes); we explicitly discuss estimation of the effective population size as it relates to our demographic models in our Conclusions (Section 4). In the main part of the article, we examine changes in size occurring on different time scales, we determine analytical expressions for the resulting sample site frequency spectra via the forward algorithm from the HMM-inspired method (Vogl and Clemente, 2012; Vogl and Bergman, 2016) (recap in Section 2.4). These enable us to assess the convergence behaviour of the sample spectra vs the expected convergence of the effective population size for these differing time scales, and discuss whether this matches expectations from past literature. Further, we propose an agnostic inference framework for inferring past (effective) population sizes from these sample spectra (see Section 3.2). Throughout, we wish to emphasise that the use of the boundary mutation model simplifies the spectral representation of the transition densities (which we discuss in Section 2.5): Changes in population size, and thus the overall scaled mutation rate, affect only the temporal component. Therefore, only a single full spectral decomposition is required rather than one per population size change as in the more general mutation models used both in the first half of Bergman et al. (2018b) and in the moment methods of Zivkovic and Stephan (2011), Zivkovic et al. (2015).

2. Methods: general mathematical framework

2.1. Outline

To start off, we provide an introductory sketch of the methods utilised in this article, and their purpose within the context of this article in particular. We will assume L genomic sites, which have evolved independently according to a biallelic mutation-drift model. Such biallelic systems are determined from the four letter DNA alphabet either by polarisation of alleles into ancestral and derived (as in the infinite sites model), or by intentional nucleotide grouping, *e.g.,* by contrasting the bases adenine and thymine vs the bases cytosine and guanine; here, we will assume reversible mutations and thus data in the form of the latter. At the extant time, which we designate as $t = 0$, a sample of K haploid individuals with these L sites is taken and partitioned into an observed site frequency spectrum $\mathbf{y} = (y_0, \dots, y_k, \dots, y_K)$ according to the number of focal alleles k in the sample, which will generally be either the ancestral type or a chosen group of nucleotides, whereby $\sum \mathbf{y} = \sum_k y_k = L$. If we denote the probability of observing k focal alleles as p_k we can write the likelihood of observing the full site frequency spectrum as the following multinomial distribution:

$$\Pr(y_0, \dots, y_K | L, p_0, \dots, p_K). \tag{1}$$

Note that the exact distribution the event probabilities p_k follow will differ according to the underlying mutation model, and need not correspond to the stationary sampling distributions; when they do, we will denote them as $\bar{p}_k$. In any case, the shape of the resulting spectrum given the event probabilities is governed by the overall scaled mutation rate θ (which is a model-dependent multiple of the (effective) population size and the overall genomic point-wise mutation rate per generation μ), as well as the mutation biases β towards the focal allele and $\alpha = 1 - \beta$ away from the focal allele. Throughout this article, we will assume that the source population of our sample has reached an evolutionary equilibrium with respect to the mutation bias. However, the (effective) population size and thus θ are assumed to have changed over time, which runs from $-\infty$ to 0, and consequently the shape of the polymorphic site frequency spectrum sampled at the extant time may differ from equilibrium expectations. Specifically, this means there is an underlying vector θ of past overall scaled mutation rates, where the entries θ_j represent a constant value that the mutation rate takes within the epoch j between the epoch break points at times t_{j-1} and t_j. In the rest of this section, we will provide a mathematical introduction to the specific continuous (diffusion) model that describes the evolution of the

source population, and then introduce the method we apply to analytically determine the discrete event probabilities p_k given the vector θ and the corresponding epoch break points; the expected observed site frequency spectrum given these event probabilities then follows. The novelty of the current article and focus of the following sections is then to vary θ and the epoch break points to model specific demographic scenarios, and then assess the effect on the expected sample spectrum.

2.2. General mutation model

Let us begin by considering the transition rate density $\phi(x, t)$ of population allele frequencies $x \in [0, 1]$ along a continuous time axis. The standard forward Kolmogorov (Fokker–Planck) diffusion equation operating on this transition density for biallelic, reversible mutations and drift at each genomic site is given by:

$$
\begin{aligned}
\frac{\partial \phi(x,t)}{\partial t} &= -\frac{\partial}{\partial x}\left(\theta\beta(1-x)\phi(x,t) - \theta\alpha x\phi(x,t)\right) + \frac{\partial^2}{\partial x^2}x(1-x)\phi(x,t) \\
&= -\frac{\partial}{\partial x}\theta(\beta - x)\phi(x,t) + \frac{\partial^2}{\partial x^2}x(1-x)\phi(x,t),
\end{aligned} \tag{2}
$$

where the first term on the right hand side describes mutation events and the second corresponds to drift. The usual boundary conditions at $x = 0$ and $x = 1$, which we will not reproduce here, ensure that there is no flow of probability outside the unit interval (*i.e.,* that the initial probability mass is conserved at all times) (see McKane and Waxman, 2007). It is well known that the stationary distribution $\phi(x, t = \infty)$ is the beta distribution $beta(\beta\theta, \alpha\theta)$ (Wright, 1931). Binomially sampling K haploid individuals results in the following beta-binomially distributed event probabilities:

$$
\begin{aligned}
\bar{p}_k &= \Pr(k \mid K, \beta, \theta) \\
&= \int_0^1 \binom{K}{k} x^k(1-x)^{K-k} \frac{\Gamma(\theta)}{\Gamma(\beta\theta)\Gamma(\alpha\theta)} x^{\beta\theta-1}(1-x)^{\alpha\theta-1}\, dx \\
&= \binom{K}{k} \frac{\Gamma(\theta)}{\Gamma(\beta\theta)\Gamma(\alpha\theta)} \frac{\Gamma(k+\beta\theta)\Gamma(K-k+\alpha\theta)}{K+\theta} \\
&= betabin(k|K, \beta, \theta).
\end{aligned} \tag{3}
$$

Substituting these $\bar{p}_k$ into Eq. (1) yields the likelihood of the observed sample site frequency spectrum in the general mutation-drift equilibrium.

2.3. Boundary mutation model

Note that if the scaled mutation rate θ is very low, most of the probability mass associated with the stationary distribution of population allele frequencies x will be concentrated at the boundaries (compare to the visualisation of the stationary distribution of a Wright–Fisher process approaching its diffusion limit in Fig. 2B of Der and Plotkin, 2014). Then the same will be true for a sample taken from this stationary distribution. The first order Taylor series expansion in the overall scaled mutation rate θ of the beta-binomial stationary sampling distributions in Eq. (3) results in the following event probabilities (Vogl, 2014b):

$$
\bar{p}_k = \Pr(k \mid K, \beta, \theta) = \begin{cases} \alpha(1 - \beta\theta H_{K-1}) + \mathcal{O}(\theta^2) & k = 0; \\ \alpha\beta\theta\dfrac{K}{k(K-k)} + \mathcal{O}(\theta^2) & 1 \le k \le K-1; \\ \beta(1 - \alpha\theta H_{K-1}) + \mathcal{O}(\theta^2) & k = K, \end{cases} \tag{4}
$$

where $H_{K-1} = \sum_k^{K-1} \frac{1}{k}$ is the harmonic number. Omitting the higher order terms, these correspond to the equilibrium distribution of the discrete boundary mutation Moran model with population size $N = K$ (Vogl and Clemente, 2012; Vogl and Mikula, 2021). We will refer to

this distribution as the following boundary binomial distribution (Vogl and Bergman, 2015, Eq. 13):

$$
boundbin(k \mid K, \beta, \theta) = \begin{cases} \alpha(1 - \beta\theta H_{K-1}) & k = 0; \\ \alpha\beta\theta \dfrac{K}{k(K-k)} & 1 \le k \le K-1; \\ \beta(1 - \alpha\theta H_{K-1}) & k = K, \end{cases}
\tag{5}
$$

due to its similarities with the beta-binomial distribution (see also Appendix B.2.4). Note that such boundary mutation models restrict mutations entirely to the monomorphic boundaries; the mutation rate is then normalised so that the expected equilibrium heterozygosity is maintained regardless of the population size (Vogl and Clemente, 2012; Vogl, 2014b). Simulations confirm that the boundary mutation Moran model approximates the general mutation Moran model well if the expected equilibrium heterozygosity of the population $\alpha\beta\theta <$ 0.025 (Vogl and Clemente, 2012), which holds for the protein coding genes of most eukaryotes (Lynch et al., 2016). Note that the boundary mutation model can be extended to multiple alleles (Schrempf et al., 2016; Burden and Griffiths, 2019a,b; Vogl et al., 2020). Samples from such multiallelic diffusion models with similarly low scaled overall mutation rates are also well approximated by the boundary mutation Moran model because the probability of sites being more than biallelic is also very low (Burden and Griffiths, 2019a,b).

Different to the case of general mutation models, the boundary mutation diffusion cannot be obtained directly by passing $N = K$ to infinity in a corresponding discrete system: The stationary probabilities of observing monomorphic sites in Eq. (4) contains the harmonic number H_{K-1}, which approaches infinity logarithmically and leads to boundary singularities; thus, the sample size must be lower than $K_{max} \approx e^{min(\alpha, \beta)/(\alpha\beta\theta)}$ (Bergman et al., 2018b) for the model to remain mathematically tractable. In Appendix A, we provide an alternative derivation of the boundary mutation diffusion model, the dynamics of which we explain here: Methodologically, the derivation reflects the logic of Kimura's infinite sites model (Kimura, 1969), where novel mutations occur at a steady but infinitesimally small genomic rate of θ and hit a previously unmutated site where "the mutant form is represented only once at the moment of its occurrence, $p = 1/(2N)$" (Kimura, 1969, immediately after Eq. (16)), and is subsequently spread by drift. Note that p is the proportion of the new allele in the population consisting of N diploid individuals. A proportion of exactly p alleles in the population can be represented via a Dirac delta function.

In contrast to Kimura, who differentiates between ancestral and derived (novel) alleles in his infinite sites model, we differentiate between mutations hitting the non-focal and focal alleles in our boundary mutation model. Let $b_0(t)$ be the proportion of the non-focal allele in the population at time t and $b_1(t) = 1 - b_0(t)$ the proportion of the focal allele in the population. The initial proportion of a new mutations is $\epsilon_x = 1/N$ or $1 - \epsilon_x = 1 - 1/N$ and with rates $b_0(t)\beta\theta$ and $b_1(t)\alpha\theta$ respectively. Convergence towards the equilibrium values $b_0(\infty) = \alpha$ and $b_1(\infty) = \beta$ occurs exponentially with rate θ. This rate is low compared to the rate of drift operating on the polymorphic region, which scales with 1. While the monomorphic and polymorphic regions exchange probability mass, no probability fluxes are assumed to cross the closed unit interval. Thus probability mass is conserved within the closed unit interval. Due to the distinct spatial and temporal dynamics of the monomorphic vs the polymorphic regions, we specifically denote the transition rate density of the polymorphic interior (i.e., the open interval $\epsilon_x \le x \le 1 - \epsilon_x$) as $\phi_I(x, t)$. Compactly, the forward boundary mutation diffusion equation can be written as (Vogl and Bergman, 2016, Eq. 41, and see Appendix A):

$$
\frac{\partial \phi_I(x, t)}{\partial t} = \lim_{\epsilon_x \to 0} \Big((\beta\theta b_0(t)/\epsilon_x)\delta(x - \epsilon_x) + (\alpha\theta b_1(t)/\epsilon_x)\delta(x - 1 + \epsilon_x) \Big)
$$
$$
+ \frac{\partial^2}{\partial x^2} x(1 - x)\phi_I(x, t),
\tag{6}
$$

where $\delta(.)$ denotes the Dirac delta function. Together with the dynamics at the boundaries, this system has a solution in the form of the limiting distribution:

$$
\phi(x, t) = \lim_{\epsilon_x \to 0} \Bigg(\Big(b_0(t) - \int_{0+\epsilon_x}^{1-\epsilon_x} (1 - x)\phi_I(x, t)\, dx \Big)\delta(x)
$$
$$
+ \Big(b_1(t) - \int_{0+\epsilon_x}^{1-\epsilon_x} x\phi_I(x, t)\, dx \Big)\delta(x - 1) \Bigg) + \phi_I(x, t),
\tag{7}
$$

which is a variant of the distribution (McKane and Waxman, 2007, Eq. 7) (noting that $\delta(x - 1) = \delta(1 - x)$):

$$
\phi(x, t) = \Pi_0(t)\delta(x) + \Pi_1(t)\delta(1 - x) + \phi_I(x, t).
\tag{8}
$$

The form of $\phi(x, t)$ (see Eq. (7)) ensures that the integral $\int_0^1 \phi(x, t)\, dx = 1$ for all possible values of t (i.e., that $\phi(x, t)$ is normed to one at all times). While probabilistic interpretation of the transition rates is not immediate because of the singularities at the boundaries, binomially sampling K haploid individuals results in regular, well defined event probabilities (which we discuss in detail in Appendix B.2.4):

$$
p_k = \Pr(k|K, \beta, \theta, t) = \int_0^1 \Pr(k|K, x)\phi(x, t)\, dx
$$
$$
= \int_0^1 \binom{K}{k} x^k (1 - x)^{K-k} \phi(x, t)\, dx.
\tag{9}
$$

Indeed, binomially sampling from the stationary limiting distribution $\phi(x) = \phi(x, t = \infty)$ yields stationary event probabilities $\bar{p}_k$ of the boundary-binomial distribution (see Eq. (5)). This validates our representation of the boundary mutation drift diffusion as the limiting approximation $\theta \to 0$ of the general mutation drift diffusion model.

2.4. Spectral representation of diffusion equations

2.4.1. General mutation model

In order to make the transition densities of diffusion equations analytically tractable, they can be represented as spectral sums; these sums are constituted by the weighted eigenfunctions and eigenvalues of the associated diffusion generator. For diffusion models with pure drift (Kimura, 1955b) and with parent-independent mutation (Shimakura, 1977; Griffiths, 1979), the eigendecomposition was found as a solution to the respective differential equations themselves. Much more recently, Song and Steinrücken (2012) introduced a method for directly determining the spectral decomposition of the diffusion generator for models with recurrent mutation and selection. Their result for the transition density of biallelic reversible mutation models is utilised by Bergman et al. (2018b), who construct a forward–backward algorithm for calculating expected allele proportions across time. Differently to Song and Steinrücken (2012), they must therefore model an explicit ancestral population distribution of allele proportions (compare Lukic et al., 2011) as well as an explicit sampling distribution (rather than simply starting from a Dirac delta function and waiting for an infinite amount of time), and furthermore must use both the forward and backward diffusion equations (rather than just the latter). We recapitulate the required results here (details in Section 3.4 of Bergman et al., 2018b, for the remainder of this section, unless otherwise cited):

The spectral expansion of the biallelic reversible mutation model clearly depends on both the mutation bias β as well as the scaled mutation rate θ. We will assume that at a time $t_s < 0$ in the past, the distribution of ancestral population allele proportions can be represented by a density $\rho(x)$. This density must be defined on the unit interval and assumed to integrate to unity, but is otherwise arbitrary i.e. can be chosen according to the context. It can be expanded into a series of (modified) Jacobi polynomials, which are defined by the recursion (compare Eq. 22.3.2 in Abramowitz and Stegun (1970))

$$
R_n^{(\beta, \theta)}(x) = \sum_{l=0}^m (-1)^l \frac{\Gamma(m - 1 + l + \theta)\Gamma(m + \beta\theta)}{\Gamma(m - 1 + \theta)\Gamma(l + \beta\theta)l!(m - l)!} x^l.
\tag{10}
$$

These polynomials satisfy the orthogonality relation:

$$\int_0^1 R_m^{(\beta,\theta)}(x) R_n^{(\beta,\theta)}(x) w^{(\beta,\theta)}(x)\, dx = \delta_{n,m} \Delta_n^{(G;\beta,\theta)}, \tag{11}$$

with respect to the beta-distributed weight function $w^{(\beta,\theta)}(x) = beta(\beta\theta, \alpha\theta)$, whereby $\delta_{n,m}$ is Kronecker's delta, implying that the constant

$$\Delta_n^{(G;\beta,\theta)} = \frac{\Gamma(n+\beta\theta)\Gamma(n+\alpha\theta)}{(2n+\theta-1)\Gamma(n+\theta-1)\Gamma(n+1)},$$

is non-zero only for $m = n$. Knowing this, the ancestral distribution can now be written as:

$$\rho(x) = \phi(x, t=t_s) = \sum_{n=0}^{\infty} \rho_n w^{(\beta,\theta)}(x) R_n^{(\beta,\theta)}(x),$$

where the coefficients ρ_n can be determined via

$$\rho_n = \frac{1}{\Delta_n^{(G;\beta,\theta)}} \int_0^1 R_n^{(\beta,\theta)}(x) \rho(x)\, dx.$$

Thus, the forward diffusion equation representing the trajectory of the population allele proportions for $t_s \leq t \leq 0$, i.e., starting from this ancestral distribution and extending forwards in time, can be represented uniquely as the following spectral sum:

$$\phi(x,t) = \sum_{n=0}^{\infty} \rho_n w^{(\beta,\theta)}(x) R_n^{(\beta,\theta)}(x) e^{-\lambda_n(t-t_s)},$$

where $\lambda_0 = \lambda_1 = 0$ and $\lambda_n = n(n+\theta-1)$ for $n \geq 2$ are the corresponding eigenvalues. Note that this assumes that the mutational parameters β and θ remain constant over time.

As before, we wish to draw a binomially distributed sample of size K from this population at the extant time $t = 0$. In order to do so, we must first more closely consider this binomial sampling scheme, i.e., the distribution:

$$\Pr(k|K,x) = \binom{K}{k} x^k (1-x)^{K-k}. \tag{12}$$

It can equivalently be seen as a polynomial of order K with coefficients $a_{k+i} = \binom{K}{k}\binom{K-k}{i}(-1)^i$. Let $\mathbf{a}(k,K)$ then be the vector of coefficients $a_n(k,K)$, and let $\mathbf{R}^{(\beta,\theta)}$ be the lower triangular matrix of polynomial coefficients of the Jacobi polynomials $R_n^{(\beta,\theta)}(x)$. The binomial distribution can then be expressed in matrix form via the following linear algebraic equation:

$$\mathbf{d}^{(\beta,\theta)}(K,k) = \mathbf{a}(k,K)\mathbf{R}^{(\beta,\theta)} \tag{13}$$

Note that the triangular structure of $\mathbf{R}^{(\beta,\theta)}$ obviates matrix inversion. The entries of the vector on the left hand side, i.e., $\mathbf{d}^{(\beta,\theta)}(K,y)$, can also be obtained via:

$$d_n(K,k) = \int_0^1 \binom{K}{k} x^k (1-x)^{K-k} \mathbf{R}^{(\beta,\theta)} w(x)^{(\beta,\theta)}\, dx. \tag{14}$$

It follows that the binomial sampling distribution from Eq. (12) can be uniquely expanded into Jacobi polynomials:

$$\Pr(k|K,x) = \binom{K}{k} x^k (1-x)^{K-k} = \sum_{n=0}^{K} d_n^{(\beta,\theta)}(K,y) R_n^{(\beta,\theta)}(x). \tag{15}$$

Utilising the previous results, the marginal likelihood of the sample, i.e., the event probability p_k of observing precisely k focal alleles within our binomial sample of size K drawn at time $t = 0$, can be evaluated via the following (wherein we recall the orthogonality relationship to

simplify calculations):

$$
\begin{aligned}
p_k = \Pr(k|K,\beta,\theta,t=0) &= \int_0^1 \binom{K}{k} x^k (1-x)^{K-k} \phi(x,t=0)\, dx \\
&= \int_0^1 \left(\sum_{n=0}^{K} d_n(k,K) U_n^{(\beta,\theta)}(x) \right) \\
&\quad \times \left(\sum_{m=0}^{\infty} \rho_m e^{\lambda_m t_s} w^{(\beta,\theta)}(x) U_m^{(\beta,\theta)}(x) \right) dx \\
&= \sum_{n=0}^{K} d_n(k,K) \rho_n^{(\beta,\theta)} \Delta_n^{(G;\beta,\theta)} e^{\lambda_n t_s}.
\end{aligned}
\tag{16}
$$

Note that this is technically a forward pass of the forward–backward algorithm, which we discuss in more detail in Appendix C.1. Substituting these p_k into Eq. (1) again yields the likelihood of the observed sample site frequency spectrum.

2.4.2. Boundary mutation model

As previously mentioned, the spectral decomposition of the transition density of the pure drift diffusion equation has been known for the good part of a century (Kimura, 1955a), and the eigenvectors and eigenvalues can be recovered via the formalised methodology presented by Song and Steinrücken (2012). However, both boundaries in these approaches are considered exit boundaries, and only the polymorphic region is thus explicitly modelled by the spectral representation. To include specific boundary terms, the spectral representation must be augmented by appropriate boundary conditions related to mutation and/or fixation rates (McKane and Waxman, 2007; Tran et al., 2014, 2013; Vogl and Bergman, 2016). In Appendix B, we provide a full derivation of the spectral representation of the forward and backward boundary mutation diffusion equations that follows the unified approach to solutions of differential equations proposed by McKane and Waxman (2007), and thus clarify the reasoning in the derivation originally provided by Vogl and Bergman (2016). We will cover the main results here.

Traditionally, the eigenfunctions of the diffusion generators of pure drift models (considering only the polymorphic region) are constituted by Gegenbauer polynomials; we here use the modified Gegenbauer polynomials $U_n(x)$ (see Olver et al., 2022, Eq. 18.5.77):

$$U_n(x) = \sum_{l=0}^{n-2} (-1)^{-1-l} \binom{n+l}{l} \binom{n-1}{l+1} x^l \quad \text{for } n \geq 2. \tag{17}$$

These are orthogonal with respect to the weight function $w(x) = x(1-x)$ for all orders of expansion $n, m > 2$:

$$\int_0^1 U_n(x) U_m(x) w(x)\, dx = \delta_{n,m} \Delta_n^{(B)}, \tag{18}$$

where $\delta_{n,m}$ is Kronecker's delta, and

$$\Delta_n^{(B)} = \frac{(n-1)}{n(2n-1)} \tag{19}$$

with $n \geq 2$ is the proportionality constant. The corresponding eigenvalues determining the rate of drift are $\lambda_{n \geq 2} = n(n-1)$. The augmentation of this system required to include fixation at the boundaries yields the following system of forward eigenfunctions with the corresponding eigenvalues $\lambda_0 = 0$, $\lambda_1 = 0$ and $\lambda_{n \geq 2} = n(n-1)$ (see McKane and Waxman (2007), but also Tran et al. (2013), Bergman et al. (2018b)):

$$
\begin{cases}
\mathcal{F}_0^{(\beta)}(x) &= \alpha\delta(x) + \beta\delta(x-1) \\
\mathcal{F}_1^{(\beta)}(x) &= -\delta(x) + \delta(x-1) \\
\mathcal{F}_{n \geq 2}^{(\beta)}(x) &= -\frac{(-1)^n}{n}\delta(x) + U_n(x) - \frac{1}{n}\delta(x-1),
\end{cases}
\tag{20}
$$

where $\delta(.)$ is Dirac's delta. Observe that the point masses at the boundaries of the eigenfunctions $n \geq 2$ balance the probability weight in the polymorphic interior with the probability mass point masses at the boundaries. Importantly for this article, these forwards eigenfunctions

are identical to those of the diffusion generator of the boundary mutation model (Vogl and Bergman, 2016, Eqs. 32, 37) (see the derivation in Appendix B). However, the first-order eigenvalue λ_1 in the boundary mutation diffusion model is given by θ so as to explicitly model the influx of new mutations occurring at either boundary (with the other eigenvalues remaining the same). As covered in Section 2.3, θ also determines the rate of convergence to the equilibrium boundary values associated with the eigenvalue $\lambda_0 = 0$ (see also Eq. (78)), which is slow compared to the rate of drift.

A set of backward eigenfunctions can be constructed to be orthogonal to the forward eigenfunctions above; these are the following polynomials of ascending order (McKane and Waxman, 2007; Tran et al., 2014, 2013; Bergman et al., 2018b) (see also Appendix B):

$$\begin{cases} \mathcal{B}_0^{(\beta)}(x) & = 1 \\ \mathcal{B}_1^{(\beta)}(x) & = x - \beta \\ \mathcal{B}_{n\geq 2}^{(\beta)}(x) & = x(1-x)U_n(x) . \end{cases} \tag{21}$$

This construction can be confirmed by setting $\Delta_0^{(B)} = \Delta_1^{(B)} = 1$, and checking that an analogous orthogonality relation to before holds:

$$\int_0^1 \mathcal{B}_n^{(\beta)}(x)\mathcal{F}_m^{(\beta)}(x)\,dx = \delta_{nm}\Delta_n^{(B)}$$

for a nonzero $\Delta_n^{(B)} = \frac{n-1}{(2n-1)n}$ only if $m = n$.

As with the general mutation model in the previous subsection, the spectral representation of the boundary mutation diffusion can be used within the context of a forward–backward algorithm (Bergman et al., 2018b, details in Section 4.2). This again requires the specification of a distribution of ancestral allele proportions $\rho(x)$ at time $t_s < 0$. As before, this is an arbitrary distribution that we here obtain a spectral representation for by expanding the limiting distribution that represents the solution of the boundary mutation diffusion model (see Eq. (7)) into orthogonal polynomials. Specifically:

$$\rho(x) = \phi(x, t_s) = \tau_0(t_s)\mathcal{F}_0^{\beta}(x) + \tau_1(t_s)\mathcal{F}_1^{\beta}(x) + \sum_{n=2}^{\infty} \tau_n(t_s)\mathcal{F}_n^{\beta}(x) , \tag{22}$$

where $\tau_0(t_s) = 1$, and $\tau_1(t_s) = \beta - b_1(t_s)$ represents the proportion of the focal allele in the population at time t_s. To correctly model the temporal change in focal allele proportions due to mutational input between $t_s < t \leq 0$, we must set $\tau_0(t) = 1$ and $\tau_1(t) = (b_1(t_s) - \beta)e^{-\lambda_1(t-t_s)}$ (compare Eq. (85)). For the polymorphic interior, the changes in population allele proportions over time $t_s < t \leq 0$ due to drift can be found by substituting the spectral sum for the polymorphic transition rate density $\phi_I(x, t) = \sum_{n=2}^{\infty} \tau_n(t)F_n^{\beta}(x)$ into the forward diffusion equation Eq. (6). This induces the following system of inhomogeneous differential equations for the temporal coefficients (Vogl and Bergman, 2016, Eq. 43) (see also Appendix B.2, Eq. (84)):

$$\frac{d}{dt}\tau_n(t) = -\lambda_n\left(\tau_n(t) - \alpha\beta\theta\frac{E_n}{\lambda_n}\tau_0(t) - \theta\frac{O_n}{\lambda_n}\tau_1(t)\right) . \tag{23}$$

for $n \geq 2$, with:

$$E_n = \frac{U_n(0) + U_n(1)}{\Delta_n^{(B)}} = -(2n-1)n\frac{(-1)^n + 1}{\Delta_n^{(B)}}$$

and

$$O_n = \frac{U_n(0)\beta + U_n(1)\alpha}{\Delta_n^{(B)}} = -(2n-1)n\frac{(-1)^n\beta - \alpha}{\Delta_n^{(B)}} .$$

Importantly, this eigensystem can be diagonalised to a time-homogeneous form (which we show in Appendix B.2.2). However, we will proceed with the above system for this article (details follow in Section 2.5).

