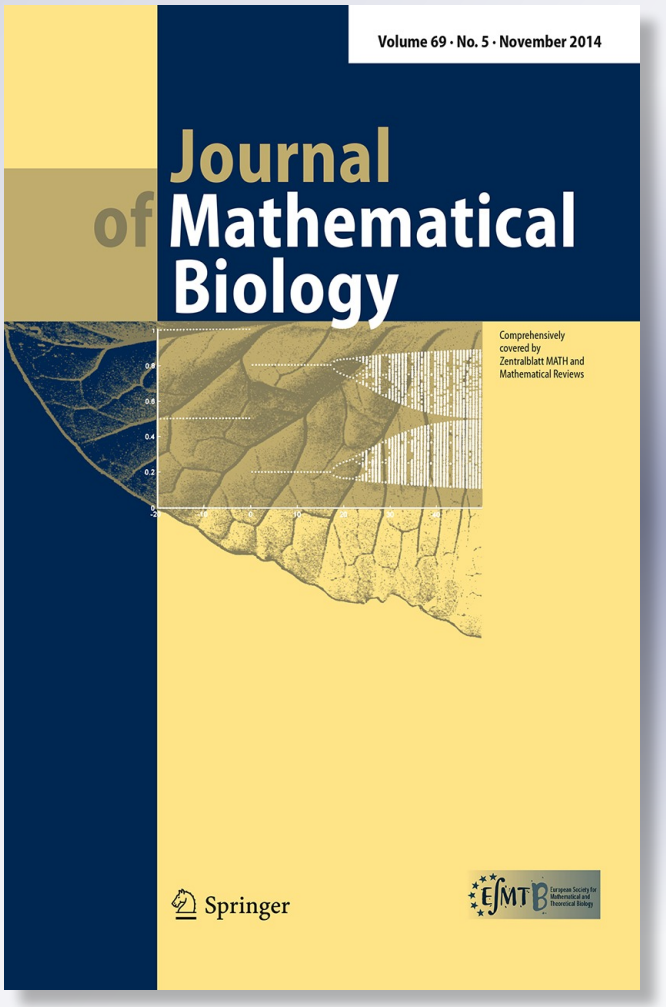
# *Quasi equilibrium, variance effective size and fixation index for populations with substructure*

**Ola Hössjer & Nils Ryman**

**Journal of Mathematical Biology**

ISSN 0303-6812
Volume 69
Number 5

J. Math. Biol. (2014) 69:1057-1128
DOI 10.1007/s00285-013-0728-9




Springer

J. Math. Biol. (2014) 69:1057–1128
DOI 10.1007/s00285-013-0728-9

**Mathematical Biology**

# Quasi equilibrium, variance effective size and fixation index for populations with substructure

**Ola Hössjer · Nils Ryman**

Received: 22 October 2012 / Revised: 11 September 2013 / Published online: 15 October 2013


**Abstract** In this paper, we develop a method for computing the variance effective size $N_{eV}$, the fixation index $F_{ST}$ and the coefficient of gene differentiation $G_{ST}$ of a structured population under equilibrium conditions. The subpopulation sizes are constant in time, with migration and reproduction schemes that can be chosen with great flexibility. Our quasi equilibrium approach is conditional on non-fixation of alleles. This is of relevance when migration rates are of a larger order of magnitude than the mutation rates, so that new mutations can be ignored before equilibrium balance between genetic drift and migration is obtained. The vector valued time series of subpopulation allele frequencies is divided into two parts; one corresponding to genetic drift of the whole population and one corresponding to differences in allele frequencies among subpopulations. We give conditions under which the first two moments of the latter, after a simple standardization, are well approximated by quantities that can be explicitly calculated. This enables us to compute approximations of the quasi equilibrium values of $N_{eV}$, $F_{ST}$ and $G_{ST}$. Our findings are illustrated for several reproduction and migration scenarios, including the island model, stepping stone models and a model where one subpopulation acts as a demographic reservoir. We also make detailed comparisons with a backward approach based on coalescence probabilities.

O. Hössjer (✉)
Division of Mathematical Statistics, Department of Mathematics,
Stockholm University, Stockholm, Sweden
e-mail: ola@math.su.se

N. Ryman
Division of Population Genetics, Department of Zoology,
Stockholm University, Stockholm, Sweden
e-mail: nils.ryman@popgen.su.se

**Keywords** Autoregressive time series · Island model · Quasi equilibrium · Stepping stone models · Spatial allele frequency fluctuations · Structured populations · Temporal allele frequency fluctuations

**Mathematics Subject Classification (2000)** 92D25 · 60F99

## 1 Introduction

Most populations exhibit some degree of spatial heterogeneity. This structure is often quite complicated and influences not only the spatial distribution, but also the time dynamics of allele frequencies. It is therefore helpful to extract a few summary statistics that broadly characterize the main properties of the population. This includes the effective population size (Wright 1931, 1938), defined as the size of an ideal population exhibiting the same rate of genetic drift as the focal population, and the fixation index (Wright 1951) which quantifies the degree of heterogeneity among subpopulations.

Many closely related versions of the effective population size and fixation index exist, see for instance Orive (1993), Caballero (1994), Wang and Caballero (1999), Waples (2002), Ewens (2004), Ryman and Leimar (2008), Hössjer et al. (2013). We will focus on the variance effective population size and a fixation index defined in terms of variance of subpopulation allele frequencies.

Ewens (1982) noted that the variance effective size is analytically intractable for structured populations. It is still of practical importance, since one of the most common procedures for assessing effective size from real data, the temporal method, estimates the variance effective size (Nei and Tajima 1981; Waples 1989; Jorde and Ryman 2007; Ryman et al. 2013). Procedures for estimating the fixation index from real data have been developed by several authors. One of the most widely used estimators is due to Weir and Cockerham (1984), see also Leviyang and Hamilton (2011) and references therein.

Our main interest is short time scales, so that the mutation rate is assumed small in comparison to the genetic drift and migration rate. We therefore disregard occurrence of *new* mutations and focus instead on equilibrium between genetic drift and migration. Half the inverse effective size then equals the rate of loss of heterozygosity, which is of great importance for quantifying population inbreeding and risk of extinction in conservation biology (Jamieson and Allendorf 2012). See also Palstra and Ruzzante (2008) and Hare et al. (2011) for recent reviews for wildlife populations.

Equilibrium between migration and genetic drift typically exists for infinite populations. For instance, values of the fixation index for the infinite island model can be derived from results of Sved and Latter (1977). For finite populations, no equilibrium value exists in absence of new mutations, since some allele will eventually become fixed at any locus. It is then of interest to study the quasi equilibrium behaviour of the system, conditional on non-fixation, see for instance Cattiaux et al. (2009) and Collet and Martinez (2013) for general results on quasi stationarity. Hössjer et al. (2013) formalized ideas of Nei et al. (1977) and derived quasi equilibrium values of the fixation index for the finite island model, conditionally on no fixation. We now extend this approach to joint computation of the fixation index and effective population size

at quasi equilibrium for a large class of finite population models. This class includes general migration and reproduction schemes, subject only to a constraint of constant subpopulation sizes over time.

In more detail we introduce in Sect. 2 the dynamics of how genes are transmitted from one generation to the next. In Sect. 3 we define the effective population size, the fixation index and their quasi equilibrium limits for a single biallelic locus and arbitrary schemes of weighting subpopulations. To this end, we introduce the vector valued, time inhomogeneous process of all subpopulations' allele frequencies. In Sect. 4 we demonstrate how it can be decomposed into two parts, one corresponding to genetic drift of the whole population and the other to fluctuations of allele frequencies among subpopulations. After a simple standardization, the latter becomes a (quasi) stationary process. It is proved in Sect. 5 that the first two conditional moments of this process and its innovations converge for large populations as long as no allele gets fixed.

In general, it is difficult to obtain closed form expressions, even for approximations of the quasi equilibrium effective population size and fixation index, although the island model is an exception. Instead, we present in Sect. 6 a general algorithm for computing approximations of these two quantities. In particular, we show that they simplify considerably when alleles are weighted proportionally to the reproductive values of their subpopulations. This weighting scheme was proposed by Felsenstein (1971) in the context of age-structured models and investigated by means of simulations in Waples and Yokota (2007). To illustrate the generality of our approach, several reproduction and migration examples are presented in Sects. 7 and 8. In Sect. 9 we consider multiallelic and multilocus extensions of the variance effective population size and Nei's (1973) coefficient of gene differentiation ($G_{ST}$), and give upper bounds for how well their quasi equilibrium limits are approximated by the formulas of Sect. 6. Numerical results are presented in Sect. 10. In Sect. 11 we compare our forward approach with the backward approach of Slatkin (1991), where the (nucleotide) effective size and fixation index are computed from expected coalescence times of pairs of genes under a small mutation rate assumption. In this process, we generalize Slatkin's results to our framework of general migration patterns and subpopulation weights, and give conditions under which the two approaches are asymptotically equivalent for large populations. Finally, we discuss possible extensions in Sect. 12, and present proofs and longer mathematical calculations in Appendices A–F.

## 2 Migration model

Consider a population that evolves in discrete non-overlapping generations $t = 0, 1, \ldots$. It is assumed to have a size of $N$ diploid individuals that is constant over time and can be divided into $s \geq 2$ subpopulations of sizes $Nu_1, \ldots, Nu_s$, where $\sum_{i=1}^{s} u_i = 1$. We will mainly think of these subpopulations as distinct geographical sites, although other interpretations are possible (see Sect. 12). To begin with, we consider a single polymorphic locus (referred to as a gene) with two alleles. Assuming that each individual carries two copies of the gene, subpopulation $i$ will consist of $2Nu_i$ genes at any generation $t$. Table 1 contains a list of the most important symbols used in the paper.

**Table 1** Notation for selected quantities in the paper

| Symbols | Definitions |
|---|---|
| $s$ | Number of subpopulations |
| $N$ | Total census size |
| $\boldsymbol{u} = (u_i)$ | Vector with relative sizes of all subpopulations $i$ |
| $Nu_i$ | Local census size of subpopulation $i$ |
| $\nu^l_{t,ki} = \nu^l_{ki}$ | Number of offspring in subpopulation $i$ of $l$:th gene of subpopulation $k$ and generation $t$ |
| $\boldsymbol{M} = (m_{ki})$ | Matrix with migration rates between all pairs of subpopulations $k$ and $i$ |
| $m'$ | Overall migration rate = proportion of offspring genes residing in subpopulations different from that of parental gene |
| $m$ | For the island model, proportion of migrants from the whole population ($m = sm'/(s-1)$) |
| ARNP | Average relative neighbourhood proportion |
| $\boldsymbol{v} = (v_k)$ | Vector of long term reproductivities of genes in all subpopulations $k$ |
| $\boldsymbol{r} = (r_k)$ | Vector of average nr of offspring of genes in all subpopulations $k$ |
| $\boldsymbol{B} = (b_{ik})$ | Backward matrix with probabilities that subpopulation $i$ genes' parents originate from subpopulation $k$ |
| $\boldsymbol{\gamma} = (\gamma_i)$ | Vector of equilibrium probabilities of distant ancestors to come from various subpopulation $i$ |
| $\boldsymbol{F} = (f_{ki})$ | Forward matrix with probabilities that offspring of subpopulation $k$ genes end up in subpopulation $i$ |
| $\boldsymbol{w} = (w_i)$ | Vector of weights of all subpopulations $i$ (default canonical choice is $\boldsymbol{\gamma}$) |
| $\boldsymbol{W}_T = \mathrm{vec}((w_i w_j))^T$ | Row vector of probabilities that a pair of genes of the total pop. belongs to subpopulation $i$ and $j$ |
| $\boldsymbol{W}_S = \mathrm{vec}((1_{\{i=j\}} w_i))^T$ | Row vector of prob that a pair of genes within the same subpopulation belongs to subpopulation $i$, $j$ |
| $\boldsymbol{P}_t = (P_{ti})$ | Vector of frequencies of Allele 1 (biallelic case) in generation $t$ in all subpopulations $i$ |
| $P^w_t$ | Overall frequency of Allele 1 in population when subpopulations are weighted as $\boldsymbol{w}$ |
| $\boldsymbol{\varepsilon}_t = (\varepsilon_{ti})$ | Vector of random drift of Allele 1 frequency in all subpopulation $i$ between generations $t-1$ and $t$ |
| $\boldsymbol{\Omega}(\boldsymbol{P}_t) = (\Omega(\boldsymbol{P}_t)_{ij})$ | Covariance matrix of genetic drift between generations $t$ and $t+1$ |
| $\boldsymbol{U} = (U_{ij,kl})$ | Coefficients in quadratic expansion of genetic drift covariance matrix $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ |
| $\boldsymbol{\Sigma}_t = (\Sigma_{tij})$ | Standardized genetic drift covariance matrix between generations $t$ and $t+1$ |
| $\boldsymbol{\Sigma} = (\Sigma_{ij})$ | Quasi equilibrium approximation of $\boldsymbol{\Sigma}_t$ |
| $\boldsymbol{V}_t = (V_{tij})$ | Standardized spatial covariance matrix for allele frequency fluctuations in generation $t$ |
| $\boldsymbol{V} = (V_{ij})$ | Quasi equilibrium approximation of $\boldsymbol{V}_t$ |
| $N^w_{eV,t}$ | Variance effective size of total (global) population in generation $t$ using weight vector $\boldsymbol{w}$ |
| $N^{\mathrm{eq},w}_{eV}$ | Quasi equilibrium limit of $N^w_{eV,t}$ |
| $N^{\mathrm{appr},w}_{eV}$ | Approximation of $N^{\mathrm{eq},w}_{eV}$ |
| $N_{ei}$ | Local effective size of subpopulation $i$ |
| $N_e$ | Local effective size of each subpopulations when they are all equal ($N_{ei} = N_e$) |

**Table 1** continued

| Symbols | Definitions |
|---|---|
| $\alpha_i$ | Dirichlet parameter of subpopulation $i$ reflecting amount of variation in migrant proportions |
| $\alpha$ | Dirichlet parameter of each subpopulation when they are all equal ($\alpha_i = \alpha$) |
| $\boldsymbol{H}_t = (H_{tij})$ | Matrix with gene diversities of all pairs of subpopulations in generation $t$ |
| $H_{Tt}^{w}$ | Gene diversity of total population in generation $t$ when subpopulations are weighted as $\boldsymbol{w}$ |
| $H_{St}^{w}$ | Gene diversity within subpopulations in generation $t$ when subpop. are weighted as $\boldsymbol{w}$ |
| $\bar{h}_{ij}$ | Predicted gene diversity of subpopulations $i$ and $j$, for two *distinct* genes |
| $\bar{h}_{T}^{w}$ | Predicted gene diversity of the total population, for two *distinct* genes, using weights $\boldsymbol{w}$ |
| $\bar{h}_{S}^{w}$ | Predicted gene diversity of subpopulations, for two *distinct* genes, using weights $\boldsymbol{w}$ |
| $F_{ST,t}^{w}$ | Fixation index of population in generation $t$ using weight vector $\boldsymbol{w}$ |
| $F_{ST}^{\mathrm{eq},w}$ | Quasi equilibrium limit of $F_{ST,t}^{w}$ |
| $F_{ST}^{\mathrm{appr},w}$ | Approximation of $F_{ST}^{\mathrm{eq},w}$ |
| $\bar{f}_{ST}^{w}$ | Predicted fixation index for two *distinct* genes, using weights $\boldsymbol{w}$ |
| $G_{ST,t}^{w}$ | Coefficient of gene differentiation of population in generation $t$ using weight vector $\boldsymbol{w}$ |
| $G_{ST}^{\mathrm{eq},w}$ | Quasi equilibrium limit of $G_{ST,t}^{w}$ |
| $G_{ST}^{\mathrm{appr},w}$ | Approximation of $G_{ST}^{\mathrm{eq},w}$ |
| $n$ | Number of loci (multilocus case) |
| $x$ | Order number of a specific locus |
| $n(x)$ | Number of alleles at locus $x$ |
| $\boldsymbol{q}_t = (q_{tij})$ | Matrix of probabilities that genes from subpop. $i$ and $j$ have not coalesced in $t$ generations |
| $\boldsymbol{T} = (T_{ij})$ | Matrix of coalescence times for two distinct genes from all pairs of subpopulations $i$, $j$ |

We number the genes of subpopulation $k$ in generation $t$ as $l = 1, \ldots, 2Nu_k$. In order to describe how these genes are transmitted to the next generation $t+1$, we introduce $\nu_{t,ki}^{l} = \nu_{ki}^{l}$ as the number of copies a particular gene $l$ of subpopulation $k$ and generation $t$ passes on to subpopulation $i$ in the next generation. The total number of genes transmitted from $k$ to $i$ is then $\sum_{l=1}^{2Nu_k} \nu_{ki}^{l}$. Since we assume a constant number of genes of subpopulation $i$ at level $2Nu_i$ in generation $t+1$, when summing the contributions from all source populations $k$ of the previous generation, we must have

$$\sum_{k=1}^{s} \sum_{l=1}^{2Nu_k} \nu_{ki}^{l} = 2Nu_i, \quad i = 1, \ldots, s. \tag{1}$$

The migration rate $m_{ki}$ between subpopulations $k$ and $i$ is defined as the average number of copies each gene in subpopulation $k$ passes on to $i$. We write this as

$$\frac{1}{2Nu_k} \sum_{l=1}^{2Nu_k} \nu_{ki}^{l} = m_{ki}(1 + o(1)), \tag{2}$$

where $o(1)$ is a remainder term that tends to zero as the population size grows. If $\{v_{ki}^l\}_{l=1}^{2Nu_k}$ are exchangeable random variables (Cannings 1974; Möhle 2010), the expected values $E(v_{ki}^l)$ are all the same for $l = 1, \ldots, 2Nu_i$. It then follows from (2) that

$$E(v_{ki}^l) = m_{ki}(1 + o(1)).$$

Hence, $m_{ki}$ is essentially the expected number of copies that each gene of subpopulation $k$ passes on to $i$.

Combining (1) and (2), we find that the requirement of constant subpopulation sizes amounts to assuming

$$\sum_{k=1}^{s} u_k m_{ki} = u_i \tag{3}$$

for all $i = 1, \ldots, s$. This implies that $\boldsymbol{u} = (u_1, \ldots, u_s)$ is a left eigenvector of the migration matrix $\boldsymbol{M} = (m_{ki})$ with eigenvalue 1. The corresponding right eigenvector $\boldsymbol{v} = (v_1, \ldots, v_s)$ with eigenvalue 1 contains the long term reproductive values. If these are normalized so that

$$\sum_{i=1}^{s} u_i v_i = 1, \tag{4}$$

$v_i$ can be interpreted as the average number of descendants of the genes of subpopulation $i$ at a fixed but large number of generations later, see for instance Fisher (1958) and Caswell (2001).

Reversing time, the probability that the parent of a gene of subpopulation $i$ originates from subpopulation $k$ is of order

$$b_{ik} = \frac{u_k m_{ki}}{u_i}. \tag{5}$$

Because of (3), it is easy to see that $\boldsymbol{B} = (b_{ik})$ is the transition matrix of a Markov chain with row sums 1. This Markov chain is assumed to be irreducible and aperiodic, with an asymptotic distribution $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s)$ that is the unique probability vector satisfying $\boldsymbol{\gamma} = \boldsymbol{\gamma B}$. In general $\boldsymbol{\gamma}$ differs from $\boldsymbol{u}$, although

$$\boldsymbol{\gamma} = \boldsymbol{u} \tag{6}$$

holds for conservative migration models, when the number of immigrants to each subpopulation equals the number of emigrants from it (Nagylaki 1980). It is easy to see that

$$\gamma_i = u_i v_i, \quad i = 1, \ldots, s, \tag{7}$$

since

$$\sum_{i=1}^{s}(u_i v_i)b_{ik} = \sum_{i=1}^{s} u_i v_i \frac{u_k m_{ki}}{u_i} = u_k \sum_{i=1}^{s} m_{ki} v_i = u_k v_k$$

for $k = 1, \ldots, s$, using (5) in the first equality and that $\boldsymbol{v}$ is a right eigenvector of $\boldsymbol{M}$ with eigenvalue 1 in the fourth equality.

The global migration rate

$$m' = 1 - \sum_{i=1}^{s} u_i m_{ii} = 1 - \sum_{i=1}^{s} u_i b_{ii} \tag{8}$$

is defined as the fraction of offspring genes of the whole population in each generation that do not remain in the same subpopulation as their parents, but rather migrate to other subpopulations.

## 3 Effective population size, fixation index and quasi equilibrium

To start with, we assume that the gene is biallelic, with alleles 0 and 1, and let $P_{ti}$ denote the fraction of one of the two alleles (say 1) in subpopulation $i$ at generation $t$. We let

$$P_t^w = \sum_{i=1}^{s} w_i P_{ti}$$

refer to the total frequency of allele 1 in the whole population in generation $t$ when subpopulations are assigned non-negative weights as determined by $\boldsymbol{w} = (w_1, \ldots, w_s)$, with $\sum_{i=1}^{s} w_i = 1$. Apart from these restrictions, $\boldsymbol{w}$ is arbitrary. When $\boldsymbol{w} = \boldsymbol{u}$, we get the usual definition of allele frequency, with subpopulations weighted proportionally to their sizes and hence all genes in the population are given equal weights. When $\boldsymbol{w} = \boldsymbol{\gamma}$, genes are weighted proportionally to their long term reproductive values. We will refer to these weights as canonical and write

$$P_t = P_t^{\boldsymbol{\gamma}} = \sum_{i=1}^{s} \gamma_i P_{ti} = \boldsymbol{\gamma} \boldsymbol{P}_t, \tag{9}$$

with the convention of dropping superscript $\boldsymbol{w}$, for any quantity, whenever $\boldsymbol{w} = \boldsymbol{\gamma}$.

In order to describe the time dynamics of the allele frequencies, we notice that $P_{t+1,i}$ would equal $\sum_{k=1}^{s} b_{ik} P_{tk}$ if individuals of subpopulation $i$ received genes in exact proportions $b_{i1}, \ldots, b_{is}$ from the subpopulations of the parental generation, and in addition the allele frequencies of genetic material transmitted from subpopulation $k$ to $i$ was identical to $P_{tk}$, the allele frequency of the parental subpopulation $k$. However, the subpopulation proportions of parental origin as well as transmitted allele frequencies

from the various subpopulations will vary randomly, and hence $P_{t+1,i}$ will exhibit some random fluctuation $\varepsilon_{t+1,i}$ around $\sum_{k=1}^{s} b_{ik} P_{tk}$. This can be expressed as a recursion

$$\boldsymbol{P}_{t+1} = \boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1} \tag{10}$$

for the vector $\boldsymbol{P}_t = (P_{t1}, \ldots, P_{ts})^T$ of subpopulation allele frequencies, where $T$ is the Hermitian (conjugate transpose) operator of a matrix and $\boldsymbol{\varepsilon}_{t+1} = (\varepsilon_{t+1,1}, \ldots, \varepsilon_{t+1,s})^T$ is a random error vector, assumed to satisfy

$$E(\boldsymbol{\varepsilon}_{t+1}|\boldsymbol{P}_t) = 0,$$

corresponding to a selectively neutral allele, with covariance matrix

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{t+1}|\boldsymbol{P}_t) = \boldsymbol{\Omega}(\boldsymbol{P}_t), \tag{11}$$

the form of which will depend on the reproduction model. Hence $\boldsymbol{P}_t$ is a multivariate autoregressive time series (Brockwell and Davis 1987) with some degree of heteroscedasticity, since $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ depends on $\boldsymbol{P}_t$.

The two quantities of main interest are the variance effective size of the total (global) population

$$N^w_{eV,t} = \frac{P^w_t(1-P^w_t)}{2E\left((P^w_{t+1}-P^w_t)^2|P^w_t\right)} \overset{w=\gamma}{=} N_{eV,t}, \tag{12}$$

and a version

$$F^w_{ST,t} = \frac{\sum_{i=1}^{s} w_i (P_{ti}-P^w_t)^2}{P^w_t(1-P^w_t)} \overset{w=\gamma}{=} F_{ST,t} \tag{13}$$

of the fixation index. The numerator of (13) is a variance that describes variation in allele frequencies among subpopulations, when these are assigned weights $w_i$, and the denominator normalizes this variance to a number between 0 (identical allele frequencies) and 1 (complete isolation). When subpopulations are weighted equally ($w_i = 1/s$), this gives the usual definition of the fixation index, see for instance Wright (1951), equation 3.12.3 of Crow and Kimura (1970) and equation (12.13) of Nei and Kumar (2000). More general weighting schemes are treated, for instance, by Nei (1977).

The denominator of (12) is the variance of the change in the allele frequency of the total populations from one generation ($t$) to the next ($t+1$), when subpopulations are assigned weights $w_i$. For equal weights ($w_i = 1/s$) this gives the usual definition of the variance effective size, see for instance equation 7.6.3.25 of Crow and Kimura (1970) or the introductory section of Waples (1989).

Ewens (1982) noted that the variance effective size is difficult to analyze for structured populations. The reason is that $N^w_{eV,t}$, and also $F^w_{ST,t}$, are functions of $\boldsymbol{P}_t$, and therefore exhibit some random variation. We will approach this difficulty by studying

**Table 2** Mathematical notation

| Object | Operation | Definition |
|---|---|---|
| $\boldsymbol{x} = (x_i)^T$ | $\lvert\boldsymbol{x}\rvert_1$ | $\sum_i \lvert x_i\rvert$ |
| | $\lvert\boldsymbol{x}\rvert_\infty$ | $\max_i \lvert x_i\rvert$ |
| | $\boldsymbol{x}^0$ | $\boldsymbol{x} - \boldsymbol{\gamma x 1}$ (when $\boldsymbol{x}$ is $s \times 1$) |
| $\boldsymbol{C} = (C_{ij})$ | $\lvert\boldsymbol{C}\rvert_1$ | $\sum_{i,j} \lvert C_{ij}\rvert$ |
| | $\lvert\boldsymbol{C}\rvert_\infty$ | $\max_{i,j} \lvert C_{ij}\rvert$ |
| | $\lVert\boldsymbol{C}\rVert$ | $\max_i \sum_j \lvert C_{ij}\rvert = \sup_{x\neq 0} \lvert\boldsymbol{Cx}\rvert_\infty / \lvert\boldsymbol{x}\rvert_\infty$ |
| | $\boldsymbol{C}^0$ | $\boldsymbol{C} - \boldsymbol{1\gamma}$ (when $\boldsymbol{C}$ is $s \times s$) |
| | | or Kronecker product of such matrices |
| $\boldsymbol{\varepsilon} = (\varepsilon_i)^T$ | $E(\boldsymbol{\varepsilon})$ | $(E(\varepsilon_i))^T$ |
| | $\mathrm{Cov}(\boldsymbol{\varepsilon})$ | $(\mathrm{Cov}(\varepsilon_i, \varepsilon_j))$ |
| $x = x_N$ | $o(x)$ | $o(x)/x \to 0$ as $N \to \infty$ |
| | $O(x)$ | $O(x)/x$ stays bounded as $N \to \infty$ |

the long run behaviour of the expected values of these two quantities, given that none of the two alleles is fixed in all subpopulations. We write non-fixation as

$$\boldsymbol{P}_t \notin \mathcal{A} = \{\boldsymbol{0} = (0, \ldots, 0), \boldsymbol{1} = (1, \ldots, 1)\}, \tag{14}$$

where $\mathcal{A}$ is the set of two absorbing states. Conditionally on (14), the distributions of $N^w_{eV,t}$ and $F^w_{ST,t}$ will converge to quasi equilibrium distributions, and in particular their expected values converge to limits

$$\begin{aligned} F^{\mathrm{eq},w}_{ST} &= \lim_{t\to\infty} E_c(F^w_{ST,t}), \\ N^{\mathrm{eq},w}_{eV} &= \lim_{t\to\infty} E_c(N^w_{eV,t}), \end{aligned} \tag{15}$$

where $E_c(X_t) = E\,(X_t | \boldsymbol{P}_t \notin \mathcal{A})$ and index $c$ is short for conditioning on non-fixation. We refer to $F^{\mathrm{eq},w}_{ST}$ and $N^{\mathrm{eq},w}_{eV}$ as quasi equilibrium values of $F^w_{ST,t}$ and $N^w_{eV,t}$, as indicated by superscripts eq.

Nei et al. (1977) calculated Monte Carlo approximations of $F^{\mathrm{eq},w}_{ST}$ for the island model (see their Table 2), when all subpopulations are assigned equal weights $w_i = 1/s$. They concluded that $F^{\mathrm{eq},w}_{ST}$ is well approximated by

$$E\left(F^w_{ST,t}\right)^* = \frac{\sum_{i=1}^s w_i E\left((P_{ti} - P^w_t)^2\right)}{E\left(P^w_t(1 - P^w_t)\right)}, \tag{16}$$

a quantity introduced by Nei (1975). In Sect. 6 we will define approximations of $F^{\mathrm{eq},w}_{ST}$ and $N^{\mathrm{eq},w}_{eV}$, closely related to (16) and

$$E\left(N^w_{eV,t}\right)^* = \frac{E\left(P^w_t(1 - P^w_t)\right)}{2E\left(E\left((P^w_{t+1} - P^w_t)^2 | P^w_t\right)\right)} \tag{17}$$

respectively. Formally, we say that $F_{ST}^{\text{appr},w}$ and $N_{eV}^{\text{appr},w}$ approximate the quasi equilibrium values in (15) with errors

$$\begin{aligned} \left|F_{ST}^{\text{eq},w} - F_{ST}^{\text{appr},w}\right| &= \lim_{t\to\infty} \left|E_c(F_{ST,t}^w) - F_{ST}^{\text{appr},w}\right|, \\ \left|N_{eV}^{\text{eq},w} - N_{eV}^{\text{appr},w}\right| &= \lim_{t\to\infty} \left|E_c(N_{eV,t}^w) - N_{eV}^{\text{appr},w}\right|. \end{aligned} \tag{18}$$

Equation (15) can be understood more generally in terms of the quasi equilibrium distribution of $\boldsymbol{P}_t$ conditionally on non-fixation. Suppose for a moment that $\boldsymbol{P}_t$ can be modeled (more detailed than in (10)) as a Markov chain with state space

$$\{0,\dots,2Nu_1\}(2Nu_1)^{-1} \times \cdots \times \{0,\dots,2Nu_s\}(2Nu_s)^{-1}.$$

This Markov chain has no asymptotic distribution, but rather two absorbing states $\mathcal{A}$. The quasi equilibrium distribution of $\boldsymbol{P}_t|\boldsymbol{P}_t \notin \mathcal{A}$ is related to the left eigenvector of the transition matrix corresponding to the largest non-unit eigenvalue, see for instance Chapter 13 of Karlin (1966) or Chapter 1 of Ewens (2004) when $s = 1$. For a large population, this can also be deduced from a series expansion of a diffusion approximation of the unconditional distribution of $\boldsymbol{P}_t$. The quasi equilibrium distribution is proportional to the eigenfunction corresponding to the largest negative eigenvalue of a Kolmogorov forward partial differential equation, see for instance Kimura (1955, 1964), Chapter 8 in Crow and Kimura (1970) or Section 5.2.2 in Nei (1975) when $s = 1$.

Ethier and Nagylaki (1980) have studied diffusion approximations with two time scales. Nagylaki (1980) applied this theory to structured populations ($s > 1$) with strong migration, so that migration rates in $\boldsymbol{M}$ (or equivalently in $\boldsymbol{B}$) are kept fixed while the population size $N \to \infty$. Then asymptotically, $\boldsymbol{P}_t$ behaves as a one-dimensional diffusion with allele frequency $P_t = P_t^{\boldsymbol{\gamma}}$ in all subpopulations and hence the quasi equilibrium behaviour of $\boldsymbol{P}_t$ is determined by that of $P_t$.

Let $H_t = f(\boldsymbol{P}_t)$ be some genetic variable of interest. Then, conditionally on non-fixation, we can generalize (15) and define the quasi equilibrium limit

$$H^{\text{eq}} = \lim_{t\to\infty} E_c(H_t). \tag{19}$$

A possible choice is the gene diversity of the total ($T$) population

$$H_t = H_{Tt}^w = 2P_t^w(1 - P_t^w), \tag{20}$$

i.e. the probability that two genes have different alleles when drawn from the whole population with replacement, given that the probabilities are $w_1,\dots,w_s$ of choosing the genes from subpopulations $1,\dots,s$. When $w_i = u_i$, this would equal the heterozygosity of the total population under an idealized assumption of Hardy–Weinberg equilibrium. The probability $1 - H_{Tt}^w$ that the two genes have the same allele is referred to as the gene identity in the total population, see for instance Chapter 6 of Nei (1975) for more details.

Slatkin (1981, 1985) and Barton and Slatkin (1986) have studied the quasi equilibrium of two other (multiallelic) functions, the number of subpopulations where a (rare) allele is present, and the average frequency of this allele among the subpopulations where it is present.

For most choices of $f$, the convergence in (19) is extremely slow, unless the population is very small, since the size of the largest non-unit eigenvalue of the transition matrix for the Markov chain $\boldsymbol{P}_t$, is $1 - O(N^{-1})$. However, Crow and Aoki (1984) and Crow (2004) have noted that $F^w_{ST,t}$ is an exceptionally stable function of $\boldsymbol{P}_t$, and the same is true for $N^w_{eV,t}$. The implication is that for migration rates $m'$ such that $Nm' \gg 1$, both $E_c(F^w_{ST,t})$ and $E_c(N^w_{eV,t})$ will convergence more rapidly in (15) than does $\boldsymbol{P}_t$, and attain values close to their asymptotic limits. Indeed, we will show that the rate of convergence is determined by the second largest eigenvalue of the backward migration matrix $\boldsymbol{B}$, which is not close to unity unless $m'$ is small. It is this fact that motivates us to find explicit approximations $F^{\text{appr},w}_{ST}$ and $N^{\text{appr},w}_{eV}$ of these limits, as well as upper bounds for the errors in (18). Due to the rapid convergence of $E_c(F^w_{ST,t})$ and $E_c(N^w_{eV,t})$, the quasi equilibrium distribution of $\boldsymbol{P}_t$ will not enter into our formulas for $F^{\text{appr},w}_{ST}$ and $N^{\text{appr},w}_{eV}$.

## 4 Orthogonal decomposition of allele frequency process

The properties of the vector valued time series (10) of allele frequencies are crucial for our subsequent development. Since $\boldsymbol{B}$ has largest eigenvalue 1, this time series is non-stationary, and therefore no equilibrium solution exists. However, as in Nagylaki (1980), we decompose $\boldsymbol{P}_t$ into a sum

$$\boldsymbol{P}_t = P_t \mathbf{1} + \boldsymbol{P}^0_t \tag{21}$$

of two parts $P_t\mathbf{1}$ and $\boldsymbol{P}^0_t = \boldsymbol{P}_t - P_t\mathbf{1} = (P^0_{t1}, \ldots, P^0_{ts})^T$ that are orthogonal in a sense specified in Appendix A. Analogously we decompose the error term of (10) as $\boldsymbol{\varepsilon}_t = \varepsilon_t\mathbf{1} + \boldsymbol{\varepsilon}^0_t$, with $\varepsilon_t = \sum_{i=1}^s \gamma_i \varepsilon_{ti} = \boldsymbol{\gamma}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}^0_t = \boldsymbol{\varepsilon}_t - \varepsilon_t\mathbf{1}$. As shown in Appendix A, the recursion formula (10) can be reformulated as

$$\begin{aligned} P_{t+1} &= P_t + \varepsilon_{t+1}, \\ \boldsymbol{P}^0_{t+1} &= \boldsymbol{B}^0\boldsymbol{P}^0_t + \boldsymbol{\varepsilon}^0_{t+1}, \end{aligned} \tag{22}$$

with $\boldsymbol{B}^0 = (b^0_{ik})$ obtained by changing the largest eigenvalue $\lambda_1 = 1$ of $\boldsymbol{B}$ from 1 to 0.

The first part of (22) describes the dynamics of the genetic drift of the allele frequency of the whole population when using canonical weights (9). It has no systematic drift, since

$$\begin{aligned} E(P_{t+1}|P_t) &= E(\boldsymbol{\gamma}\boldsymbol{P}_{t+1}|P_t) = E\left(\boldsymbol{\gamma}E(\boldsymbol{P}_{t+1}|\boldsymbol{P}_t)|P_t\right) = E\left(\boldsymbol{\gamma}\boldsymbol{B}\boldsymbol{P}_t|P_t\right) \\ &= E\left(\boldsymbol{\gamma}\boldsymbol{P}_t|P_t\right) = E(P_t|P_t) = P_t, \end{aligned}$$

and therefore $\{P_t\}$ will behave as a non-stationary random walk. In contrast, $\{P^w_t\}$ will typically have some systematic drift when the weight vector $\boldsymbol{w}$ differs from $\boldsymbol{\gamma}$.