At the extant time $t = 0$, we once again draw a binomial sample of size K from the population that has evolved according to the above. The binomial likelihood itself can, similarly to the previous subsection, be

uniquely expanded into a series of appropriate backward polynomials at $t = 0$:

$$\begin{aligned} \Pr(k|x, K) &= \binom{K}{k} x^k(1-x)^{K-k} \\ &= \sum_{n=0}^{K} d_n(k, K)\mathcal{B}_n^{(\beta)}(x) , \end{aligned} \tag{24}$$

where $d_n(k, K)$ are again coefficients depending on k and K that can be obtained analogously to either Eq. (13) via a matrix multiplication or to Eq. (14) via integration:

$$d_n(K, k) = \int_0^1 \binom{K}{k} x^k(1-x)^{K-k}\mathcal{F}_n^{\beta}(x)\,dx . \tag{25}$$

Note that the coefficients $d_n(k, K)$ of the boundary mutation expansion are of course different from those of the general mutation expansion, but we choose to not distinguish them in their notation since they are always associated with the appropriate eigenfunctions $R_n^{(\beta,\theta)}$ vs $\mathcal{B}_n^{(\beta)}(x)$ or with the constants $\Delta_n^{(G;\dots)}$ or $\Delta_n^{(B)}$, from which context it is obvious which model is assumed. Either way, Eq. (24) represents a straightforward linear transformation, since a binomial sample can be represented as polynomial of order K, just like the backwards eigenfunction $\mathcal{B}_n^{(\beta)}(x)$.

Overall, we can thus determine the likelihood of the sample, *i.e.*, the event probability p_k of observing precisely k focal alleles within our binomial sample of size K drawn at time $t = 0$, via the following (wherein we again use the orthogonality relationship of the orthogonal polynomials to simplify calculations):

$$\begin{aligned} p_k = \Pr(k|K, \beta, \theta, t = 0) &= \int_0^1 \Pr(k|K, x)\phi(x, 0)\,dx \\ &= \int_0^1 \binom{K}{k} x^k(1-x)^{K-k}\phi(x, 0)\,dx \\ &= \int_0^1 \left(\sum_{n=0}^{K} d_n(k, K)\mathcal{B}_n^{(\beta)}(x)\right) \\ &\quad \times \left(\sum_{m=0}^{\infty} \tau_m(0)\mathcal{F}_m^{(\beta)}(x)\right)dx \\ &= \sum_{n=0}^{K} d_n(k, K)\tau_n(0)\Delta_n^{(B)} . \end{aligned} \tag{26}$$

Clearly, this is again a forward pass of the forward–backward algorithm (see Appendix C.1), and substituting these p_k into Eq. (1) yields the likelihood of the observed sample site frequency spectrum.

Remark. In the formulae for the event probabilities p_k assuming either the general or boundary mutation diffusion for the population allele frequencies (Eq. (16) and Eq. (26) respectively), polynomial expansions only up to the sample size K rather than to infinity are required. This is due to our application of the HMM forward–backward scheme to a system of orthogonal polynomials: the sample size K at time $t = 0$ determines (backwards in time) the order of the expansion needed for the forward eigensystem. In principle, this approach to determining expected sample site frequency spectra therefore has a low computational burden. Directly implemented as above, the event probabilities p_k can be evaluated in a matter of seconds in the open source statistics software R (R. Core Team, 2017); however, exceeding a sample size of $K = 24$ incurs numerical instability. With the high precision floating point library MPFR (Fousse et al., 2007), reliable results can be obtained up to $K = 37$. Clearly, specialised programs are still lacking for application to larger samples. However, direct implementation of these formulae suffices for model-validating simulations and application to small or down-sampled data sets as required for this article.

Remark. In Section 2.3, we show that the spectral representation of the backwards diffusion rate densities of both the general mutation and boundary mutation models has a dual representation as a time

dependent pure-death jump process on the space of binomial distributions with respect to their respective transition rate distributions $\psi(x,t)$. Specifically, at each time point $t_s < t \leq 0$ and each order of expansion $0 < n < K$, the expected event probabilities p_k can be evaluated to the same result. In the case of the boundary mutation diffusion model, $\psi(x,t)$ is the limiting stationary distribution but not a density in the traditional sense. The application of the HMM framework and the interpretation of the dual process are still possible, however, precisely because the sampling distribution at each time point $t_s < t \leq 0$ and each order of expansion $0 < n < K$ is a regular discrete probability distribution even in non-equilibrium (which can be seen in Section 2.3, compare Section 4.2 in Papaspiliopoulos and Ruggiero (2014)).

Remark. Recall that the forward eigenfunctions of the boundary mutation diffusion are given by $\mathcal{F}_{n\geq2}^{(\beta)}(x)$. Interestingly, these can also be considered a zeroth-order Taylor series expansion in the scaled mutation rate θ of the weighted Jacobi polynomials $w^{(\beta,\theta)}R_{n\geq2}^{(\beta,\theta)}(x)$ (Vogl and Bergman, 2016, Appendix A.1.). This is of note because analogously, the discrete boundary mutation Moran model (Vogl and Clemente, 2012) was first derived as a first-order expansion in the overall scaled mutation rate θ of the general biallelic Moran model with separately parameterised drift and mutation terms (Baake and Bialowons, 2008).

2.5. Piecewise constant (effective) population size

Recall that our main aim is to analytically determine the expected site frequency spectra under a variety of demographic scenarios. So far, we have determined that utilising the spectral representation of diffusion models within a forward pass of a HMM forward–backward algorithm yields a computationally efficient framework for analytically calculating expected site frequency spectra, but have not explicitly introduced demographic changes into these models. In this section, however, we will introduce a time axis that is partitioned into a sequence of epochs between which the population undergoes changes in (effective) sizes; this will be reflected in a change of the overall scaled mutation rate θ and re-scaling of time between individual epochs. Incorporating this into the general mutation model requires that any change in the scaled mutation rate θ or the mutation bias β must be matched by a change in the base of the Jacobi polynomials $R_n^{(\beta,\theta)}$. This is cumbersome and complicates the treatment of the demographic models. We will therefore proceed with the spectral representation of the boundary mutation diffusion model for the remainder of this article. As long as we use the un-diagonalised spectral representation of this model, demographic changes do not affect the base of the Gegenbauer polynomials within the eigensystem. These can thus be considered the purely spatial component of the spectral representation of the transition rate density, depending on the allele frequencies but not on the overall scaled mutation rate or mutation bias. Changes in the mutation parameters only require appropriate modification of the temporal coefficients that solve the time-inhomogeneous system of equations from Eq. (23).

A Simplification

As a further simplification within our modelling approach, we will assume that the mutation bias β is evolutionarily stable, *i.e.,* does not change between epochs. This enables us to examine the effect of demography on the allele frequencies via the overall scaled mutation rate in isolation from any other evolutionary force. We can then assume that the proportion of the focal allele within the population has already converged to β, *i.e.,* that $\tau_1(t)$ has converged to 0. Indeed $\tau_1(t)$ converges to 0 at rate θ even with changing population sizes as long as the mutation bias β remains constant (recall that the diffusion time is scaled by the population sizes). Thus, the only requirement for convergence of τ_1 to zero is that our process starts at $t = -\infty$.

Note that only a further change in β, which we do not permit, but not a change in θ can then induce $\tau_1(t) \neq 0$. Hence, we set $\tau_1(t) = 0$ and thus $b_0(t) = \alpha$ and $b_1(t) = \beta$. As a result, the time-inhomogeneous

system for the temporal coefficients from Eq. (23) can be simplified to (see Appendix B.2.4):

$$\frac{d}{dt}\tau_n(t) = -\lambda_n\tau_n(t) + \alpha\beta\theta(2n-1)n((-1)^n + 1), \text{ for } n \geq 2. \quad (27)$$

Since $\tau_1(t) = 0$ and $\lambda_{n\geq2} = n(n-1)$ are much greater than $\lambda_1 = \theta$, we deduce that all odd $\tau_{n\geq2}(t)$ converge to zero very rapidly. This implies that we need only consider the even temporal coefficients in our modelling approach, *i.e.,* that the polymorphic spectrum remains symmetric regardless of potential demographic changes because the mutation bias undergoes no further shifts. This symmetry does not pertain to the monomorphic boundaries, which converge to a symmetric equilibrium only if $\beta = \alpha = \frac{1}{2}$.

Introducing Demography

From now on, we will assume that time is segmented into epochs indexed by $j \in \{1, \ldots, J\}$. Each epoch starts with time point t_{j-1} and ends with time point t_j. We will usually end the last epoch at the extant time $t_J = 0$, since our samples are assumed to come from the present. Let us define the epoch lengths as $s_j = t_j - t_{j-1}$. We will assume that the (effective) population size remains constant at N_j within epoch j and then instantaneously switches to N_{j+1} at the onset of epoch $j + 1$. This translates to corresponding changes in what we will call the bias-complemented overall scaled mutation rate, defined per epoch j as $\theta_j^* = \alpha\beta\theta_j$. Note that θ_j^* also corresponds to the equilibrium solution of the boundary mutation diffusion model. Since we only consider the polymorphic part of the sample spectrum in our later demographic models, we will specify only the values of θ_j^* in our examples/simulations rather than any component values.

Some caution is required around our treatment of the first epoch: In our model where we assume a single change in (effective) population size (Section 3.1), as well as in our inference approach (Section 3.2), we set the beginning of the first epoch to $t_{j=0} = -\infty$, and thus the initial interval s_1 is assumed to be infinitely long. In the boom-bust (Section 3.3) and stochastic fluctuation models (Section 3.4, and Section 3.5), we obtain convergence results for an infinite number of epochs J, which means that the initial epoch has no real impact on the outcome. Therefore, we can readily assume that temporal functions are equal to their equilibrium values at the end of the first epoch, so at t_1 (having either converged there, or being generally negligible). These values are:

$$\begin{cases} \tau_0 = 1 \\ \tau_1 = 0 \\ \tau_n(t) = \Xi_{n\geq2}\theta_1^*, \text{ for } t = t_1, \end{cases} \quad (28)$$

with $\Xi_n = -\frac{4n-2}{n-1} = \frac{E_n}{\lambda_n}$.

Moving forwards in time, the bias-complemented overall scaled mutation rate θ_j^* changes from epoch to epoch; within each, the population allele frequencies converge towards their current expected equilibrium. This convergence is governed by the time-inhomogeneous equation for the odd temporal coefficients from Eq. (27), which can now be re-parameterised again. Specifically, for even n, *i.e.,* $n \in 2\mathbb{N}$, and $j \geq 2$, the system within each epoch is given by:

$$\frac{d}{dt}\tau_n(t)\Xi_n^{-1} = -\lambda_n\left(\tau_n(t)\Xi_n^{-1} + \theta_j^*\right), \text{ for } t_{j-1} < t \leq t_j. \quad (29)$$

This applies to all models within this paper. Setting the starting condition specifically as the epoch start point $\tau_n(t_{j-1})$, the solution of this differential equation is:

$$\tau_n(t)\Xi_n^{-1} = \theta_j^* - \left(\theta_j^* - \tau_n(t_{j-1})\Xi_n^{-1}\right)e^{-\lambda_n(t-t_{j-1})}, \text{ for } t_{j-1} < t \leq t_j. \quad (30)$$

Note that each epoch with $j \geq 2$ now technically has an ancestral allele distribution $\rho(x) = \phi(x, t = t_{j-1})$ that is precisely the final allele configuration of the previous epoch; these are determined by substituting the temporal coefficients of the two equation systems immediately above into the representation of $\rho(x)$ as a spectral sum

using Gegenbauer polynomials given in Eq. (22). Explicit modelling of the system therefore actually begins at the end of the first epoch, which represents an equilibrium state.

At the extant time, which is also the end of the last epoch, we will as always assume that a binomial sample of haploid size K is drawn; more explicitly, we assume that the binomial sampling probabilities are expanded into a series of backward polynomials as in Eq. (24). Then the event probability p_k of observing precisely k focal alleles within this sample, after the population has undergone a series of piecewise constant demographic scenarios, can be determined via Eq. (26):

$$
\begin{aligned}
p_k &= \Pr(k|K,\beta,\theta,t=0) \\
&= -\sum_{n=2}^{K} \frac{2}{n} d_n(k,K)\theta_1^* \qquad \text{for } k \in \{1,\dots,K-1\},
\end{aligned}
\tag{31}
$$

for the polymorphic spectrum. The expected event probabilities at the boundaries follow accordingly:

$$
\begin{aligned}
p_0 &= \Pr(k=0|K,\beta,\theta,t=0) = \alpha - \sum_{l=1}^{K-1} \frac{1}{2} p_l \\
p_K &= \Pr(k=K|K,\beta,\theta,t=0) = \beta - \sum_{l=1}^{K-1} \frac{1}{2} p_l .
\end{aligned}
\tag{32}
$$

Importantly, recall that only an expansion of the temporal coefficients up to the sample size K is required in order to determine the event probabilities given the sample at the extant time. Further, note that although we have assumed that $b_0(t)$ and $b_1(t)$ remain constant at α and β respectively, the frequencies of the monomorphic classes in the sample may nevertheless vary over time with the changes in the mutation rate.

It should be noted that Eqs. (30)–(31) will form the basis for the analyses of specific demographic models in Section 3, where we look at the effect of these models on the polymorphic sample spectrum.

3. Theory: demographic models

3.1. Single shift in population size (model 1)

Now consider a population that undergoes a single change in (effective) population size, meaning there are only two epochs: The ancestral epoch runs between $t_0 = -\infty$ and $t_1 < 0$ with the overall bias-complemented ancestral mutation rate of θ_1^*; the overall bias-complemented mutation rate switches instantaneously to θ_2^* at t_1 and remains constant until the sample is taken at $t_2 = 0$. The population allele frequencies thus evolve from the old equilibrium attained before t_1 towards a new one according to Eq. (30) after this point, with $\tau_n(t_1)\Xi_n^{-1} = \theta_1^*$ for $n \in 2\mathbb{N}$:

$$
\tau_n(t)\Xi_n^{-1} = \theta_2^* - \left(\theta_2^* - \theta_1^*\right) e^{-\lambda_n \frac{\theta_1^*}{\theta_2^*}(t-t_1)}, \text{ for } t_1 < t \le 0..
\tag{33}
$$

The marginal probabilities of sample allele frequencies can then be determined as outlined in Section 2.5. Note that if $\theta_1^* \ll \theta_2^*$, the solution above can be approximated by the solution of a simple one-parameter model (see Appendix D.1).

3.1.1. Results and discussion for model 1

By calculating the marginal probabilities for each possible segregating value of the focal allele, an expected biallelic polymorphic sample site frequency spectrum can be constructed and assessed for departure from equilibrium. Visual representations that contrast an observed or simulated distribution of segregating sites with a neutral spectrum are simple and informative (Nawa and Tajima, 2008; Achaz, 2009): Fig. 1 shows a series of expected polymorphic sample spectra drawn from a growing (panel 1 A) and a shrinking (panel 2 A) population at successive time points after the shift in (effective) population size. The log ratio of these expected sample spectra vs a sample assumed to

be in equilibrium at the ancestral overall-biased scaled mutation rate (see Vogl and Clemente, 2012, Eq. 13) are depicted in Fig. 1 panels 1B and 2B; these second plots emphasise the distinct demographic signatures. In Fig. 1 panels 1 A and 1B, which show the samples from a growing population, the presence of excess rare alleles is recognisable. Conversely, Fig. 1 panels 2 A and 2B show an increased proportion of intermediate frequency alleles for the samples from a shrinking population. These are well-established phenomena in population genetics and a battery of neutrality tests have been constructed to detect them (Tajima, 1983; Fu and Li, 1993; Fu, 1995; Fay and Wu, 2000; Korneliussen et al., 2013), among them Tajima's D. For such tests, the site frequency spectrum is construed as the minor allele frequency distribution and is given by either the number of polymorphic sites ζ_k at frequency $\frac{k}{K}$, where $k \in [1, K-1]$ when an outgroup is available, or by the number of polymorphic sites $\zeta_{k,K-k}$ at frequencies $\frac{k}{K}$ and $\frac{K-k}{K}$ when ancestral and derived alleles cannot be distinguished. The classic neutrality tests contrast two estimators for the overall scaled mutation rate θ that are sensitive to deviations from neutrality in different regions of the allele frequency spectrum. These estimators are constructed by differently weighted linear combinations of $\hat{\theta}_k = k\zeta_k$ (Achaz, 2009). In Fig. 1 panels 1C and 2C, a version of Tajima's D modified for the boundary mutation model, which we will call the D-statistic (see Appendix D.2 for its derivation), is inferred for the series of samples: This clearly shows how the signal of the demographic event first becomes clearer in the samples over time, and then fades to undetectable as a new equilibrium is approached (compare also the change in deviation from equilibrium over time in Fig. 1 panels 1 A, 1B and 2 A, 2B. It is also apparent that populations lose polymorphism much faster than they accrue it (compare Nei et al., 1975), because the drift time scale is much faster than the mutation time scale.

3.2. An agnostic inference approach

Let us now assume a population that undergoes multiple shifts in (effective) population size over time, yielding a history of J demographic epochs, indexed with $1 \le j \le J$, and let us assume the time points t_j as given, with $t_0 = -\infty$, $t_J = 0$ and $t_{j-1} < t_j$. The general solution of the time-inhomogeneous equations at $t = t_J$ is then for $n \in 2\mathbb{N}$:

$$
\begin{aligned}
\tau_n(t_J)\Xi_n^{-1} = {}& \theta_J^*\left(1 - e^{-\lambda_n r_J}\right) + \sum_{j=2}^{J-1} \theta_j^*\left(e^{-\lambda_n r_J} - e^{-\lambda_n \sum_{j=1}^{J} r_j}\right) \\
& \times \theta_1^*\left(e^{-\lambda_n \sum_{j=1}^{J} r_j}\right),
\end{aligned}
\tag{34}
$$

where r_j is shorthand for the scaled epoch length $r_j = \frac{\theta_{j-1}^*}{\theta_j^*} s_j$.

It has previously been noted that sampled spectra are not fully informative of a population's demographic history (Myers et al., 2008). With orthogonal polynomial expansions, the demographic history is mapped onto a function space spanned by $e^{-\lambda_n r_j}$ with $1 \le j \le J$, which implies several things: Because the exponential function decays rapidly, increasing orders of expansion have a decreasing influence on the shape of the spectrum. Furthermore, demographic histories orthogonal to the function space remain "hidden" - in effect, this means past events can cancel each other out and produce a 'simplified' historical trajectory. In other words, an infinite number of demographic histories can actually produce the same set of spatial and temporal coefficients and therefore the same observed spectrum. However, it was later shown that the expected sample spectra results uniquely from demographic models as long as these models are defined piecewise using population size functions that do not oscillate/switch signs too often, and as long as the samples are sufficiently large (Bhaskar and Song, 2014); note, however, that fixed epoch start and end times were assumed. Generally, these past discussions have also always assumed an infinite-sites mutation scheme and corresponding either polarised or unpolarised (i.e., folded)
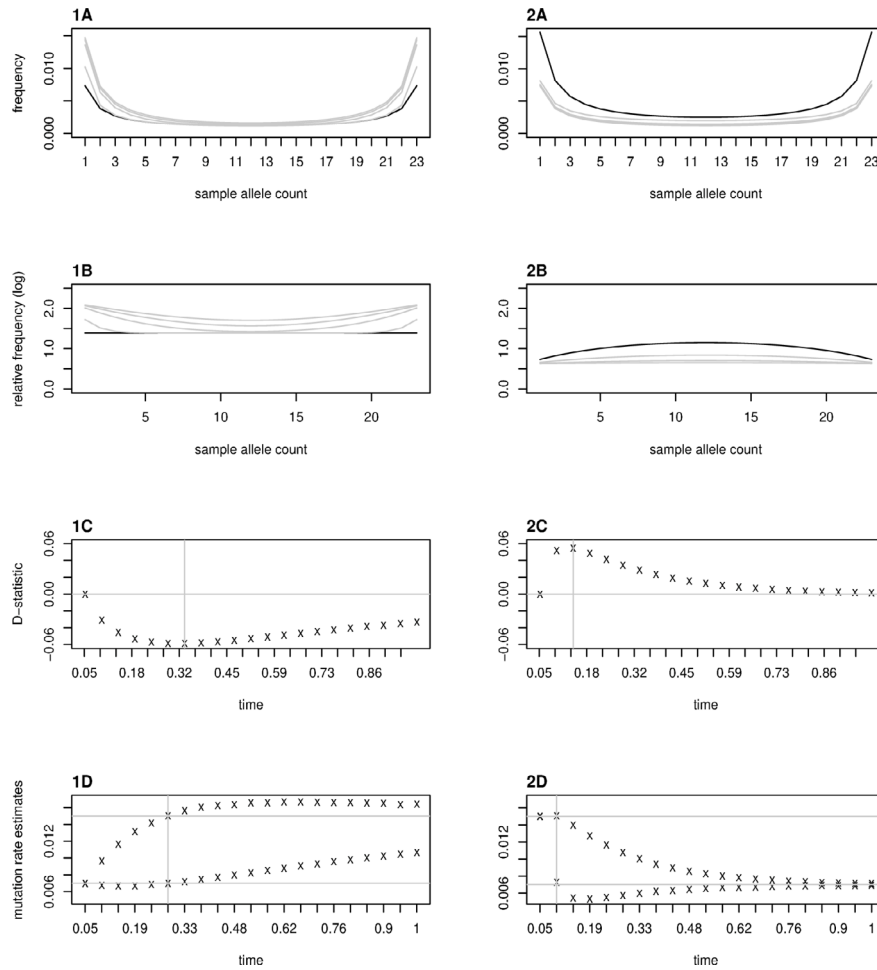
Fig. 1. These figures examine the effect of an increase (1A-1D) and a decrease (2A-2D) in the overall bias-complemented scaled mutation rate of the polymorphic site frequency spectrum. Specifically, we examine samples of size $K = 24$ from a population whose mutation rate has increased from $\theta_1^* = 0.007$ to $\theta_2^* = 0.015$ (1A-1D) or decreased in the reverse (2A-2D) at times $t_1 = c(-0.083, -0.292, -0.500, -0.708)$ back from the time of sampling (see Section 3.1). In (A), the sampled polymorphic site frequency spectra are shown with the first sample in black and the succeeding in grey; in (B) the log ratio between the sampled polymorphic spectra and the equilibrium distribution of the boundary mutation Moran model with the ancestral overall bias-complemented scaled mutation rate is shown in a similar way. The figures (C) show the value of the D-statistic inferred from each sample at the times since the shift, with the vertical grey line marking the most extreme value. The plots in (D) show our inferred values of both the ancestral and the current bias-complemented overall scaled mutation rate from each sample across time determined via Eq. (34) and the procedure outlined beneath it, with the true values indicated by the horizontal grey lines.

sample site frequency spectra. Here, we motivate a new analytic inference approach for piecewise constant demographic histories with unknown epoch break points, where the starting point is a biallelic sample spectrum.

To start, let us return to Eq. (34). In order to infer all θ_j for $1 \leq j \leq J$ and therefore all the past (effective) population sizes attained at the end of each epoch, (i) the left hand side of the equation, *i.e.,* the temporal dynamics, must be determined from the sampled spectrum, and (ii) the r_j, *i.e.,* the scaled time points of demographic events, must also be specified in some way. Consider a sampled spectrum of size K comprised of a total of L loci, with $L_{(0)}$ and $L_{(1)}$ loci fixed for the respective alleles as well as $L_{(0,1)}$ polymorphic loci, so that $L = \sum (L_{(0)} + L_{(0,1)} + L_{(1)})$. In the context of the biallelic boundary mutation Moran model, these observed counts of loci are sufficient statistics for the probabilities of the respective monomorphic and polymorphic events, meaning that they contain all the information from the data that can be used to construct estimators of these events (Vogl et al., 2020). It follows that $\hat{\alpha} = \frac{L_1 + (L_{(0,1)}/2)}{L}$ is the minimum variance, unbiased maximum likelihood estimator for the mutation bias (Vogl, 2014c; Vogl et al., 2020). Hence, the mutation bias can be immediately estimated from the observed spectrum. Then the polymorphic spectrum can be

symmetrised, yielding unbiased sample allele occupancy probabilities for $k = 1, \ldots, K - 1$. These can be plugged into the left side of the polymorphic equation for expected sample allele frequencies given the past population history from Eq. (31). The purely spatial binomial coefficients on the right hand side of this equation can also be readily determined, so only the temporal coefficients $\tau_n(t_J)$ remain unknown. Seen for all the inferred polymorphic allele frequencies simultaneously, Eq. (31) clearly describes a consistent system of $K - 1$ equations that can be uniquely solved for the $K - 1$ temporal coefficients. These are required for the left hand side of the system of equations in Eq. (34).

As the number and duration of the past population epochs is generally unknown, specification of the scaled epoch lengths determined by the r_j must be *ad hoc*. Recall that only the even-order solutions of the temporal equations are non-zero since the mutation bias is assumed constant. For example, a sample spectrum of size $K = 6$ is shaped by the polynomials of degree $n = \{2, 4, 6\}$. Furthermore, for Eq. (34) to constitute a unique solution, the number of epochs J plus one (the additional one being for the ancestral state) must be equal to the number of informative polynomial coefficients. In our example with $K = 6$, this means we can uniquely infer the current as well as two past overall bias-complemented overall mutation rates. A convenient

agnostic placement of the time points at which to make these inferences is inspired by the half-life of exponential decay: Setting $\bar{r}_j = \frac{\log(2)}{\lambda_{2j+1}}$, each epoch is placed where half the change in overall-scaled mutation compared to the preceding epoch is expected to have occurred based on the eigenvalues of the observed spectrum. And within epoch, the amount of change captured by each order of expansion is proportional to the corresponding eigenvalue (see Fig. 2). There is no guarantee that these r_j are the true scaled time points at which the demographic events occurred; they are simply "well placed" for characterising the effect of exponential change. Note that only the lowest (or to a lesser extent the second lowest) order of expansion from the first epoch significantly influences the shape of the spectrum at the extant time, whereby all (or almost all) orders of expansion from later epochs influence the shape of the spectrum at the extant time, albeit less drastically.

Altogether, we can conclude that *if* we assume the same number of (effective) population sizes – current and historical – as there are even coefficients in the polynomial expansion of the observed site frequency spectrum, *and if* the demographic events that caused the changes in size are assumed to be well placed in the above sense, unique solutions for all current and historical bias-complemented overall mutation rates θ_j can be found from the system of equations in Eq. (34). These resulting estimators for all the θ_j are derived from sufficient statistics via two consistent systems of equations with unique solutions (one-to-one mappings of parameter spaces); therefore the resulting estimators for the θ_j are unique minimum unbiased estimators (see Hogg and Craig, 1995, Chapter 10).

Both the accuracy and potential for wider application of this method are, as yet, limited by numerical instability: For our previous example with sample size $K = 6$, a total number of loci in the order of 10^8 is required for the accuracy of all three estimators to be adequate (see Fig. 3). This is, however, the upper limit of loci in a site frequency spectrum that can be simulated using the open software program R. It is clear that the feasibility of large-scale use of this method hinges not only on specialised implementation being developed in the future, but also on the availability of sets with a large enough number of loci.

In Fig. 1 panels 1D and 2D, the sample allele occupancy probabilities from the single-shift examples of the previous subsection are hypergeometrically downsampled to $K = 4$ and used to generate spectra of $2 \cdot 10^8$ independent loci, from which both the extant and the ancestral overall bias-complemented scaled mutation rate are inferred. This novel inference approach infers the true values most accurately a touch sooner after the change in (effective) population size than the D-statistic detects the greatest signal of departure from equilibrium (compare Fig. 1 panels 1C and 2C). Note the pattern of inferring slightly more extreme values for the ancestral mutation rate before and for the current mutation rate after this point to compensate for the positioning of the hypothetical epoch split time.