The second term of (21) is also non-stationary, and corresponds to local subpopulation allele frequency fluctuations around $P_t$. Its dynamics is governed by the largest eigenvalue of $\boldsymbol{B}^0$, which equals the second largest eigenvalue $\lambda_2$ of $\boldsymbol{B}$. Since $\boldsymbol{B}$ corresponds to an irreducible and aperiodic Markov chain, $|\lambda_2| < 1$, so that $\boldsymbol{P}_t^0$ does not drift away from zero. We will find below that a simple normalization by $(P_t(1-P_t))^{-1/2}$ makes $\boldsymbol{P}_t^0$ close to a quasi stationary process, although $\boldsymbol{P}_t$ itself has not yet reached quasi equilibrium. Hence it is crucial to distinguish between quasi equilibrium of $\boldsymbol{P}_t$, with slow convergence, and quasi equilibrium of $\boldsymbol{P}_t^0/\sqrt{P_t(1-P_t)}$, with a more rapid convergence. It is the latter phenomenon that makes the convergence in (15) quite fast and enables us to find a closed form procedure for computing $N_{eV}^{\text{appr},w}$ and $F_{ST}^{\text{appr},w}$ that doesn't require the full quasi equilibrium distribution of $\boldsymbol{P}_t$.

Interestingly, the decomposition (21) has some resemblance with cointegration of time series in econometrics (Granger 1981; Engle and Granger 1987). In our case there are $s$ non-stationary time series $P_{ti}$, one for each subpopulation. If they are combined by means of a linear combination, the resulting cointegrated time series is close to stationary provided the vector of weights is orthogonal to $\mathbf{1}$ and that the above mentioned normalization $(P_t(1-P_t))^{-1/2}$ is applied.

## 5 Standardized genetic drift and spatial covariance matrices

It will be convenient to introduce the quadratic matrix $\boldsymbol{H}_t = (H_{tij})_{i,j=1}^s$ of gene diversities

$$H_{tij} = P_{ti}(1-P_{tj}) + P_{tj}(1-P_{ti}) \tag{23}$$

of all pairs of subpopulations $i$ and $j$ at time $t$, defined as the probability that two genes have different alleles when chosen randomly with replacement from subpopulations $i$ and $j$. We notice for instance that the gene diversity

$$H_{Tt}^w = \sum_{i,j} w_i w_j H_{tij} = \boldsymbol{W}_T \operatorname{vec}(\boldsymbol{H}_t) \tag{24}$$

of the total population in (20) is a weighted average of all $H_{tij}$. In the last step of (24), we introduced vec, the vectorization operator that converts a quadratic matrix of order $s$ into a column vector of length $s^2$ by stacking the columns on top of each other, and $\boldsymbol{W}_T = \operatorname{vec}((w_i w_j)_{i,j})^T$, a row vector of length $s^2$ containing probabilities of all pairs of subpopulations when two genes are drawn from the *T*otal population.

The following result reveals that the covariance matrix (21) is a quadratic function of the allele frequencies if and only if the gene diversities form a multivariate autoregressive process:

**Proposition 1** *Suppose there exist constants $U_{ij,kl}$ such that the entries of the covariance matrix $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ in (11) satisfy*

$$\Omega(\boldsymbol{P}_t)_{ij} = \frac{1}{2} \sum_{1 \le k,l \le s} U_{ij,kl} H_{tkl}, \tag{25}$$

*where $U_{ij,kl} = U_{ij,lk}$ can be assumed, without loss of generality. Then (25) is equivalent to the $s^2$ gene diversities $\{H_{tij};\ 1 \le i, j \le s\}$ forming a multivariate autoregressive process, so that*

$$E(H_{t+1,ij}|\boldsymbol{P}_t) = \sum_{1\le k,l\le s} A_{ij,kl} H_{tkl} \tag{26}$$

*for other constants $A_{ij,kl}$, related to $U_{ij,kl}$ through*

$$A_{ij,kl} = b_{ik}b_{jl} - U_{ij,kl}. \tag{27}$$

*Moreover, (25) holds if and only if $\boldsymbol{P} \to \Omega(\boldsymbol{P})$ is a quadratic and symmetric function of the allele frequency vector $\boldsymbol{P}$, which vanishes at the boundary of fixation, i.e.*

$$\begin{aligned} &\boldsymbol{\Omega}(\boldsymbol{0}) = \boldsymbol{0}, \\ &\boldsymbol{\Omega}(\boldsymbol{P}) = \boldsymbol{\Omega}(\boldsymbol{1} - \boldsymbol{P}), \textit{ for all } \boldsymbol{P}. \end{aligned} \tag{28}$$

In Sect. 11, it will be useful to rewrite (26) in matrix form as

$$E\ (\mathrm{vec}(\boldsymbol{H}_{t+1})|\boldsymbol{P}_t) = \boldsymbol{A}\mathrm{vec}(\boldsymbol{H}_t), \tag{29}$$

where $\boldsymbol{A} = (A_{ij,kl};\ 1 \le i, j, k, l \le s)$ is a quadratic matrix of order $s^2$, with $ij$ short hand notation for the row number $(j-1)s + i$, which equals the column number that the vectorization operator assigns to the $i,\ j$:th element of a quadratic matrix of order $s$. Similarly, the column number $(l-1)s + k$ of $\boldsymbol{A}$ is abbreviated as $kl$.

In the sequel, we will assume that (25) holds, and it is shown in Table 3 that all reproduction scenarios of Sect. 7 satisfy this equation. Decomposition (21) can be used to rewrite (25) in a more compact form

$$\mathrm{vec}(\boldsymbol{\Omega}(\boldsymbol{P}_t)) = \boldsymbol{U}\underline{\boldsymbol{1}}P_t(1-P_t) - \boldsymbol{U}\mathrm{vec}(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T) + (1-2P_t)\boldsymbol{U}\mathrm{vec}(\boldsymbol{P}_t^0\boldsymbol{1}^T). \tag{30}$$

where $\underline{\boldsymbol{1}}$ and $\boldsymbol{1}$ are column vectors of lengths $s^2$ and $s$. We will see below that an implication of (30) is that $\boldsymbol{\varepsilon}_{t+1}/\sqrt{P_t(1-P_t)}$ and $\boldsymbol{P}_t^0/\sqrt{P_t(1-P_t)}$ will converge *jointly*

**Table 3** Values of covariance coefficients $U_{ij,kl}$ in (30) for Reproduction scenarios 1–3

| Scenario | $U_{ij,kl}$ |
|---|---|
| 1 | $1_{\{k=l\}}\left(\frac{1_{\{i=j\}}m_{ki}u_k}{2Nu_i^2} + \frac{u_k^2 m_{ki}m_{kj}}{u_i u_j}\left(\frac{1}{2N_{ek}} - \frac{1}{2Nu_k}\right)\right)$ |
| 2 | $1_{\{k=l\}}\frac{\mathrm{Cov}(\nu_{ki}^l, \nu_{kj}^l)u_k}{2Nu_i u_j}$ |
| 3 | $1_{\{k=l\}}\left(1_{\{i=j\}}\left(\frac{b_{ik}^2}{2Nu_i} - \frac{b_{ik}(1-b_{ik})}{\alpha_i+1}\right)\right.$ |
| | $\left. + b_{ik}b_{jk}\left(\frac{1}{2N_{ek}} - \frac{1}{2Nu_k}\right)\right)$ |
| | $+ 1_{\{k\neq l\}}1_{\{i=j\}}b_{ik}b_{il}\left(\frac{1}{2Nu_i} + \frac{1}{\alpha_i+1}\right)$ |

to a quasi equilibrium distribution. The two most important quantities of this quasi equilibrium distribution are the *standardized genetic drift covariance matrix*

$$\boldsymbol{\Sigma}_t = (\Sigma_{tij}) = \frac{E_c(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T)|P_t)}{P_t(1-P_t)} \tag{31}$$

and the *standardized spatial covariance matrix*

$$\boldsymbol{V}_t = (V_{tij}) = \frac{E_c(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|P_t)}{P_t(1-P_t)} \tag{32}$$

at time $t$. The word spatial is used since the diagonal entries $V_{tii}$ of $\boldsymbol{V}_t$ quantify how much the allele frequencies vary in space, i.e. among subpopulations. Moreover, $V_{tij}$ is closely related to a conditional kinship coefficient between subpopulations $i$ and $j$, see Hardy and Vekemans (1999) and Hössjer (2013).

For any index pair $i$ and $j$, $V_{tij}$ and $\Sigma_{tij}$ are functions of $P_t$, and hence stochastic processes whose quasi equilibrium behaviour is of fundamental importance in order to understand the quasi equilibrium behaviour of $N_{eV,t}^w$ and $F_{ST,t}^w$. We will approximate $V_{tij}$ and $\Sigma_{tij}$ by $V_{ij}$ and $\Sigma_{ij}$ for all pairs of subpopulations $i, j$. The corresponding matrices $\boldsymbol{V} = (V_{ij})$ and $\boldsymbol{\Sigma} = (\Sigma_{ij})$ are defined very generally, as a solution of the linear system of equations

$$\begin{aligned} \text{vec}(\boldsymbol{\Sigma}) &= \boldsymbol{U}\underline{\mathbf{1}} - \boldsymbol{U}\text{vec}(\boldsymbol{V}), \\ \text{vec}(\boldsymbol{V}) &= \boldsymbol{\Pi}\text{vec}(\boldsymbol{\Sigma}) + \boldsymbol{G}^0\text{vec}(\boldsymbol{V}), \end{aligned} \tag{33}$$

with $s(s+1)$ unknown parameters in $\text{vec}(\boldsymbol{\Sigma})$ and $\text{vec}(\boldsymbol{V})$. The matrix $\boldsymbol{\Pi} = (\Pi_{ij,kl})$ has elements

$$\Pi_{ij,kl} = 1_{\{(k,l)=(i,j)\}} - \gamma_k 1_{\{j=l\}} - \gamma_l 1_{\{i=k\}} + \gamma_k\gamma_l, \tag{34}$$

and $\boldsymbol{\Pi}\text{vec}(\boldsymbol{\Sigma})$ projects the covariance matrix $\boldsymbol{\Sigma}$, so that global changes along $\mathbf{1}$, which is common to all subpopulations, is removed, whereas variation among subpopulations is retained. The other matrix $\boldsymbol{G}^0 = \boldsymbol{B}^0 \otimes \boldsymbol{B}^0 = (G_{ij,kl}^0)$ is the Kronecker product of $\boldsymbol{B}^0$ with itself, having elements

$$G_{ij,kl}^0 = b_{ik}^0 b_{jl}^0. \tag{35}$$

A crucial point is that (33) can be solved explicitly once all migration and reproduction parameters of the model have been defined, since $\boldsymbol{U}$, $\boldsymbol{\Pi}$ and $\boldsymbol{G}^0$ are all functions of them. In Sect. 6 it is explained in detail how (33) is used for computing $N_{eV}^{\text{appr},w}$ and $F_{ST}^{\text{appr},w}$, and Appendix C motivates further how (33) is attained.

As discussed in Sects. 3 and 4, when migration is faster than genetic drift, $\boldsymbol{\varepsilon}_{t+1}/\sqrt{P_t(1-P_t)}$ and $\boldsymbol{P}_t^0/\sqrt{P_t(1-P_t)}$ will converge fast to a quasi equilibrium distribution, whereas the convergence of $\boldsymbol{P}_t$ is slow. For this reason, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ will *not* involve the quasi equilibrium distribution of $\boldsymbol{P}_t$. Some of the most important properties

of these two matrices are summarized in the following two propositions, whose proofs can be found in Appendix B (see Table 2 for vector and matrix norm notation):

**Proposition 2** *The solution of the system of equations (33), which approximates the standardized genetic drift and spatial covariance matrices* $\mathbf{\Sigma}_t$ *and* $\boldsymbol{V}_t$*, can be written as*

$$\begin{aligned} vec(\mathbf{\Sigma}) &= \boldsymbol{U}\underline{\mathbf{1}} - \boldsymbol{U}\sum_{\tau=0}^{\infty}(\boldsymbol{G}^0 - \mathbf{\Pi}\boldsymbol{U})^{\tau}\,\mathbf{\Pi}\boldsymbol{U}\underline{\mathbf{1}}, \\ vec(\boldsymbol{V}) &= \sum_{\tau=0}^{\infty}(\boldsymbol{G}^0 - \mathbf{\Pi}\boldsymbol{U})^{\tau}\,\mathbf{\Pi}\boldsymbol{U}\underline{\mathbf{1}}, \end{aligned} \tag{36}$$

*provided that the series converges. The maximal standardized amount of local allele frequency fluctuations,*

$$|\boldsymbol{V}_t|_{\infty} = \max_{1\le i\le s}\frac{E_c\left((P_{ti}^0 - P_t)^2|P_t\right)}{P_t(1-P_t)} \tag{37}$$

*can be upper bounded, according to approximation (36), as*

$$|\boldsymbol{V}|_{\infty} \le Mixtime\,\|\mathbf{\Pi}\|\,|\boldsymbol{U}\underline{\mathbf{1}}|_{\infty}, \tag{38}$$

*where* $\|\mathbf{\Pi}\| \le 4$ *and*

$$Mixtime = \sum_{\tau=0}^{\infty}\|(\boldsymbol{G} - \mathbf{\Pi}\boldsymbol{U})^{\tau}\|.$$

*The maximal amount of standardized genetic drift among all subpopulations,*

$$|\mathbf{\Sigma}_t|_{\infty} = \max_{1\le i\le s}\frac{E_c((\varepsilon_{t+1,i}^0)^2|P_t)}{P_t(1-P_t)}, \tag{39}$$

*can be upper bounded, according to approximation (36), as*

$$|\mathbf{\Sigma}|_{\infty} \le (1 + \|\boldsymbol{U}\|Mixtime\|\mathbf{\Pi}\|)\,|\boldsymbol{U}\underline{\mathbf{1}}|_{\infty}. \tag{40}$$

Mixtime can be interpreted as an upper bound for the time it takes for the subpopulations to mix. Intuitively, a higher migration rate $m'$ implies a lower Mixtime and hence a smaller amount of allele frequency fluctuation among subpopulations, as quantified by $|\boldsymbol{V}|_{\infty}$.

The following proposition shows that Mixtime is strongly related to the second largest eigenvalue $\lambda_2$ of $\boldsymbol{B}$:

**Proposition 3** *It holds that*

$$Mixtime \leq \frac{\sum_{\tau=0}^{\infty} \|(\boldsymbol{B}^0)^\tau\|^2}{1 - \|\boldsymbol{\Pi}\|\|\boldsymbol{U}\| \sum_{\tau=0}^{\infty} \|(\boldsymbol{B}^0)^\tau\|^2}. \tag{41}$$

*In addition, for each* $0 < \epsilon < 1 - |\lambda_2|$ *there exists a constant* $C = C(\epsilon)$ *such that*

$$\|(\boldsymbol{B}^0)^\tau\| \leq C(|\lambda_2| + \epsilon)^\tau \tag{42}$$

*for all* $\tau = 0, 1, \ldots$, *so that*

$$Mixtime \leq \frac{\frac{C^2}{1-(|\lambda_2|+\epsilon)^2}}{1 - \|\boldsymbol{\Pi}\|\|\boldsymbol{U}\| \frac{C^2}{1-(|\lambda_2|+\epsilon)^2}}. \tag{43}$$

It follows from Proposition 3 that the closer $\lambda_2$ is to 1, the slower is the mixing rate of subpopulations, and the larger is Mixtime.

In turns out that the accuracy of $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ as approximations of $\boldsymbol{V}_t$ and $\boldsymbol{\Sigma}_t$ depends not only on Mixtime, but also on three remainder terms, the first of which is the $s \times 1$ column vector

$$\boldsymbol{\mu}_t = (\mu_{ti}) = \frac{E_c(\boldsymbol{P}_0|P_t)}{P_t(1 - P_t)}. \tag{44}$$

This term will vanish whenever allele frequencies and migration patterns between subpopulations exhibit spatial symmetry, such as for the island and circular stepping stone models treated in Sect. 8.1. The second remainder term, the $s \times s$ matrix $\boldsymbol{\xi}_t$, is defined in the recursive equation (117) for the standardized spatial covariance matrix $\boldsymbol{V}_t$. The third term $\boldsymbol{\zeta}_t$ is caused by ascertainment bias when conditioning on non-fixation. It is also an $s \times s$ matrix, as defined in (119). With a slight abuse of notation compared to (19), we then write

$$|\boldsymbol{\mu}|^{\text{eq}} = \lim_{t\to\infty} E_c(|\boldsymbol{\mu}_t|_\infty), \tag{45}$$

$$|\boldsymbol{\xi}|^{\text{eq}} = \lim_{t\to\infty} E_c(|\boldsymbol{\xi}_t|_\infty) \tag{46}$$

and

$$|\boldsymbol{\zeta}|^{\text{eq}} = \lim_{t\to\infty} E_c(|\boldsymbol{\zeta}_t|_\infty) \tag{47}$$

respectively, for the corresponding quasi equilibrium limits of the max norms of all three remainder terms. This enables us to formulate the following result, which is proved in Appendix C.

**Theorem 1** *The asymptotic quasi equilibrium size of the approximation errors* $\Delta\boldsymbol{V}_t = \boldsymbol{V}_t - \boldsymbol{V}$ *and* $\Delta\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}$ *of the standardized genetic drift and spatial covariance matrices, have upper bounds*

$$|\Delta \boldsymbol{V}|^{\mathrm{eq}} := \lim_{t\to\infty} E_c(|\Delta \boldsymbol{V}_t|_\infty) \leq \mathit{Mixtime}\left(\|\boldsymbol{\Pi}\|(|\boldsymbol{\zeta}|^{\mathrm{eq}} + \|\boldsymbol{U}\||\boldsymbol{\mu}|^{\mathrm{eq}}) + |\boldsymbol{\xi}|^{\mathrm{eq}}\right) \tag{48}$$

*and*

$$|\Delta \boldsymbol{\Sigma}|^{\mathrm{eq}} := \lim_{t\to\infty} E_c(|\Delta \boldsymbol{\Sigma}_t|_\infty) \leq (|\boldsymbol{\zeta}|^{\mathrm{eq}} + \|\boldsymbol{U}\||\boldsymbol{\mu}|^{\mathrm{eq}}) + \|\boldsymbol{U}\|\mathit{Mixtime}\left(\|\boldsymbol{\Pi}\|(|\boldsymbol{\zeta}|^{\mathrm{eq}} + \|\boldsymbol{U}\||\boldsymbol{\mu}|^{\mathrm{eq}}) + |\boldsymbol{\xi}|^{\mathrm{eq}}\right) \tag{49}$$

*respectively.*

## 6 Defining and computing approximations of $N_{eV}^{\mathbf{eq},w}$ and $F_{ST}^{\mathbf{eq},w}$

We can use the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ in (33) to compute approximations

$$N_{eV}^{\mathrm{appr},w} = \frac{1-(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}(\boldsymbol{w}-\boldsymbol{\gamma})^T}{2\left(\boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{V}(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T + \boldsymbol{w}\boldsymbol{\Sigma}\boldsymbol{w}^T\right)} \overset{\boldsymbol{w}=\boldsymbol{\gamma}}{=} N_{eV}^{\mathrm{appr}} = \frac{1}{2\boldsymbol{\gamma}\boldsymbol{\Sigma}\boldsymbol{\gamma}^T}, \tag{50}$$

and

$$F_{ST}^{\mathrm{appr},w} = \frac{\sum_{i=1}^s w_i V_{ii} - (\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}(\boldsymbol{w}-\boldsymbol{\gamma})^T}{1-(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}(\boldsymbol{w}-\boldsymbol{\gamma})^T} \overset{\boldsymbol{w}=\boldsymbol{\gamma}}{=} F_{ST}^{\mathrm{appr}} = \sum_{i=1}^s \gamma_i V_{ii} \tag{51}$$

of the quasi equilibrium effective population size and fixation index, with $\boldsymbol{I}$ the identity matrix of order $s$.