3.3. Deterministically alternating population sizes (model 2)

In this subsection, we construct a caricature model of boom-bust population size dynamics: To do so, we assume a population that has haploid (effective) size N_u for an epoch of length s_u (where u stands for the b*u*st epoch), and then switches instantaneously to having haploid (effective) size N_o for an epoch of length s_o (where o stands for the b*o*om epoch). Together, these two epochs form a cycle that repeats indefinitely.

Classic population genetic results apply to the overall (effective) size of this population: If the epochs are short compared to generation length (which is the inverse population size in the diffusion setting), the long-term effective size can be approximated by the harmonic mean across the two epochs (Wright, 1938; Kimura, 1963). The heterozygosity of populations undergoing deterministic, cyclical changes in (effective) size has been studied extensively for various trajectories of rapid population growth followed by instantaneous collapse; the harmonic mean approximation holds both without mutations and when
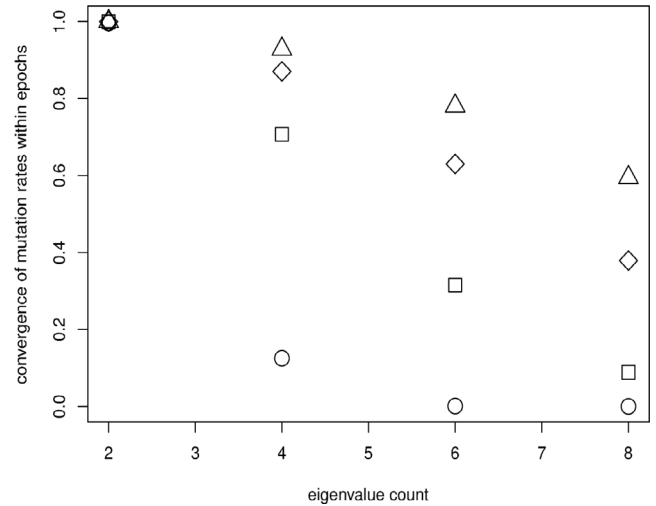

Fig. 2. Here we visualise the function space that demographic histories are mapped onto for our inference approach: Specifically, this space is spanned by the values of the exponential $e^{-\lambda_n \bar{r}_j} = e^{-\lambda_n \frac{\log(2)}{\lambda_{2j+1}}}$ (on the y-axis), which we evaluate at the orders of expansion $n = 2, 4, 6, 8$ (on the x-axis) for every epoch $j = 1, 2, 3, 4$, which are represented by circles, squares, diamonds, and triangles respectively. Assuming the epoch break points $\bar{r}_j = \frac{\log(2)}{\lambda_{2j+1}}$, the expression on the y-axis informs us of the amount of change (or the rate of convergence) in the overall scaled bias-complemented mutation rate that is captured by each order of expansion for every epoch.

novel mutations enter at a low rate per locus per generation (Nei et al., 1975; Motro and Thomson, 1982). Both the ratio between the maximum and minimum heterozygosity within each cycle as well as the ratio between the maximum heterozygosity and the harmonic average heterozygosity within each cycle increase with the severity and duration of the population collapse (Nei et al., 1975; Motro and Thomson, 1982). We will examine both the expected sample polymorphic site frequency spectrum of populations undergoing boom-bust life cycles as well as the temporal part of their spectral decomposition.

The harmonic mean of the (effective) population size across each cycle is:

$$N_h = \frac{s_u + s_o}{\frac{s_u}{N_u} + \frac{s_o}{N_o}}.$$

Diffusion time within each epoch can be scaled relative to this harmonic mean so that the scaled duration of the epochs become $s_u \frac{N_h}{N_u}$ and $s_o \frac{N_h}{N_o}$, *i.e.*, in this subsection only, time is scaled in units of the harmonic mean effective population size. Similarly, the harmonic mean of the overall bias-complemented scaled mutation rates is defined as:

$$\theta_h^* = \mu N_h = \frac{s_u + s_o}{\frac{s_u}{\theta_{s_u,n}^*} + \frac{s_o}{\theta_{s_o,n}^*}}.$$

Then the solutions to the linear differential equations for the temporal dynamics of the boom and bust phases can be written as the following, introducing the dummy index for the even epochs $l \in 2\mathbb{N}$:

$$\tau_n(t)\Xi_n^{-1} = \theta_u^* - \left(\theta_u^* - \tau_n(t_{l-1})\Xi_n^{-1}\right) e^{-\lambda_n s_u \frac{\theta_h^*}{\theta_u^*}(t-t_{l-1})}, \quad \text{for } t_{l-1} < t \leq t_l.$$

$$\tau_n(t)\Xi_n^{-1} = \theta_o^* - \left(\theta_o^* - \tau_n(t_l)\Xi_n^{-1}\right) e^{-\lambda_n s_o \frac{\theta_h^*}{\theta_o^*}(t-t_l)}, \qquad \text{for } t_l < t \leq t_{l+1}.$$

$$(35)$$

As the number of cyclical iterations increases with $l \to \infty$, the temporal dynamics of the even and odd epochs evolve towards two
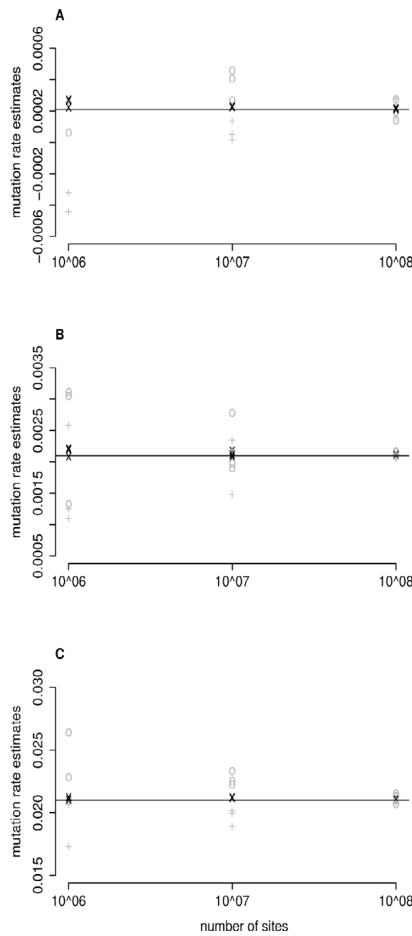
Fig. 3. The above figures assess the numerical accuracy of the proposed inference framework by attempting to infer historic overall bias-complemented scaled mutation rates from equilibrium distributions; any deviation in the estimated rates from their equilibrium values must therefore be due to computational inaccuracy. We vary the magnitude of the overall scaled mutation rate and the number of sampled loci to assess which parameter ranges yield the most accurate results. To this end, the polymorphic equilibrium distribution of the discrete boundary-mutation Moran model of size $N = 6$ is calculated (using Eq. (4)) assuming $\theta^* = 0.0021$ for panel (A), $\theta^* = 0.021$ for panel (B), and $\theta^* = 0.21$ for panel (C). Three corresponding sample spectra comprised of 10^6, 10^7, and 10^8 loci respectively are then simulated according to the allele occupancy probabilities $\bar{p}_k$ of each equilibrium distribution from (A)-(C). Then, the current as well as two past overall bias-complemented scaled mutation rates (which should be $\theta_1^* = \theta_2^* = \theta_3^* = \theta^*$) are inferred (via the above described method) from each of these three sample spectra for every equilibrium value of the overall scaled mutation rate (A)-(C). This procedure is repeated three times. For each value of the overall scaled mutation rate in (A)-(C), the estimators of the historical overall scaled mutation rates (y-axis) are represented in dependence of the number of loci (x-axis), where we use the symbols "+", "×", and "○" to denote estimates of θ_1^*, θ_2^*, and θ_3^* respectively. The horizontal lines in (A)-(C) mark the true/equilibrium values of the overall scaled bias-complemented mutation rates.

separate solutions at t_l and t_{l+1}, which we will denote as $\vartheta_{s_u,n} = \lim_{l \to \infty} \tau_n(t_l) \Xi_n^{-1}$ and $\vartheta_{s_o,n} = \lim_{l \to \infty} \tau_n(t_{l-1}) \Xi_n^{-1}$, respectively. These solutions follow the system of equations:

$$
\begin{aligned}
\vartheta_{s_u,n} &= \theta_u^* - \left(\theta_u^* - \vartheta_{s_o,n}\right) e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u} \\
\vartheta_{s_o,n} &= \theta_o^* - \left(\theta_o^* - \vartheta_{s_u,n}\right) e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o} .
\end{aligned}
\tag{36}
$$

This system is easily solved:

$$
(\theta_u^* - \theta_o^*) = -\left(\vartheta_{s_o,n} - \theta_u^*\right) e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u} + \left(\vartheta_{s_o,n} - \theta_o^*\right) e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o}
$$

$$
\vartheta_{s_o,n} = \frac{\theta_u^*(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}) + \theta_o^*(e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o} - 1)}{(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}) + (e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o} - 1)}
$$

$$
\vartheta_{s_o,n} = \frac{\theta_u^*(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}) - \theta_o^*(1 - e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o})}{(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}) - (1 - e^{\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o})} ,
\tag{37}
$$

and

$$
\vartheta_{s_u,n} = \frac{\theta_o^*(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o}) - \theta_u^*(1 - e^{\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u})}{(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_o^*} s_o}) - (1 - e^{\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u})} .
\tag{38}
$$

Thus, analytical expressions for the long-term solutions of the temporal dynamics of populations of deterministically alternating (effective) population size have been found. When the epoch lengths are relatively short, so that $s_{max} \lambda_n \ll 1$ with $s_{max} = \max(s_u \frac{\theta_u^*}{\theta_h^*}, s_o \frac{\theta_o^*}{\theta_h^*})$, the exponential decay of relative mutation effects in each epoch can be approximated by first order Taylor series expansions:

$$
\begin{aligned}
\vartheta_{s_o,n} &= \frac{\theta_u^* \lambda_n \frac{\theta_h^*}{\theta_u^*} s_u + \theta_o^* \lambda_n \frac{\theta_h^*}{\theta_o^*} s_o}{\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u + \lambda_n \frac{\theta_h^*}{\theta_o^*} s_o} + \mathcal{O}(s_{max}^2) \\
&= \frac{s_u + s_o}{\frac{s_u}{\theta_{s_u,n}^*} + \frac{s_o}{\theta_{s_o,n}^*}} + \mathcal{O}(s_{max}^2) \\
&= \theta_h^* + \mathcal{O}(s_{max}^2) ,
\end{aligned}
\tag{39}
$$

and equivalently:

$$
\vartheta_{s_u,n} = \theta_h^* + \mathcal{O}(s_{max}^2) .
\tag{40}
$$

Essentially, rapidly alternating boom and bust phases ultimately result in the temporal coefficients of both epochs converging towards the harmonic mean of the individual overall bias-complemented scaled mutation rates. The effective population size is then accurately captured by N_h, as anticipated.

Conversely, it follows immediately from Eq. (36) that for long epoch lengths $s_{min} \lambda_n \gg 1$ where $s_{min} = \min(s_u \frac{\theta_u^*}{\theta_h^*}, s_o \frac{\theta_o^*}{\theta_h^*})$, the temporal coefficients of the epochs converge to their individual equilibrium solutions:

$$
\begin{aligned}
\vartheta_{s_o,n} &= \theta_o^* \\
\vartheta_{s_u,n} &= \theta_u^* .
\end{aligned}
\tag{41}
$$

Then the arithmetic mean of $\vartheta_{s_u,n}$ and $\vartheta_{s_o,n}$ is the appropriate approximation of the expected long-term effective population size, again as anticipated.

3.3.1. Results and discussion for model 2

In Fig. 4 panels 1A-1C, the expected polymorphic sample spectra from boom and bust epochs are plotted relative to the harmonic mean for relatively long (A), intermediate (B), and short (C) epoch lengths compared to the average overall bias-complemented scaled mutation rate. Recall that this is easily achieved by plugging the long-term solution of the temporal dynamics of the boom epochs from Eq. (37) and bust epochs from Eq. (38) into Eq. (31). Technically, according to the aforementioned criteria, both scenarios A and B qualify as intermediate: In A, only the expansion order $n = 2$ can be approximated by the harmonic mean with the remainder likely well-approximated by the arithmetic mean. Meanwhile in B, only the expansion order $n = 2$ fulfils the approximation criteria of the harmonic mean but several higher order expansion are no longer accurately represented by
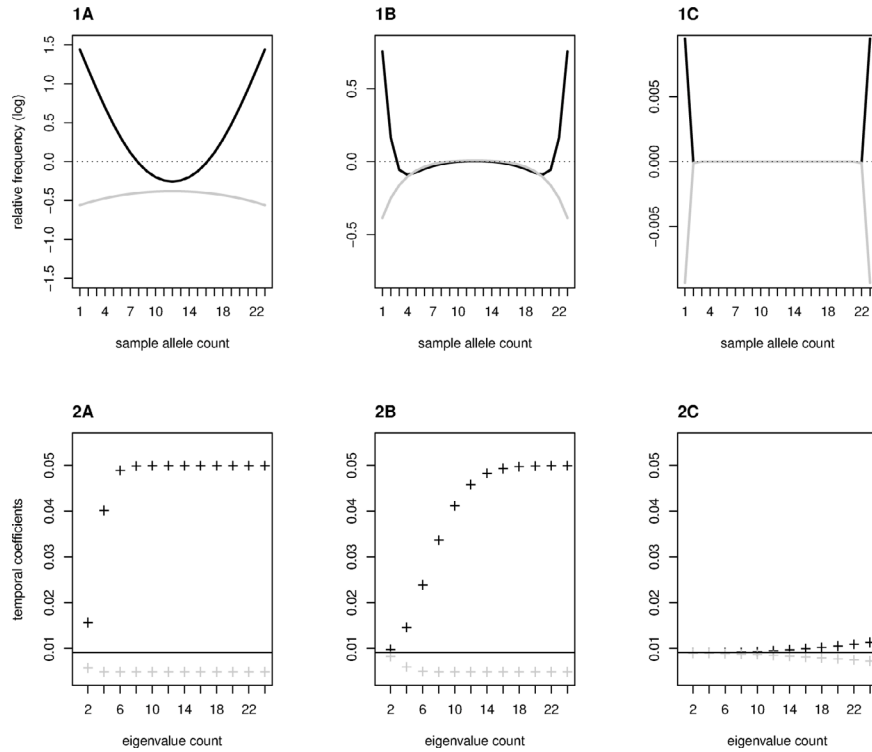
Fig. 4. Samples of size $K = 24$ drawn from a population evolving according to the caricature boom-bust model with $\theta_o^* = 0.05$ in the boom epoch and $\theta_u^* = 0.005$ in the bust epoch. In (1), the log-ratio of the expected sample site frequency spectra during the boom and bust epochs vs the equilibrium distribution at the harmonic mean of the overall bias-complemented scaled mutation rates are shown in black and grey respectively. The epoch lengths are set to $s_u = s_o = 7 \cdot 10^{-1}$ (A), $s_u = s_o = 1 \cdot 10^{-1}$ (B), and $s_u = s_o = 1 \cdot 10^{-3}$ (C). In (2), the temporal coefficients are plotted for both boom (Eq. (37); in black) and bust (Eq. (38); in grey) epochs for each order of expansion.

the arithmetic mean. In panel 1 A, the sample spectra from the boom epochs show the u-shape characteristic of population growth with an excess of low and dearth of high frequency alleles; similarly, the bust epochs show a flat, inverted u-shape that consistently lies below the harmonic equilibrium, characteristic of population contraction. Only the lowest order temporal coefficients have started converging towards the harmonic equilibrium (see panel 2 A). The result of the D-statistics reflect these patterns (see Table 1). In our intermediate scenario B, the sample spectrum of the boom epoch has a clear w-shape; the sample spectrum of the bust epoch simply shows a more pronounced inverse u-shape with the peak hitting the harmonic mean. For these epoch lengths, the lowest order temporal coefficients are almost equal to the harmonic mean, and some but not all the higher order coefficients have started converging (see panel 2B). In Nawa and Tajima (2008), visual assessment of the spectrum of segregating sites reveals w-shapes for populations recovering from a bottleneck: the number of intermediate alleles increase faster than the number of minor alleles, leading to confused signals in the D-statistic (see Nawa and Tajima (2008, Table 2) alongside our Table 1). This w-shape becomes more pronounced with bottlenecks that are further back in the population history, are comparatively short, or are less extreme in terms of collapse in population size. Bottlenecks with the opposite characteristics lead to such an extreme w-shape that it becomes an inverted u-shape. The bust epoch can therefore be considered a mild, recurrent bottleneck. In scenario C, where the epoch lengths are relatively short, all but the highest orders of expansion can be approximated by the harmonic mean. Both sample spectra deviate from the harmonic mean only by a small enrichment or diminution in the proportion of singletons respectively (see Table 1).

In Table 2, the results of applying our agnostic inference approach from Section 3.2 to sample spectra drawn from both the boom and the bust phases are presented for each of the previous epoch length

Table 1
D-statistic for sampled spectra from Fig. 4.

	boom epoch ($\theta_1^* = 0.05$)	bust epoch ($\theta_0^* = 0.005$)
(A) $s_u = s_o = 7 \cdot 10^{-1}$	−0.2140	0.01351
(B) $s_u = s_o = 10^{-1}$	0.0008	0.03521
(C) $s_u = s_o = 10^{-3}$	−0.0008	0.0008

scenarios. In order to do this, the spectra are hypergeometrically down-sampled to size $K = 4$ so that the current and one past overall-biased mutation rate can be inferred. The results gained from the sample spectra drawn from the bust phase are reasonably informative of the underlying boom-bust population cycle: The results for long epoch lengths are near the individual equilibrium overall bias-complemented epoch lengths, and convergence towards the harmonic mean is noticeable as epoch lengths decrease. The results from the bust phase are less informative: Because the collapse in (effective) population size is so rapid (with $\exp\left(-\lambda_n \frac{\log(2)}{\lambda_{2j+1}}\right) \ll \exp\left(-\lambda_n s_u \frac{\theta_h^*}{\theta_u^*} t\right)$ for all j when $n > 2$), the spectrum contains little to no historical information. So while the estimators for the current overall bias-complemented scaled mutation rates are reasonable, the past estimates (inferred from not very far back in time) are no more than an indication that the (effective) population size was formerly larger.

Generally, deterministic boom-bust type models are applicable to organisms that experience major, semi-regular environmental events on a time scale that is long compared to their generation lengths. For example, the El Nino currents in the equatorial pacific occur roughly every four years and are known to impact heterozygosity in some insect populations (Franca et al., 2020). With the annual growth and subsequent collapse of naturally occurring populations of *Drosophila* in temperate climates, however, generation times are short compared to

Table 2

Inference of temporal coefficients for sampled spectra from Fig. 4.

sample from boom epoch	$\widehat{\theta^*}_1$	$\widehat{\theta^*}_0$
(A) $s_u = s_o = 7 \cdot 10^{-1}$	0.00648	0.05151
(B) $s_u = s_o = 1 \cdot 10^{-1}$	0.00805	0.01692
(C) $s_u = s_o = 1 \cdot 10^{-3}$	0.00908	0.00918
sample from bust epoch	$\widehat{\theta^*}_0$	$\widehat{\theta^*}_1$
(A) $s_u = s_o = 7 \cdot 10^{-1}$	0.00463	0.00617
(B) $s_u = s_o = 1 \cdot 10^{-1}$	0.00503	0.00926
(C) $s_u = s_o = 1 \cdot 10^{-3}$	0.00898	0.00911

the scale of seasonal variation in the environment, making seasonal adaptation possible (Machado et al., 2021). This comparatively rapid turnover in generations implies that the difference between the (effective) sizes of boom and bust epochs will often be of several orders of magnitude, so $N_u/N_o \ll 1$. Realistic values may be $N_u/N_o \leq 10^{-4}$ with $\theta_h^* = 0.01$ and epoch lengths of roughly $\frac{10}{N_h}$. From such large ratios between $N_u : N_o$, it follows that $\frac{\theta_h^*}{\theta_o^*} \approx 0$. Note that then the long-term temporal coefficients of the boom and bust epochs (respectively Eq. (37) and Eq. (38)) can be approximated by:

$$\vartheta_{s_o,n} = \frac{\lambda_n \theta_h^* s_o + \theta_u^* \left(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}\right)}{\left(1 - e^{-\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}\right)} + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right) \quad (42)$$

and

$$\vartheta_{s_u,n} = \frac{\lambda_n \theta_h^* s_o - \theta_u^* \left(1 - e^{\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}\right)}{-\left(1 - e^{\lambda_n \frac{\theta_h^*}{\theta_u^*} s_u}\right)} + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right). \quad (43)$$

Furthermore, observe that $\theta_h^* = \theta_{s_u,n}^* \frac{s_u + s_o}{s_u} + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right)$. Substituting this, the difference between the long-term dynamics of the epochs becomes:

$$\begin{aligned}
\vartheta_{s_o,n} - \vartheta_{s_u,n} &= \frac{\lambda_n \theta_u^* \frac{s_u + s_o}{s_u} s_o + \theta_u^* \left(1 - e^{-\lambda_n(s_u + s_o)}\right)}{\left(1 - e^{-\lambda_n(s_u + s_o)}\right)} \\
&\quad - \frac{-\lambda_n \theta_u^* \frac{s_u + s_o}{s_u} s_o + \theta_u^* \left(1 - e^{\lambda_n(s_u + s_o)}\right)}{\left(1 - e^{\lambda_n(s_u + s_o)}\right)} + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right) \\
&= \lambda_n \theta_u^* \frac{s_u + s_o}{s_u} s_o \left(\frac{1}{1 - e^{-\lambda_n(s_u + s_o)}} + \frac{1}{1 - e^{\lambda_n(s_u + s_o)}}\right) + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right) \\
&= \lambda_n \theta_u^* \frac{s_u + s_o}{s_u} s_o + \mathcal{O}\left(\frac{\theta_h^*}{\theta_o^*}\right).
\end{aligned} \quad (44)$$

The difference between the long-term temporal mutation dynamics of the boom and bust epochs is therefore independent of the overall bias-complemented scaled mutation rate of the boom epoch. Rather, for each order n of the temporal expansion, the corresponding eigenvalue is scaled by a term comprised of the overall bias-complemented scaled mutation rate of the bust epoch, the cycle length, and the duration of the boom epoch relative to the bust epoch. While this effect was apparent from past simulations (Motro and Thomson, 1982), we have now shown it analytically. The accuracy of this final expression depends on the conditions ahead of Eqs. (42) and (43). If necessary, upper bounds for accuracy can be determined via the error expansion of the Taylor series. Note that one can see from these last calculations that unrealistically large sample sizes may be necessary to be able to detect

the signatures of the boom-bust dynamics in the spectra of *Drosophila*, since the number of generations per year will be very much lower than N_h.

3.4. Serially correlated population sizes (model 3)

3.4.1. Autoregression model for temporal dynamics

In Section 3.1 and Section 3.3, populations are subject to major and predictable changes in (effective) population size. However, most fluctuations in population size are responses to demographic (Lee et al., 2011) or environmental (Lande, 2014) changes that are not necessarily regular over long time periods. While these fluctuations can be crucial to the fate of the populations, they are generally not easily quantifiable. In the following, stochastic variation in the overall bias-complemented scaled mutation rate is incorporated into the system of temporal equations of the forwards population process from Eq. (28).

Let us denote the long-term solution for the temporal coefficient of $\tau_n(t)$ for each epoch $t_{j-1} < t \leq t_j$ as $\vartheta_{j,n} \Xi_n$. Furthermore, let us fix the relative scaled epoch length $c = s_j 4 N_j \alpha \beta \mu$. This relative scaled epoch length is independent of j because the epoch lengths s_j scale with $1/N_j$ on the continuum time scale and we assume that the number of generations (and thus the mutation parameters) is the same in each epoch. The resulting long-term temporal coefficients solve the following equation for $n \in 2\mathbb{N}$ and $j > 0$ (compare Eq. (28) and Eq. (36)):

$$\vartheta_{j,n} = \vartheta_{j-1,n} e^{-\lambda_n \frac{c}{\theta^*}} + \theta_j^* \left(1 - e^{-\lambda_n \frac{c}{\theta^*}}\right). \quad (45)$$

This equation has the general form $\gamma_j = \gamma_{j-1} m_j + \epsilon_j(1 - m_j)$ and is therefore a first order autoregression model. Because $|m_j| < 1$, the process is also stationary, *i.e.,* its moments are independent of j (Johnston, 1960, Chapter 7). In the classic case, ϵ_j enters the process without a coefficient and is a serially uncorrelated, zero-mean and constant variance stochastic process (*i.e.,* white noise). Here, it corresponds to θ_j^*, and we must characterise its distribution in order to make statements about $\vartheta_{j,n}$. Note that the occurrence of the overall bias-complemented scaled mutation rates θ_j^* in Eq. (45) is reminiscent of an inverse gamma distributed variable. Therefore, we assume for $j > 0$:

$$\theta_j^* \overset{iid}{\sim} invgamma(a, b) \text{ with } a > 2, b > 0$$

$$p(\theta^*) = \Pr(\theta_j^* = \theta^* | a, b) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\theta^*}\right)^{a+1} e^{\frac{-b}{\theta^*}} \quad (46)$$

The range of the hyperparameter a is restricted to ensure existence of the mean and the variance.

3.4.2. Characterising the solutions to the temporal dynamics

Next, we will use the distribution of the overall scaled mutation parameter θ_j^* together with the properties of the first-order autoregression processes to determine the mean, variances, and covariances of the solutions of the long-term temporal coefficients $\vartheta_{j,n}$.

Mean

Returning to the autoregression process in Eq. (45) and invoking stationarity, we have for $n \in 2\mathbb{N}$ and $j > 0$:

$$\mathrm{E}_\vartheta(\vartheta_{j,n}) - \mathrm{E}_\vartheta(\vartheta_{j-1,n}) = 0$$

$$\mathrm{E}_\vartheta \, \mathrm{E}_{\theta^*}(\vartheta_{j,n}) - \mathrm{E}_\vartheta \, \mathrm{E}_{\theta^*}(\vartheta_{j-1,n}) = \mathrm{E}_\vartheta \, \mathrm{E}_{\theta^*}\left((\theta^* - \vartheta_{j,n})(1 - e^{-\lambda_n \frac{c}{\theta^*}})\right) = 0 \quad (47)$$

$$\mathrm{E}_\vartheta(\vartheta_{j,n}) \, \mathrm{E}_{\theta^*}\left(1 - e^{-\lambda_n \frac{c}{\theta^*}}\right) = \mathrm{E}_{\theta^*}(\theta^*) \, \mathrm{E}_{\theta^*}\left(1 - e^{-\lambda_n \frac{c}{\theta^{*\prime}}}\right)$$

Recall that $\mathrm{E}_{\theta^*} = \int_0^\infty p(\theta^*) d\theta^*$, where $p(\theta^*)$ is the density of the inverse gamma distribution from Eq. (46). By straightforward application of the substitution and integration by parts rules for evaluation of each

integral, we obtain for $n \in 2\mathbb{N}$ and $j > 0$:

$$E_\vartheta(\vartheta_{j,n})\left(1 - \frac{b^a}{(b + \lambda_n c)^a}\right) = \frac{b}{(a-1)} - \frac{1}{(a-1)}\frac{b^a}{(b + \lambda_n c)^{a-1}}$$

$$E_\vartheta(\vartheta_{j,n})\frac{(b + \lambda_n c)^a - b^a}{(b + \lambda_n c)^a} = \frac{b}{(a-1)}\left(\frac{(b + \lambda_n c)^{a-1} - b^{a-1}}{(b + \lambda_n c)^{a-1}}\right)$$

$$E_\vartheta(\vartheta_{j,n}) = \frac{b}{(a-1)}\left(\frac{(b + \lambda_n c)^a - (b + \lambda_n c)b^{a-1}}{(b + \lambda_n c)^a - b^a}\right)$$

$$= \frac{b}{(a-1)}\left(1 - \frac{\lambda_n c b^{a-1}}{(b + \lambda_n c)^a - b^a}\right) \quad (48)$$

Using L'Hôpital's rule, the limiting results for short and long relative epoch lengths (scaled by the eigenvalues) can be determined:

$$\lim_{\lambda_n c \to 0} E_\vartheta(\vartheta_{j,n}) = \lim_{\lambda_n c \to 0} \frac{b}{a-1}\left(1 - \frac{b^{a-1}}{a(b + \lambda_n c)^{a-1}}\right) = \frac{b}{a-1}\left(1 - \frac{1}{a}\right) = \frac{b}{a}$$

$$\lim_{\lambda_n c \to \infty} E_\vartheta(\vartheta_{j,n}) = \lim_{\lambda_n c \to \infty} \frac{b}{a-1}\left(1 - \frac{b^{a-1}}{a(b + \lambda_n c)^{a-1}}\right) = \frac{b}{a-1}.$$
$$(49)$$

When epochs are relatively short, the expected value of the long-term temporal coefficients converges towards the harmonic mean of the inverse gamma distribution that was chosen to model the distribution of the overall bias-complemented scaled mutation parameter. When epochs are relatively long, convergence is towards the arithmetic mean of this inverse gamma distribution. Similar results have been obtained for the heterozygosity and the effective size of populations with stochastic cycles of varying size within the framework of the (Wright–Fisher) diffusion (Iizuka, 2001; Iizuka et al., 2002; Iizuka, 2010): There, predetermined potential population sizes can easily be modelled as finite states of a Markov Chain. Both the effective size of the population and its heterozygosity generally converge towards the harmonic means when the autocorrelation of transitions between population size states is low. In contrast, high (positive) autocorrelation generates convergence to the arithmetic mean (which is always greater than the harmonic mean), and, when the scenario permits negative autocorrelation, this can lead to results that lie below the harmonic mean.

Variance

The method for calculating the variance of the solutions to the long-term temporal coefficients $\vartheta_{j,n}$ is analogous to that for determining the mean. Starting again from Eq. (45) and using the stationarity of the process for $n \in 2\mathbb{N}$ and $j > 0$:

$$E_\vartheta(\vartheta_{j,n}^2)E_{\theta^*}\left(1 - e^{-\lambda_n \frac{c}{\theta^*}}\right) = 2E_{\theta^*}\left((1 - e^{-\frac{c}{\theta^*}})e^{-\frac{c}{\theta^*}}\theta^*\right)E_\vartheta(\vartheta_{j,n})$$

$$+ E_{\theta^*}(\theta^{*2})E_{\theta^*}\left(1 - 2e^{-\frac{c}{\theta^*}} + e^{-\lambda_n \frac{c}{\theta^*}}\right).$$
$$(50)$$

Evaluating integrals as before, one obtains for $n \in 2\mathbb{N}$ and $j > 0$:

$$E_\vartheta(\vartheta_{j,n}^2) = \frac{1}{1 - \frac{b^a}{(b+2\lambda_n c)^a}}\left(\frac{2}{(a-1)}\left(\frac{b^a}{(b + \lambda_n c)^{a-1}} - \frac{b^a}{(b + 2\lambda_n c)^{a-1}}\right)E_\vartheta(\vartheta_{j,n})\right.$$

$$+ \frac{b^a}{(a-1)(a-2)} - \frac{2b^a}{(a-1)(a-2)(b + \lambda_n c)^{a-2}}$$

$$\left. + \frac{b^a}{(a-1)(a-2)(b + 2\lambda_n c)^{a-2}}\right).$$
$$(51)$$

Clearly, the variance can then be determined by the law of total variance; its limiting values for both short and long relative scaled epoch lengths follow directly, for $n \in 2\mathbb{N}$ and $j > 0$:

$$\text{Var}_\vartheta(\vartheta_{j,n}) = E_\vartheta(\vartheta_{j,n}^2) - E_\vartheta(\vartheta_{j,n})^2$$

$$\lim_{\lambda_n c \to 0} \text{Var}_\vartheta(\vartheta_{j,n}) = 0$$
$$(52)$$
$$\lim_{\lambda_n c \to \infty} \text{Var}_\vartheta(\vartheta_{j,n}) = \frac{b^2}{(a-1)(a-2)} - \left(\frac{b}{(a-1)}\right)^2 = \frac{b^2}{(a-1)^2(a-2)}.$$

Correlations and Covariances

Note that we have implicitly assumed that the long-term coefficients $\vartheta_{j+r,n}$ are pairwise-uncorrelated for $r > 0$, since there is serial independence between the overall bias-complemented scaled mutation parameters θ_j^* of successive epochs. Using these properties together with the law of total variance, the first order covariances (*i.e.*, those between $\vartheta_{j+1,n}$ and $\vartheta_{j,n}$) are easily determined, for $n \in 2\mathbb{N}$ and $j > 0$:

$$\Phi(r = 1) = \text{Cov}_\vartheta(\vartheta_{j+1,n}, \vartheta_{j,n})$$

$$= \text{Cov}_\vartheta(\vartheta_{j,n}e^{-\lambda_n \frac{c}{\theta_j^*}} + \theta_j^*(1 - e^{-\lambda_n \frac{c}{\theta_j^*}}), \vartheta_{j,n})$$

$$= \text{Cov}_\vartheta(\vartheta_{j,n}e^{-\lambda_n \frac{c}{\theta_j^*}}, \vartheta_{j,n}) + \text{Cov}_\vartheta((1 - e^{-\lambda_n \frac{c}{\theta_j^*}})\theta_j^*, \vartheta_{j,n})$$

$$= \text{Cov}_\vartheta(\vartheta_{j,n}e^{-\lambda_n \frac{c}{\theta_j^*}}, \vartheta_{j,n}) + 0$$

$$= E_\vartheta E_{\theta^*}(e^{-\lambda_n \frac{c}{\theta^*}}\vartheta_{j,n}^2) - E_\vartheta E_{\theta^*}(e^{-\lambda_n \frac{c}{\theta^*}}\vartheta_{j,n})E_\vartheta(\vartheta_{j,n})$$

$$= \frac{b^a}{(b + \lambda_n c)^a}\text{Var}_\vartheta(\vartheta_{j,n}).$$
$$(53)$$

Note this is simply the variance times the correlation coefficient. Higher orders of covariance are similarly, for $n \in 2\mathbb{N}$ and $j > 0$:

$$\Phi(r > 1) = \text{Cov}_\vartheta(\vartheta_{j+r,n}, \vartheta_{j,n}) = \left(\frac{b^a}{(b + \lambda_n c)^a}\right)^r \text{Var}_\vartheta(\vartheta_{j,n}). \quad (54)$$

Clearly, the limits for short and long scaled relative epoch lengths are:

$$\lim_{\lambda_n c \to 0} \Phi(r > 0) = 1 \times \lim_{\lambda_n c \to 0} \text{Var}_\vartheta(\vartheta_{j,n}) = 0$$
$$(55)$$
$$\lim_{\lambda_n c \to \infty} \Phi(r > 0) = 0.$$

3.4.3. Results and discussion for model 3

In Fig. 5, the behaviour of the theoretical mean, variance, and first order covariance of the temporal population coefficients is exemplified for varying relative epoch lengths c and select expansion orders n. For low orders of expansion n, the moments of the temporal mutation coefficients are close to the lower limits (*i.e.*, the harmonic means) even for rather long relative epoch lengths c; convergence towards the lower limits is then very rapid when $c \to 0$. For the higher expansion orders n, the trajectory of the moments of the temporal coefficients shows a more gradual convergence from the higher to lower limits (*i.e.*, the arithmetic to the harmonic mean) as the relative epoch lengths decrease. Note that the analytical solutions examined here correspond almost precisely to results from simulations; snippets of these comparisons are presented in Appendix A and Appendix D.3.

Also apparent in each panel of Fig. 5 are three little 'x' marks: at these epoch lengths, as well as at one additional epoch length that exceeds the limit of the *x*-axis, we determined the expected polymorphic allele frequencies of a sample of size $K = 24$ according to procedure described in Section 2.5 and using the expected value of the temporal coefficients from Eq. (48). These are shown relative to both the arithmetic and harmonic equilibrium spectra in Fig. 6. Scenario A corresponds to relatively long epoch lengths; at this point the moments of the temporal mutation coefficients have not yet fully converged to the harmonic mean even for the expansion order $n = 2$ (see Fig. 5) and the polymorphic sample spectrum cannot be distinguished from the arithmetic equilibrium spectrum. Scenarios B and C show epoch lengths between those at which expansion orders $n = 2$ and $n = 4$ fully converge to the harmonic mean but are close to these respective required lengths (see Fig. 5); the expected polymorphic sample spectra
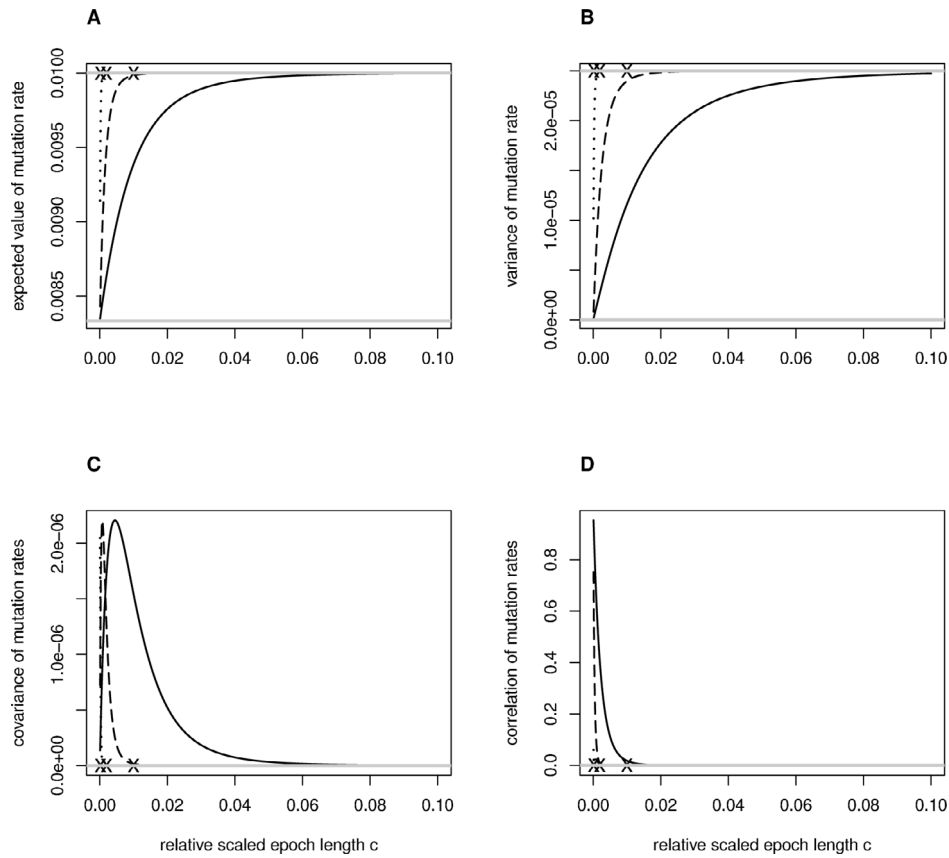
Fig. 5. Let the overall bias-complemented scaled mutation rate be distributed according to $\theta_j^* \overset{iid}{\sim} IG(a = 6, b = 0.01 \cdot (a - 1))$. The theoretical mean from Eq. (48), variance from Eq. (52), and covariance and correlation Eq. (53) dependence of the relative scaled epoch length c. In each panel, the solutions for expansion orders $n = c(2, 4, 12)$ are shown in decreasing line strength and the horizontal lines indicate the values of the statistics in the limits $\lambda_n c \to 0$ and $\lambda_n c \to \infty$ respectively. The symbols 'x' in each panel mark the epoch lengths chosen for Fig. 6(A), (B), and (C), while the epoch length from (D) is beyond the limit of the x-axis.

for both scenarios clearly lie below the arithmetic equilibrium spectrum – in fact, the intermediate frequency alleles have almost (B) or fully (C) reached harmonic mean frequency while the high/low frequency alleles exceed it. With the very short epoch length of scenario D, the moments have fully converged to the harmonic mean for almost all orders of expansion (note that the highest are not shown for the sake of visibility) and the expected polymorphic allele spectrum is almost identical to the harmonic equilibrium spectrum except for excess singletons. Note that scenario B has epoch lengths of almost the same magnitude as the average overall bias-complemented scaled mutation rate, and scenario D has epoch lengths of the same magnitude as the square of the average overall bias-complemented scaled mutation rate. In general, the variance of the expected polymorphic sample allele spectra visibly decreases with relative epoch length.

Interestingly, the D-statistic picks up a clear signature of population collapse for scenario A with the relatively long comparative epoch lengths, and equally strong signatures of population expansion for the increasingly short relative epoch lengths of scenarios B-D (see Table 3). Overall, it is apparent that compared to samples of the same size drawn from a population evolving according to a boom-bust model with the same harmonic mean of the overall mutation parameter across epochs, the samples from populations subject to stochastic changes in size generally harbour signatures of more distinct and consistent demographic change. Traces of this are still present even for very short epoch lengths. In fact, this lingering excess of singletons for very short epoch lengths is particularly interesting: Empirically, excess singletons are often observed and have recently been shown to be well-modelled by multiple-merger coalescents (Freund et al., 2023). We will address this further in our Conclusions (Section 4).

Table 3
D-statistic for the expected sample spectra from Fig. 6.

(A) $c = 10^{-1}$	(B) $c = 2 \cdot 10^{-2}$	(C) $c = 10^{-3}$	(D) $c = 10^{-4}$
0.1323629	−0.01156199	−0.01160323	−0.01160785

Table 4
Inference of temporal coefficients for sampled spectra from Fig. 6. The "–" in the naming of epochs indicates one step back in time.

	(A) $c = 1 \cdot 10^{-1}$	(B) $c = 2 \cdot 10^{-2}$	(C) $c = 1 \cdot 10^{-3}$	(D) $c = 1 \cdot 10^{-4}$
$\hat{\theta}_{s_J}^*$	0.010009	0.008603	0.008292	0.008295
$\hat{\theta}_{s_{J-}}^*$	0.009938	0.010889	0.008925	0.008681
$\hat{\theta}_{s_{J--}}^*$	0.010132	0.009646	0.009922	0.008254

After hypergeometrically downsampling all the sample spectra to $K = 6$, our agnostic inference approach from Section 3.2 can be applied to infer the current and two historical overall bias-complemented scaled mutation rates (see Table 4). Generally, these vary around the expected means per scenario.

3.5. Serially correlated population size and stochastic epoch lengths (model 4)

3.5.1. Rewriting the autoregression model

In Section 3.4, the relative epoch length $c = s_j 4\alpha\beta\mu$ is assumed constant. Next, we wish to model the relative epoch lengths as stochastic. Doing so is decently straightforward; for each epoch $j > 0$, the distribution of the bias-complemented overall scaled mutation rate θ_j^* is generalised from the inverse gamma distribution with constant
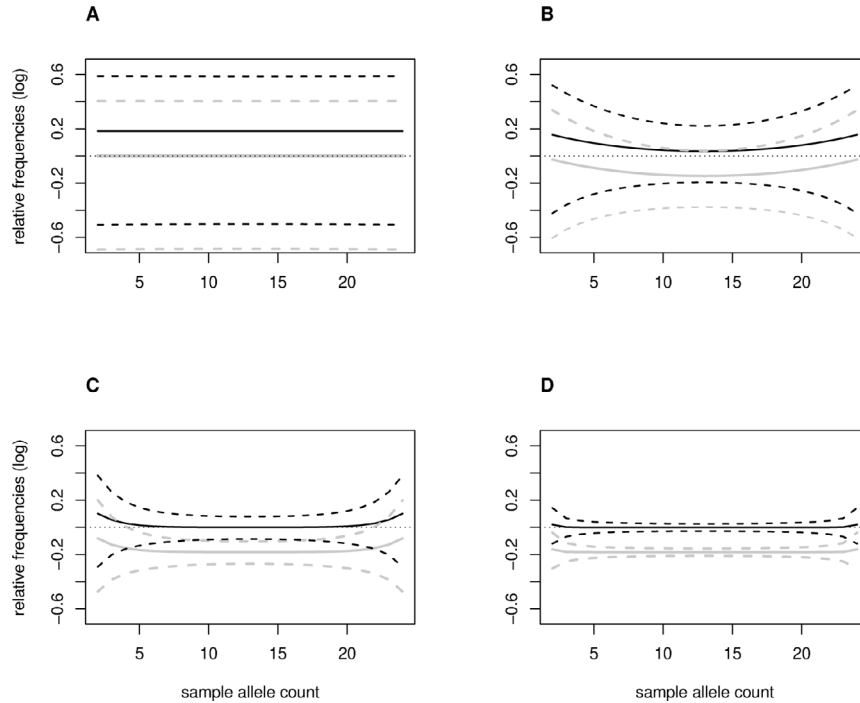
Fig. 6. Assuming $\theta_j^* \overset{iid}{\sim} IG(a=6, b=0.01 \cdot (a-1))$, $J = 1 \cdot 10^6$ epochs of the autoregression model for the long-term temporal coefficients $\vartheta_{j,n}$ in Eq. (45) were simulated for relative epoch lengths $c = 10^{-1}$ (A), $c = 2 \cdot 10^{-2}$ (B), $c = 1 \cdot 10^{-3}$ (C), and $c = 1 \cdot 10^{-4}$ (D). Expected sample allele spectra of size $K = 24$ were generated for each scenario at the end of epoch J. Visualised here are the log-ratios of the expected sample site frequency spectra vs the spectrum generated by the harmonic means of θ_j^* in black, and the log-ratio of the expected sample site frequency spectra vs the spectrum determined by the arithmetic mean of θ_j^* in grey. The dashed lines represent the spectra obtained by adding/subtracting one standard deviation of $\vartheta_{J,n}$ to the mean and calculating the respective expected sample spectra and ratios again.

parameters in Eq. (46) to the following:

$$\theta_j^* \overset{iid}{\sim} invgamma(a, bv_j) \text{ with } a > 2, bv_j > 0, \tag{56}$$

where v_j are unit mean exponential random variables:

$$v_j \overset{iid}{\sim} exp(1)$$
$$p(v) = \Pr(v_j = v | 1) = e^{-v}. \tag{57}$$

Therefore, the constant relative epoch length in the autoregression process for the solution to the long-term temporal dynamics from Eq. (45) can be replaced with the random relative epoch length cv_j; note that this means that we no longer consider the number of generations per epoch to be the same (compare to the reasoning at the beginning of Section 3.4). Further setting $\eta_j = \frac{1}{\theta_j^*}$, the autoregression process for the long-term temporal dynamics becomes:

for $n \in \{(1, \ldots, N) : 2n\}$ and $j > 0$:

$$\vartheta_{j,n} = \vartheta_{j-1,n} e^{-\lambda_n c v_j \eta_j} + \eta_j^{-1}\left(1 - e^{-\lambda_n c v_j \eta_j}\right). \tag{58}$$

Again, we wish to characterise the distribution of the driving mutation parameters in order to study the dynamics of the long-term temporal coefficients $\vartheta_{j,n}$; we therefore determine the distribution of the η_j via:

$$p(\eta) = \int_0^\infty \frac{(bv)^a}{\Gamma(a)} \eta^{a-1} e^{-(bv)\eta} e^v \, dv$$
$$= \frac{b^a}{\Gamma(a)} \eta^{a-1} \int_0^\infty v^a e^{-(b\eta+1)v} \, dv \tag{59}$$
$$= \frac{b^a}{\Gamma(a)} \eta^{a-1} \frac{\Gamma(a+1)}{(b\eta+1)^{a+1}},$$

which can be re-stated as:

$$\eta_j \overset{iid}{\sim} genbeta(a, 1, 1, 1/b), \tag{60}$$

where genbeta(.) is the generalised beta prime distribution. Note that the mean of $p(\eta)$ is naturally the inverse of the mean of $p(\theta^*)$.

Moments

As in Section 3.4, the moments of the temporal mutation parameters, which correspond to the moments of the autoregression process in Eq. (58), can readily be determined: Once again, stationarity of the process can be invoked and then the expectations can be taken with respect to first v and then η. It is easily seen from there that $E_\vartheta(\vartheta_{j,n})$ is identical to before. $E_\vartheta(\vartheta_{j,n})^2$ differs only slightly in that the final three summands of the numerator in Eq. (51) are each multiplied by factor 2 (observe that this corresponds with the terms multiplied with $E_\eta(\eta^2 p(\eta|v))$ from Eq. (59)). The variance, covariance and correlation are therefore inflated compared to the model with fixed relative epoch lengths (see also Appendix D.3).

4. Conclusions

In this article, a spectral representation of the transition density of the forward diffusion approximation of the biallelic boundary mutation Moran model is utilised to describe population allele trajectories evolving forwards in time. This model separates the dynamics of mutations, which are assumed to occur at a very low overall scaled rate and are thus restricted to the monomorphic boundaries, from those of drift, which determine the polymorphic interior. The eigenfunctions of the approximating diffusion process can be represented using a combination of (i) Gegenbauer polynomials, a class of orthogonal polynomials that describe the spatial component of polymorphic allele frequency trajectories subject to only drift, (ii) temporal coefficients that scale the polymorphic spatial component by the effect induced by the boundary mutations, and (iii) specialised boundary terms that balance the probability flux between the monomorphic boundaries and polymorphic interior, taking into account that mutations enter the polymorphic

region at a constant rate during an epoch defined by unchanging (effective) population size. The corresponding eigenvalues of the system are $\lambda_0 = 0$, which is associated with the equilibrium distribution, $\lambda_1 = \theta$, which is associated with the proportion of the focal allele present within the population, and $\lambda_n = n(n-1)$ for $n \geq 2$, which are associated with the polymorphic spectrum. Hence the spectral decomposition comprises both the monomorphic and polymorphic parts of the allele frequency spectrum (compare Song and Steinrücken, 2012; Zivkovic et al., 2015); note that classic diffusion and coalescent models only explicitly describe the polymorphic interior. Incorporating piecewise changes in (effective) population size into the forward diffusion of the boundary mutation diffusion model by proxy of changes in the overall bias-complemented scaled mutation rate requires only one full spectral decomposition to determine the spatial component of the process, which remains unchanged over time. One must further solve a system of equations for the temporal coefficients, which differ by time epoch. This is simpler and computationally advantageous to changing the basis function for the spatial component for every piecewise change in (effective) population size, as has been done for spectral representations of general mutation (Wright–Fisher) diffusion models in the past (Lukic et al., 2011; Zivkovic and Stephan, 2011; Zivkovic et al., 2015).

In this article, the exact form of expected sample site frequency spectra was determined by coupling the spectral representation of the diffusion equation with a forward pass of a forward–backward algorithm (Bergman et al., 2018b). We note here that a backward pass of the same forward–backward algorithm could have been employed with the same numerical burden; for a fuller understanding of this, we provided a brief exposition of the backward diffusion equation of the boundary mutation model and its corresponding backward eigensystem in Appendix B. For now, recall that the spectral representations of the forward and backward processes involving orthogonal polynomials are defined on the space of allelic proportions. In Appendix C), we had briefly shown that a dual process to the spectral representation of the backwards diffusion equation can be found in an embedded jump process that runs backwards in time and operates in the space of binomial distributions. From those derivations, it can be seen that the rate of the jump process is intimately related to the rate of Kingman's block counting process (see also Papaspiliopoulos and Ruggiero, 2014). As of now, it is unclear whether coalescent dual or orthogonal polynomial approaches are numerically preferable within the context of the forward–backwards algorithm. Similarly, the forwards-backwards algorithm approaches have yet to be compared to the methods based on the "matrix coalescent" for efficiency (Wooding and Rogers, 2002; Polanski and Kimmel, 2003; Bhaskar et al., 2015). Doing so would also enable an explicit comparison of our results to past literature on the effect of demography on the branch lengths of sample genealogies (see for example Eriksson et al., 2010). Overall, however, we hope to have at least established an understanding of how the approaches relate to each other as much as the scope of the current article allows.

Returning to the main part of the article: We examine the temporal coefficients of the forward pass of the spectral forward–backward algorithm as well as the exact sample allele frequency spectra for populations with (i) single deterministic shifts in (effective) population size, (ii) deterministic boom-bust life cycles, and (iii) two models of stochastic changes in (effective) population size. To the best of our knowledge, the stochastic models are novel. Throughout this article, we present conjoint assessment of the temporal population dynamics and the shape of the polymorphic sample spectra (more specifically, the shape of the log ratio of the sample spectrum vs an equilibrium distribution). Other authors have promoted visual inspection of sample spectra has over summary statistics for a clearer understanding of which regions of sample spectra deviate from equilibrium and to what extent (Nawa and Tajima, 2008; Achaz, 2009). We see the advantages of the visual approach ourselves: In our evaluation of the boom-bust model with intermediate epoch lengths in Fig. 4, the polymorphic sample spectrum

from the boom phase has a w-shape that the D-statistic interprets as a collapsing (effective) population size (see Table 1). However, this signal is actually caused by the rapid accrual of intermediate frequency alleles vs high/low frequency alleles after recurrent, intermediately spaced bust phases (*i.e.,* mild-effect bottlenecks) that fail to eradicate all the population's standing variation (compare Nawa and Tajima, 2008, Table 2). In our evaluation of sample spectra from populations with stochastic (effective) population size, we see that the D-statistic is blind to the differing effect of the time scale on which the changes occur (Tables 3 and 6). It picks up on the excess high/low frequency alleles vs intermediate as soon as the epochs become shorter and interprets this as population growth, but not on the decrease of excess intermediate alleles for shortening epoch lengths. In both cases, the figures are more informative than the summary statistic.

As informative as the polymorphic sample spectra are per se, we argue that assessing the level of convergence of the temporal coefficients of the source populations leads to a more nuanced understanding of the population dynamics. In practical terms, carrying out forward population simulations and spectral decompositions as part of experimental planning can help inform what sample sizes/orders of expansion are needed to detect departure from equilibrium or, conversely, for what sample size/orders of expansion equilibrium can safely be assumed. Even retrospectively, decomposition of estimated transition rate matrices from population samples can be informative of precisely how far which regions of the sample spectrum are from equilibrium.

Recall that the effective size of a population in time- and space-discrete (pure drift) Wright–Fisher and Moran models without mutation is often defined via the first non-unit eigenvalue λ_1 of the transition probability matrix for the respective neutral models. This is equal to $1 - \frac{1}{2N} \approx e^{-\frac{1}{2N}}$ or $1 - \frac{2}{N^2} \approx e^{-\frac{2}{N^2}}$, respectively (Ewens, 2004, p. 127). This eigenvalue determines the rate of decline in population heterozygosity (H) over time: In the Wright–Fisher model, heterozygosity at the discrete time point T is $H_T = \lambda_1^T H_0$, which can be approximated by $H_T = H_0 e^{-\frac{T}{2N}}$ in the limit of large population sizes (Ewens, 1982) (but see also Crow, 1954). In the Moran model, the equivalent approximation is $H_\iota = H_0 e^{-\frac{2\iota}{N^2}}$ per Moran step ι. In the limit of large population sizes, the genealogical tree of a sample drawn from a diploid Wright–Fisher or a haploid Moran model is given by Kingman's coalescent (Kingman, 1982) (with equilibrium coalescent times scaled by $2N$ and $\frac{N}{2}$ generations, respectively). Arguably the most prevalent definition of the effective size of a population is through the linear rescaling of the observed/census population size required for the limiting genealogical tree to still correspond to Kingman's coalescent even when there are temporal or spatial variations in population size (Nordborg and Krone, 2002). In the limit of large N, Wright–Fisher/Moran models are approximated by diffusion models; in the case of pure drift, the first non-zero eigenvalue of the corresponding transition rate densities converge towards the coalescent effective population size. If in addition to drift, mutation is also considered, the first non-zero eigenvalue is $\lambda_1 = \theta$, and thus only influenced by mutation. As long as $\theta \ll 1$, the second non-zero eigenvalue λ_2 then corresponds to λ_1 of the pure drift model.