It follows from Lemma 1 in Appendix E that $E(F_{ST,t}^w)^*$ and $E(N_{eV,t}^w)^*$ in (16) and (17) are closely related to (50) and (51), with $\boldsymbol{V}_t$ and $\boldsymbol{\Sigma}_t$ instead of $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$.

The accuracy of (50) and (51), in the sense of (18), will be established in Sect. 9 in a multilocus setup. Here, we will rather discuss the interpretation and computation of (50)–(51).

We notice that both $N_{eV}^{\mathrm{appr},w}$ and $F_{ST}^{\mathrm{appr},w}$ depend on the subpopulation weights $w_i$, and they simplify substantially when the canonical weights (9) are used. This agrees well with Waples and Yokota (2007), who concluded, by means of simulations, that the variance effective population size for age-structured models is much more stable for these weights.

Since $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ are symmetric matrices, there is a redundancy of parameters in (33). For this reason, we will rewrite this equation as

$$\begin{aligned}\mathrm{vech}(\boldsymbol{\Sigma}) &= \boldsymbol{U}^h\underline{\boldsymbol{1}}^h - \boldsymbol{U}^h\mathrm{vech}(\boldsymbol{V}),\\ \mathrm{vech}(\boldsymbol{V}) &= \boldsymbol{\Pi}^h\mathrm{vech}(\boldsymbol{\Sigma}) + (\boldsymbol{G}^0)^h\mathrm{vech}(\boldsymbol{V}),\end{aligned} \tag{52}$$

where vech is the half vectorization operator, that converts the diagonal and subdiagonal elements of a quadratic matrix of order $s$ to a column vector of length $s(s+1)/2$. Moreover, $\underline{\boldsymbol{1}}^h$ is a column vector of $s(s+1)/2$ ones, and for any matrix $\boldsymbol{U} = (U_{ij,kl})$

of order $s^2$, $\boldsymbol{U}^h = (U^h_{ij,kl};\ 1 \le j \le i \le s, 1 \le l \le k \le s)$ is a quadratic matrix of order $s(s+1)/2$ obtained when changing from vec to vech format. Its elements are defined as

$$U^h_{ij,kl} = \begin{cases} U_{ij,kk}, & 1 \le j \le i \le s, k = l, \\ U_{ij,kl} + U_{ij,lk} & 1 \le j \le i \le s, l < k. \end{cases}$$

Based on (52), we get the following algorithm for computing $N_{eV}^{\text{appr},w}$ and $F_{ST}^{\text{appr},w}$:

1. Specify the total population size $N$, weight vector $\boldsymbol{w}$ and spatial structure, i.e. number of subpopulations $s$ and migration matrix $\boldsymbol{M}$.
2. Make sure that $\boldsymbol{M}$ is a migration matrix by computing $\boldsymbol{u}$ as the left eigenvector of $\boldsymbol{M}$ corresponding to the largest eigenvalue $\lambda_1$, normalized so that $\sum_{i=1}^s u_i = 1$. If $\lambda_1 \neq 1$, replace $\boldsymbol{M}$ by $\boldsymbol{M}/\lambda_1$ (cf. (3)).
3. Compute the transition matrix $\boldsymbol{B}$ of the backward Markov chain from (5).
4. Evaluate the equilibrium distribution $\boldsymbol{\gamma}$ of $\boldsymbol{B}$ and compute $\boldsymbol{\Pi}$ from (34).
5. Evaluate $\boldsymbol{B}^0$ from the Jordan decomposition of $\boldsymbol{B}$, as shown in Appendix A (cf. (106)) and then compute $\boldsymbol{G}^0$ from (35).
6. Select reproduction scenario, compute $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ and find $\boldsymbol{U}$ from (25).
7. Find $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ by solving (52).
8. Evaluate $N_{eV}^{\text{appr},w}$ and $F_{ST}^{\text{appr},w}$ from (50) and (51).

The only input parameters of this algorithm are the population size $N$, the weight vector $\boldsymbol{w}$, the number of subpopulations $s$ and the migration matrix $\boldsymbol{M}$ of step 1 and then the reproduction scenario parameters of step 6.

In the following two sections, we will illustrate the usefulness of (50) and (51) for several reproduction and migration models.

## 7 Reproduction scenarios

We will consider three reproduction scenarios. The first and third one of these generalize those considered by Hössjer et al. (2013) for the island model.

**Reproduction scenario 1** (*Fertilization precedes migration*) In order to describe reproduction from generation $t$ to $t+1$, we assume that an infinitely large gamete pool is formed for each subpopulation of generation $t$. Only a subset of $N_{ek} \le Nu_k$ breeders in subpopulation $k$ contribute with the same (large) number of gametes to the gamete pool. Here $N_{ek}$ represents the local effective size of subpopulation $k$. Gamete formation is followed by fertilization, when the gametes within each pool pair up and form offspring, which finally migrate to various subpopulations. Mathematically, we describe allele frequency change from one generation to the next by means of the three step procedure

$$\begin{aligned} \tilde{P}_{tk} | P_{tk} &\sim \text{Hyp}(2Nu_k, 2N_{ek}, P_{tk})/(2N_{ek}), \text{ gamete formation}, \\ P^*_{tki} | \tilde{P}_{tk} &\sim \text{Bin}(2Nu_k m_{ki}, \tilde{P}_{tk})/(2Nu_k m_{ki}), \text{ fertilization}, \\ P_{t+1,i} &= \sum_{k=1}^s b_{ik} P^*_{tki}, \text{ migration}, \end{aligned} \tag{53}$$

where $\tilde{P}_{tk}$ is the allele frequency of gamete pool $k$ of generation $t$, obtained by drawing genes of the breeders according to a hypergeometric distribution. In the second step $P^*_{tki}$ is the allele frequency of the $Nu_k m_{ki}$ offspring that are formed in subpopulation $k$ and then migrate to subpopulation $i$. This is achieved by drawing $2Nu_k m_{ki}$ genes binomially from gamete pool $k$. Notice that the offspring of various subpopulations migrate in exact proportions, so that $i$ receives exactly a fraction $b_{ik}$ of its genes from subpopulation $k$.

It is shown in Appendix D that the genetic drift covariance matrix of this reproduction scenario has entries

$$\begin{aligned}\boldsymbol{\Omega}(\boldsymbol{P}_t)_{ij} = 1_{\{i=j\}} \sum_{k=1}^{s} \frac{m_{ki} u_k}{2N u_i^2} P_{tk}(1 - P_{tk}) \\ + \sum_{k=1}^{s} \frac{u_k^2 m_{ki} m_{kj}}{u_i u_j} \left( \frac{1}{2N_{ek}} - \frac{1}{2Nu_k} \right) P_{tk}(1 - P_{tk}),\end{aligned} \tag{54}$$

a special case of (25), with $U_{ij,kl}$ as in Table 3. □

**Reproduction scenario 2** (*Exact migration proportions*) We can generalize reproduction scenario 1, still requiring that (2) holds exactly, so that $2Nu_k m_{ki}$ copies of genes of subpopulation $k$ end up in subpopulation $i$ in the next generation. However, we will not specify in detail how this is achieved in terms of breeders and gamete pools, and therefore replace (53) by the more general

$$\begin{aligned} P^*_{tki} | P_{tk} &= \sum_{l=1}^{2Nu_k P_{tk}} \nu^l_{ki} / (2Nu_k m_{ki}), \\ P_{t+1,i} &= \sum_{k=1}^{s} b_{ik} P^*_{tki}. \end{aligned} \tag{55}$$

In the first part, the allele frequency $P^*_{tki}$ of the migrant genes from subpopulation $k$ to $i$ is determined by the total number of copies $\nu^l_{ki}$ of the $2Nu_k P_{tk}$ genes $l$ that have the specified allele. Without loss of generality we can assume that these genes are numbered as $l = 1, \ldots, 2Nu_k P_{tk}$, if $\{\boldsymbol{\nu}^l_k = (\nu^l_{k1}, \ldots, \nu^l_{ks})\}_{l=1}^{2Nu_k}$ is a collection of exchangeable random vectors for each subpopulation $k$.

It turns out that the only further properties of $\nu^l_{ki}$ needed for computing the genetic drift covariance matrix $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ is independence of $\{\boldsymbol{\nu}^l_k\}_{l=1}^{2Nu_k}$ for different $k$ and the covariances $\mathrm{Cov}(\nu^l_{ki}, \nu^l_{kj})$, which do not depend on $l$ because of the exchangeability assumption. Indeed, it is shown in Appendix D that the entries of $\boldsymbol{\Omega}(\boldsymbol{P}_t)$ are

$$\boldsymbol{\Omega}(\boldsymbol{P}_t)_{ij} = \sum_{k=1}^{s} \frac{\mathrm{Cov}(\nu^l_{ki}, \nu^l_{kj}) u_k}{2Nu_i u_j} P_{tk}(1 - P_{tk}). \tag{56}$$

This equation is of form (25), with $U_{ij,kl}$ as in Table 3. We notice in particular that (54) is a special case of (56), with

$$\mathrm{Cov}(\nu_{ki}^l, \nu_{kj}^l) \sim 1_{\{i=j\}} m_{ki} + m_{ki} m_{kj} \left( \frac{N u_k}{N_{ek}} - 1 \right).$$

□

**Reproduction scenario 3** (*Migration precedes fertilization*) For pollination of plants, it is of interest to reverse the order of migration and reproduction, so that gametes rather than individuals migrate. We then divide the reproduction cycle into three steps;

$$\begin{aligned} \tilde{P}_{tk} | P_{kt} &\sim \mathrm{Hyp}(2Nu_k, 2N_{ek}, P_{kt})/(2N_{ek}), \text{ gamete formation}, \\ \check{P}_{ti} &= \sum_{k=1}^{s} B_{ik} \tilde{P}_{tk}, \text{ migration}, \\ P_{t+1,i} | \check{P}_{ti} &\sim \mathrm{Bin}(2Nu_i, \check{P}_{ti})/(2Nu_i), \text{ fertilization}, \end{aligned} \tag{57}$$

where $\tilde{P}_{tk}$ is the allele frequency of gamete pool $k$, as before. Then gametes mix, so that $\check{P}_{ti}$ is the allele frequency of gamete pool $i$ after migration, and $B_{ik}$ the fraction of its alleles that originates from gamete pool $k$, so that $\sum_{k=1}^{s} B_{ik} = 1$. Finally, reproduction in subpopulation $i$ occurs by drawing $2Nu_i$ genes binomially from post-migration gamete pool $i$.

When $B_{ik} = b_{ik}$ the gamete pools mix in exact proportions. More generally, we can allow for random mixing, where the rows

$$(B_{i1}, \ldots, B_{is}) \sim \mathrm{Dir}\,(\alpha_i (b_{i1}, \ldots, b_{is})) \tag{58}$$

have independent Dirichlet distributions for $i = 1, \ldots, s$, so that $E(B_{ik}) = b_{ik}$ and

$$\mathrm{Cov}(B_{ik}, B_{il}) = \begin{cases} b_{ik}(1 - b_{ik})/(\alpha_i + 1), & k = l, \\ -b_{ik} b_{il}/(\alpha_i + 1), & k \neq l. \end{cases} \tag{59}$$

The parameter $\alpha_i \geq 0$ quantifies the amount of random variability of the mixing proportions when the contents of gamete pools from various subpopulations $k$ migrate to subpopulation $i$. When $\alpha_i = 0$, all parents of subpopulation $i$ are selected from the same randomly chosen subpopulation, with probabilities $b_{ik}$, whereas $\alpha_i = \infty$ gives exact mixing proportions $B_{ik} = b_{ik}$. The latter reproduction scheme has been studied by Nagylaki (1980) when $N_{ek} = Nu_k$, but with additional steps of selection and mutation added. See also Sampson (2006) for a related reproduction scenario in which subpopulation sizes are allowed to vary rapidly as well.

Notice that this model is not a special case of exact migration proportions, since the remainder term of (2) is not zero, although it is asymptotically negligible when $N$ and $\alpha_i$ are *both* large. However, exact mixing proportions alone ($\alpha_i = \infty$) are not sufficient for a vanishing remainder term in (2).

It is shown in Appendix D that the genetic drift covariance matrix has entries

$$\begin{aligned}\boldsymbol{\Omega}(\boldsymbol{P}_t)_{ij} = &\frac{1_{\{i=j\}}}{2Nu_i}(\boldsymbol{B}\boldsymbol{P}_t)_i(1-(\boldsymbol{B}\boldsymbol{P}_t)_i)\\ &+\sum_{k=1}^{s} b_{ik}b_{jk}\left(\frac{1}{2N_{ek}}-\frac{1}{2Nu_k}\right)P_{tk}(1-P_{tk})\\ &+\frac{1_{\{i=j\}}}{\alpha_i+1}\sum_{k=1}^{s} b_{ik}\,(P_{tk}-(\boldsymbol{B}\boldsymbol{P}_t)_i)^2\\ &+O\left(\frac{1}{N(1+\min\{\alpha_i\})}\right),\end{aligned} \tag{60}$$

where the remainder term is asymptotically negligible when $\min(\alpha_1,\ldots,\alpha_s)\to\infty$ as $N\to\infty$. It can be shown that (60) agrees with (25), with $U_{ij,kl}$ as in Table 3. □

## 8 Spatial structures

The reproductive fitness of subpopulation $k$ is defined as the average number of offspring

$$r_k=\sum_{i=1}^{s} m_{ki}$$

of its members in the *next* generation. Since it only involves the subpopulation it is a purely demographic parameter, not reflecting genetic differences among subpopulations. It should not be confused with $v_k$ in (4), another demographic parameter that quantifies average number descendants *many* generations ahead.

It follows from (3) that the average reproductive fitness of all populations, weighted according to their sizes, is

$$\sum_{k=1}^{s} u_k r_k=\sum_{k=1}^{s} u_k\sum_{i=1}^{s} m_{ki}=\sum_{i=1}^{s}\sum_{k=1}^{s} u_k m_{ki}=\sum_{i=1}^{s} u_i=1, \tag{61}$$

as a consequence of the constant total population size.

### 8.1 Conservative migration

Conservative migration (6) is equivalent to all subpopulations being equally productive. Indeed, if $r_k$ does not depend on $k$, it follows from (61) that $r_k=1$, so that the row sums of $\boldsymbol{M}$ are 1. But then $\boldsymbol{v}=(1,\ldots,1)$ is a right eigenvector of $\boldsymbol{M}$, and therefore (7) implies (6), or equivalently, that $\boldsymbol{u}$ is a left eigenvector of $\boldsymbol{B}$ with eigenvalue 1, When subpopulations are weighted proportionally to their census sizes, i.e. $\boldsymbol{w}=\boldsymbol{u}$, it therefore follows from (50) and (51) that

$$N_{eV}^{\text{appr}} = \frac{1}{2\boldsymbol{u}\boldsymbol{\Sigma}\boldsymbol{u}^T} \tag{62}$$

and

$$F_{ST}^{\text{appr}} = \sum_{i=1}^{s} u_i V_{ii}. \tag{63}$$

We can simplify (62) and (63) further for reproduction scenario 1, if all local effective sizes equal the local census sizes ($N_{ek} = Nu_k$). After some computations, it then follows from upper part of (33), (54) and Table 3 that

$$\Sigma_{ii} = \frac{1}{2Nu_i}\left(1 - \sum_{k=1}^{s} b_{ik} V_{kk}\right).$$

Inserting this expression into (62) and using (6), we obtain an approximation

$$N_{eV}^{\text{appr}} = \frac{1}{2\sum_{i=1} u_i^2 \Sigma_{ii}} = \frac{N}{1 - \sum_{k=1}^{s} u_k V_{kk}} = \frac{N}{1 - F_{ST}^{\text{appr}}} \tag{64}$$

of the variance effective size of the total population. This well known formula was originally derived by Wright (1951) for the island model. Wang and Caballero (1999, eqn. 15) showed the validity of (64) more generally when fertilization precedes migration.

For reproduction scenario 3, it is shown in Appendix D that a similar calculation when $N_{ek} = Nu_k$ yields

$$N_{eV}^{\text{appr}} = \frac{N}{1 - \sum_{i=1}^{s} u_i (\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T)_{ii} + \sum_{i=1}^{s} \frac{2Nu_i^2}{1+\alpha_i}\left(\sum_{k=1}^{s} b_{ik} V_{kk} - (\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T)_{ii}\right)}. \tag{65}$$

However, neither (64) nor (65) hold when $N_{ek} \neq Nu_k$, since then $\boldsymbol{\Sigma}$ is no longer diagonal.

**Spatial structure 1** (*Island model*) The most well known population genetic model with spatial structure is the island model (Wright 1943; Maruyama 1970b). All subpopulations (or islands) have equal size, $u_k = 1/s$, and migration is symmetric, so that a fraction $m$ of the offspring select island uniformly (including its present island), and the remaining fraction $1 - m$ of offspring never migrate. This corresponds to

$$\boldsymbol{M} = \boldsymbol{B} = (1 - m)\boldsymbol{I} + \frac{m}{s}\boldsymbol{1}\boldsymbol{1}^T. \tag{66}$$

It is seen in Appendix D that the second largest eigenvalue of this matrix is $\lambda_2 = 1 - m = 1 - m'/(1 - 1/s)$, cf. Table 4, so that a smaller migration rate implies that $\lambda_2$ is closer to 1.