For populations of short-scale, cyclical variations in size (Wang and Pollak, 2000a,b; Pollak, 2002) or for subdivided populations with high migration rates and large demes (Hössjer, 2015), convergence of the first eigenvalue informative of drift is towards the harmonic mean across time intervals or demes. This eigenvalue is typically λ_1 in pure drift models, but λ_2 in the McKane and Waxman (2007) pure drift model and in typical in mutation-drift models. When deterministic or stochastic changes in population size or migration rates occur on a shorter time scale than the coalescent events, the coalescent effective population size corresponds to the harmonic mean across epochs or demes (Jagers and Sagitov, 2004; Sjödin et al., 2005; Hössjer, 2011). However, when changes occur on the same time scale as

coalescent events, the limiting ancestral genealogy of the sample is a time-stochastic version of Kingman's coalescent (Kaj and Krone, 2003); and changes on longer time scales than coalescent events are negligible in terms of the sample genealogy. In the latter two cases, the coalescent effective population size is usually considered undefined (Sjödin et al., 2005) (but see Sano et al., 2004, where non-linear scaling across time is permitted so that the effective population size lies between the harmonic and arithmetic means and converges towards the arithmetic mean in these cases, respectively). Regardless of whether the effective population size is assessed via heterozygosity, leading eigenvalue, or re-scaling to Kingman's coalescent rate, only the variation in a sample of size $K = 2$ is typically considered.

In contrast, the temporal coefficients considered in this article are coefficients of all even orders $n \geq 2$ of expansion of the population transition rate density into orthogonal polynomials (as the odd ones are zero when mutation bias is constant). They are of the form $c_1 + c_2 e^{-\lambda_n f(n,N)}$, where n is the order of the eigenvalue and N the effective population size, and describe the decline of the effect of changes in overall-biased mutation rate on polymorphism over time. Clearly, the temporal coefficient of order $n = 2$ converges to its equilibrium value more slowly than the higher order coefficients, which substantiates the reasoning behind using it to estimate the effective population size. However, when past demographic changes are cyclical or fluctuating, the faster-responding higher-order temporal coefficients collect more short-term information and are perturbed relatively further from the harmonic means with every transition to a new epoch. It is therefore not uncommon for the sample allele frequency spectra to deviate from its equilibrium shape even after the temporal coefficient of order $n = 2$ has converged, as shown repeatedly in this article (specifically, for the boom-bust model see Fig. 4 and for the stochastic model see Fig. 6). The deviations from equilibrium in sample spectra drawn from populations of stochastically varying (effective) population size across shorter epoch lengths are of particular interest (*i.e.,* Fig. 6, scenarios B, C, and D): The spectra exhibit excess high and low frequency alleles for epoch lengths of the same magnitude as the average overall bias-complemented scaled mutation rate and below (scenarios B and C), until only excess low frequency alleles/singletons remain when the epochs decrease to roughly the same magnitude as the square of the average overall bias-complemented scaled mutation rate (scenario D). While epoch lengths below the same order of magnitude as the overall-biased mutation rate and especially on the order of magnitude of its square may conventionally be considered "safely small" to assume equilibrium, it is apparent in our analyses that this is not the case (see Fig. 6). Fluctuations in population size therefore strongly impact the shape of sample spectra, even when limiting results may conventionally be assumed to hold.

Deviations from neutrality are also commonly observed empirically, when sample allele frequency spectra are polarised into ancestral and derived alleles using outgroup data (Freund et al., 2023). In particular, patterns of excess high and low frequency derived alleles as well as excess singletons have recently been noted in genome-wide data of many species. These patterns are often considered more suitably modelled by multiple-merger coalescents than the classic Kingman coalescent (Freund et al., 2023). Multiple-merger coalescents reflect skewed offspring distributions across generations, whether these are caused by population substructure, demography, recurrent selection, biased gene conversion, *etc*. Note that misleading signals of multiple-merger genealogies can be caused by poor sampling coverage and erroneous allele polarisation. It is well-known that extreme discrete changes in (effective) population size, such as population growth after a bottleneck, can also generate sample spectra with excess singletons; there is some indication that inference procedures can discriminate between these and multiple-merger generated sample spectra since the entire spectrum will not exhibit the same shape between the models for a comparable level of singletons (Eldon et al., 2015). In this article, we

did not assume polarisation of alleles via an outgroup, but arbitrarily assigned one of two alleles to be the focal allele. Therefore, the shape of any of our sample spectra cannot immediately be related to this body of literature. In particular, information on high frequency derived alleles is not available with our approach. The main outcome from our article in relation to this is that one cannot discount stochastic changes in population size, especially relatively rapid ones, from being responsible for a part of the empirical observations without further investigation. In fact, studies such as that of Eldon et al. (2015) could be conducted to address this in the future, alongside more theoretical treatment of the relationship between our model and general coalescent models allowing for simultaneous- and multiple- mergers (Spence et al., 2016).

In terms of demographic inference alone, we have shown that inference of population history from sample spectra can be performed using sufficient statistics within the framework of the boundary mutation diffusion model and the spectral forward–backward algorithm. By placing hypothetical demographic events conveniently with respect to the time scales defined by the eigenvalues $\lambda_n = n(n - 1)$, we can very efficiently infer population history without specific assumptions or knowledge about past demographic events (including their timing), other than their maximum number (compare to Myers et al., 2008; Bhaskar and Song, 2014). This could be a starting point for future computational implementation of demographic inference procedures.

CRediT authorship contribution statement

Lynette Caitlin Mikula: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Claus Vogl:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Derivation of the forward partial differential diffusion equation (FPDE) for biallelic boundary mutations

In this Appendix, we derive the forward Kolmogorov forward (Fokker–Planck) diffusion equation for boundary-mutations from the transition rates of the discrete boundary mutation Moran model. We do this by first recapitulating how the general biallelic mutation diffusion can be obtained from the transition rates of the general decoupled Moran model (Etheridge and Griffiths, 2009a); this is largely reproduced from Appendix 7.1 of Bergman et al. (2018b). Then we modify this approach for the case of boundary-mutations.

A.1. General biallelic FPDE

In a discrete time and space Moran model, each reproduction step (which we index by ι) constitutes a randomly chosen individual haploid individual being replaced by the offspring of another randomly chosen individual. The expected lifetime of an individual therefore corresponds to the generation time/population size N.

Forward in time, the difference in the probability of observing the focal allele at proportion i at every genomic site per Moran step may be written as (Bergman et al., 2018b, Eq. 76):

$$\Pr(x_{\iota+1} = i) - \Pr(x_\iota = i) =$$
$$\frac{\beta\theta}{N^2}\left((N - i + 1)\Pr(x_\iota = i - 1) - (N - i)\Pr(x_\iota = i)\right)$$
$$+ \frac{\alpha\theta}{N^2}\left((i + 1)\Pr(x_\iota = i + 1) - i\Pr(x_\iota = i)\right)$$
$$+ \frac{1}{N^2}\Big((i - 1)(N - i + 1)\Pr(x_\iota = i - 1) + (i + 1)(N - i - 1)\Pr(x_\iota = i + 1)$$
$$- 2i(N - i)\Pr(x_\iota = i)\Big),$$

(61)

where the first and second line after the equality sign correspond to mutation events that increase or decrease the proportion of the focal allele respectively, and the final lines corresponds to the effect of drift.

Let us now introduce infinitesimal steps $\epsilon_t = 1/N^2$ and $\epsilon_x = 1/N$. We use this to re-scale time via $t = \iota\epsilon_t$ and space, *i.e.,* the allele proportions, via $x = i\epsilon_x$ so that $\phi(x, t)\epsilon_x\epsilon_t = \Pr(x_s = i)$. Then, the infinitesimal transition probabilities can be written, using Eq. (61), as (compare Bergman et al., 2018b, Eq. 77) (and further compare Etheridge, 2012, Definition 2.33):

$$\frac{\phi(x, t + \epsilon_t) - \phi(x, t)}{\epsilon_t} =$$
$$\frac{\beta\theta}{\epsilon_x}\left((1 - x + \epsilon_x)\phi(x - \epsilon_x, t) - (1 - x)\phi(x, t)\right)$$
$$+ \frac{\alpha\theta}{\epsilon_x}\left((x + \epsilon_x)\phi(x + \epsilon_x, t) - x\Pr(\phi(x, t))\right)$$
$$+ \frac{1}{\epsilon_x^2}\Big((x - \epsilon_x)(1 - x + \epsilon_x)\phi(x - \epsilon_x, t) + (x + \epsilon_x)(1 - x - \epsilon_x)\phi(x + \epsilon_x, t)$$
$$- 2x(1 - x)\phi(x, t)\Big)$$

(62)

Taking the limit $N \to \infty$, which implies $\epsilon_x, \epsilon_t \to 0$, and applying the standard definitions for the first and second symmetric derivatives recovers the forward partial differential diffusion equation (FPDE) (Bergman et al., 2018b, Eq. 78):

$$\frac{\partial\phi(x, t)}{\partial t} = -\frac{\partial}{\partial x}\left(\theta\beta(1 - x)\phi(x, t) - \theta\alpha x\phi(x, t)\right) + \frac{\partial^2}{\partial x^2}x(1 - x)\phi(x, t)$$
$$= -\frac{\partial}{\partial x}\theta(\beta - x)\phi(x, t) + \frac{\partial^2}{\partial x^2}x(1 - x)\phi(x, t).$$

(63)

A.2. Biallelic boundary mutation FPDE

In the boundary mutation model, we differentiate between the polymorphic interior, *i.e.,* the interval $[\epsilon_x \leq x \leq 1 - \epsilon_x]$, where we denote the infinitesimal transition rate density by $\phi_I(x, t)$, and boundary singularities, where alleles are fixed or lost. Recall that $b_0(t)$ is the proportion of the non-focal allele at time t, and $b_1(t)$ that of the focal allele. In equilibrium, b_0 and b_1 represent the stationary probability that a sample of size one corresponds to the non-focal ($b_0 = \alpha$) and focal

($b_1 = \beta$) alleles respectively. Convergence towards these equilibrium values occurs exponentially with rate θ, which is low compared to the rate of drift that governs the polymorphic interior (*i.e.,* the time sites spend in the latter region is negligible). For some starting value $b_0(t_s)$ at time $t_s < 0$, we then have $b_0(t) = \alpha - (b_0(t_s) - \alpha)e^{-\theta(t - t_s)}$ and $b_1(t) = 1 - b_0(t)$.

At the boundary $x = 0$, probability mass is lost to the interior $\phi_I(x, t)$ at a constant rate due to mutation and gained through drift from the polymorphic region adjoining the boundary at $x = \epsilon_x$. The infinitesimal transition probabilities then are:

$$\frac{\phi(x = 0, t + \epsilon_t) - \phi(x = 0, t)}{\epsilon_t} = \frac{-\beta\theta b_0(t) + 0 \cdot \alpha\theta}{\epsilon_x}$$
$$+ \frac{\epsilon_x(1 - \epsilon_x)\phi_I(\epsilon_x, t)\epsilon_x - 0}{\epsilon_x^2}$$
$$= \frac{-\beta\theta b_0(t) + \epsilon_x(1 - \epsilon_x)\phi_I(\epsilon_x, t)}{\epsilon_x}.$$

(64)

The zero in the first (mutation) term on the right hand side is due to the absence of mutations in the interior (which is the essence of the boundary mutation model), and the zero in the second (drift) term reflects the absence of drift in the mononorphic state. The other boundary $x = 1$ follows analogously:

$$\frac{\phi(x = 1, t + \epsilon_t) - \phi(x = 1, t)}{\epsilon_t} = \frac{-\alpha\theta b_1(t) + (1 - \epsilon_x)\epsilon_x\phi_I(1 - \epsilon_x, t)}{\epsilon_x}.$$

(65)

Conversely, the same probability masses are gained and lost from and to the boundaries by the polymorphic interior $\phi_I(x, t)$ at the adjacent regions $x = \epsilon_x$ and $x = 1 - \epsilon_x$. Overall, the infinitesimal transition probabilities in the interior of the boundary mutation model can be written as:

$$\frac{\phi_I(x, t + \epsilon_t) - \phi_I(x, t)}{\epsilon_t} =$$
$$\begin{cases} \frac{1}{\epsilon_x^2}\beta\theta b_0(t) + \frac{1}{\epsilon_x^2}\left(0 + (2\epsilon_x)(1 - 2\epsilon_x)\phi_I(2\epsilon_x, t) - 2\epsilon_x(1 - \epsilon_x)\phi_I(\epsilon_x, t)\right) \\ \qquad \text{for } x = \epsilon_x; \\ \frac{1}{\epsilon_x^2}\Big((x - \epsilon_x)(1 - x + \epsilon_x)\phi_I(x - \epsilon_x, t) + (x + \epsilon_x)(1 - x - \epsilon_x)\phi_I(x + \epsilon_x, t) \\ \qquad - 2x(1 - x)\phi_I(x, t)\Big) \qquad \text{for } 2\epsilon_x \leq x \leq 1 - 2\epsilon_x; \\ \frac{1}{\epsilon_x^2}\alpha\theta b_1(t) + \frac{1}{\epsilon_x^2}\Big((1 - 2\epsilon_x)(2\epsilon_x)\phi_I(1 - 2\epsilon_x, t) + 0 \\ \qquad - 2(1 - \epsilon_x)\epsilon_x\phi_I(1 - \epsilon_x, t)\Big) \qquad \text{for } x = 1 - \epsilon_x. \end{cases}$$

(66)

The zeros in the top and bottom rows above signify that drift does not act on the boundaries to move probability mass into the polymorphic region. Taking the limit $N \to \infty$ yields the forward boundary mutation diffusion equation presented in Eq. (6) of the main text, as long as $\phi_I(x, t)$ is twice differentiable and remains finite approaching the boundaries.

Remark. Note that we divide by ϵ_x in Eqs. (64) and (65) and by ϵ_x^2 in Eq. (66) due to the difference in scaling between the interior and the boundary.

Remark. If $\phi_I(x, t)$ does not remain finite approaching the boundaries, the infinitesimal transition probabilities can still be written as in Eqs. (64), (65), (66), but the forward boundary mutation diffusion equation (Eq. (6)) does not hold. In particular, substituting $\phi_I(x) = \alpha\beta\theta\frac{1}{x(1 - x)}$, which approaches infinity towards the boundaries, into Eq. (66) shows it to be the stationary solution. On the other hand, substituting $\phi_I(x)$ into the diffusion equation Eq. (6), we find that the

second derivative $\frac{\partial^2}{\partial x^2} x(1-x)\phi_I(x) = 0$ for all $x \in]0,1[$ and particularly for $x = \epsilon_x$ and $x = 1 - \epsilon_x$.

Appendix B. Spectral representation of diffusion models

B.1. Eigensystem for the pure drift diffusion equation

B.1.1. Forwards equation

Consider the familiar forward Komolgorov (Fokker–Planck) diffusion equation for pure drift (Kimura, 1955a):

$$\frac{\partial \phi_I(x,t)}{\partial t} = \frac{\partial^2}{\partial x^2} x(1-x)\phi_I(x,t). \tag{67}$$

defined in the open interval $x \in]0,1[$. An analytically tractable form for the transition rate density can be found by decomposing $\phi_I(x,t)$ into eigenfunctions constituted by orthogonal polynomials and corresponding eigenvalues. Specifically, it can be expanded into a series of (modified) Gegenbauer polynomials $U_n(x)$ with temporally dependent coefficients $\tau_n(t)$: $\phi_I(x,t) = \sum_{n=2}^{\infty} \tau_n(t) U_n(x)$ (Kimura, 1955a; Song and Steinrücken, 2012; Vogl and Bergman, 2016). Substituting this into the above differential equation induces the following system of ordinary homogeneous differential equations:

$$\frac{d}{dt}\tau_n(t)U_n(x) = \frac{\partial^2}{\partial x^2} x(1-x)\tau_n(t)U_n(x)$$

$$\frac{d}{dt}\tau_n(t)U_n(x) = -\lambda_n \tau_n(t)U_n(x) \tag{68}$$

$$\frac{d}{dt}\tau_n(t) = -\lambda_n \tau_n(t).$$

In order to solve this system, an initial condition must be defined. At time $t = t_s$ in the past, let a function $\rho(x)$ defined within the polymorphic region represent an ancestral state of the population that can also be represented as a series of (modified) Gegenbauer polynomials $U_n(x)$ with coefficients ρ_n:

$$\rho_n = \frac{1}{\Delta_n} \int_0^1 x(1-x)U_n(x)\rho(x)dx. \tag{69}$$

Following Kimura (Kimura, 1955a), $\rho(x)$ is commonly taken to be a unit probability mass concentrated at a point p (which is often taken to represent a single mutant allele in the population, $p = 1/N$ (Kimura and Ohta, 1969)) represented as $\delta(x - p)$ (*e.g.*, McKane and Waxman, 2007, Section 4). This probability mass is then expanded into (modified) Gegenbauer polynomials $\phi_I(x, t = t_s) = \sum_{n=2}^{\infty} \rho_n(p) U_n(x)$ with coefficients:

$$\begin{aligned}\rho_n(p) &= \frac{1}{\Delta_n} \int_0^1 x(1-x)U_n(x)\delta(x-p)\,dx \\ &= \frac{1}{\Delta_n} p(1-p)U_n(p).\end{aligned} \tag{70}$$

For $t_s \leq t \leq 0$, the temporal component then correspond to $\tau_n(t_s, \rho) = \rho_n(p)e^{-\lambda_n(t-t_s)}$, yielding the overall solution:

$$\phi_I(x, t_s) = \sum_{n=2}^{\infty} \rho_n(p)e^{-\lambda_n(t-t_s)}U_n(x).$$

Extending time t into the future, *i.e.*, to values greater than zero, the system eventually converges to the trivial solution $\tau_n(\infty) = 0$, from which it follows that $\int_0^1 \phi_I(x,t)\,dx$ also converges to zero as the entire probability accumulates at boundary singularities. Modelling this accumulation in a way that appropriately conserves probability requires the addition of two eigenfunctions that are linear combinations of delta functions, as in McKane and Waxman (2007, Appendix C, last paragraph) and see also Tran et al. (2013). Given the initial allele frequency p, and ensuring that the average frequency of the mutant allele must coincide with it at all times, makes the following full system of forward eigenvalues appropriate:

$$\begin{cases} \mathcal{F}_0^{(p)}(x) &= (1-p)\delta(x) + p\delta(x-1) \\ \mathcal{F}_1^{(p)}(x) &= -\delta(x) + \delta(x-1) \\ \mathcal{F}_{n\geq 2}^{(p)}(x) &= -\frac{(-1)^n}{n}\delta(x) + U_n(x) - \frac{1}{n}\delta(x-1). \end{cases} \tag{71}$$

The corresponding eigenvalues are $\lambda_0 = \lambda_1 = 0$ and $\lambda_n = n(n-1)$ for $n \geq 2$ (McKane and Waxman, 2007, Appendix C, last paragraph). Note that the same spectral representation of the forward pure drift diffusion equation has been obtained from the equivalent representation of the forward biallelic general mutation diffusion equation model via a zeroth order Taylor series expansion in θ (Bergman et al., 2018b, Eq. 56).

B.1.2. Backwards equation

Let us now consider the backward Kolmogorov diffusion equation for pure drift:

$$-\frac{\partial \psi_I(x,t)}{\partial t} = x(1-x)\frac{\partial^2}{\partial x^2}\psi_I(x,t), \tag{72}$$

with $t_s \leq t \leq 0$. The complete set of backward eigenfunctions, including appropriate boundary conditions, can be constructed to be orthogonal to the forward eigenfunctions $\mathcal{F}_n^{(p)}$ (compare McKane and Waxman, 2007, Appendix C, last paragraph):

$$\begin{cases} \mathcal{B}_0^{(p)}(x) &= 1 \\ \mathcal{B}_1^{(p)}(x) &= x - p \\ \mathcal{B}_{n\geq 2}^{(p)}(x) &= x(1-x)U_n(x), \end{cases} \tag{73}$$

but again compare to the alternative derivation in Bergman et al. (2018b, Eq. 56). Note that only $\mathcal{B}_1^{(p)}(x)$ actually depends on p.

The orthogonality relation can be explicitly stated as:

$$\int_0^1 \mathcal{B}_n^{(p)}(x)\mathcal{F}_m^{(p)}(x)\,dx = \delta_{nm}\Delta_n^{(B)} \tag{74}$$

for a nonzero $\Delta_n^{(B)} = \frac{n-1}{(2n-1)n}$ only if $m = n$.

The initial condition for the backwards equation is generally a prespecified sampling distribution at the extant time. Recall that in this article, we consider a binomial sample of haploid size K conditional on the population allele frequencies x, *i.e.*, $\Pr(k|K,x)$, drawn at time $t = 0$. Any binomial likelihood $\Pr(k|K,x)$ can be expressed as a sum of $\mathcal{B}_n^{(p)}(x)$ up to order $n = K$ with coefficients $d_n(k,K)$. Note that the probabilities of drawing monomorphic samples can only be represented because the spectral representation was intentionally augmented with the boundary terms. Therefore, the backwards diffusion equation can be written as the spectral sum:

$$\psi(x,t) = \sum_{n=0}^{K} d_n(k,K)e^{\lambda_n t}\mathcal{B}_n^{(p)}(x).$$

B.1.3. Sample probabilities

Now that explicit analytic representations of the forward and backward diffusion equations have been found, the probability of observing k focal loci in our binomial sample can be determined as the following marginal likelihood:

$$p_k = \Pr(k|K,\rho,t_s) = \int_0^1 \psi(x,t)\phi(x,t)\,dx. \tag{75}$$

This corresponds to the forward–backward algorithm introduced in Bergman et al. (2018b, Section 2.2), and can be evaluated at any time $t_s \leq t \leq 0$. Using the results from the previous subsections, particularly that the orthogonality condition simplifies computation,

we can analytically determine:

$$p_k = \Pr(k|K, \rho, t_s)$$

$$= \int_0^1 \psi(x,t)\phi(x,t)\,dx$$

$$= \int_0^1 \left(\sum_{m=0}^K d_m(k,K)e^{\lambda_n t}\mathcal{B}_m^{(p)}(x) \right)\left(\sum_{n=0}^\infty \rho_n e^{-\lambda_n(t-t_s)}\mathcal{F}_n^{(p)}(x) \right)dx \tag{76}$$

$$= \sum_{n=0}^K \int_0^1 d_n(k,K)e^{\lambda_n t}\mathcal{B}_n^{(p)}(x)\rho_n e^{-\lambda_n(t-t_s)}\mathcal{F}_n^{(p)}(x)\,dx$$

$$= \sum_{n=0}^K d_n(k,K)\rho_n(x)\Delta_n^{(B)}e^{\lambda_n t_s}.$$

Note that, in the main text, we choose to evaluate the probabilities p_k at the extant time $t = 0$. In the above equation, $\psi(x, t = 0)$ then equates to the expansion of the binomial sampling scheme $\sum_{n=0}^K d_n(k,K)\mathcal{B}_n^{(p)}(x)$, and therefore the calculation as a whole reduces to a forward pass of the forward–backward algorithm.

B.2. Eigensystem for the boundary mutation diffusion equation

B.2.1. Derivation of the forward system

Formally, the boundary mutation diffusion model resembles a pure drift diffusion model that explicitly takes mutations at the boundaries into account. In this subsection, we will discuss the modifications to the previous results required to represent the transition rate density of the boundary mutation diffusion equation as a spectral sum.

Let us now begin with the forward diffusion equation of the polymorphic interior, which we here rewrite from Eq. (6) to the more convenient form:

$$\frac{\partial \phi_I(x,t)}{\partial t} = \lim_{\epsilon_x \to 0}\left(\frac{\beta\theta b_0(t)}{\epsilon_x}\delta(x-\epsilon_x) + \frac{\alpha\theta b_1(t)}{\epsilon_x}\delta(x-1+\epsilon_x) \right) + \frac{\partial^2}{\partial x^2}x(1-x)\phi_I(x,t). \tag{77}$$

The main novelty here compared to the pure drift model is that there are probability masses coming in from the boundaries due to the input of biased mutation: Importantly, $b_0(t)$ and $b_1(t)$ have an equivalent representation in terms of the eigenfunctions $\mathcal{F}_0^{(\beta)}$ and $\mathcal{F}_1^{(\beta)}$, where the mutation biases β and α determine the equilibrium proportion of alleles of the focal and non-focal type in the population, and thus the probability mass concentrated at each boundary in a stationary system. These eigenfunctions are adapted from Eq. (71) to account for the fact that the average proportion of focal and non-focal alleles may change over time (which p in the pure drift system does not). By comparing to the beginning of Appendix A.2), we state that if the system starts at time $t_s < 0$ in the past, convergence towards the equilibrium boundary values must follow according to:

$$b_1(t)\delta(x) + b_1(t)\delta(x-1) = \mathcal{F}_0^{(\beta)}(x) + (b_1(t_s) - \beta)e^{-\lambda_1(t-t_s)}\mathcal{F}_1^{(\beta)}(x), \tag{78}$$

where the eigenvalue λ_1 is equal to the scaled mutation rate θ.

Furthermore, the Dirac delta function $\delta(x - \epsilon_x)$ in Eq. (77), which defines the point of entry of a mutation into the polymorphic region, can be expressed as the polynomial expansion:

$$\lim_{\epsilon_x \to 0}\frac{\delta(x-\epsilon_x)}{\epsilon_x} = \sum_{n=2}^\infty c_n U_n(x) \tag{79}$$

with

$$c_n = \lim_{\epsilon_x \to 0}\frac{1}{\Delta_n}\int_{\epsilon_x}^{1-\epsilon_x} x(1-x)U_n(x)\frac{\delta(x-\epsilon_x)}{\epsilon_x}\,dx$$

$$= \lim_{\epsilon_x \to 0}\frac{1}{\Delta_n^{(B)}}\epsilon_x(1-\epsilon_x)U_n(\epsilon_x)/\epsilon_x \tag{80}$$

$$= \frac{U_n(0)}{\Delta_n^{(B)}} = -(-1)^n(2n-1)n.$$

Analogous operations can of course be performed on $\delta(x - 1 + \epsilon_x)$ (Vogl and Bergman, 2016, Eq. 44); these are simplified by noting that $U_n(0) = (-1)^n U_n(1)$.

Observing the above, we can rewrite the diffusion equation for the polymorphic interior from Eq. (77); not that the last step below substitutes in $\phi_I(x,t) = \sum_{n=2}^\infty \tau_n(t)U_n(x)$:

$$\frac{\partial \phi_I(x,t)}{\partial t} = \lim_{\epsilon_x \to 0}\left(\frac{\beta\theta b_0(t)}{\epsilon_x}\delta(x-\epsilon_x) + \frac{\alpha\theta b_1(t)}{\epsilon_x}\delta(x-1+\epsilon_x) \right)$$
$$+ \frac{\partial^2}{\partial x^2}x(1-x)\phi_I(x,t)$$

$$\frac{\partial \phi_I(x,t)}{\partial t} = \sum_{n=2}^\infty U_n(x)\left(-\beta\theta b_0(t)\frac{U_n(0)}{\lambda_n \Delta_n^{(B)}} - \alpha\theta b_1(t)\frac{U_n(1)}{\lambda_n \Delta_n^{(B)}} \right)$$
$$+ \frac{\partial^2}{\partial x^2}x(1-x)\phi_I(x,t)$$

$$\frac{\partial}{\partial t}\sum_{n=2}^\infty \tau_n(t)U_n(x) = \sum_{n=2}^\infty -\lambda_n U_n(x)\left(-\beta\theta b_0(t)\frac{U_n(0)}{\lambda_n \Delta_n^{(B)}} - \alpha\theta b_1(t)\frac{U_n(1)}{\lambda_n \Delta_n^{(B)}} + \tau_n(t) \right). \tag{81}$$

This induces a system of equations for the temporal coefficients:

$$\frac{d}{dt}\tau_n(t) = -\lambda_n\left(-\beta\theta b_0(t)\frac{U_n(0)}{\lambda_n \Delta_n^{(B)}} - \alpha\theta b_1(t)\frac{U_n(1)}{\lambda_n \Delta_n^{(B)}} + \tau_n(t) \right). \tag{82}$$

The $b_0(t)$ and $b_1(t)$ here can be expressed using $\tau_0(t)$ and $\tau_1(t)$; to this end, define:

$$E_n = \frac{U_n(0) + U_n(1)}{\Delta_n^{(B)}} = -(n-1)\frac{(-1)^n + 1}{\Delta_n^{(B)}} = -(2n-1)n((-1)^n + 1)$$

$$O_n = \frac{U_n(0)\beta + U_n(1)\alpha}{\Delta_n^{(B)}} = -(n-1)\frac{(-1)^n\beta - \alpha}{\Delta_n^{(B)}} = -(2n-1)n((-1)^n\beta - \alpha). \tag{83}$$

Then, we obtain:

$$\frac{d}{dt}\tau_n(t)\mathcal{F}_n^{(\beta)}(x) = -\lambda_n\left(-\alpha\beta\theta\frac{E_n}{\lambda_n}\tau_0(t) - \theta\frac{O_n}{\lambda_n}\tau_1(t) + \tau_n(t) \right)\mathcal{F}_n^{(\beta)}(x). \tag{84}$$

This system of forward temporal differential equations for the polymorphic interior is complemented by the following equations of the same form for the boundaries (compare Eq. (78)):

$$\frac{d}{dt}\tau_0(t)\mathcal{F}_0^{(\beta)}(x) = 0$$
$$\frac{d}{dt}\tau_1(t)\mathcal{F}_1^{(\beta)}(x) = -\theta\tau_1(t)\mathcal{F}_1^{(\beta)}(x) \tag{85}$$

B.2.2. Diagonalisation of the forward system

The system of temporal equations we have just derived can also be diagonalised. This will be demonstrated below, in part in order to rectify an erroneous equation in Bergman et al. (2018a) and in part as a preparation for the derivations in Appendix C, which require a diagonalised eigensystem.

Let us proceed by setting $\tau_0(t) = 1$ and observing that $\tau_1(t) = \rho_1 e^{-\theta(t-t_s)}$, where $\rho_1 = b_1(t_s) - \beta$ (see Eq. (78)). Then the solution to the inhomogeneous differential equation from Eq. (84) can be written as:

$$\tau_n(t) = \alpha\beta\theta\frac{E_n}{\lambda_n} + \theta\frac{O_n}{\lambda_n - \theta}\rho_1 e^{-\theta(t-t_s)} + \rho_n^{(d)}e^{-\lambda_n(t-t_s)}, \tag{86}$$

where $\rho_n^{(d)}$ is set to

$$\rho_n^{(d)} = \rho_n - \alpha\beta\theta\frac{E_n}{\lambda_n} - \theta\frac{O_n}{\lambda_n - \theta}\rho_1$$

to satisfy the starting the condition $\tau_n(t_s) = \rho_n$.

Recall that $\phi(x,t) = \sum_{n=2}^\infty \tau_n(t)\mathcal{F}_n^{(\beta)}$. The eigenfunctions $\mathcal{F}_n^{(\beta)}(x)$ in $\phi(x,t)$ associated with $\tau_0(t) = 1$, $\tau_1(t) = \rho_1 e^{-\theta t}$ and each $\rho_n e^{-\lambda_n t}$ can therefore be collected to obtain a diagonalised system of forward eigenfunctions, which induces the homogeneous system of temporal

differential equations:

$$
\begin{cases}
\mathcal{F}_0^{(\beta,\theta)}(x) &= \mathcal{F}_0^{(\beta)}(x) + \alpha\beta\theta \sum_{n=2}^{\infty} \frac{E_n}{\lambda_n} \mathcal{F}_n^{(\beta)}(x) \\
\mathcal{F}_1^{(\beta,\theta)}(x) &= \mathcal{F}_1^{(\beta)}(x) + \theta \sum_{n=2}^{\infty} \frac{O_n}{\lambda_n-\theta} \mathcal{F}_n^{(\beta)}(x) \\
\mathcal{F}_{n\geq2}^{(\beta,\theta)}(x) &= \mathcal{F}_{n\geq2}^{(\beta)}(x).
\end{cases}
\tag{87}
$$

This corrects an error in the denominator of the summation in the $\mathcal{F}_1^{(\beta,\theta)}$ given by Bergman et al. (2018b, Eqs. 63-65). Note that the eigenvalues of the diagonalised system are the same as for the original temporal system, *i.e.,* correspond to $\lambda_0 = 0$, $\lambda_1 = \theta$, and $\lambda_n = n(n-1)$ for $n \geq 2$ (Vogl, 2014a; Bergman et al., 2018b).

This diagonalised system for the eigenfunctions can easily be written in matrix form: Defining

$$
\mathbf{A} = \begin{pmatrix}
1 & 0 & \alpha\beta\theta\frac{E_2}{\lambda_2} & \alpha\beta\theta\frac{E_3}{\lambda_3} & \alpha\beta\theta\frac{E_4}{\lambda_4} & \cdots \\
0 & 1 & \theta\frac{O_2}{\lambda_2-\theta} & \theta\frac{O_3}{\lambda_3-\theta} & \theta\frac{O_4}{\lambda_4-\theta} & \cdots \\
0 & 0 & 1 & 0 & 0 & \cdots \\
0 & 0 & 0 & 1 & 0 & \cdots \\
0 & 0 & 0 & 0 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},
\tag{88}
$$

and the column vectors

$$
\boldsymbol{\mathcal{F}}^{(\beta)\prime} = (\mathcal{F}_0^{(\beta)}(x), \mathcal{F}_1^{(\beta)}(x), \mathcal{F}_2^{(\beta)}(x), \mathcal{F}_3^{(\beta)}(x), \dots)'
$$

as well as an analogous column vector for the diagonalised eigenfunctions $\boldsymbol{\mathcal{F}}^{(\beta,\theta)\prime}$, the homogeneous system is given by:

$$
\boldsymbol{\mathcal{F}}^{(\beta,\theta)\prime} = \mathbf{A}\boldsymbol{\mathcal{F}}^{(\beta)\prime}.
\tag{89}
$$

Since the inverse of $\mathbf{A}$ is quite simply:

$$
\mathbf{A}^{-1} = \begin{pmatrix}
1 & 0 & -\alpha\beta\theta\frac{E_2}{\lambda_2} & -\alpha\beta\theta\frac{E_3}{\lambda_3} & -\alpha\beta\theta\frac{E_4}{\lambda_4} & \cdots \\
0 & 1 & -\theta\frac{O_2}{\lambda_2-\theta} & -\theta\frac{O_3}{\lambda_3-\theta} & -\theta\frac{O_4}{\lambda_4-\theta} & \cdots \\
0 & 0 & 1 & 0 & 0 & \cdots \\
0 & 0 & 0 & 1 & 0 & \cdots \\
0 & 0 & 0 & 0 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},
\tag{90}
$$

the diagonalising transformation can be reversed easily.

The system of inhomogeneous temporal equations for the temporal coefficients alone (*i.e.,* Eq. (86)) can also be expressed in matrix form. To this end, define the row vector of inhomogeneous temporal coefficients

$$
\boldsymbol{\tau} = (\tau_0(t), \tau_1(t), \tau_2(t), \dots)
$$

and the equivalent row vector of the temporal coefficients of the homogeneous system

$$
\boldsymbol{\tau}^{(d)} = (1, \rho_1^{(d)} e^{-\theta(t-t_s)}, \rho_2^{(d)} e^{-2(t-t_s)}, \dots).
$$

The system of equations from Eq. (86) can then be stated compactly as:

$$
\boldsymbol{\tau} = \boldsymbol{\tau}^{(d)} \mathbf{A};
\tag{91}
$$

note that the matrix transformation can once again be reversed easily.

B.2.3. Derivation of the backwards system

A backwards system of diagonalised eigenfunctions for the boundary mutation model, denoted as $\mathcal{B}_n^{(\beta,\theta)}(x)$ for $n \geq 0$, should fulfil the following orthogonality relation for any $n, m > 0$:

$$
\int_0^1 \mathcal{B}_n^{(\beta,\theta)}(x) \mathcal{F}_m^{(\beta,\theta)}(x)\, dx = \delta_{nm}\Delta_n^{(B)},
$$

with $\Delta_0^{(B)} = \Delta_1^{(B)} = 1$ and $\Delta_n^{(B)}$ as before. Such a system can be achieved by performing the following diagonalising transformation concurrently

to the diagonalisation of the forward eigensystem (Bergman et al., 2018b, Eq. 68):

$$
\begin{cases}
\mathcal{B}_0^{(\beta,\theta)}(x) &= \mathcal{B}_0^{(\beta)}(x) \\
\mathcal{B}_1^{(\beta,\theta)}(x) &= \mathcal{B}_1^{(\beta)}(x) \\
\mathcal{B}_{n\geq2}^{(\beta,\theta)}(x) &= \mathcal{B}_{n\geq2}^{(\beta)}(x) - \alpha\beta\theta\frac{E_n\Delta_n^{(B)}}{\lambda_n}\mathcal{B}_0^{(\beta)}(x) - \theta\frac{O_n\Delta_n^{(B)}}{\lambda_n-\theta}\mathcal{B}_1^{(\beta)}(x).
\end{cases}
\tag{92}
$$

The above equations correct an error in the $\mathcal{B}_{n\geq2}^{(\beta,\theta)}(x)$ given in Bergman et al. (2018b, Eq. 68); aside from this, the orthogonality relationship can be checked precisely as demonstrated in Bergman et al. (2018b, Eq. 69).

The system of diagonalised backwards eigenfunctions can also be written in matrix form. Specifically, defining the matrix

$$
\mathbf{B} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \cdots \\
0 & 1 & 0 & 0 & 0 & \cdots \\
-\alpha\beta\theta\frac{E_2\Delta_2^{(B)}}{\lambda_2} & -\theta\frac{O_2\Delta_2^{(B)}}{\lambda_2-\theta} & 1 & 0 & 0 & \cdots \\
-\alpha\beta\theta\frac{E_3\Delta_3^{(B)}}{\lambda_3} & -\theta\frac{O_3\Delta_3^{(B)}}{\lambda_3-\theta} & 0 & 1 & 0 & \cdots \\
-\alpha\beta\theta\frac{E_4\Delta_4^{(B)}}{\lambda_4} & -\theta\frac{O_4\Delta_4^{(B)}}{\lambda_4-\theta} & 0 & 0 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},
\tag{93}
$$

the column vector

$$
\boldsymbol{\mathcal{B}}^{(\beta)\prime} = (\mathcal{B}_0^{(\beta)}(x), \mathcal{B}_1^{(\beta)}(x), \mathcal{B}_2^{(\beta)}(x), \dots)',
$$

and an analogous column vector $\boldsymbol{\mathcal{B}}^{(\beta,\theta)\prime}$, we obtain:

$$
\boldsymbol{\mathcal{B}}^{(\beta,\theta)\prime} = \mathbf{B}\boldsymbol{\mathcal{B}}^{(\beta)\prime}.
\tag{94}
$$

The matrix $\mathbf{B}$ can be inverted similarly to the matrix $\mathbf{A}$ above, and therefore the diagonalising transformation can again be reversed as easily.

Expressions for the temporal coefficients can also be found via the backwards system. Let us expressly define the row vector

$$
\bar{\boldsymbol{\tau}}^{(d)} = (1, d_1(k, K)e^{-\theta t}, d_2(k, K)e^{-\theta(t-t_s)}, \dots).
$$

The system of equations from Eq. (86) can then be stated compactly as:

$$
\bar{\boldsymbol{\tau}} = \bar{\boldsymbol{\tau}}^{(d)} \mathbf{B}.
\tag{95}
$$

We can use these relationships to write the solution of the inhomogeneous temporal equations as:

$$
-\frac{\partial}{\partial t}\bar{\tau}_0(t)\mathcal{B}_0^{(\beta)}(x) = \alpha\beta\theta \sum_{n=2}^{\infty} E_n\Delta_n\tau_n(t)\mathcal{B}_0^{(\beta)}(x)
$$

$$
-\frac{\partial}{\partial t}\bar{\tau}_1(t)\mathcal{B}_1^{(\beta)}(x) = -\theta\bar{\tau}_1(t)\mathcal{B}_1^{(\beta)}(x) + \theta\sum_{n=2}^{\infty} O_n\Delta_n\bar{\tau}_n(t)\mathcal{B}_1^{(\beta)}(x)
$$

$$
-\frac{\partial}{\partial t}\bar{\tau}_n(t) = -\lambda_n\bar{\tau}_n(t)
$$

<div align="right">(96)</div>

Thus a homogeneous temporal system can be obtained for $n \geq 2$ from the last line above: Consider $t_s \leq t \leq 0$, which is solved by $\tau_{2\leq n\leq K}(x) = d_n(k, K)e^{\lambda_n t}$ if one starts with a binomial sampling scheme at $t = 0$. Substituting these temporal functions into the differential equations for $\bar{\tau}_0(t)$ and $\bar{\tau}_1(t)$ results in inhomogeneous equations analogous to those running forward in time. In particular, we have:

$$
-\frac{\partial}{\partial t}\bar{\tau}_0(t)\mathcal{B}_0^{(\beta)} = \alpha\beta\theta \sum_{n=2}^{K} d_n(k, K)E_n\Delta_n e^{\lambda_n t}\mathcal{B}_0^{(\beta)},
\tag{97}
$$

for which

$$
\bar{\tau}_0(t) = d_0(k, K) + \alpha\beta\theta \sum_{n=2}^{K} \frac{E_n\Delta_n}{\lambda_n}d_n(k, K)(1 - e^{\lambda_n t})
\tag{98}
$$

is a solution that satisfies the starting condition $\bar{\tau}_0(0) = d_0(k, K)$.

Proceeding similarly for $\tau_1(t)$, we have:

$$-\frac{\partial}{\partial t}\bar{\tau}_1(t)\mathcal{B}_1^{(\beta)} = -\theta\left(\bar{\tau}_1(t) - \sum_{n=2}^{\infty} O_n \Delta_n d_n(k,K)e^{\lambda_n t}\right)\mathcal{B}_1^{(\beta)}, \tag{99}$$

for which the solution satisfying the starting condition is:

$$\bar{\tau}_1(t) = d_1(k,K)e^{\theta t} + \theta\sum_{n=2}^{K} d_n(k,K)\frac{O_n \Delta_n}{\lambda_n - \theta}(e^{\theta t} - e^{\lambda_n t}). \tag{100}$$

This can be repeated again for higher order temporal coefficients.

Backward Partial Differential Equation

The results presented above can be used to obtain a system of backwards PDEs corresponding to the forward PDEs of Eq. (81). Let us first reflect on the dynamics of mutations in the boundary mutation diffusion model and how these are analytically represented in the spectral representation of the transition rate densities: Forwards in time, the mutation terms appearing in the first line of Eq. (81) represent the mutational flow from the monomorphic boundaries into the polymorphic interior. The delta functions are then expanded into the orthogonal polynomials (compare Eqs. (79) and (80)) and appear in the diffusion operator, *i.e.,* the polymorphic region. Backwards in time, the mutation terms in the PDE will have to represent the conditioning on the origin of the polymorphism-causing mutation, *i.e.,* either the non-focal or the focal allele, which will be represented as the backwards eigenfunctions of order $n = 0$ and $n = 1$. In particular, the mutation term for the non-focal allele is a product of the mutation rate, $\beta\theta$, the backward function at ϵ_x, $\psi_I(x = \epsilon_x, t)$, and the scaling constant, $1/\epsilon_x$. Taking the limit, we obtain (compare Eqs. (79) and (80)):

$$\lim_{\epsilon_x \to 0} \beta\theta \cdot \psi_I(x = \epsilon_x, t) \cdot 1/\epsilon_x = \lim_{\epsilon_x \to 0}\beta\theta\sum_{n=2}^{\infty}U_n(\epsilon_x)\epsilon_x(1-\epsilon_x)/\epsilon_x$$

$$= \beta\theta\sum_{n=2}^{\infty}U_n(0). \tag{101}$$

The analogous operation can be performed for the focal allele. We must now consider only the spatial dynamics of the system represented by the eigenfunctions. Recall that analogously to the transformation in Eq. (78), $1 - x$ and x can be represented by the eigenfunctions $\mathcal{B}_0^{(\beta)} = 1$ and $\mathcal{B}_0^{(\beta)} = -\beta + x = -\beta(1-x) + \alpha x$, so that we can overall express the boundary dynamics as:

$$\beta U_n(0)(1-x) + \alpha U_n(1)x = \alpha\beta E_n\Delta_n\mathcal{B}_0^{(\beta)}(x) + O_n\Delta_n\mathcal{B}_1^{(\beta)}(x),$$

with E_n and O_n as in Eq. (83). Note that for $K = 1$, only the first two eigenfunctions need to be considered, yielding a homogeneous temporal system with rates $\lambda_0 = 0$ and $\lambda_1 = 1$; for $K \geq 2$, the system is inhomogeneous. Conveniently, for polynomials of order $n \geq 2$ the operator is identical to the pure drift operator. Dynamics are represented via the eigenfunctions $\mathcal{B}_{n\geq 2}(x)$ resulting in a homogeneous system for $n \geq 2$ (see Eq. (96)).

Remark. Note that in the backward system immediately above, the conditioning on x is taken from $t = 0$ to earlier times t. This means that although we consider time to be moving backwards since it is the backwards system, the direction of conditioning over time is the same as in the forward system. Thus in Eq. (101), the scaling constant of the mutational terms enters as $1/\epsilon_x$, *i.e.,* identical in the forward and backward systems, rather than as the inverse. (Compare the explanation for the direction of the arrows in Fig. 8.)

Remark. Our diagonalising transformations are similar in spirit to that in Song and Steinrücken (2012); however while these authors diagonalise the spatial component of their spectral decompositions via a linear transformation, we diagonalise the temporal component via a linear transformation.

B.2.4. Simplifying assumptions and convergence

In the main part of this article, we consider modifications of the forwards system of inhomogeneous differential equations for the temporal coefficients from Eq. (84). In this part, however, we assume that the system has reached an evolutionary equilibrium with respect to the mutation bias at time $t = \infty$, which means that $\tau_1(\infty) = 0$, and $b_0(\infty) = \alpha$ and $b_1(\infty) = \beta$.

Unlike the pure drift system, this system has a non-trivial stationary solution (Vogl, 2014a, Appendix A2):

$$\tau_n(\infty) = \alpha\beta\theta\frac{E_n}{\lambda_n} = -\alpha\beta\theta\frac{(2n-1)n((-1)^n + 1)}{\lambda_n}, \tag{102}$$

whereby one can differentiate between $\tau_n(\infty) = -\alpha\beta\theta\frac{4n-2}{n-1}$ for $n \in 2\mathbb{N}$, and $\tau_n(\infty) = 0$ otherwise. The full forwards diffusion equation therefore has a limiting stationary distribution at $t = \infty$ that can be represented as:

$$\phi(x,\infty) = \sum_{n=0}^{\infty}\tau_n(\infty)\mathcal{F}_n^{(\beta)}(x). \tag{103}$$

Importantly, this limiting distribution has singularities at the boundaries. We still call it a distribution since it can easily be seen from Eq. (71) that $\phi(x,\infty)$ integrates to one over the closed interval $[0,1]$: Specifically, only $\mathcal{F}_0^{(\beta)}(x)$ integrates to one while all other $\mathcal{F}_n^{(\beta)}$ integrate to zero.

The stationary event probabilities $\bar{p}_k$ of observing exactly k focal alleles in a sample of size K can be obtained by expanding the binomial sampling scheme into the backward boundary polynomials given in Eq. (24), multiplying the result with $\phi(x,\infty)$, and then integrating:

$$\bar{p}_k = \Pr(k \mid K, \beta, \theta) = \int_0^1 \Pr(k|x,K)\phi(x,\infty)\,dx$$

$$= \sum_{n=0}^{K}d_n(k,K)\mathcal{B}_n^{(\beta)}(x)\tau_n(\infty)\mathcal{F}_n^{(\beta)}(x)$$

$$= \begin{cases} \alpha(1 - \beta\theta H_{K-1}) & k = 0 \\ \alpha\beta\theta\dfrac{K}{k(K-k)} & 1 \leq k \leq K-1 \\ \beta(1 - \alpha\theta H_{K-1}) & k = K \end{cases} \tag{104}$$

$$= bound bin(k \mid K, \beta, \theta).$$

This is precisely the boundary-binomial distribution from Eq. (5), which describes a sample of size K from the stationary distribution of the discrete boundary mutation Moran model with $K \leq N$, where N is the population size (Vogl and Bergman, 2015, Eq. 13). Note that while the limiting stationary distribution $\phi(x,\infty)$ has singularities at the boundaries and therefore does not conform to a proper density, the equilibrium distribution of a sample of size K from this limiting distribution conforms to a regular probability distribution as long as the sample size is lower than $K_{max} \approx e^{min(\alpha,\beta)/(\alpha\beta\theta)}$ (since the boundary values otherwise become negative). This restriction on the sample size follows immediately from the fact that the stationary boundary mutation Moran model corresponds to the first order expansion in the overall mutation rate θ of the beta-binomial sampling distribution of the general mutation drift Moran model from Eq. (4).

Overall, recovering the stationary distribution of the discrete boundary mutation Moran model as the limiting stationary sampling distribution here validates our representation of the boundary mutation diffusion model as the limiting $\theta \to 0$ approximation of the general mutation diffusion model: Considering this limit in a diffusion setting typically eradicates mutation and leads to a pure drift model on only the polymorphic allele frequency space, but in the boundary mutation

diffusion model we then re-instate mutation via constructed boundary terms at the monomorphic states (recall derivations in Appendix B.2 and Vogl and Bergman (2016), as well as the final remark in Section 2.4.2). While this appears indirect, we can thereby see that this approach yields consistent results. Most importantly, we have therefore also shown that the discrete sampling distribution from the boundary mutation diffusion is not only a proper distribution but also a regular density for even moderately large sample sizes as long as the appropriate assumptions on the upper limits of the mutation rate and sample sizes combined are met. Therefore, it is easily tractable as a starting point for development of inferential procedures in suitable settings.

Remark. While it is beyond the scope of this paper to investigate this matter in detail, the polynomial expansion of the binomial sampling scheme acts as a test function in Eq. (104), against which the limiting stationary distribution integrates to the equilibrium sampling distribution. Although the limiting stationary distribution itself diverges near the boundaries, making the full integral over $\phi(x, \infty)$ on $[0, 1]$ improper, it appears that this integral can still be assigned a value: Partitioning the integral into interior and boundary regions in the examination of the transition rates in Eq. (66) and Eqs. (64) and (65), it can be seen that the 'breakpoints' between the interior and the boundaries can be approached smoothly from either side and that the terms that diverge for $N \to \infty$ near either edge of the interior as well as those that diverge at the boundaries themselves balance precisely. In such cases, analytical considerations can be applied that lead to the divergent terms cancelling out, thus assigning the improper integral a finite value Harris and Stocker (1998, p.552, see the Cauchy principal value). The derivation of the stationary sampling distribution via the orthogonal polynomials in Eq. (104), however, circumvents such analytical difficulties so that we need not expand on them further here.

Remark. Burden and Tang (2017, Eq. 35) provide the analogous discrete distribution to Eqs. (5) or equivalently Eq. (104) for the multiallelic boundary mutation model, however they only show that the equality holds up to order $1/N$ (see their Eq. 41).

Remark. Substituting $\mathcal{F}_0(x)g(x, t)$, *i.e.,* the limiting stationary distribution times a test function, into the interior forward PDE (see Eq. (77)) results in the pure drift backward PDE (up to a change in the direction of time) (Bergman et al., 2018b, Section 3.2), with the exception of the mutation terms, which must be added to first two eigenfunctions in the backward system, as explained in previous sections.

Appendix C. Embedded dual jump processes

In this section, we aim to provide some understanding of the dual jump processes embedded in the general biallelic reversible mutation and the biallelic boundary mutation diffusion models. Our approach is motivated by both Etheridge and Griffiths (2009a), who derive a coalescent dual process for a stationary multiallelic Moran model with selection from the generator of its limiting diffusion approximation, as well as Papaspiliopoulos and Ruggiero (2014), who present a comprehensive method for identifying a dual processes for stationary, reversible diffusion models from their diffusion generator. This latter method is particularly relevant, since (similarly to us) the authors view the continuous diffusion process as a hidden process within the framework of a Hidden-Markov-Model, and the sampling distribution as the emission density. In this setting, the identified dual process is the computationally optimal filter for the diffusion. Here, we will present embedded dual jump processes for the general biallelic reversible mutation and the biallelic boundary mutation diffusion models, specifically pure death jump processes that describe the ancestry of

population samples. These are comparable to the dual processes of the above articles, but we will focus specifically on their probabilistic interpretation in the context of the forward–backward algorithm.

C.1. Recap of the forward-backward algorithm

Recall that in Appendix B.1.3, we determined p_k, the probability of observing k focal alleles in a binomial sample of size K drawn from a pure drift diffusion model at time $t = 0$, as the marginal likelihood of the product of the forward and backward transition rate densities, both represented as a spectral sum, via a forward–backward algorithm. In the main text, specifically Sections 2.4, 2.5, and 3, we use a forward pass of the algorithm to determine the same probability for diffusion equations with mutational input in the form of general biallelic reversible mutations and boundary-mutations respectively; this means that we choose to evaluate the probability of p_k at $t = 0$. In this first subsection, our main aim is to establish a notation for p_k that holds for both of these patterns of mutation, allowing us to evaluate it at any time $t_s \leq t \leq 0$ via a combination of forward and backward passes of the forward–backward algorithm.

In order to do so, we will again assume a population allele proportion that evolves forward in time via a diffusion equation; specifically, it evolves from an ancestral distribution $\rho(x)$ at time $t_s < 0$ up to the extant time $t = 0$, at which point we draw a binomial sample of haploid size K. We will treat this as a single time epoch, assuming constancy of mutation parameters for its duration. Note that this could be because we do not wish to model changes in these parameters, or because we assume that we are considering the last of a series of epochs with piecewise constant mutation parameters as in the main text.