In general it is not possible to find explicit expressions for $N_{eV}^{\text{appr}}$ and $F_{ST}^{\text{appr}}$. However, due its symmetry, the island model is an exception, as we now illustrate for reproduction scenarios 1 and 3:

**Table 4** Values of the total migration rate $m'$ and Average Relative Neighbourhood Proportion (ARNP) for the island and various stepping stone models

| Model | $m'$ | ARNP |
|---|---|---|
| Island | $m(s-1)/s$ | 1 |
| Torus stepping stone | $m$ | $(4-1_{\{\lvert s_1\rvert=2\}}-1_{\{\lvert s_2\rvert=2\}})/(s-1)$ |
| Rectangular stepping stone | $m(1-0.5(s_1^{-1}+s_2^{-1}))$ | $(4-2(s_1^{-1}+s_2^{-1}))/(s-1)$ |
| Circular stepping stone | $m$ | $2/s$ |
| Linear stepping stone | $m(s-1)/s$ | $\min(1, 2/(s-1))$ |
| Genetic reservoir | $(s-1)\,(u_1 r_1(\gamma+\delta)+u_s r_s \beta)$ | $u_1 1_{\{\beta>0\}}+(1-u_1)\left(u_1 1_{\{\gamma>0\}}+2u_s 1_{\{\delta>0\}}\right)/(1-u_s)$ |

We first consider reproduction scenario 1 with a constant local effective population size $N_{ek} = N_e$ for all islands. It is shown in Appendix D that (50) and (51) then simplify to

$$N_{eV}^{\text{appr}} = \frac{s N_e}{1 - F_{ST}^{\text{appr}}} \tag{67}$$

and

$$F_{ST}^{\text{appr}} = \frac{1}{\frac{s}{s-1} 2\tilde{N}\left(1-(1-m)^2\right)+1} \tag{68}$$

respectively, with

$$\frac{1}{\tilde{N}} = \frac{(1-m)^2}{N_e} + \frac{2m-m^2}{N/s}$$

a weighted harmonic average of $N_e$ and $N/s$. Formula (68) was derived in Hössjer et al. (2013) by other methods, exploiting that $F_{ST,t}$ can be well approximated by a univariate autoregressive time series for the island model. Related expressions for $F_{ST}^{\text{eq}}$ can also be found in Wright (1943), Nei (1975), Takahata (1983), Crow and Aoki (1984), Takahata and Nei (1984) and Ryman and Leimar (2008).

For reproduction scenario 3, with the local effective population sizes $N_{ek} = N_e$ and the Dirichlet parameters $\alpha_k = \alpha$ is the same for all subpopulations $k$, the corresponding formulas

$$N_{eV}^{\text{appr}} = \frac{s N_e}{1 - F_{ST}^{\text{appr}} + \left(\frac{N_e}{N/s} + \frac{2N_e}{\alpha+1}\right)\left(1-(1-m)^2\right) F_{ST}^{\text{appr}}} \tag{69}$$

and

$$F_{ST}^{\text{appr}} = \frac{1}{(\frac{s}{s-1} - \frac{1}{\alpha+1}) 2\tilde{N}\left(1-(1-m)^2\right) + \frac{\tilde{N}}{N_e}(1-m)^2} \tag{70}$$

are valid when $\alpha$ is large or $N_e = N/s$. Formula (69) is, to the best of our knowledge, new, and (70) extends a formula of Hössjer et al. (2013) to $\alpha \neq 0$. A smaller $\alpha$ implies that the proportion of migrants that each subpopulation receives from the other subpopulations, varies more. As a consequence, $N_{eV}^{\text{appr}}$ decreases and $F_{ST}^{\text{appr}}$ increases. □

**Spatial structure 2** (*Stepping stone models*) In natural populations, migration is often restricted to neighbouring subpopulations. Kimura (1953) proposed a class of stepping stone models that reflect this behaviour. Its first mathematical treatment appeared in Kimura and Weiss (1964) and Weiss and Kimura (1965) and a more recent one in Durrett (2008). When $s \geq 3$, the one-dimensional circular version of the stepping stone model has all subpopulations distributed on a circle with

$$m_{ki} = \begin{cases} 1-m, & i = k, \\ m/2, & i = k+1 \text{ or } k-1 \text{ modulo } s, \\ 0, & \text{otherwise.} \end{cases} \tag{71}$$

When $s = 2$, the nonzero offdiagonal elements are instead $m_{12} = m_{21} = m$, since $k + 1 = k - 1$ modulo 2. Maruyama (1970a) considered the eigenvalue effective population size $N_{eE}^{\text{eq}}$ when either $m \to 0$ or $N \to \infty$. Based on his asymptotic results, Wang and Caballero (1999) suggested the approximation

$$N_{eE}^{\text{appr}} = sN_e + \frac{s^2}{2m\pi^2}, \tag{72}$$

for all values of $m$, when $N_{ek} = N_e = N/s$ and $s$ is an even integer.

A linear stepping stone model can be defined when $s \geq 3$ that differs from (71) in that no direct communication occurs between the end populations 1 and $s$, i.e. $m_{1s} = m_{s1} = 0$. All other off-diagonal elements of $\boldsymbol{M}$ are given by (71), whereas the diagonal elements satisfy $m_{11} = m_{ss} = 1 - m/2$ and $m_{ii} = 1 - m$ for $2 \leq i \leq s - 1$.

A two-dimensional stepping stone model has $s = s_1 s_2$ subpopulations $\{i = (i_1, i_2);\ 1 \leq i_1 \leq s_1, 1 \leq i_2 \leq s_2\}$ positioned on a rectangular grid, with $s_1 \geq 2$ and $s_2 \geq 2$. In order to avoid edge effects we identify islands along the left edge as neighbours to those along the right edge, provided their second coordinates agree. Similarly, we identify islands along the bottom and top edges as neighbours whenever their first coordinates agree. Consequently, the islands can be thought of as uniformly positioned along a torus with migration intensities

$$m_{(k_1,k_2),(i_1,i_2)} = \begin{cases} 1-m, & (i_1, i_2) = (k_1, k_2), \\ m/4, & i_2 = k_2 \text{ and } i_1 = k_1 + 1 \text{ or } k_1 - 1 \text{ modulo } s_1, \\ m/4, & i_1 = k_1 \text{ and } i_2 = k_2 + 1 \text{ or } k_2 - 1 \text{ modulo } s_2, \\ 0, & \text{otherwise,} \end{cases} \tag{73}$$

when $s_1 \geq 3$ and $s_2 \geq 3$. When $s_1$ and/or $s_2$ equals 2, the off-diagonal entries $m_{(k_1,k_2),(i_1,i_2)}$ are added for those $(i_1, i_2)$ that correspond to the same subpopulation in (73).

When $s_1 \geq 3$ and $s_2 \geq 3$, a rectangular version without edge effects is defined by putting those off-diagonal elements in $\boldsymbol{M}$ to zero that correspond to transitions between opposite edges. Hence, the diagonal elements are $1 - m$ for interior subpopulations, $1 - 3m/4$ along edges and $1 - m/2$ at corners.

It is easily verified that $\boldsymbol{B} = \boldsymbol{M}$ and $\boldsymbol{u} = \boldsymbol{\gamma} = \mathbf{1}^T/s$ for all four stepping stone models. □

## 8.2 Varying reproductive fitness

In general the single generation reproductive fitness $r_k$ varies between subpopulations. The migration matrix may then be expressed as

$$\boldsymbol{M} = \boldsymbol{RF},$$

with $\boldsymbol{R} = \text{diag}(r_1, \ldots, r_s)$ a diagonal matrix of fitness coefficients, $\boldsymbol{F} = (f_{ki})_{k,i=1}^{s}$ the transition matrix of a forward Markov chain and $f_{ki}$ the probability that a copy of a subpopulation $k$ gene that is passed on to the next generation ends up in subpopulation $i$.

**Spatial structure 3** (*Demographic reservoir*) Suppose $s \geq 2$ and that one subpopulation $s$ acts as a demographic reservoir (or source population) for the other subpopulations $1, \ldots, s-1$. The reservoir is located in the center of a circle and the other subpopulations symmetrically along its perimeter. We assume that the reproductive fitness coefficients $r_1 = \cdots = r_{s-1}$ and sizes $u_1 = \cdots = u_{s-1}$ of the perimeter populations are identical, where

$$0 < r_1 < 1 < r_s \tag{74}$$

quantifies how much more productive the reservoir is compared to the other subpopulations. Since the relative subpopulation sizes $u_i$ sum to one, it follows from (61) that

$$u_1 = \frac{1}{s-1}\frac{r_s - 1}{r_s - r_1} \quad \text{and} \quad u_s = \frac{1 - r_1}{r_s - r_1}.$$

When $s \geq 4$, the entries of the forward transition matrix are specified as

$$f_{ki} = \begin{cases} \beta, & k = s, i = 1, \ldots, s-1, \\ 1 - (s-1)\beta, & k = i = s, \\ \gamma, & k = 1, \ldots, s-1, i = s, \\ \delta/2, & k = 1, \ldots, s-1, i = k-1 \text{ or } k+1 \text{ modulo } s-1, \\ 1 - \gamma - \delta, & k = i = 1, \ldots, s-1, \\ 0, & \text{otherwise.} \end{cases}$$

When $s = 3$, the migration rates for $i = k - 1$ and $i = k + 1$ modulo $s - 1$ are added together, since they correspond to the same subpopulation $i$. When $s = 2$, we put $\delta = 0$. The perimeter islands act as a circular stepping stone model with internal probability $\delta$ that any gene migrates to a neighbouring perimeter island. However, there is also communication with the reservoir and the other two parameters $\gamma$ and $\beta$ control the degree of gene flow between the reservoir and the perimeter populations respectively. It turns out that in order for (3) to hold, $\beta$ and $\gamma$ must be related as

$$(s-1)u_s r_s \beta = u_s(r_s - 1) + (s-1)u_1 r_1 \gamma,$$

where the left hand side is the total migration rate from the reservoir to the perimeter, the first term on the right hand side is the migration rate from the reservoir needed to keep its population size constant in absence of immigration, and the last term is the total migration rate from the perimeter to the reservoir.

Hence there are four parameters; $r_1$, $r_s$, $\delta$ and $\gamma$ that can be varied in this model subject to constraints (74), $0 \leq \delta \leq 1$ and $0 < \gamma \leq \min(1 - \delta, (1 - r_1)/(r_1(r_s - 1)))$. Given $\delta$, the upper bound on $\gamma$ gives the maximal possible gene flow between the reservoir and perimeter, and it assures that all diagonal elements of $\boldsymbol{F}$ are non-negative. Because of (74), we have non-conservative migration $\boldsymbol{u} \neq \boldsymbol{\gamma}$. □

## 9 Multilocus and multiallelic extension

Assume there are $n$ genetic markers, with the $x$:th marker having alleles $a = 1, \ldots, n(x)$ for $x = 1, \ldots, n$. This generalizes the one locus, biallelic scenario $n = 1$ and $n(1) = 2$ treated so far.

Let $P_{ti}(x, a)$ refer to the frequency of allele $a$ at marker $x$ in subpopulation $i$ and let $P_t^w(x, a) = \sum_{i=1}^s w_i P_{ti}(x, a)$ denote the frequency of the same allele in the whole population when subpopulations are weighted according to $\boldsymbol{w}$.

We argued in Sect. 3 that for most functions $H_t$ of allele frequencies at a single locus, the convergence in (19) towards the equilibrium expected value is very slow. If $H_t = f\left(\{P_{ti}(x, a)\}_{i,x,a}\right)$ is instead a function of allele frequencies at several loci, the convergence in (19) is usually faster, at least if the loci are in linkage equilibrium. On the other hand, we motivated in Sect. 5 that when $H_t$ equals the fixation index or the variance effective population size, the convergence in (19) is faster in spite of the fact that these are single locus functions, with a limit well approximated by quantities that do not include the quasi equilibrium distribution of the allele frequencies. In this section, we will show that the same is true for the multilocus extensions of the fixation index and variance effective size.

For the variance effective population size, we generalize a multilocus and multiallelic expression suggested by Jorde and Ryman (2007) for a single panmictic population ($s = 1$). The idea is to sum separately over loci and alleles in the numerator and denominator, so that

$$N_{eV,t}^w = \frac{\sum_{x=1}^n \sum_{a=1}^{n(x)} P_t^w(x, a)(1 - P_t^w(x, a))}{\sum_{x=1}^n \sum_{a=1}^{n(x)} 2E\left(\left(P_{t+1}^w(x, a) - P_t^w(x, a)\right)^2 | P_t^w(x, a)\right)}. \tag{75}$$

The extension of $F^w_{ST,t}$ to multilocus and multialleles is a bit more involved. The gene diversity in generation $t$ can be written as

$$H^w_{Tt} = \frac{1}{n}\sum_{x=1}^{n}\left(1 - \sum_{a=1}^{n(x)} P_t^w(x,a)^2\right) = \frac{1}{n}\sum_{x=1}^{n}\sum_{a=1}^{n(x)} P_t^w(x,a)(1 - P_t^w(x,a)), \quad (76)$$

which generalizes (20) to multiple loci and alleles. On the other hand, the gene diversity within subpopulations in generation $t$, is

$$H^w_{St} = \sum_{i=1}^{s} w_i \frac{1}{n}\sum_{x=1}^{n}\left(1 - \sum_{a=1}^{n(x)} P_{ti}(x,a)^2\right). \quad (77)$$

We then introduce the coefficient of gene differentiation for weight vector $\boldsymbol{w}$ and time $t$ as

$$G^w_{ST,t} = \frac{H^w_{Tt} - H^w_{St}}{H^w_{Tt}} = \frac{\sum_{x=1}^{n}\sum_{a=1}^{n(x)}\sum_{i=1}^{s} w_i (P_{ti}(x,a) - P_t^w(x,a))^2}{\sum_{x=1}^{n(x)} P_t^w(x,a)(1 - P_t^w(x,a))}. \quad (78)$$

When subpopulations are weighted equally ($w_i = 1/s$), the first equality of (78) gives the original definition of the coefficient of gene differentiation due to Nei (1973). It can be shown that the second equality of (78) implies that $G^w_{ST,t}$ is a weighted average of fixation indeces $F^w_{ST,t}(x,a)$ (cf. (13)) for all loci $x$ and alleles $a$. The weights are proportional to $P_t^w(x,a)(1 - P_t^w(x,a))$, so that alleles close to extinction or fixation have less influence. In the unweighted case $w_i = 1/s$, this has been discussed by several authors, for instance Nei (1975), Wright (1978, Chapter 3) and Chakraborty and Leimar (1987).

We will now use Theorem 1 in order to investigate how accurate the approximations of the quasi equilibrium values of (75) and (78) are. In order to simplify the proof (see Appendix E), we only consider biallelic markers.

**Theorem 2** *Assume $n$ biallelic markers ($n(x) \equiv 2$) in linkage (gametic) equilibrium, so that allele transmission is independent across loci. Let $N_{eV}^{\mathrm{eq},w}$ be the quasi equilibrium limit, defined in the lower equation of (15), based on (75), and analogously put*

$$G_{ST}^{\mathrm{eq},w} = \lim_{t\to\infty} E_c(G^w_{ST,t}).$$

*Define $N_{eV}^{\mathrm{appr},w}$ as in (50) and $G_{ST}^{\mathrm{appr},w}$ by the same formula as for $F_{ST}^{\mathrm{appr},w}$ in (51). Then there exist constants $C'_1, \ldots, C'_9$ such that*

$$\left|G_{ST}^{\mathrm{eq},w} - G_{ST}^{\mathrm{appr},w}\right| \le C'_1|\Delta \boldsymbol{V}|^{\mathrm{eq}} + C'_2|\boldsymbol{\mu}|^{\mathrm{eq}} + \frac{C'_3}{n}. \quad (79)$$

*and*

$$\left|N_{eV}^{\mathrm{eq},w} - N_{eV}^{\mathrm{appr},w}\right| \le C_4'|\Delta \boldsymbol{V}|^{\mathrm{eq}} + C_5'|\Delta \boldsymbol{\Sigma}|^{\mathrm{eq}} + C_6'|\boldsymbol{\mu}|^{\mathrm{eq}} + C_7'|\boldsymbol{\varsigma}|^{\mathrm{eq}} + C_8'|\boldsymbol{\zeta}|^{\mathrm{eq}} + \frac{C_9'}{n}, \tag{80}$$

*with* $|\boldsymbol{\mu}|^{\mathrm{eq}}$ *and* $|\boldsymbol{\zeta}|^{\mathrm{eq}}$ *as defined in (45) and (47),* $|\boldsymbol{\varsigma}|^{\mathrm{eq}}$ *another ascertainment bias term defined in (140) and (148), and upper bounds for* $|\Delta \boldsymbol{V}|^{\mathrm{eq}}$ *and* $|\Delta \boldsymbol{\Sigma}|^{\mathrm{eq}}$ *are provided in Theorem 1.*

Combining Theorems 1 and 2, we notice that the accuracy of the quasi equilibrium approximations (79) and (80) are upper bounded by a linear combination of $|\boldsymbol{\xi}|^{\mathrm{eq}}$ in (46), $|\boldsymbol{\mu}|^{\mathrm{eq}}$, $|\boldsymbol{\varsigma}|^{\mathrm{eq}}$, $|\boldsymbol{\zeta}|^{\mathrm{eq}}$ and $n^{-1}$. Moreover, if follows from the proof in Appendix E that $C_2'$, $C_4'$, $C_6'$ and $C_7'$ all vanish for the canonical weighting scheme $\boldsymbol{w} = \boldsymbol{\gamma}$. The impact of the ascertainment error terms $|\boldsymbol{\varsigma}|^{\mathrm{eq}}$ and $|\boldsymbol{\zeta}|^{\mathrm{eq}}$ should be small, since loci close to fixation of one allele are downweighted.

## 10 Numerical results

We will illustrate how the approximate quasi equilibrium values (50) and (51) depend on various parameters, including the total migration rate $m'$. However, it is also important to know whether migration in one step from each subpopulation is local (reaches neighbouring subpopulations only) or global (reaches all other subpopulations). This can either be assessed using migration conversion factors between the island and other spatial models, see for instance Malécot (1946), Kimura and Weiss (1964) and Crow and Aoki (1982, 1984) . We will use a somewhat different approach, and define the one step neighborhood $\mathcal{N}_k = \{i;\ i \neq k, m_{ki} > 0\}$ of each subpopulation $k$ as well as the Average Relative Neighbourhood Proportion

$$\mathrm{ARNP} = \sum_{k=1}^{s} u_k \frac{\sum_{i \in \mathcal{N}_k} u_i}{1 - u_k},$$

which should not be confused with the neighborhood size $NS$ of Wright (1946), developed for continuous spatial isolation by distance models. Provided migration rates to neighbouring subpopulations are of the same order, a larger ARNP indicates more global migration whenever it takes place. Table 4 lists values of $m'$ and ARNP for the models introduced in Sect. 8. Among the island and stepping stone models, it can be seen that the island model has the most global and the linear stepping stone model the most local migration, with the other stepping stone models in between.