The core principle of the forward–backward algorithm is that the marginal probabilities of the observations, here the event probabilities p_k, depend on the length of the interval between $t = t_s$ and $t = 0$ and can be evaluated to the same result at any point t within this interval. This is achieved by a combination of forward probabilities of population allele proportions covering the events from t_s up to t, and backward probabilities that evaluate the likelihood of the sampling distribution from t to $t = 0$. Implicitly, this requires an adjoint relationship between the forward and backward diffusion generators (Bergman et al., 2018b, Section 3.1., specifically Eq. 32). It is more straightforward to see that this holds in time homogeneous systems; therefore, we will from here on use the spectral decomposition of the general mutation diffusion model from Section 2.4 (adjoint relationship verified in Bergman et al., 2018b, Section 3.4) as well as the diagonalised eigensystem of the boundary mutation diffusion from Appendix B.2.3 (Bergman et al., 2018b, Section 4.1).

Following on from this, we expand the distribution that describes the distribution of ancestral population allele proportions at time t_s into a series of polynomials $\rho(x) = \phi(x, t = t_s) = \sum_{m=0}^{\infty} \rho_m^{(\beta, \theta)} \widetilde{\mathfrak{F}}_m^{(\beta, \theta)}(x)$, and then similarly expand the binomial sampling distribution into a complementary series of orthogonal polynomials $\Pr(k|K, x) = \psi(x, t = 0) = \sum_{n=0}^{K} d_n(k, K) \mathfrak{B}_n^{(\beta, \theta)}(x)$. The functions of the indeterminate x in these spectral sums are constituted by the orthogonal polynomials themselves, which here represent the eigenfunctions of *either* time homogeneous eigensystem with mutational input, and the coefficients of these function in the spectral sums therefore depend on both the mutation rate and bias.

Within either homogeneous eigensystem, the probability of observing k focal alleles in the sample drawn at the extant time $t = 0$ can then be determined by evaluating the following marginal likelihood at any

time t within $t_s \leq t \leq 0$ (Bergman et al., 2018b, Section 3.1):

$$p_k = \Pr(k|K, \beta, \theta, \rho, t_s)$$

$$= \int_0^1 \psi(x, t)\phi(x, t)\, dx$$

$$= \int_0^1 \left(\sum_{n=0}^{K} d_n(k, K) e^{\lambda_n t} \mathfrak{B}_n^{(\beta,\theta)}(x) \right)$$

$$\times \left(\sum_{m=0}^{\infty} \rho_m^{(\beta,\theta)} e^{-\lambda_m(t - t_s)} \mathfrak{F}_m^{(\beta,\theta)}(x) \right) dx \qquad (105)$$

$$= \sum_{n=0}^{K} \int_0^1 d_n(k, K) e^{\lambda_n t} \mathfrak{B}_n^{(\beta,\theta)}(x) \rho_n^{(\beta,\theta)} e^{-\lambda_n(t - t_s)} \mathfrak{F}_n^{(\beta,\theta)}(x)\, dx$$

$$= \sum_{n=0}^{K} d_n(k, K) \rho_n^{(\beta,\theta)} \Delta_n^{(\beta,\theta)} e^{\lambda_n t_s}.$$

Note that the above distribution is discrete, *i.e.,* $0 \leq p_k \leq 1$ for $0 \leq k \leq K$, whereby $\sum_k p_k = 1$. In the remainder of this Appendix, we will show that the transition rates $\psi(x, t)$ of the backwards process can be interpreted probabilistically for $t_s \leq t \leq 0$. Specifically, we will rewrite its representation as a spectral sum into temporally weighted mixtures of binomial samples across time, thereby tracing the genealogy of the observed sample back through the population history via an embedded time-dependent jump process.

Remark. In this Section, we considered a single interval between $t = t_s$ and $t = 0$; however, multiple epochs can easily be incorporated, *i.e.,* the backward process can be extended to additional epochs going further back in time. At the time where the mutation parameters change, which we denote as t_j (which may coincide with t_s), $\phi(x, t_j)$ is represented by a weighted sum of binomial distributions. This can be used as the initial distribution, *i.e.,* $\phi(x, t)$, in determining the backwards process for epochs further back in time; note that previously, $\phi(x, t)$ was set to a single binomial distribution $\Pr(k|K, x)$ at time $t = 0$.

C.2. General description of dual embedded pure death jump processes

Let us begin by providing an outline of the embedded pure death jump processes we will derive: To start, denote the observed sample allele configuration at the extant time $t = 0$ as (k, K), which simply means that k of the total K sampled alleles are of the focal type. In the history of this sample, we can "lose" alleles of both the focal or non-focal type as they coalesce with a common ancestor as well as through parent-independent mutation. Looking back in time, any reduction in sample size due to these events will be considered a jump step. After $K - n$ of these, we will denote our allele configuration as (i, n), meaning that we have i (where $\max(0, k + n - K) \leq i \leq \min(k, n)$) focal alleles among our remaining total of $n \leq K$ from the original sample. The transition probabilities between configurations of different historical sample sizes will be denoted as $\Pr(n|K, t)$, and they determine the temporal aspect of the sample genealogy. Subordinated to this is a spatial process that governs the number of focal alleles i across this genealogy, whereby we take "spatial" to mean "dependent on the population allele proportion x": This process is constituted by a sequence of pure death jump steps on the space of binomial distributions $\Pr(i|n, x) = \binom{n}{i} x^i (1 - x)^{n-i}$ with $0 \leq i \leq k$ and $0 \leq (n - i) \leq (K - k)$, with coefficients $e_{i,n}$ recording the probability of observing k focal alleles in the sample at the extant time across all possible sample paths from configuration (i, n) to (k, K):

$$e_{i,n} = \Pr(k|K, i, n, \beta, \theta).$$

Putting these two processes together, the transition rates $\psi(x, t)$ of the backwards process can be alternatively expressed as the following time-dependent pure death jump process:

$$\psi(x, t) = \Pr(k|K, \beta, \theta, \rho, x, t)$$

$$= \sum_{n=0}^{K} \Pr(n|K, t) \sum_{i=0}^{n} e_{i,n} \Pr(i|n, x). \qquad (106)$$

We will now proceed to derive this representation of the backwards transition rate density for the general biallelic reversible mutation diffusion model, and explicitly show that it is an embedded dual process with direct connections to the spectral sum we have typically expanded the transition rate density into in this article. Then, we will appropriately modify the derivation to obtain corresponding results for the boundary mutation model.

Remark. Note that terms analogous to our Eq. (106) appear in an expression for the stationary transition distribution of the backward multiallelic Moran diffusion equation in Etheridge and Griffiths (2009b, Eq. 2).

C.3. Derivation of dual embedded pure death jump processes

C.3.1. General biallelic reversible mutation model

In order to separate the temporal and spatial processes, we take an approach comparable to Etheridge and Griffiths (2009b, Eq. 7). Specifically, we substitute the binomial likelihood $\Pr(i|n, x) = \binom{n}{i} x^i (1 - x)^{n-i}$ into the backwards diffusion equation of the general biallelic reversible mutation model (see Eq. (2)):

$$\mathcal{L}^* \Pr(i|n, x) = \mathcal{L}^* \binom{n}{i} x^i (1 - x)^{n-i}$$

$$= i(i - 1 + \beta\theta) \binom{n}{i} x^{i-1} (1 - x)^{n-i+1}$$

$$+ (n - i)(n - i - 1 + \alpha\theta) \binom{n}{i} x^i (1 - x)^{n-i-1}$$

$$- (i(n - i + \alpha\theta) + (i + \beta\theta)(n - i)) \binom{n}{i} x^i (1 - x)^{n-i} \qquad (107)$$

$$= \lambda_n \left(\frac{i - 1 + \beta\theta}{n - 1 + \theta} \Pr(i - 1|n - 1, x) \right.$$

$$\left. + \frac{n - i - 1 + \alpha\theta}{n - 1 + \theta} \Pr(i|n - 1, x) - \Pr(i|n, x) \right).$$

Recall that the eigenvalues of this system, which govern its temporal aspect, are given by $\lambda_n = n(n - 1 + \theta)$. Regarding the number of focal alleles rather than the population allele proportion x as the dependent variable, the last line of the above equation can be seen as an infinitesimal generator of the spatial process within the large parentheses (compare Etheridge and Griffiths, 2009b, Eq. 7). Setting this spatial process zero yields a recursive set of equations on the probability of observing i focal alleles within historical samples of size n:

$$\Pr(i|n, x) = \frac{i - 1 + \beta\theta}{n - 1 + \theta} \Pr(i - 1|n - 1, x) + \frac{n - i - 1 + \alpha\theta}{n - 1 + \theta} \Pr(i|n - 1, x), \qquad (108)$$

which can also be written as:

$$\Pr(i|n, x) = \frac{i}{n} \frac{betabin(i|n, \beta, \theta)}{betabin(i - 1|n - 1, \beta, \theta)} \Pr(i - 1|n - 1, x)$$

$$+ \frac{n - i}{n} \frac{betabin(i|n, \beta, \theta)}{betabin(i|n - 1, \beta, \theta)} \Pr(i|n - 1, x). \qquad (109)$$

This recursion defines a time-independent, pure death jump process: Since it goes backwards in time, it starts from the allele configuration $(i = k, n = K)$ at the extant time $t = 0$, proceeds via configurations $(i = k - 1, n = K - 1)$ and $(i = k, n = K - 1)$ at the next lower possible sample size n, and down to the trivial empty sample $(i = 0, n = 0)$, which has $\Pr(i = 0|n = 0, x) = 1$, since the only possible value of i after K steps is zero. Let $e_{i,n} \geq 0$ be the solution of this recursive system; it is given by:

$$e_{i,n} = \Pr(k|K, i, n, \beta, \theta) = \binom{K - n}{k - i} \frac{\Gamma(\theta + n)}{\Gamma(\theta + K)} \frac{\Gamma(\beta\theta + k)}{\Gamma(\beta\theta + i)} \frac{\Gamma(\alpha\theta + K - k)}{\Gamma(\alpha\theta + n - i)}$$

$$= hyper(i|K, n, k) \frac{betabin(k|K, \beta, \theta)}{betabin(i|n, \beta, \theta)}. \qquad (110)$$
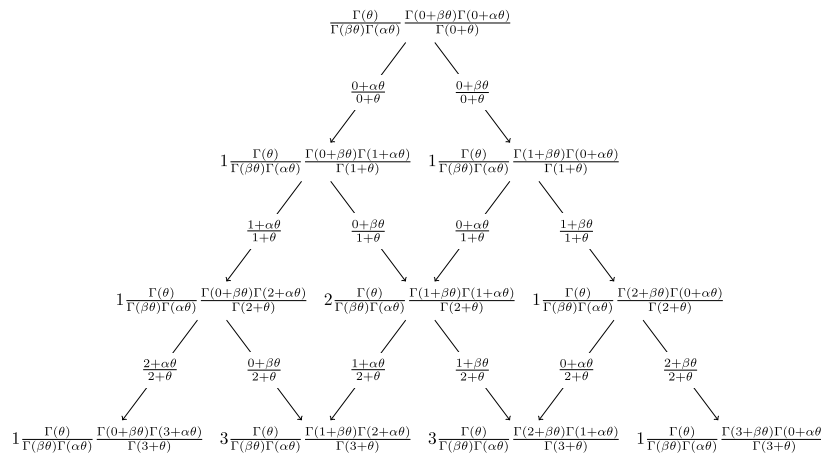
Fig. 7. This is a schematic for the Polya-type urn scheme that is reminiscent of Pascal's triangle. Specifically, we trace the history of a sample allele configuration from an ancestral state to the observed configuration (*i.e.,* forwards in time) under the assumption of an underlying general reversible biallelic mutation model. The entry at the tip of the triangle represents the trivial sample ($i = 0, n = 0$) (which has a probability of 1), and as we move down the triangle we proceed forwards in time in the urn process. The edges of the arrows in this scheme denote the transition probabilities from Eq. (111). The entries in the nth row from the top correspond to the increasing values of the historic sample size n, while the rows represent the possible values i of the focal allele within each row, *i.e.,* a historic sample. Note that the expressions at each node are the stationary beta-binomial sampling distributions from Eq. (3) for the appropriate sample configuration.

Note that $e_{0,0}$, which is the solution for the trivial historic sample configuration ($i = 0, n = 0$), corresponds to the stationary beta-binomial sampling distribution for the configuration (k, K) from Eq. (3). From this, it is easy to see that the $e_{i,n}$ are indeed the probabilities of observing k focal alleles in a sample configuration (k, K) at the extant time $t = 0$ given the historic sample configuration (i, n) at a time $K - n$ jump steps in the past; note also that this process is stationary for constant β and θ.

Connections to well-established models can be made by reversing the dependency relationships in this jump process: By beginning from the trivial historic sample configuration $\Pr(i = 0|n = 0) = 1$, we can construct a trivially modified Polya-type urn process for the probabilities of past sample configurations as though we were moving forwards in time (Hoppe, 1987, compare Eqs. (1),(16)) and Stephens and Donnelly (2000, Algorithm 2.1). The process starts from an urn containing balls of two colours, one focal and the other non-focal, with respective weights of $\beta\theta$ and $\alpha\theta$. At each jump step, we randomly draw a ball from the urn and then return it with a ball of the same type that has unit weight. Expressly, let the indicator variable z record whether the next ball drawn is of the focal colour. The transition probabilities from the sample configuration (i, n) to the sample configuration of $(i + z, n + 1)$ are proportional to the weight of the focal and non-focal balls, and can be written as:

$$\Pr(z|i, n, \beta, \theta) = p^z (1-p)^{1-z} \tag{111}$$

with

$$p = \frac{i + \beta\theta}{n + \theta} = \frac{betabin(i+1|n+1, \beta, \theta)}{betabin(i|n, \beta, \theta)}.$$

For further clarification of the spatial process, we provide a visualisation for both the urn model (see Fig. 7) and the jump process determining the $e_{i,n}$ (see Fig. 8).

We have now characterised the spatial component of the backwards diffusion equation operating on a binomial sampling function given in Eq. (107). This leaves the temporal component: Looking backwards in time, a jump step in the sample genealogy corresponds to a reduction in the historical sample size n by either a coalescent or a parent-independent mutation event. Let $\Pr(n|K, \theta, t)$, with $0 \leq n \leq K$, denote the probability of a realised historic sample of size n at time t. Considering this as the dependent variable in the backwards diffusion equation from Eq. (107), we can re-write the latter as the following system of

temporal linear differential equations:

$$-\frac{d}{dt} \Pr(n|K, \theta, t) = \lambda_K \Pr(n|K, \theta, t), \quad \text{for } n = K, \text{ and}$$
$$-\frac{d}{dt} \Pr(n|K, \theta, t) = -\lambda_{n+1} \Pr(n+1|K, \theta, t)$$
$$+ \lambda_n \Pr(n|K, \theta, t), \quad \text{for } K > n \geq 0. \tag{112}$$

Imposing the starting conditions $\Pr(n = K|K, \theta, t = 0) = 1$, and $\Pr(K > n \geq 0|K, \theta, t = 0) = 0$, the solution of this system is given by the standard solution for a pure death process, except that our historical time t is negative (Taylor and Karlin, 1998, chapter 6, Eq. 2.2) (and compare Tavare, 1984, Eq. 4.7):

$$\Pr(n|K, \theta, t) = e^{\lambda_K t}, \quad \text{for } n = K$$
$$\Pr(n|K, \theta, t) = \left(\prod_{j=n+1}^{K} \lambda_j \right) \sum_{i=n}^{K} \frac{e^{\lambda_i t}}{\prod_{j=n, j \neq i}^{K} (\lambda_j - \lambda_i)}, \quad \text{for } K - 1 \geq n \geq 0. \tag{113}$$

Clearly, this solution requires the eigenvalues $\lambda_n = n(n - 1 + \theta)$ to be distinct, which they are in this model for realistic values of θ, *i.e.,* values distinct from zero and infinity. Further, all $\Pr(n|K, \theta, t)$ are positive and it is easily seen that $\sum_{n=0}^{K} \Pr(n|K, \theta, t) = 1$ for $t_s \leq t \leq 0$, *i.e.,* probability must be conserved across the sample lineages at all times.

Combining the temporal pure death jump process and the subordinated spatial pure death jump process yields a time-dependent pure death jump process representation of the transition rates $\psi(x, t)$ of the backwards process:

$$\psi(x, t) = \sum_{n=0}^{K} \Pr(n|K, t) \sum_{i=0}^{n} e_{i,n} \Pr(i|n, x). \tag{114}$$

If we rewrite this equation to make the exponential function visible, we obtain the following:

$$\psi(x, t) = \sum_{n=0}^{K} \Pr(n|K, t) \sum_{i=0}^{n} e_{i,n} \Pr(i|n, x)$$
$$= \sum_{n=0}^{K} \left(\sum_{j=n}^{K} f_{n,j} e^{\lambda_j t} \right) \sum_{i=0}^{n} hyper(i|K, n, k) g_{i,n} x^i (1-x)^{n-i} \tag{115}$$

where we use $f_{j,n}$ and $g_{i,n}$ as placeholders for the remaining coefficients.
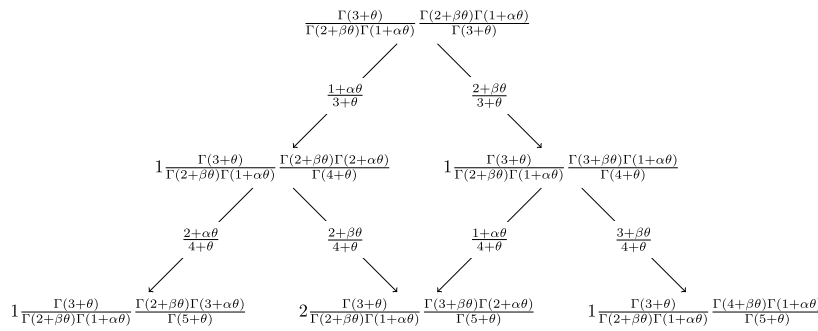
Fig. 8. This illustration elucidates the jump process determined by the probabilities $e_{i,n}$ from Eq. (110). We assume that a sample of size $K = 5$ is drawn at the extant time $t = 0$. The bottom row in the triangle shows the probabilities $e_{k,K}$ for the number of focal alleles $k = (2, 3, 4)$, assuming that this configuration has evolved from the historical sample configuration ($i = 2, n = 3$) at the top of the triangle two jump steps previously. While this process runs backwards in time, the conditional probabilities are constructed forwards in time and this is indicated by the direction of the arrows that represent the same transition probabilities as described by the Polya-type urn process in Eq. (111). Note that the node at the tip of the triangle corresponds to a probability of 1.

Recall that the expansion of the transition rates of the backwards process into a sum of orthogonal polynomials yields uniquely:

$$\psi(x,t) = \sum_{n=0}^{K} e^{\lambda_n t} d_n(k,K) R_n^{(\beta,\theta)}(x)$$

Collecting the polynomials multiplied by each of the $e^{\lambda_n t}$ for $n = 0, \dots, K$ in both the time-dependent jump process and the spectral sum, the coefficients for every order of expansion n can be equated between the two representations as in the Appendix B.2.2).

We will not pursue this further here. Nevertheless, we hope to have established that the jump process on the space of binomial distributions and the spectral expansion on the space of orthogonal polynomials are equivalent processes for evaluating the probability of sample genealogies at each point in time. In other words, multiplication of either representations of $\psi(x,t)$ with the probability measure $\phi(x,t)$ at any time t and integration over the population allele proportion will yield the same result for the probability of the observed sample configuration; *i.e.,* these processes are both dual to the backwards diffusion equation with respect to $\phi(x,t)$. Specifically, instead of determining the probability of observing k focal in a total of K alleles in a sample drawn at the extant time via the spectral expansion in the main text, we could utilise the time-dependent jump process:

$$p_k = \Pr(k|K, \beta, \theta, \rho, t)$$

$$= \int_0^1 \psi(x,t)\phi(x,t)\,dx$$

$$= \int_0^1 \sum_{n=0}^{K} \Pr(n|K,t) \sum_{i=0}^{n} e_{i,n} \Pr(i|n,x)\,\phi(x,t)\,dx \quad (116)$$

$$= \sum_{n=0}^{K} \Pr(n|K,t) \sum_{i=0}^{n} e_{i,n} \sum_{m=0}^{n} d_m(i,n)\,\rho_m^{(\beta,\theta)}\Delta_m^{(G;\beta,\theta)} e^{-\lambda_m(t-t_s)}\ .$$

While this expression may be easier to interpret probabilistically in the light of this past subsection than the equivalent calculation involving the spectral sum, its evaluation clearly involves more intermediate summation steps. Note however, that this is specifically true for non-equilibrium scenarios since the final sum reduces to a single coefficient in equilibrium.

Remark. Our embedded time-dependent jump process bears the most similarity to Papaspiliopoulos and Ruggiero (2014, Section 3.3), where the same general result is presented as the dual process for a reversible multiallelic Wright–Fisher model; incidentally, our process corresponds to the special case of theirs with only two alleles, and thus also to the dual process derived by Chaleyat-Maurel and Genon-Catalot (2009). Both of these articles concern filtering the Wright–Fisher model on the basis of binomially distributed data, so that all tasks performed by a forward–backward algorithm involve finite mixture expansions as in the above Eq. (116). However, whilst we arrive at the same time-dependent pure death jump process as a dual process for the general biallelic reversible diffusion model as Papaspiliopoulos and Ruggiero (2014) with a superficially analogous method, we additionally elaborate on the probabilistic interpretation of its component processes and explicitly allow for non-equilibrium scenarios. We do so because the partitioning into spatial and temporal components reflects the logic of the orthogonal polynomial expansion. Note that the start of the derivation of a coalescent dual process for the Wright–Fisher diffusion requires a similar separation (Etheridge and Griffiths, 2009a). However, these authors consider a model with selection and differ methodologically in their approach in that they assume an infinite entrance boundary rather than starting from an observed sample at a specified time point.

Remark. In terms of determining the expected sample allele configuration of a population evolving according to Moran and Wright–Fisher diffusion models, a dual processes is generally found in Kingman's coalescent (Kingman, 1982) or a variant thereof (Tavare, 1984). As pointed out in Papaspiliopoulos and Ruggiero (2014), the expression $\sum_{i=n}^{K} \frac{e^{\lambda_i t}}{\prod_{j=n, j\neq i}^{K}(\lambda_j - \lambda_i)}$ with $K-1 \geq n \geq 0$, which occurs within our temporal process, is precisely the transition probability of Kingman's block-counting process with mutation. The duality to Kingman's coalescent is more explicitly shown in Tavare (1984, Eq. 7.25), who recovers an ancestral process on the number of distinct ancestors in a population immediately from the diffusion equation (this is cited as Kingman's own diffusion time scale approximation to the original discrete time and space coalescent). Further, Griffiths and Spanò (2010) have shown that the Wright–Fisher diffusion with transition rate densities expressed via expansion into Jacobi polynomials is dual to Kingman's coalescent by showing that these transition rate densities correspond to the mixture distribution tracing the non-mutant lineages in the population arising from the coalescent (whereby the latter also appear in Etheridge and Griffiths (2009a)); note that these are population level considerations involving infinite sums.

Remark. Computational methods related to the "matrix coalescent" (Wooding and Rogers, 2002) employ a spectral decomposition of the transition rate matrix of the ancestral processes of a sample (Tavare, 1984, compare Section 5.1), *i.e.,* the rate matrix describing the decrease in distinct lineages looking back in time. For Kingman's coalescent, this would be exactly the corresponding rate matrix to the transition probabilities recovered within our time-homogeneous pure death process (see previous remark). Within this framework, time-inhomogeneous coalescent processes, which include non-constant demographic scenarios

as well as multiple-merger coalescents, can be modelled by realising that such inhomogeneity affects the observed site-frequency spectrum only via the first coalescent time within each (historical) sample size and explicitly incorporating the dynamics between it and the other coalescent times into the system (Spence et al., 2016). Separating out the impact of temporal changes in this way is similar in spirit to our approach; it is simply the genealogy that is considered rather than allele proportions.

C.3.2. Boundary mutation model

Because of the unique dynamics at the boundaries, substituting the binomial likelihood representing the sampling scheme into the appropriate backwards diffusion equation for the boundary mutation model in order to derive an embedded dual process is not immediately possible. However, the recursive system for the spatial process analogous to Eq. (109) in the general mutation model can be directly given as:

$$
\Pr(i|n, x) = \frac{i}{n} \frac{bound\,bin(i|n, \beta, \theta)}{bound\,bin(i - 1|n - 1, \beta, \theta)} \Pr(i - 1|n - 1, x)
$$
$$
+ \frac{n - i}{n} \frac{bound\,bin(i|n, \beta, \theta)}{bound\,bin(i|n - 1, \beta, \theta)} \Pr(i|n - 1, x). \tag{117}
$$

Recall that the factored out eigenvalues must be $\lambda_0 = 0$, $\lambda_1 = \theta$, and $\lambda_n = n(n - 1)$ for $n \geq 2$ for the boundary mutation model. The above recursive relationship then describes a spatial process for the history of the number of focal alleles in a sample (compare Eq. (108)); its result is given by:

$$
e_{i,n} = hyper(i|K, n, k) \frac{bound\,bin(k|K, \beta, \theta)}{bound\,bin(i|n, \beta, \theta)}. \tag{118}
$$

Reversing time in this spatial process, a Polya-type urn model can again be formulated to describe the system: The transition rates are determined by a Bernoulli distributed indicator variable as in Eq. (111), but with the following success probabilities:

$$
p = \frac{bound\,bin(i + 1|n + 1, \beta, \theta)}{bound\,bin(i|n, \beta, \theta)}. \tag{119}
$$

For clarity, this urn process is visualised in Fig. 9.

Importantly, we must draw attention to the fact that this dual embedded pure death jump process is actually not equivalent to the process considered in the rest of this article. Instead, it is analogous to the considerations made by Burden and Griffiths (2019a) in their proofs of their results for the limits of low scaled mutation rates; however, to our knowledge the spectral representation of their backwards process for low scaled mutation rates is not known. Recall that the backwards system considered in the rest of this article specifically contrasts the monomorphic boundary region, consisting of the first two eigenfunctions associated with eigenvalues λ_0 and λ_1 in the spectral expansion, with the polymorphic interior. The interior eigenfunctions associated with the eigenvalues $\lambda_{n \geq 2}$ in the spectral expansion are assumed to follow pure drift dynamics. Hence, the corresponding embedded dual pure death process for $\Pr(i|n, x)$ from any starting values of n down to $n = 2$ can be readily obtained from Eq. (107) by setting $\theta = 0$ and following the pattern in the general mutation model. In contrast, to recover the dynamics of $\Pr(i|n = 1, x)$ in terms of this death process, the system from Eq. (96) would need to be expressed as a function of $\Pr(i|n, x)$ for $n = 0$ and $n = 1$ (see Appendix B.2.3). A fuller treatment of this approach, or also an in depth comparison of the two alternative approaches, would over-extend the scope of our manuscript.

Remark. Note that while the general structure of the temporal pure death process for the boundary mutation model is identical to that of the general reversible biallelic mutation model, the transition probabilities will differ slightly: Instead of the stationary distribution being a Beta distribution, the stationary distribution of the boundary mutation diffusion from Eq. (103) must be employed; thus, instead of the Beta-binomially distributed terms in Eq. (109), boundary-binomially distributed terms (see Eq. (5)) must be utilised as in Eq. (117). Note

also that the eigenvalues must also be those of the boundary mutation model.

Remark. The methodology formalised by Papaspiliopoulos and Ruggiero (2014) for identifying a dual process for diffusion models that allows for performing the calculations involved in a forward–backward algorithm via finite sums appears to require diffusion models that are reversible with respect to their stationary distribution. This is indeed the case for the general biallelic reversible mutation model (compare Song and Steinrücken (2012, p. 3) and Bergman et al. (2018b, Eq. 36)) where the stationary distribution is Beta distributed (or Dirichlet distributed when multiallelic models are considered, as in Papaspiliopoulos and Ruggiero (2014)). However, they note that the methodology can be extended to non-stationary diffusion processes that are reversible with respect to a reversible measure (which is not necessarily a well-defined distribution) as long their sampling distributions are well-defined (see Discussion in Papaspiliopoulos and Ruggiero (2014)). While we were not able to adapt their methodology to the boundary mutation diffusion process without workarounds and additional reasoning, we would like to note that it does fall into this latter category: The process is reversible with respect to the stationary distribution with boundary singularities from Eq. (103) (see Section 4.2 in Papaspiliopoulos and Ruggiero, 2014), and we have shown that, for K small enough, the sampling distribution is well-defined in the sense that it is a regular density both in equilibrium (Appendix B.2.4) and non-equilibrium (as in this section).

Appendix D. Miscellaneous additional results for the demographic models

D.1. Population explosion

Note that if $\theta_1^* \ll \theta_2^*$, the expression in the exponent of the solution to Eq. (33) tends to zero. A first order Taylor expansion around it yields for $t_1 < t \leq 0$:

$$
\tau_n(t)\, \Xi_n^{-1} \approx \theta_2^* + \left(\theta_1^* - \theta_2^*\right)\left(1 - \lambda_n \frac{\theta_1^*}{\theta_2^*} t\right)
$$
$$
= \theta_1^* + \frac{\theta_2^* - \theta_1^*}{\theta_2} \theta_1^* \lambda_n t \tag{120}
$$
$$
\approx \theta_1^*(1 + \lambda_n t).
$$

D.2. D-statistic

Neutrality tests based on the infinite sites model are pervasive in population genetics. Formulated for site frequency data, they take the general form (Korneliussen et al., 2013):

$$
D = \frac{\hat{\theta}_W - \hat{\theta}_Z}{\sqrt{\mathrm{Var}(\hat{\theta}_W - \hat{\theta}_Z)}},
$$

where $\hat{\theta}_W$ and $\hat{\theta}_Z$ are unbiased estimators for the overall scaled mutation rate that differ in their weighting of polymorphic sites assuming neutral equilibrium. In the case of Tajima's D (Tajima, 1989), $\hat{\theta}_W$ is Watterson's (or Ewens' and Watterson's) estimator (Ewens, 1972, 1974; Watterson, 1975), which counts the expected number of segregating sites, and $\hat{\theta}_Z$ is a statistic that counts the number of pairwise differences between ancestral and derived nucleotides (Tajima, 1983). The variance in the denominator of the statistic D is conventionally calculated assuming complete linkage between sites, *i.e.,* sites are not independent. Note that $\hat{\theta}_Z$ weights intermediate frequency alleles higher than $\hat{\theta}_W$. Therefore recent positive selection or population expansion, which increase the frequency of rare alleles, lead to values $D < 0$; balancing selection and population collapse elevate the relative numbers of intermediate allele frequencies and therefore lead to values $D > 0$.
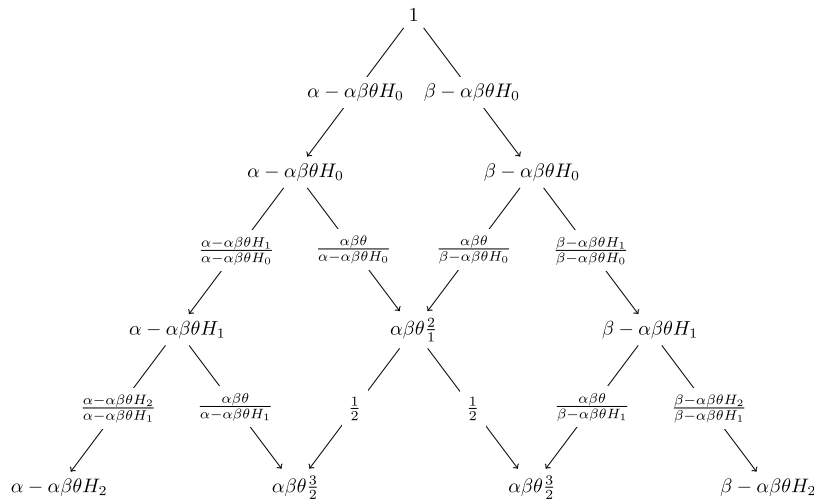
Fig. 9. This is a schematic for the Polya-type urn scheme that is reminiscent of Pascal's triangle. Specifically, we trace the history of a sample allele configuration from an ancestral state to the observed configuration (*i.e.*, forwards in time) under the assumption of an underlying boundary mutation model. The entry at the tip of the triangle represents the trivial sample ($i = 0, n = 0$) (which has a probability of 1), and as we move down the triangle we proceed forwards in time in the urn process. The edges of the arrows in this scheme denote the transition probabilities from Eq. (119). The entries in the nth row from the top correspond to the increasing values of the historic sample size n, while the rows represent the possible values i of the focal allele within each row, *i.e.*, a historic sample. Note that the expressions at each node are the stationary beta-binomial sampling distributions from Eq. (5) for the appropriate sample configuration. For the visualisation scheme to work, we defined the trivial harmonic number $H_0 = 0$ (so that the recurrence relation $H_{n+1} = H_n + \frac{1}{n+1}$ still holds).

Here, we reformulate Tajima's D for the context of the biallelic boundary mutation Moran model: Let us assume a sample site frequency spectrum of size K comprised of L loci, and recall that y_k denoted the realised count of focal alleles. We further assume recombination rates high enough for linkage to be negligible and sites therefore independent, which we in fact do implicitly throughout this article. For low overall scaled mutation rates the distribution of the counts of focal alleles y_k of the polymorphic spectrum ($k \in \{1, \ldots, K-1\}$) can be approximated by a Poisson distribution with :

$$y_k \overset{iid}{\sim} Pois\left(L\bar{p}_k \right).$$

In this case, the stationary distribution $\bar{p}_k = \alpha\beta\theta \frac{K}{k(K-k)}$ is given by the boundary binomial distribution from Eq. (5). Then,

$$\hat{\theta}_W = \frac{\sum_{k=1}^{K} y_k}{2L \sum_{k=1}^{K-1} \frac{1}{k}}$$

is the minimum variance, unbiased maximum likelihood estimator for the bias-complemented overall scaled mutation rate $\theta^* = \alpha\beta\theta$ (Vogl et al., 2020). Since we do not distinguish between ancestral and derived alleles, the average pairwise distance between alleles of a sample can also be used as an estimator of the same parameter:

$$\hat{\theta}_Z = \frac{\sum_{k=1}^{K} y_k \, k(K-k)}{LK(K-1)}.$$

Substituting the estimator $\hat{\theta}_W$ for the parameter θ^*, the estimator for the variance of the difference between the two estimators, recalling that mean and variance of the Poisson distribution are equal, is:

$$\widehat{\text{Var}}(\hat{\theta}_W - \hat{\theta}_Z) = \sum_{k=1}^{K-1} \text{E}\left(\frac{y_k}{L}\right) \left(\frac{k(K-k)}{K(K-1)} - \frac{1}{2\sum_{k=1}^{K-1} \frac{1}{k}} \right)^2$$

$$= \hat{\theta}_W \sum_{k=1}^{K-1} \frac{K}{k(K-k)} \left(\frac{k(K-k)}{K(K-1)} - \frac{1}{2\sum_{k=1}^{K-1} \frac{1}{k}} \right)^2.$$

D.3. Tabulated comparisons of analytical and simulated results for models 3 and 4

See Tables 5–7.

Table 5

A population is assumed evolve forwards in time according to the autoregression model from Eq. (45), with the overall bias-complemented scaled mutation rate θ_j^* randomly fluctuating across epochs according to $\theta_j^* \overset{iid}{\sim} IG(a = 6, b = 0.01 \cdot (a - 1))$. The arithmetic mean of θ_j^* is then 0.01 and the harmonic mean 0.008333. After $J = 10^6$ epochs, the mean, variance, covariance, and correlations for the temporal coefficients $\vartheta_{j,n}$ for j between 1 and J and select n are evaluated and compared with their theoretical counterparts (rounded to 6 post-decimal digits).

Additional Example 1	Mean	Variance	Covariance	Correlation
Model 3				
$c = 10^{-4}$				
$n = 2$ - Theoretical	0.008350	0	0	0.976332
Simulated	0.008351	0	0	0.976349
$n = 12$ - Theoretical	0.009142	0.000008	0.000002	0.245198
Simulated	0.009142	0.000008	0.000002	0.245154
$c = 10^{-2}$				
$n = 2$ - Theoretical	0.009387	0.000012	0.000002	0.132810
Simulated	0.009387	0.000012	0.000002	0.133339
$n = 12$ - Theoretical	0.01	0.000025	0	0
Simulated	0.009999	0.000025	0	−0.000345
$c = 1$				
$n = 2$ - Theoretical	0.01	0.000025	0	0
Simulated	0.009998	0.000025	0	−0.000095
$n = 12$ - Theoretical	0.01	0.000025	0	0
Simulated	0.009998	0.000025	0	0.000071
Model 4				
$c = 1$				
$n = 2$ - Theoretical	0.01	0.000150	0	0
Simulated	0.01	0.000150	0	−0.000205
$n = 12$ - Theoretical	0.01	0.000150	0	0
Simulated	0.009996	0.000150	0	−0.000332

Table 6

A population is assumed to be evolving forwards in time according to the autoregression model from Eq. (45), with the overall bias-complemented scaled mutation rate θ_j^* randomly fluctuating across epochs according to $\theta_j^* \overset{iid}{\sim} IG(a = 3, b = 0.01 \cdot (a - 1))$. The arithmetic mean of θ_j^* is then 0.01 and the harmonic mean 0.006667. After $J = 10^6$ epochs, the mean, variance, covariance, and correlations for the temporal coefficients $\vartheta_{j,n}$ for j between 1 and J and select n are evaluated and compared with their theoretical counterparts (rounded to 6 post-decimal digits).

Additional Example 2	Mean	Variance	Covariance	Correlation
Model 3				
$c = 10^{-4}$				
$n = 2$ - Theoretical	0.006700	0	0	0.970590
Simulated	0.006700	0	0	0.970490
$n = 12$ - Theoretical	0.008153	0.000015	0.000003	0.218613
Simulated	0.008153	0.000014	0.000003	0.218262
$c = 10^{-2}$				
$n = 2$ - Theoretical	0.008571	0.000020	0.000003	0.125
Simulated	0.008571	0.000020	0.000003	0.124526
$n = 12$ - Theoretical	0.009998	0.000096	0	0.000003
Simulated	0.009998	0.000096	0	−0.000071
$c = 1$				
$n = 2$ - Theoretical	0.009999	0.000097	0	0.000001
Simulated	0.009999	0.000097	0	−0.000172
$n = 12$ - Theoretical	0.01	0.0001	0	0
Simulated	0.01	0.0001	0	0.000629
Model 4				
$c = 1$				
$n = 2$ - Theoretical	0.009999	0.000294	0	0.000001
Simulated	0.009995	0.000288	0	0.000387
$n = 12$ - Theoretical	0.01	0.0003	0	0
Simulated	0.009998	0.000299	0	0.000074

Table 7

A population is assumed to be evolving forwards in time according to the autoregression model from Eq. (45), with the overall bias-complemented scaled mutation rate θ_j^* randomly fluctuating across epochs according to $\theta_j^* \overset{iid}{\sim} IG(a = 6, b = 0.005 \cdot (a - 1))$. The arithmetic mean of θ_j^* is then 0.005 and the harmonic mean 0.004167. After $J = 10^6$ epochs, the mean, variance, covariance, and correlations for the temporal coefficients $\vartheta_{j,n}$ for j between 1 and J and select n are evaluated and compared with their theoretical counterparts (rounded to 6 post-decimal digits).

Additional Example 3	Mean	Variance	Covariance	Correlation
Model 3				
$c = 10^{-4}$				
$n = 2$ - Theoretical	0.004183	0	0	0.953316
Simulated	0.004184	0	0	0.953433
$n = 12$ - Theoretical	0.004775	0.000004	0	0.078571
Simulated	0.004774	0.000004	0	0.077895
$c = 10^{-2}$				
$n = 2$ - Theoretical	0.004879	0.000004	0	0.029401
Simulated	0.004880	0.000004	0	0.029057
$n = 12$ - Theoretical	0.0050	0.000006	0	0
Simulated	0.004999	0.000006	0	0.000088
$c = 1$				
$n = 2$ - Theoretical	0.005	0.000006	0	0
Simulated	0.005001	0.000006	0	0.000084
$n = 12$ - Theoretical	0.005	0.000006	0	0
Simulated	0.005001	0.000006	0	0.0000273
Model 4				
$c = 1$				
$n = 2$ - Theoretical	0.005	0.000038	0	0
Simulated	0.004999	0.0000375	0	0.000510
$n = 12$ - Theoretical	0.005	0.000038	0	0
Simulated	0.005	0.000037	0	−0.000396

References

Abramowitz, M., Stegun, I. (Eds.), 1970. Handbook of Mathematical Functions, nineth ed. Dover, New York, USA.

Achaz, G., 2009. Frequency spectrum neutrality tests: One for all and all for one. Genetics 183, 249–258.

Baake, E., Bialowons, R., 2008. Ancestral processes with selection: Branching and Moran models. Banach Cent. Publ. 80, 33–52.

Bergman, J., Betancourt, A., Vogl, C., 2018a. Transcription-associated compositional skews in Drosophila genes. Genome Biol. Evol. 10, 269–275.

Bergman, J., Schrempf, D., Kosiol, C., Vogl, C., 2018b. Inference in population genetics using forward and backward, discrete and continuous time processes. J. Theoret. Biol. 439, 166–180.

Bhaskar, A., Song, Y., 2014. Descartes' rule of signs and the identifiability of population genetic models from genomic variation data. Ann. Statist. 42, 2469–2493.

Bhaskar, A., Wang, Y., Song, Y., 2015. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. Genome Res. 25, 268–279.

Burden, C., Griffiths, R., 2019a. The stationary distribution of a sample from the Wright-Fisher diffusion model with general small mutation rates. J. Math. Biol. 78, 1211–1224.

Burden, C., Griffiths, R., 2019b. The transition distribution of a sample from a Wright-Fisher diffusion with general small mutation rates. J. Math. Biol. 79, 2315–2342.

Burden, C., Tang, Y., 2017. Rate matrix estimation from site frequency data. Theor. Popul. Biol. 113, 23–33.

Chaleyat-Maurel, M., Genon-Catalot, V., 2009. Filtering the Wright-Fisher diffusion. ESAIM: Porbabil. Stat. 13, 196–217.

Chen, H., 2019. A computational approach for modeling the allele frequency spectrum of populations with arbitrarily varying size. Genom. Proteom. Bioinform. 17, 635–644.

Crow, J., 1954. Breeding structure of populations II. Effective population number. In: Statistics and Mathematics in Biology. Iowas State Coll. Press, Ames, Iowa, pp. 543–556.

Der, A., Plotkin, J., 2014. The equilibrium Allele frequency distribution for a population with reproductive skew. Genetics 192, 1199–1216.

Eldon, B., Birkner, M., Blath, J., Freund, F., 2015. Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents? Genetics 199, 841–856.

Eriksson, A., Mehlig, B., Rafajlovic, M., Sagitov, S., 2010. He total branch length of sample genealogies in populations of variable size. Genetics 186, 601–611.

Etheridge, A., 2012. Some Mathematical Models from Population Genetics: Lecture Notes in Mathematics. Springer Verlag, Berlin, Heidelberg.

Etheridge, A., Griffiths, R., 2009a. A coalescent dual process in a Moran model with genic selection. Theor. Popul. Biol. 75, 320–330.

Etheridge, A., Griffiths, R., 2009b. A coalescent dual process in a Moran model with genic selection. Theor. Popul. Biol. 75, 320–330.

Evans, S., Shvets, Y., Slatkin, M., 2007. Non-equilibrium theory of the Allele frequency spectrum. Theor. Popul. Biol. 71, 109–119.

Ewens, W., 1972. The sampling theory of selectively neutral Alleles. Theor. Popul. Biol. 3, 87–112.

Ewens, W., 1974. A note on the sampling theory for infinite Alleles and infinite sites models. Theor. Popul. Biol. 6, 143–148.

Ewens, W., 1982. On the concept of effective population size. Theor. Popul. Biol. 21, 373–378.

Ewens, W., 2004. Mathematical Population Genetics, second ed. Springer, New York, USA.

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

Fay, J., Wu, C.-I., 2000. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413.

Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P., Zimmermann, P., 2007. MPFR: A multiple-precision binary floating-point library with correct rounding. ACM Trans. Math. Software 33 (2), 13.

Franca, F., Ferreira, J., Vaz-de Mello, F., Maia, L., Berenguer, E., Palmeira, A., Fadini, R., Louzada, J., Braga, R., Oliveira, V., Barlow, J., 2020. El Nino impacts on human-modified tropical forests: Consequences for dung beetle diversity and associated ecological processes. Biotropica 52, 252–262.

Freund, F., Kerdoncuff, E., Matuszewskic, S., Lapierre, M., Hildebrandte, M., Jeffrey, J., Ferretti, L., Lambert, A., Sackton, T., Achaz, G., 2023. Interpreting the pervasive observation of U-shaped site frequency spectra. PLOS Genet. 19, e1010677.

Fu, Y., 1995. Statistical properties of segregating sites. Theor. Popul. Biol. 48, 172–197.

Fu, Y., Li, W., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Griffiths, R., 1979. A transition density expansion for a multi-allele diffusion model. Adv. in Appl. Probab. 11, 310–325.

Griffiths, R., 2003. The frequency spectrum of a mutation, and its age, in a general diffusion model. Theor. Popul. Biol. 64, 241–251.

Griffiths, R., Li, W.-H., 1983. Simulating Allele frequencies in a population and the genetic differentiation of populations under mutation pressure. Theor. Popul. Biol. 23 (1), 19–33.

Griffiths, R., Spanò, D., 2010. Diffusion processes and coalescent trees. In: Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman. Cambridge University Press, Cambridge, UK., pp. 358–375.

Griffiths, R., Tavaré, S., 1998. The age of a mutation in a general coalescent tree. Stoch. Models 14, 273–295.

Harris, J., Stocker, H., 1998. Handbook of Mathematics and Computational Science, first ed. Springer, New York.

Hobolth, A., Siri-Jègousse, A., Bladt, M., 2019. Phase-type distributions in population genetics. Theor. Popul. Biol. 127, 16–32.

Hogg, R.V., Craig, A.T., 1995. Introduction to Mathematical Statistics, nineth ed. Prentice Hall, New York.

Hoppe, F.M., 1987. The sampling theory of neutral Alleles and an urn model in population genetics. J. Math. Biol. 25, 123–159.

Hössjer, O., 2011. Coalescence theory for a general class of structured populations with fast migration. J. Adv. Appl. Probabil. 43, 1027–1047.

Hössjer, O., 2015. On the eigenvalue effective size of structured populations. J. Math. Biol. 71, 595–646.

Hudson, R.R., 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18, 337–338.

Iizuka, M., 2001. The effective size of fluctuating populations. Theor. Popul. Biol. 59, 281–286.

Iizuka, M., 2010. Effective population size of a population with stochastically varying size. J. Math. Biol. 61, 369–375.

Iizuka, M., Tachida, H., Matsuda, H., 2002. A neutral model with fluctuating population size and its effective size. Genetics 161, 381–388.

Jagers, P., Sagitov, S., 2004. Convergence to the coalescent in populations of substantially varying size. J. Appl. Probab. 41, 368–378.

Johnston, J., 1960. Econometric Methods. McGraw-Hill, New-York.

Johri, P., Aquadro, C., Beaumont, M., Charlesworth, B., Excoffier, L., Eyre-Walker, A., Keightley, P., Lynch, M., McVean, G., Payseur, B., Pfeifer, S., Stephan, W., Jensen, J., 2022. Recommendations for improving statistical inference in population genomics. PLoS Biol. 20 (5), e3001669.

Kaj, I., Krone, S., 2003. The coalescent process in a population with stochastically varying size. J. Appl. Probab. 40, 33–48.

Kimura, M., 1955a. Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. 41, 144–150.

Kimura, M., 1955b. Stochastic processes and the distribution of gene frequencies under natural selection. In: Cold Spring Harbour Symposium on Quantitative Biology, vol. 20, pp. 33–53.

Kimura, M., 1963. The measurement of effective population number. Evolution 17, 279–288.

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61, 893–903.

Kimura, M., Ohta, T., 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61, 763–771.

Kingman, J., 1982. On the genealogy of large populations. J. Appl. Probab. 19A, 27–43.

Korneliussen, T., Moltke, I., Albrechtsen, A., Nielsen, R., 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics 14, 289–293.

Lande, R., 2014. Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative character in a fluctuating environment. J. Evol. Biol. 27, 866–875.

Lee, A., Bernt-Erik, S., Steinar, E., 2011. Demographic stochasticity, Allee effects, and extinction: The influence of mating system and sex ratio. Amer. Nat. 177 (3), 301–313.

Lukic, S., Hey, J., Chen, K., 2011. Non-equilibrium allele frequency spectra via spectral methods. Theor. Popul. Biol. 79, 2003–2019.

Lynch, M., Ackerman, M., Gout, J., Long, H., Sung, W., Thomas, W., Foster, P., 2016. Genetic drift, selection and the evolution of the mutation rate. Nature Rev. Genet. 17, 704–714.

Machado, H., Bergland, A., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D., Flatt, T., Gonzalez, J., Karasov, T., Kim, B., Kozeretska, I., Lazzaro, B., Merritt, T., Pool, J., O'Brien, K., Rajpurohit, S., Roy, P., Schaeffer, S., Serga, S., Schmidt, P., Petrov, D., 2021. Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in Drosophila. eLife 10, e67577.

Marth, G., Czabarka, E., Murvai, J., Sherry, S., 2004. The Allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166, 351–372.

McKane, A., Waxman, D., 2007. Singular solutions of the diffusion equation of population genetics. J. Theoret. Biol. 247, 849–858.

Motro, U., Thomson, G., 1982. On heterozygosity and the effective size of populations subject to size changes. Evolution 36, 1059–1066.

Myers, S., Fefferman, C., Patterson, N., 2008. Can one learn history from the Allelic spectrum? Theor. Popul. Biol. 73, 342–348.

Nawa, N., Tajima, F., 2008. Simple method for analyzing the pattern of DNA polymorphism and its application to SNP data of humans. Genes Genet. Syst. 83, 353–360.

Nei, M., Maruyama, T., Chakraborty, R., 1975. The bottleneck effect and genetic variability in populations. Evolution 29, 1–10.

Nordborg, M., Krone, S., 2002. Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin, M., Veuille, M. (Eds.), Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malecot. Oxford University Press, pp. 194–232.

Olver, F., Olde Daalhuis, A., Lozier, D., Schneider, B., Boisvert, R., Clark, C., Miller, B., Saunders, B., Cohl, H., McClain, M., 2022. NIST digital library of mathematical functions. Release 1.1.6 of 2022-06-30 http://dlmf.nist.gov/.

Papaspiliopoulos, O., Ruggiero, M., 2014. Optimal filtering and the dual process. Bernoulli 20 (4), 1999–2019.

Polanski, A., Kimmel, M., 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. Genetics 165, 427–436.

Pollak, E., 2002. Eigenvalue effective population numbers for populations that vary cyclically in size. Math. Biosci. 177, 11–24.

R. Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rabiner, L., Juang, B., 1986. An introduction to hidden Markov models. IEEE ASSP Mag. 3, 4–16.

Sano, A., Shimzu, A., Iizuka, M., 2004. Coalescent process with fluctuating population size and its effective size. Theor. Popul. Biol. 65, 39–45.

Sawyer, S., Hartl, D., 1992. Population genetics of polymorphism and divergence. Genetics 132, 1161–1176.

Schrempf, D., Minh, B.Q., De Maio, N., von Haeseler, A., Kosiol, C., 2016. Reversible polymorphism-aware phylogenetic models and their application to tree inference. J. Theoret. Biol. 407, 362–370.

Shimakura, N., 1977. Equations différentielles provenant de la génétique des populations. Tohoku Math. J. 29, 287–318.

Sjödin, P., Kaj, I., Krone, S., Lascoux, M., Nordborg, M., 2005. On the meaning and existence of an effective population size. Genetics 169, 1061–1070.

Song, Y., Steinrücken, M., 2012. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. Genetics 190, 1117–1129.

Spence, J., Kamm, J., Song, Y., 2016. The site frequency spectrum for general coalescents. Genetics 202, 1549–1561.

Steinrücken, M., Wang, R., Song, Y., 2013. An explicit transition density expansion for a multi-allelic Wright-Fisher diffusion with general diploid selection. Theor. Popul. Biol. 83, 1–14.

Stephens, M., Donnelly, P., 2000. Inference in molecular population genetics. J. R. Stat. Soc. Ser. B Stat. Methodol. 62, 605–655.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tavare, S., 1984. Line-of-descent and genealogical processes, and their applications in population genetics models. Theor. Popul. Biol. 26, 119–164.

Taylor, H., Karlin, S., 1998. An Introduction to Stochastic Modelling. Academic Press (Elsevier).

Tran, T., Hofrichter, J., Jost, J., 2013. An introduction to the mathematical structure of the Wright-Fisher model of population genetics. Theory Biosci. 132, 73–82.

Tran, T., Hofrichter, J., Jost, J., 2014. A general solution of the Wright-Fisher model of random genetic drift.

Vogl, C., 2014a. Computation of the likelihood in biallelic diffusion models using orthogonal polynomials. Computation 2, 199–220.

Vogl, C., 2014b. Estimating the scaled mutation rate and mutation bias with site frequency data. Theor. Popul. Biol. 98, 19–27.

Vogl, C., 2014c. Estimating the scaled mutation rate and mutation bias with site frequency data. Theor. Popul. Biol. 98, 19–27.

Vogl, C., Bergman, J., 2015. Inference of directional selection and mutation parameters assuming equilibrium. Theor. Popul. Biol. 106, 71–82.

Vogl, C., Bergman, J., 2016. Computation of the likelihood of joint site frequency spectra using orthogonal polynomials. Computation 4, 6.

Vogl, C., Clemente, F., 2012. The Allele-frequency spectrum in a decoupled moran model with mutation, drift, and directional selection, assuming small mutation rates. Theoret. Popul. Genet. 81, 197–209.

Vogl, C., Mikula, L., 2021. A nearly-neutral Biallelic Moran model with biased mutation and linear and quadratic selection. Theor. Popul. Biol. 139, 1–17.

Vogl, C., Mikula, L., Burden, C., 2020. Maximum likelihood estimators for scaled mutation rates in an equilibrium mutation-drift model. Theor. Popul. Biol. 134, 106–118.

Wakeley, J., 2009. Coalescent theory, an Introduction. Roberts and Co.

Wang, Y., Pollak, E., 2000a. The effective number of a population that varies cyclically in size 1. discrete generations. Math. Biosci. 166, 1–21.

Wang, Y., Pollak, E., 2000b. The effective number of a population that varies cyclically in size 2. discrete generations. Math. Biosci. 179, 161–181.

Watterson, G., 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7 (2), 256–276.

Wooding, S., Rogers, A., 2002. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics 161, 1641–1650.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Wright, S., 1938. Size of population and breeding structure in relation to evolution. Science 13, 430–431.

Zivkovic, D., Steinrücken, M., Song, Y., Stephan, W., 2015. Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. Genetics 200, 601–617.

Zivkovic, D., Stephan, W., 2011. Analytical results on the neutral non-equilibrium Allele frequency spectrum based on diffusion theory. Theor. Popul. Biol. 79, 184–191.