In Fig. 1 we display $N_{eV}^{\mathrm{appr}}$ and $F_{ST}^{\mathrm{appr}}$ as functions of the migration rate $m'$ in (8), the local census size $Nu_k$ and the number of subpopulations $s$ for the island and various stepping stone models when fertilization precedes migration. In general, the larger ARNP is, the smaller are $N_{eV}^{\mathrm{appr}}$ and $F_{ST}^{\mathrm{appr}}$. However, $N_{eV}^{\mathrm{appr}}$ varies very little between the linear stepping stone model (with smallest ARNP) and the island model (with largest ARNP). The fixation index $F_{ST}^{\mathrm{appr}}$ varies somewhat more, at least in relative

terms. See also Figs. 1 and 2 of Crow and Aoki (1984) for similar conclusions how approximations of the fixation index depends on $s_1$ and $s_2$ for the torus stepping stone model. We also notice that although $N_{eV}^{\text{appr}}$ and $F_{ST}^{\text{appr}}$ vary somewhat with $Nu_k$ when the local effective size $N_e$ is kept fixed, this dependence is very small for $N_{eV}^{\text{appr}}$ (see also Hössjer et al. 2013).

When migration precedes fertilization, the impact of the Dirichlet parameter $\alpha_i = \alpha$ in (58) is quite dramatic. Even though subpopulation sizes are kept fixed, a smaller $\alpha$ implies a larger variability of the proportions of migrants from the various subpopulation in the parental generation. It is seen in Fig. 2 that this implies a substantially decreased $N_{eV}^{\text{appr}}$, as well as an increased $F_{ST}^{\text{appr}}$. As in Fig. 1 we also conclude that the migration structure (island or stepping stone) has less effect on $N_{eV}^{\text{appr}}$ and $F_{ST}^{\text{appr}}$.

In Fig. 3, we compare $N_{eV}^{\text{appr}}$ with the approximation $N_{eE}^{\text{appr}}$ of the eigenvalue effective population size in (72), for the circular stepping stone model. The agreement is quite good, although for small migration rates $N_{eE}^{\text{appr}}$ is somewhat larger.

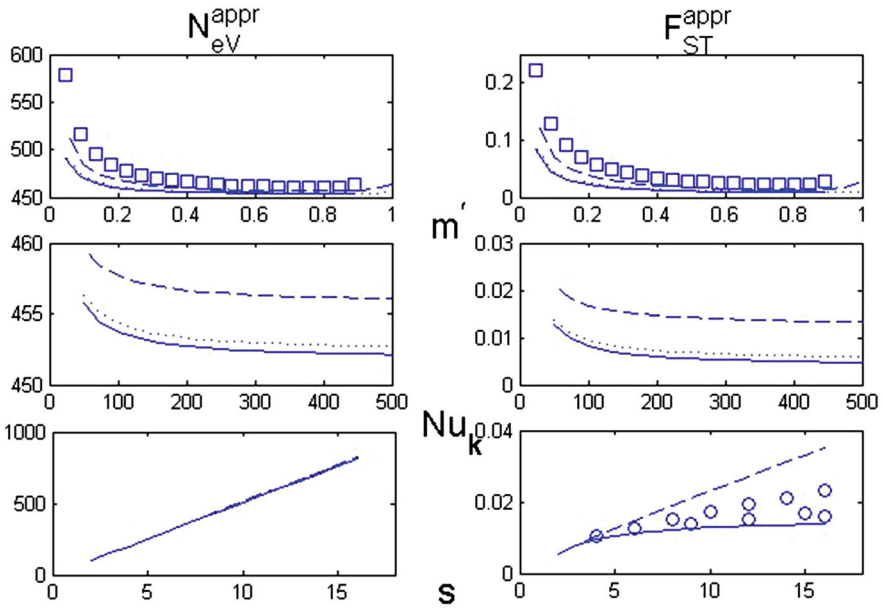



**Fig. 1** Plots of variance effective population size $N_{eV}^{\text{appr}}$ (*left*) and fixation index $F_{ST}^{\text{appr}}$ (*right*) when fertilization precedes migration for the island model (*solid*), circular (*dashed*), linear (*squares*) and torus (*dotted* and *circles*) stepping stone models. *Upper* The migration rate $m'$ in (8) is varied, while the number of subpopulations $s = 9$ (arranged as a $s_1 \times s_2 = 3 \times 3$ grid for the torus model), the total population size $N = 450$, the relative subpopulation sizes $u_k = 1/9$ and the local effective population sizes $N_e = N/9 = 50$ are kept fixed. *Middle* Local census size $Nu_k = N/9$ is varied, while $s = 9$, $m' = 0.4$ and $N_e = 50$. *Lower* $s$ is varied, while $N = 50s$, $u_k = 1/s$, $m' = 0.4$ and $N_e = Nu_k = 50$. For the torus stepping stone model (*circles*), the *upper circles* correspond to $s_1 = 2, \ldots, 8$ and $s_2 = 2$, the *middle circles* to $s_1 = 3, 4, 5$ and $s_2 = 3$, and the *lower right circle* to $s_1 = s_2 = 4$

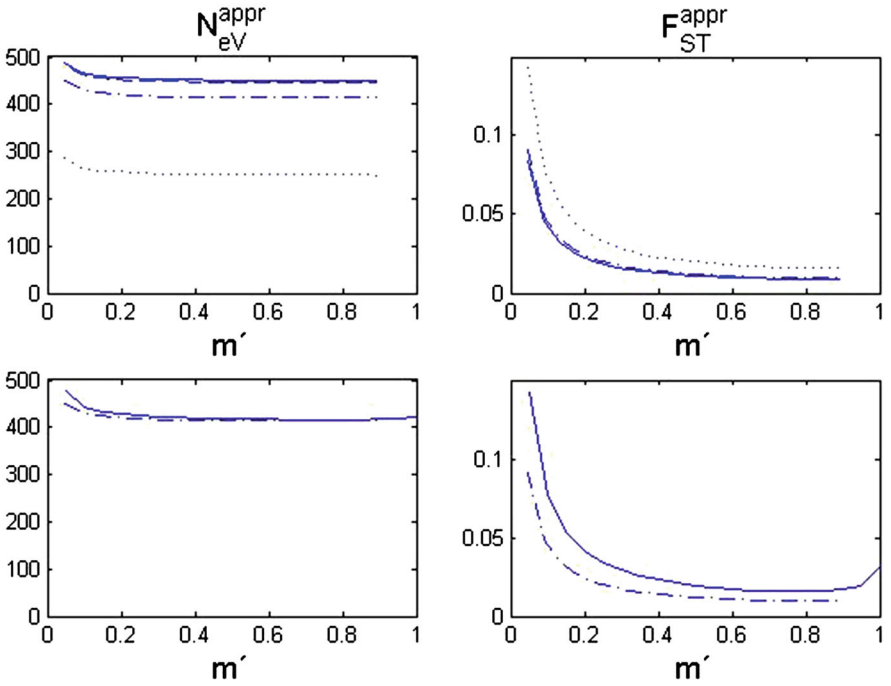



**Fig. 2** Plots of $N_{eV}^{\text{appr}}$ (*left*) and $F_{ST}^{\text{appr}}$ (*right*) as functions of $m'$ when migration precedes fertilization with Dirichlet parameters $\alpha_k = \alpha$ (cf. (58)). For all curves $s = 9$, $N = 450$, $u_k = 1/9$ and $N_e = N/9 = 50$. *Upper* plots for the island model with $\alpha = 1$ (*dotted*), $\alpha = 10$ (*dash-dotted*), $\alpha = 100$ (*dashed*) and $\alpha = \infty$ (*solid*), which corresponds to fixed migrant proportions, as in Wright (1951). *Lower* plots, with $\alpha = 10$, for the island (*dash-dotted*) and circular stepping stone models (*solid*)

The results of Sect. 5 (see Propositions 2, 3 and Theorem 1) reveal that a larger $1 - |\lambda_2|$ implies a smaller Mixtime, a smaller amount of allele fluctuations among subpopulations and hence a more accurate quasi equilibrium approximation. In Fig. 4 it is shown that $1 - |\lambda_2|$ is proportional to the migration rate $m'$ for the island and several stepping stone models. The proportionality constant varies quite a lot between models, so that a larger ARNP implies a faster convergence rate to quasi equilibrium of $\boldsymbol{P}_t^0/\sqrt{P_t(1-P_t)}$.

Figure 5 shows plots of $N_{eV}^{\text{appr},u}$ and $F_{ST}^{\text{appr},u}$ for the demographic reservoir model, where we have indicated the subpopulation weights $\boldsymbol{w} = \boldsymbol{u}$, since in this case they differ from the default choice $\boldsymbol{w} = \boldsymbol{\gamma}$. We notice that $N_{eV}^{\text{appr},u}$ and $F_{ST}^{\text{appr},u}$ are both very sensitive to varying reproductive fitness among subpopulations. The larger the fitness $r_s$ of the reservoir, the smaller are $N_{eV}^{\text{appr},u}$ and $F_{ST}^{\text{appr},u}$. For instance, when $r_s = 20$, the reservoir occupies $u_s = 4.76\%$ of the total population. Still, almost all genes are inherited from the reservoir, so that $N_{eV}^{\text{appr},u}$ approximately equals the census size $Nu_s$ of the reservoir. On the other hand, the migration rate $\delta$ between the perimeter islands and the migration rate $\gamma$ from the perimeter to the reservoir effect $N_{eV}^{\text{appr},u}$ and $F_{ST}^{\text{appr},u}$ very little.

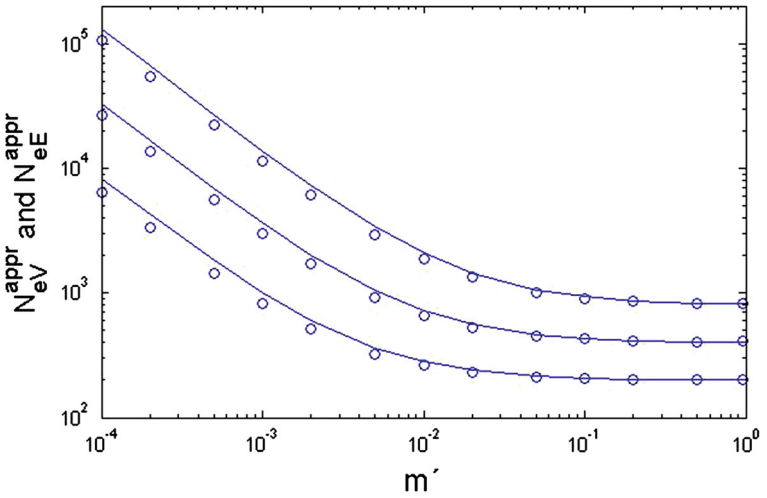



**Fig. 3** Plots of variance and eigenvalue effective population sizes $N_{eV}^{\text{appr}}$ (*dots*) and $N_{eE}^{\text{appr}}$ (*solid lines*, cf. (72)), as functions of the migration rate $m'$ for the circular stepping stone model. The *plots* are on a log–log scale and the number of subpopulations is $s = 16$ (*upper*), $s = 8$ (*middle*) and $s = 4$ (*lower*). Fertilization precedes migration, with population size $N = 50s$. For all subpopulations $k$, $u_k = 1/s$ and $N_e = Nu_k = 50$

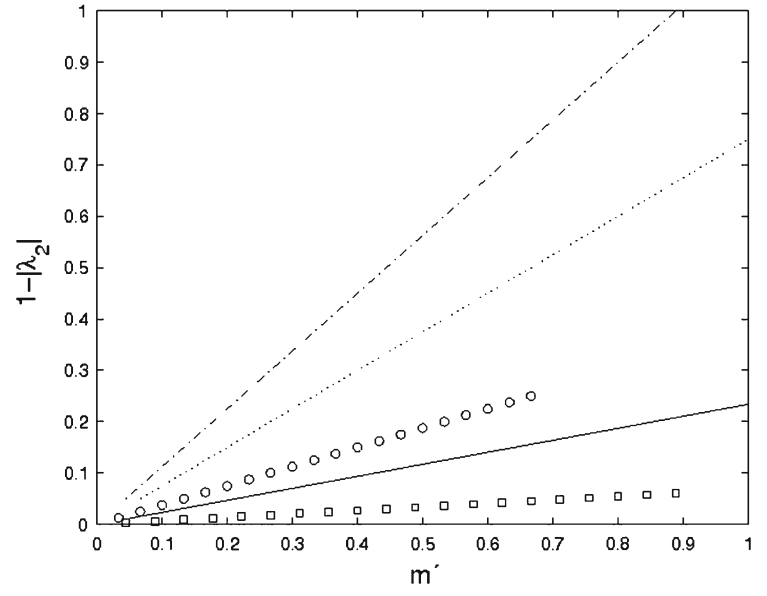



**Fig. 4** Plots of $1 - |\lambda_2|$ as function of the migration rate $m'$ for a number of models with $s = 9$, $N = 450$, $u_k = 1/9$, $N_e = N/9 = 50$ and different Average Relative Neighbourhood Proportions (ARNP): The island model (*dash-dotted*, ARNP $= 1$), the torus stepping stone model (*dotted*, ARNP $= 1/2$), the rectangular stepping stone model (*circles*, ARNP $= 1/3$), the circular stepping stone model (*solid*, ARNP $= 1/4$) and the linear stepping stone model (*squares*, ARNP $= 2/9$). For the two-dimensional (*rectangular* and *torus*) stepping stone models $s_1 = s_2 = 3$

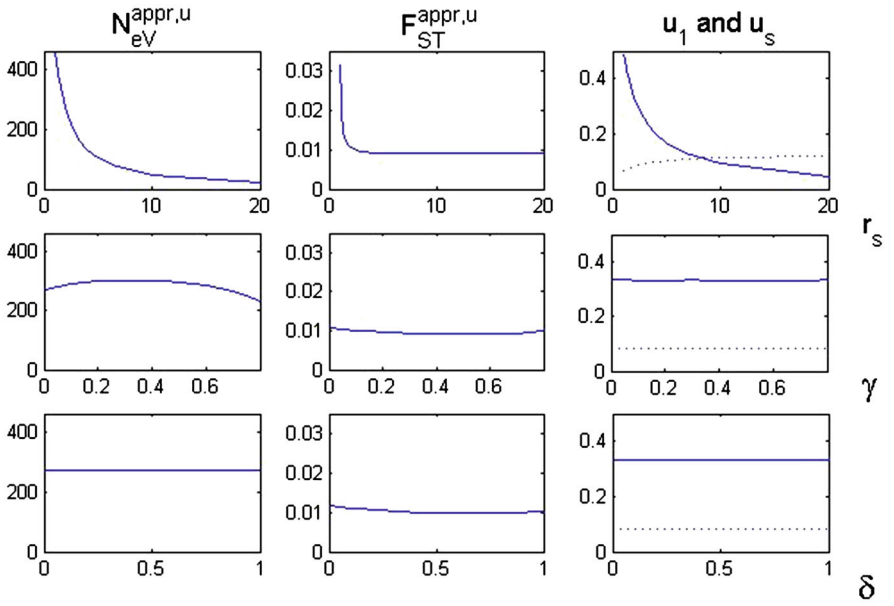



**Fig. 5** Plots, for the demographic reservoir model, of $N_{eV}^{\text{appr},u}$ (*left*), $F_{ST}^{\text{appr},u}$ (*middle*) and relative sizes (*right*) $u_1 = \ldots = u_{s-1}$ for the $s-1$ perimeter populations (*dotted*) and $u_s$ for the reservoir population (*solid*), when fertilization precedes migration, $s = 9$, $N = 450$ and $N_{ek} = Nu_k$ for $k = 1, \ldots, 9$. *Upper* The fitness $r_s$ and $r_1 = \cdots = r_{s-1} = 1/r_s$ of the reservoir and perimeter populations are varied, whereas $\delta = 0.2$ and $\gamma = 0$ are kept fixed. *Middle* migration from perimeter to reservoir, $0 \le \gamma \le 0.8$, is varied, whereas $r_1 = 0.5$, $r_s = 2$ and $\delta = 0.2$ are kept fixed. *Lower* migration rate between perimeter populations, $0 \le \delta \le 1$, is varied, whereas $r_1 = 0.5$, $r_s = 2$ and $\gamma = 0$ are kept fixed

## 11 Coalescent based criteria

### 11.1 Nucleotide diversity effective size

In this section, we compare $N_{eV}^{\text{appr}} = N_{eV}^{\text{appr},\boldsymbol{\gamma}}$ with the nucleotide diversity effective size

$$N_{e\pi} = \frac{E(T)}{2}, \tag{81}$$

where $T$ is the number of generations back to the most recent common ancestor of two different genes chosen randomly from the population, the so called coalescence time. Formula (81) implies that $N_{e\pi}$ is the size of a Wright–Fisher population with the same heterozygosity as the studied population at a site with small ($o(N^{-1})$) mutation probability, see Section 4.4 of Durrett (2008).

We will extend a formula due to Slatkin (1991) for computing $E(T)$ within our general framework of structured populations with arbitrary migration matrices and subpopulation weights. Define $q_{tij}$ as the probability that two different and randomly chosen genes from subpopulations $i$ and $j$ have *not* found their most recent common

ancestor within $t$ generations. We collect all these probabilities into a quadratic matrix $\boldsymbol{q}_t = (q_{tij})$ of order $s$ and show in Appendix F:

**Proposition 4** *The matrices $\boldsymbol{q}_t$ satisfy $\boldsymbol{q}_0 = \mathbf{1}$ and*

$$vec(\boldsymbol{q}_{t+1}) = \boldsymbol{D}vec(\boldsymbol{q}_t), \quad t = 0, 1, 2, \ldots \tag{82}$$

*where $\boldsymbol{D} = (D_{ij,kl})$ a quadratic matrix of order $s^2$, related to the matrix $\boldsymbol{A} = (A_{ij,kl})$ in (27) as*

$$A_{ij,kl} = \frac{\left(1 - \frac{1}{2Nu_i}\right)^{\{i=j\}}}{\left(1 - \frac{1}{2Nu_k}\right)^{\{k=l\}}} D_{ij,kl}. \tag{83}$$

Recursion (82) is equivalent to equation (3) of Slatkin (1991), and it differs from the gene diversity recursion (29) only by replacing the matrix $\boldsymbol{A}$ with $\boldsymbol{D}$. Whereas the elements of $\boldsymbol{A}$ are derived by drawing two genes without replacement, those of $\boldsymbol{D}$ are found by drawing them with replacement. In Proposition 7 of Appendix F, we derive a very explicit expression of $D_{ij,kl}$ in terms of the backward probabilities $b_{ik}$ and $b_{jl}$ that two genes from subpopulations $i$ and $j$ have their parents in $k$ and $l$, and the coalescence probability

$$p_{ijk} = \frac{\sigma_{ijk}(N)}{2Nu_k} + o(N^{-1}) \tag{84}$$

that the two parents are the same *given* that $k = l$. The quantity $\sigma_{ijk}(N)$ in the numerator of (84) is a coalescence rate, i.e. a coalescence probability standardized by the size of the parental subpopulation.

Let $T_{ij}$ denote the coalescence time of two genes sampled from subpopulations $i$ and $j$. It follows from (82) that

$$\begin{aligned} E(T_{ij}) &= \sum_{\tau=0}^{\infty} q_{\tau ij} \\ &= \sum_{\tau=0}^{\infty} \sum_{k,l} D_{ij,kl}^{(\tau)} q_{0kl} \\ &= \sum_{\tau=0}^{\infty} \sum_{k,l} D_{ij,kl}^{(\tau)}, \end{aligned}$$

where $D_{ij,kl}^{(\tau)}$ are the elements of $\boldsymbol{D}^{\tau}$. With $\boldsymbol{T} = (T_{ij})$ a quadratic matrix of order $s$ that contains all coalescence times $T_{ij}$, we can rewrite the last equation as

$$\text{vec}\,(E(\boldsymbol{T})) = \sum_{\tau=0}^{\infty} \boldsymbol{D}^{\tau}\underline{\mathbf{1}} = (\boldsymbol{I} - \boldsymbol{D})^{-1}\underline{\mathbf{1}}, \tag{85}$$

with $\boldsymbol{I}$ the identity matrix of order $s^2$ and $\underline{\mathbf{1}}$ a column vector of length $s^2$. This corresponds to equation (5) of Slatkin (1991).

Suppose two different genes are drawn randomly, by selecting subpopulations independently with probabilities $w_i$, and let $T$ be their coalescence time. Then

$$E(T) = \sum_{i,j=1}^{s} w_i w_j E(T_{ij}) = \boldsymbol{W}_T E(\boldsymbol{T}), \tag{86}$$

with $\boldsymbol{W}_T$ the same vector of weights for pairs of subpopulations as defined below (24). Combining (81), (85) and (86), we find that

$$N_{e\pi} = N_{e\pi}^{w} = \frac{\boldsymbol{W}_T (\boldsymbol{I} - \boldsymbol{D})^{-1} \underline{\mathbf{1}}}{2}. \tag{87}$$

The following result gives the asymptotic relation between the nucleotide diversity and variance effective population sizes (see Appendix F for a proof):

**Theorem 3** *Suppose reproduction scenario 2 or 3 (with $\alpha_i \equiv \infty$) is used and strong migration, i.e. that $N \to \infty$, while the backward migration matrix $\boldsymbol{B}$ is kept fixed. Suppose further that the coalescence rates $\sigma_{ijk}(N)$ in (84) converge to limits $\sigma_{ijk}$ as $N \to \infty$. Then, regardless of the choice of subpopulation weights $\boldsymbol{w}$, it holds that*

$$N_{e\pi}^{w} = \frac{N}{C} + o(N) \text{ as } N \to \infty, \tag{88}$$

*with*

$$C = \sum_{i,j,k=1}^{s} u_k^{-1} \gamma_i \gamma_j b_{ik} b_{jk} \sigma_{ijk}. \tag{89}$$

*Moreover, if $\boldsymbol{w} = \boldsymbol{\gamma}$, it holds that*

$$N_{eV}^{\text{appr}} = N_{eV}^{\text{appr},\boldsymbol{\gamma}} = \frac{N}{C} + o(N) \text{ as } N \to \infty, \tag{90}$$

*with the same constant $C$ as in (89).*

An important special case is a generalized Wright–Fisher model, i.e. a population in which each subpopulation behaves locally as a Wright–Fisher model before migration. This corresponds to reproduction scenario 3 with $N_{ek} = N u_k$ and $\alpha_i \equiv \infty$, and it follows from Appendix F that $\sigma_{ijk} = 1$. Since $\boldsymbol{\gamma}$ is a left eigenvector of $\boldsymbol{B}$ with eigenvalue 1, this leads to

$$C = \sum_{i,j,k=1}^{s} u_k^{-1} \gamma_i \gamma_j b_{ik} b_{jk} = \sum_{k=1}^{s} u_k^{-1} \gamma_k^2. \tag{91}$$

Inserting (91) into (88), we get an expression for $N_{e\pi}$ due to Nagylaki (1982, 1998). For the models treated in this paper $\sigma_{ijk}$ only depends on the parental subpopulation $k$, whereas for age-structured models, it will also depend on the subpopulations $i$ and $j$ of the offspring (Hössjer 2011).

We see from Theorem 3 that the nucleotide diversity effective size is robust towards choice of subpopulation weights $\boldsymbol{w}$. This is much in contrast to the variance effective size, which depends on the subpopulations weights, even asymptotically. It is only for the canonical weights $\boldsymbol{w} = \boldsymbol{\gamma}$ that $N_{eV}^{\text{appr},w}$ is asymptotically equivalent to $N_{e\pi}$. The reason is that $N_{eV}^{\text{appr},\boldsymbol{\gamma}}$ quantifies the long term genetic changes (see the discussion), and for large populations $N_{e\pi}$ does the same. The coalescence effective size is related to, but more stringent than $N_{e\pi}$, requiring that the whole ancestral tree of an arbitrary but finite number of genes converges to Kingman's coalescent (Kingman 1982) as the population size grows. Hössjer (2011) proved that the coalescence effective size equals (88)–(89), which is not surprising, since it quantifies long term genetic changes in the population (Sjödin et al. 2005).

## 11.2 Fixation index

For a single biallelic locus we can rewrite the fixation index (13) as

$$F_{ST,t}^{w} = \frac{H_{Tt}^{w} - H_{St}^{w}}{H_{Tt}^{w}}, \tag{92}$$

with $H_{Tt}^{w}$ the gene diversity of the total population defined in (24) (or more generally (76)), and

$$H_{St}^{w} = \sum_{i=1}^{s} w_i H_{tii}, \tag{93}$$

the corresponding gene diversity within subpopulations, a special case of (77).

Both (92) and (93) are defined in terms of drawing two genes with replacement. If instead of (23) we draw two *distinct* genes from subpopulation $i$ and $j$ and generation $t$, the probability that they have different alleles,

$$\bar{H}_{tij} = \begin{cases} P_{ti}(1 - P_{tj}) + P_{tj}(1 - P_{ti}), & i \neq j, \\ 2P_{ti}(1 - P_{ti})/(1 - (2Nu_i)^{-1}), & i = j, \end{cases} \tag{94}$$

is denoted with a bar, to signify sampling *without* replacement. This yields a modified version

$$\bar{F}_{ST,t}^{w} = \frac{\bar{H}_{Tt}^{w} - \bar{H}_{St}^{w}}{\bar{H}_{Tt}^{w}} \tag{95}$$

of the fixation index (92), where $\bar{H}^w_{St}$ and $\bar{H}^w_{Tt}$ are the gene diversities of the total population and subpopulations, defined as in (24) and (93), but with $\bar{H}_{tij}$ in place of $H_{tij}$.

Slatkin (1991) investigated a time independent gene diversity $\bar{h}_{ij}$, within the framework of an infinite alleles model (Kimura 1971), where mutations occur with a probability $\mu > 0$ per gene and generation, and each new mutation creates a new allele. If two genes are picked randomly without replacement, they are not IBS if the tree with their most recent common ancestor as root and $2T_{ij}$ edges, contains at least one mutation, with a probability

$$\bar{h}_{ij} = 1 - E\left((1-\mu)^{2T_{ij}}\right) \approx 2\mu E(T_{ij}) \tag{96}$$

if $\mu$ is small. The gene diversities in (96) can easily be estimated from genetic data using the expected number of (selectively neutral) segregating sites between DNA sequences from subpopulations $i$ and $j$ (Durrett 2008, Chapter 1). The corresponding gene diversities

$$\begin{aligned} \bar{h}^w_S &= \sum_{i=1}^{s} w_i \bar{h}_{ii} \approx 2\mu \boldsymbol{W}_S E(\boldsymbol{T}), \\ \bar{h}^w_T &= \sum_{i,j=1}^{s} w_i w_j \bar{h}_{ij} \approx 2\mu \boldsymbol{W}_T E(\boldsymbol{T}), \end{aligned} \tag{97}$$

of the total population and subpopulations are defined analogously to (24) and (93), with $\boldsymbol{W}_T$ defined below (24), and $\boldsymbol{W}_S = \text{vec}\left((w_i 1_{\{i=j\}})^s_{i,j=1}\right)^T$ a weight vector for pairs of genes belonging to the same *S*ubpopulation. Nagylaki (1982, 1998) analyzed properties of $\bar{h}^w_T$ for the canonical weights $\boldsymbol{w} = \boldsymbol{\gamma}$.

Combining (85), (96) and (97), we follow Slatkin (1991) and define the fixation index

$$\bar{f}^w_{ST} := \frac{\bar{h}^w_T - \bar{h}^w_S}{\bar{h}^w_T} \approx \frac{(\boldsymbol{W}_T - \boldsymbol{W}_S)(\boldsymbol{I} - \boldsymbol{D})^{-1}\underline{\mathbf{1}}}{\boldsymbol{W}_T(\boldsymbol{I} - \boldsymbol{D})^{-1}\underline{\mathbf{1}}}, \tag{98}$$

where the approximation is exact in the limit of vanishingly small mutation rates. Wilkinson-Herbots (1998) derived expressions for $\bar{f}^w_{ST}$ that are valid for general structured models, without taking the $\mu \to 0$ limit in (98), using either uniform ($w_i = 1/s$) or population size ($w_i = u_i$) weights.

In analogy with (96)–(98), we define quantities $h_{ij}$, $h^w_T$ and $h^w_S$ by drawing two genes with replacement. The corresponding fixation index satisfies

$$f^w_{ST} := \frac{h^w_T - h^w_S}{h^w_T} \approx \frac{(\boldsymbol{W}_T - \boldsymbol{W}_S)(\boldsymbol{I} - \boldsymbol{A})^{-1}\underline{\mathbf{1}}}{\boldsymbol{W}_T(\boldsymbol{I} - \boldsymbol{A})^{-1}\underline{\mathbf{1}}}, \tag{99}$$

using that the matrices of non-coalescence probabilities and expected coalescence times for two genes drawn without replacement, satisfy recursions similar to (82) and (85), with $\boldsymbol{A}$ in place of $\boldsymbol{D}$.

We found in Sect. 6 that $F_{ST}^{\mathrm{appr},w}$ is closely related to $E(F_{ST,t}^{w})^*$. On the other hand, we prove in Proposition 8 in Appendix F that $f_{ST}^{w}$ is a weighted average of $E_0(F_{ST,t}^{w})^*$ for varying $t$, with weights corresponding to a distribution of $t$ with mean $O(N)$. The quantity $E_0(F_{ST,t}^{w})^*$ is a prediction of $F_{ST,t}^{w}$, given information from the founder generation that all subpopulations have the same allele frequency. We therefore expect $E_0(F_{ST,t}^{w})^*$ to be slightly smaller than $E(F_{ST,t}^{w})^*$ for small $t$, before $F_{ST,t}^{w}$ has attained quasi equilibrium, and the time required for this is $O\left((1-\lambda_2)^{-1}\right)$. The conclusion is that $F_{ST}^{\mathrm{appr},w}$ and $f_{ST}^{w}$ should be close for strong migration, when $N(1-\lambda_2) \gg 1$. This was confirmed by Hössjer (2013) for models with translationally invariant migration (e.g. island and circular stepping stone models), for which $w_i = \gamma_i = 1/s$. The following result holds more generally for canonical weights $\boldsymbol{w} = \boldsymbol{\gamma}$:

**Theorem 4** *Assume strong migration, i.e. that $N \to \infty$ while the migration matrices $\boldsymbol{M}$ and $\boldsymbol{B}$ are kept fixed. If also $\boldsymbol{U} = \boldsymbol{B} \otimes \boldsymbol{B} - \boldsymbol{A} = O(N^{-1})$, the fixation index in (99) satisfies*

$$f_{ST} = f_{ST}^{\boldsymbol{\gamma}} = (\boldsymbol{W}_S - \boldsymbol{W}_T)(\boldsymbol{I} - \boldsymbol{G}^0)^{-1}\boldsymbol{U}\underline{\boldsymbol{1}} + o(N^{-1}), \tag{100}$$

*for the canonical weighting scheme $\boldsymbol{w} = \boldsymbol{\gamma}$, and $F_{ST}^{\mathrm{appr}} = F_{ST}^{\mathrm{appr},\boldsymbol{\gamma}}$ has the same asymptotic expansion. On the other hand, the fixation index (98) has an expansion*

$$\bar{f}_{ST} = \bar{f}_{ST}^{\boldsymbol{\gamma}} = (\boldsymbol{W}_S - \boldsymbol{W}_T)(\boldsymbol{I} - \boldsymbol{G}^0)^{-1}\bar{\boldsymbol{U}}\underline{\boldsymbol{1}} + o(N^{-1}), \tag{101}$$

*that is asymptotically equivalent to a version $\bar{F}_{ST}^{\mathrm{appr}} = \bar{F}_{ST}^{\mathrm{appr},\boldsymbol{\gamma}}$ of $F_{ST}^{\mathrm{appr}}$ in (51), for which $\boldsymbol{V}$ is modified in (36) so that $\bar{\boldsymbol{U}} = \boldsymbol{B} \otimes \boldsymbol{B} - \boldsymbol{D}$ replaces $\boldsymbol{U}$.*

We conjecture that Theorem 4 does not hold for non-canonical weighting schemes, since the denominator of (99) is asymptotically independent of weights (cf. Theorem 3), and therefore the elements of $\boldsymbol{W}_T$ and $\boldsymbol{W}_S$ will enter linearly in $f_{ST}^{w}$, in the limit of large populations. On the other hand, it can be seen from (51) that $F_{ST}^{\mathrm{appr},\boldsymbol{w}}$ is a non-linear function of the elements of $\boldsymbol{W}_T$ and $\boldsymbol{W}_S$, even asymptotically.

It turns out that $f_{ST}$ is always a bit larger than $\bar{f}_{ST}$, although the difference $O(N^{-1})$ is small for large populations, see Proposition 9 in Appendix F for details. Indeed, if two genes that are drawn from subpopulation $i$, the probability that they have different alleles is decreased by replacing the first before drawing the second, so that $h_{ii} < \bar{h}_{ii}$ for all $i$ and hence $h_T - h_S > \bar{h}_T - \bar{h}_S$. whenever at least two subpopulations have positive weights.

## 12 Discussion

We have developed a general methodology for computing the fixation index, the coefficient of gene differentiation and the variance effective size for a large class of populations exhibiting substructure. Conditioning on that no allele becomes fixed in the

population, we decompose the vector valued time series of subpopulation allele frequencies into two parts corresponding to genetic drift of the whole population and genetic differentiation among subpopulations respectively. The latter turns out to be approximately in quasi equilibrium modulo a simple standardization. As a result, approximate quasi equilibrium values of the fixation index and variance effective population size are derived, with greatly simplified formulas when alleles are weighted proportionally to their subpopulations' long term reproductive values. We also perform a detailed comparison between the quasi equilibrium approach and one based on coalescence probabilities and vanishingly small mutation rates.

The numerical illustrations reveal that a few parameters seem to characterize the variance effective size and fixation index quite well, such as number of subpopulations, the reproductive fitness pattern $\boldsymbol{r}$ of subpopulations, random variation of migration proportions, overall migration rate $m'$, local effective size and average relative neighbourhood proportion (ARNP).

We regard alleles as neutral indicators of genetic drift and subpopulation structure, and as such they are only of interest before fixation. Once a mutation at a given locus has occurred, we ignore the possibility of new mutations at the same locus, before quasi equilibrium of $\boldsymbol{P}_t^0/\sqrt{P_t(1-P_t)}$ is attained. We motivated in Sects. 5 and 10 that this occurs at a rate proportional to $m'$, with a proportionality constant depending on how local migration is, as quantified by ARNP. The conclusion is that the mutation rate has to be of smaller order than $m'$. Since mutation rates are thought to be of the order of $10^{-4}$–$10^{-3}$ for highly polymorphic markers such as microsatellites and $10^{-7}$–$10^{-6}$ for allozymes and single nucleotide polymorphisms (Waples and Gaggiotti 2006; Allendorf and Luikart 2007), this puts very mild restrictions on the overall migration rate. For instance, for fish genetic data, the median fixation index was estimated to low values; 0.02, 0.081 and 0.144 by Ward et al. (1994) for a number of marine, anadromous and freshwater populations. This indicates that the migration rates are considerably higher than the abovementioned mutation rates, making the quasi equilibrium approach feasible.

A number of extensions are possible. Firstly, subpopulations can be interpreted more broadly as referring not only to geographical sites, but also age classes, social or ethnic groups of combinations thereof, see for instance Nordborg and Krone (2002), Sagitov and Jagers (2005), Hössjer (2011) and references therein. If generations are non-overlapping, $t$ is not defined as generations, but more generally as points in time, and the number of copies of a gene that are passed on to the next time point should be interpreted broadly, including the possibility that the individual carrying the gene survives to the next generation.

Secondly, the assumption (1) of constant subpopulation sizes could be dropped, allowing local as well as census sizes to fluctuate in time. For instance, Whitlock and Barton (1997) and Nunney (1999) derive inbreeding effective population sizes and Sampson (2006) coalescence effective sizes under such assumptions. In our setting of a variance effective population size, one could allow $\boldsymbol{N}_t = (N_{t1}, \ldots, N_{ts})$, the vector of subpopulation sizes at time $t$, to satisfy a recursion

$$\boldsymbol{N}_{t+1} = \boldsymbol{N}_t \boldsymbol{M}_t \tag{102}$$

where $\boldsymbol{M}_t$ is a stochastic Leslie matrix with $E(\boldsymbol{M}_t) = \boldsymbol{M}$ involving demographic, genetic as well as environmental effects, see Engen et al. (2005a, b). In Olsson et al. (2013) we derive approximate quasi equilibrium formulas for the variance effective size when the population satisfies (102) for an age structured model. In particular, we investigate the impact of the subpopulation weight vector $\boldsymbol{w}$ for time lags of different size (not necessarily one time step), and achieve good concordance with the simulation results of Waples and Yokota (2007). The main finding is that the variance effective size is much less dependent on the choice of weights for large time lags, corresponding to the long term genetic drift of the population. In particular, this long term drift is predicted from the variance effective size *regardless* of the time lag when the canonical weights $\boldsymbol{w} = \boldsymbol{\gamma}$ are used. This is consistent with the fast migration results of Nagylaki (1980). On the other hand, for conservation biology applications, the transient behaviour of the allele frequency drift is sometimes of more interest for short term protection of populations (Allendorf and Ryman 2002). In Olsson et al. (2013) we demonstrate that this transient behaviour is captured when subpopulations are weighted proportionally to sizes, $\boldsymbol{w} = \boldsymbol{u}$, but *not* when the canonical weights $\boldsymbol{w} = \boldsymbol{\gamma}$ are used.

Thirdly, Step 7 of the numerical algorithm in Sect. 6 involves solving the linear system (52) of equations with $s(s+1)$ unknown variables, which is computationally feasible only for small $s$.

In Hössjer (2013) a faster algorithm based on Fourier transforms is defined when migration is spatially invariant between subpopulations, as for the circular and torus stepping stone models.

Fourthly, the standardized spatial covariance matrix $\boldsymbol{V}$ is interesting in its own right. For instance, it enables approximate quasi equilibrium autocorrelations

$$\rho_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}} \tag{103}$$

to be computed analytically between all pairs of subpopulations $i$ and $j$. Sokal et al. (1997) and Hardy and Vekemans (1999, 2002) discuss the use of Moran's spatial autocorrelation function $I$ for fitting genetic models to data. It is shown in Hössjer (2013) that (103) is closely related to $E(I|P_t)$, thereby being less sensitive to various sources of noise than $I$ itself, as discussed by Slatkin and Arter (1991). It is also possible to use $\boldsymbol{V}$ for other measures of subpopulation differentiation that for instance quantify the genetic differentiation between two groups of subpopulations.

Fifthly, whereas we analyzed approximations of a prediction $E(F^w_{ST,t})^*$ of the fixation index $F^w_{ST,t}$, and argued that this approach is reasonable for multiple loci, it is also of interest to analyze the *distribution* of $F^w_{ST,t}$. Rottenstreich et al. (2007) and Leviyang (2011) have employed coalescence methods for this purpose, as in Sect. 11. In more detail, they consider a scenario when the number of subpopulations $s$ *and* the local census effective size tend to infinity, dividing the coalescent tree into scattering and collecting phases (Wakeley 1999). This work has been focused on island and two-dimensional stepping stone models, and it would be of interest to study more general migration schemes.

Sixthly, Wakeley and Takahashi (2004) have studied diffusion approximations with two time scales (Ethier and Nagylaki 1980) for the allele frequency process of an island model with a large number of subpopulations $s$. Due to the spatial symmetry of the island model, it suffices to study the frequency spectrum of a biallelic marker in all subpopulations. This can be viewed as a condensed representation of the allele frequency vector $\boldsymbol{P}_t$. Similarly as in (21), the frequency spectrum can be divided into a slowly varying part, the allele frequency of the whole population, and a more rapidly varying part, fluctuations of the spectrum around the asymptotic limit. It would be of interest to extend these results to more general migration patterns.

**Acknowledgments** Ola Hössjer's research was financially supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine. Nils Ryman's research was supported by grants from the Swedish Research Council, the BONUS Baltic Organisations' Network for Funding Science EEIG (the BaltGene research project), and through a grant to his colleague Linda Laikre from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas). The authors want to thank an associate editor, two referees, Anders Martin-Löf, and Fredrik Olsson for valuable comments on the work.

## Appendix A: Orthogonal decomposition of allele frequency process

**Jordan canonical form of $\boldsymbol{B}$ and motivation of (22).** Let $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$ be the Jordan canonical form of $\boldsymbol{B}$, with

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{\Lambda}_r \end{pmatrix}$$

a block diagonal matrix containing the (possibly complex-valued) eigenvalues of $\boldsymbol{B}$ along the diagonal. For each $l = 1, \dots, r$, the square matrix

$$\boldsymbol{\Lambda}_l = \begin{pmatrix} \lambda_l & 1 & 0 & \dots & & \\ 0 & \lambda_l & 1 & 0 & \dots & \\ 0 & 0 & \lambda_l & 1 & 0 & \dots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{104}$$

occupies rows and columns $j_{l-1}+1, \dots, j_l$ of $\boldsymbol{\Lambda}$, with diagonal entries equal to $\lambda_l$, all entries along the superdiagonal equal to 1 and all other entries of $\boldsymbol{\Lambda}_l$ equal 0. Hence $\lambda_l$ is an eigenvalue of $\boldsymbol{B}$ which appears $j_l - j_{l-1}$ times along the diagonal of $\boldsymbol{\Lambda}_l$, with $0 = j_0 < j_1 < \cdots < j_r = s$. In particular, $\boldsymbol{\Lambda}$ is diagonal when all eigenvalues of $\boldsymbol{B}$ are distinct and $r = s$. Then the rows of $\boldsymbol{Q}^{-1}$ contain the left eigenvectors of $\boldsymbol{B}$ and the columns $\boldsymbol{q}_1, \dots, \boldsymbol{q}_s$ of $\boldsymbol{Q}$ the right eigenvectors. See for instance Cox and Miller (1965).

Regardless of whether $\boldsymbol{\Lambda}$ is diagonal or not, since $\boldsymbol{B}$ is a transition matrix of a Markov chain, $\boldsymbol{q}_1 = \boldsymbol{1}$ is a right eigenvector with eigenvalue $\lambda_1 = 1$. By the assumed irreducibility and aperiodicity of this Markov chain, it follows from the Perron Frobe-

nius Theorem that $|\lambda_l| < 1$ for $l = 2, \ldots, r$, and without loss of generality, we may assume $|\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_r| \geq 0$.

Introduce the inner product

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{s} \gamma_i \bar{x}_i y_i \tag{105}$$

for possibly complex-valued column vectors $\boldsymbol{x} = (x_i)$ and $\boldsymbol{y} = (y_i)$ of length $s$, with $\bar{x}_i$ the complex conjugate of $x_i$. Then, we have the following result:

**Proposition 5** *The columns $\boldsymbol{q}_2, \ldots, \boldsymbol{q}_s$ of $\boldsymbol{Q}$ are all orthogonal to $\boldsymbol{q}_1 = \mathbf{1}$ with respect to inner product (105), i.e.*

$$(\mathbf{1}, \boldsymbol{q}_j) = 0, \quad j = 2, \ldots, s.$$

*Proof* We have that

$$\begin{aligned}(\mathbf{1}, \boldsymbol{q}_j) &= \sum_{i=1}^{s} \gamma_j q_j \\ &= \langle \boldsymbol{\gamma}, \boldsymbol{q}_j \rangle,\end{aligned}$$

where $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{j=1}^{s} x_j y_j$ is the standard inner product. The result follows since $\boldsymbol{\gamma}$ is the first row of $\boldsymbol{Q}^{-1}$ and $\boldsymbol{q}_j$ row number $j$ (with $j \geq 2$) of $\boldsymbol{Q}$. □

Define $\boldsymbol{\Lambda}^0 = \mathrm{diag}(0, \boldsymbol{\Lambda}_2, \ldots, \boldsymbol{\Lambda}_s)$ as the block diagonal matrix obtained by replacing $\Lambda_1 = \lambda_1 = 1$ in $\boldsymbol{\Lambda}$ by 0 (or any other with modulus less or equal to $|\lambda_2|$), and put

$$\boldsymbol{B}^0 = \boldsymbol{Q}\boldsymbol{\Lambda}^0\boldsymbol{Q}^{-1}. \tag{106}$$

It then follows that $\boldsymbol{B}^0$ has largest eigenvalue $|\lambda_2| < 1$, and it enters into the time dynamics of the allele frequency process as follows:

**Proposition 6** *The recursive autoregressive equation (10) for $\boldsymbol{P}_t$ can be decomposed into one genetic drift term for the overall allele frequency of the whole population, and one recursion part for the allele frequency fluctuations among subpopulations, as*

$$\begin{aligned}P_{t+1} &= P_t + \varepsilon_{t+1}, \\ \boldsymbol{P}^0_{t+1} &= \boldsymbol{B}\boldsymbol{P}^0_t + \boldsymbol{\varepsilon}^0_{t+1} = \boldsymbol{B}^0\boldsymbol{P}^0_t + \boldsymbol{\varepsilon}^0_{t+1},\end{aligned} \tag{107}$$

*with $\boldsymbol{\varepsilon}^0_{t+1}$ as defined in (22).*

*Proof* The upper part of (107) follows immediately from (10), since

$$P_{t+1} = (\mathbf{1}, \boldsymbol{P}_{t+1}) = (\mathbf{1}, \boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1}) = P_t + \varepsilon_{t+1}.$$

Define, for any vector $\boldsymbol{x} = (x_1, ..., x_s)$, $\boldsymbol{x}^0 = \boldsymbol{x} - (\mathbf{1}, \boldsymbol{x})\mathbf{1}$. Then, since $(\boldsymbol{x}+\boldsymbol{y})^0 = \boldsymbol{x}^0 + \boldsymbol{y}^0$, we have that

$$\boldsymbol{P}_{t+1}^0 = (\boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1})^0 = (\boldsymbol{B}\boldsymbol{P}_t)^0 + \boldsymbol{\varepsilon}_{t+1}^0 = \boldsymbol{B}\boldsymbol{P}_t^0 + \boldsymbol{\varepsilon}_{t+1}^0 = \boldsymbol{B}^0\boldsymbol{P}_t^0 + \boldsymbol{\varepsilon}_{t+1}^0. \quad (108)$$

The third equality of (108) follows since

$$\begin{aligned}(\boldsymbol{B}\boldsymbol{P}_t)^0 &= \left(\boldsymbol{B}(P_t\mathbf{1} + \boldsymbol{P}_t^0)\right)^0 \\ &= \left(P_t\mathbf{1} + \boldsymbol{B}\boldsymbol{P}_t^0\right)^0 \\ &= P_t\mathbf{1}^0 + (\boldsymbol{B}\boldsymbol{P}_t^0)^0 \\ &= 0 + \boldsymbol{B}\boldsymbol{P}_t^0 \\ &= \boldsymbol{B}\boldsymbol{P}_t^0,\end{aligned}$$

where in the second last step we used that since $\boldsymbol{P}_t^0$ is a linear combination of $\boldsymbol{q}_2, \ldots, \boldsymbol{q}_s$, so is $\boldsymbol{B}\boldsymbol{P}_t^0$, and hence orthogonal to $\mathbf{1}$ by Proposition 5, so that $(\boldsymbol{B}\boldsymbol{P}_t^0)^0 = \boldsymbol{B}\boldsymbol{P}_t^0$.

The fourth equality of (108) follows since $\boldsymbol{Q}^{-1}\boldsymbol{P}_t^0$ is a linear combination of $\boldsymbol{e}_2, \ldots, \boldsymbol{e}_s$, where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ has 1 in position $i$ and zeros elsewhere. Hence $\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}\boldsymbol{P}_t^0 = \boldsymbol{\Lambda}^0\boldsymbol{Q}^{-1}\boldsymbol{P}_t^0$ and $\boldsymbol{B}\boldsymbol{P}_t^0 = \boldsymbol{B}^0\boldsymbol{P}_t^0$. □

## Appendix B: Proofs from Sect. 5

*Proof of Proposition 1* We notice that

$$\begin{aligned}E(H_{t+1,ij}|\boldsymbol{P}_t) &= E\left(P_{t+1,i}(1 - P_{t+1,j}) + P_{t+1,j}(1 - P_{t+1,i})|\boldsymbol{P}_t\right) \\ &= E(P_{t+1,i}|\boldsymbol{P}_t)\left(1 - E(P_{t+1,j}|\boldsymbol{P}_t)\right) \\ &\quad + E(P_{t+1,j}|\boldsymbol{P}_t)\left(1 - E(P_{t+1,i}|\boldsymbol{P}_t)\right) - 2\mathrm{Cov}(P_{t+1,i}, P_{t+1,j}|\boldsymbol{P}_t) \\ &= (\boldsymbol{B}\boldsymbol{P}_t)_i\left(1 - (\boldsymbol{B}\boldsymbol{P}_t)_j\right) + (\boldsymbol{B}\boldsymbol{P}_t)_j\left(1 - (\boldsymbol{B}\boldsymbol{P}_t)_i\right) - 2\Omega(\boldsymbol{P}_t)_{ij} \\ &= \sum_{k,l=1}^{s} b_{ik}b_{jl}\left(P_{tk}(1 - P_{tl}) + (1 - P_{tk})P_{tl}\right) - 2\Omega(\boldsymbol{P}_t)_{ij},\end{aligned}$$

from which it easily follows that the two recursions in (25) and (26) are equivalent, with $A_{ij,kl}$ and $U_{ij,kl}$ related as in (27).

Next we will show that (25) and (28) are equivalent. Clearly (25) implies (28), so it remains to establish the reverse implication. Hence we assume that (28) is satisfied and we want to show that (25) holds for a unique square matrix $\boldsymbol{U} = (U_{ij,kl})$ of order $s^2$ with $U_{ij,kl} = U_{ij,lk}$. Indeed, since $\boldsymbol{\Omega}(\boldsymbol{P})$ is a quadratic function of $\boldsymbol{P}$ with $\boldsymbol{\Omega}(\boldsymbol{0}) = \boldsymbol{0}$, there is a unique such matrix $\boldsymbol{U}$ and a unique set of coefficients $c_{ij,k}$ satisfying

$$\Omega(\boldsymbol{P})_{ij} = \sum_k c_{ij,k} P_k - \sum_{k,l} U_{ij,kl} P_k P_l \tag{109}$$

for all $i$, $j$. On the other hand, according to lower part of (28),

$$\Omega(\mathbf{1} - \boldsymbol{P})_{ij} = \sum_k c_{ij,k}(1 - P_k) - \sum_{k,l} U_{ij,kl}(1 - P_k)(1 - P_l) \tag{110}$$

should agree with (109). The quadratic terms of (109) and (110) are clearly identical, but in order for the linear and constant terms to agree as well,

$$c_{ij,k} = \sum_l U_{ij,kl} \tag{111}$$

must hold for all $k$ (recall that $U_{ij,kl} = U_{ij,lk}$). On the other hand, we can add and subtract linear terms in (109) according to

$$\Omega(\boldsymbol{P})_{ij} = \frac{1}{2} \sum_{k,l} U_{ij,kl} \left(P_k(1 - P_l) + P_l(1 - P_k)\right) + \sum_k (c_{ij,k} - d_{ij,k}) P_k, \tag{112}$$

where

$$d_{ij,k} = \sum_l U_{ij,kl}$$

for all $k$. But $d_{ij,k} = c_{ij,k}$ according to (111), so that the second sum in (112) vanishes, and the proposition is proved. □

*Proof of Proposition 2* First of all, since $\sum_{\tau=0}^{\infty} (\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau$ is assumed to converge, it can be seen by insertion that (36) provides a solution to (33).

In order to prove (37), we get from the Cauchy–Schwarz inequality

$$|V_{t,ij}| \le \sqrt{V_{t,ii} V_{t,jj}} \le \max(V_{t,ii}, V_{t,jj}),$$

for all pairs $i$, $j$. This implies

$$|\boldsymbol{V}_t|_\infty = \max_{1 \le i,j \le s} |V_{t,ij}| = \max_{1 \le i \le s} V_{t,ii} = \max_{1 \le i \le s} \frac{E_c\left((P_{ti}^0 - P_t)^2 | P_t\right)}{P_t(1 - P_t)}.$$

We then use the definitions of $|\cdot|_\infty$ and $\|\cdot\|$ in Table 2, the triangle inequality and the matrix norm inequality $\|(\boldsymbol{G} - \boldsymbol{\Pi U})^\tau \boldsymbol{\Pi}\| \le \|(\boldsymbol{G} - \boldsymbol{\Pi U})^\tau\| \|\boldsymbol{\Pi}\|$ in order to prove (38), since

$$\begin{aligned} |\boldsymbol{V}|_\infty &= |\mathrm{vec}(\boldsymbol{V})|_\infty \\ &= \left| \sum_{\tau=0}^{\infty} (\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau \boldsymbol{\Pi U} \underline{\mathbf{1}} \right|_\infty \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\tau=0}^{\infty} \left| (\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau \boldsymbol{\Pi U} \underline{\mathbf{1}} \right|_\infty \\
&\leq \sum_{\tau=0}^{\infty} \|(\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau\| \|\boldsymbol{\Pi}\| |\boldsymbol{U}\underline{\mathbf{1}}|_\infty \\
&= \text{Mixtime} \|\boldsymbol{\Pi}\| |\boldsymbol{U}\underline{\mathbf{1}}|_\infty .
\end{aligned}
$$

We also have that

$$
\begin{aligned}
\|\boldsymbol{\Pi}\| &= \max_{i,j} \sum_{1 \leq k,l \leq s} |\Pi_{ij,kl}| \\
&\leq \max_{i,j} \sum_{1 \leq k,l \leq s} |1_{\{(k,l)=(i,j)\}} - \gamma_k 1_{\{j=l\}} - \gamma_l 1_{\{i=k\}} + \gamma_k \gamma_l| \\
&\leq \max_{i,j} \left( 1 + 2 \sum_k \gamma_k + \sum_{k,l=1}^{s} \gamma_k \gamma_l \right) \\
&= 4.
\end{aligned}
$$

Finally, (39)–(40) are proved in the same way as (37)–(38). □

*Proof of Proposition 3* In order to prove (41), we introduce for each pair of integers $\tau, \alpha$ with $0 \leq \alpha \leq \tau$ the set $\mathcal{N}_{\tau\alpha} = \{\boldsymbol{n} = (n_0, n_1, \ldots, n_{\alpha+1})\}$ of $\binom{\tau}{\alpha}$ sequences $\boldsymbol{n}$ such that $0 = n_0 < n_1 < \cdots < n_\alpha < n_{\alpha+1} = \tau + 1$. Then

$$
(\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau = \sum_{\alpha=0}^{\tau} (-1)^\alpha \sum_{n \in \mathcal{N}_{\tau\alpha}} \left( \prod_{i=1}^{\alpha+1} (\boldsymbol{U}^{\{i>1\}} (\boldsymbol{G}^0)^{n_i - n_{i-1} - 1} \boldsymbol{\Pi}^{\{i<\alpha+1\}}) \right), \tag{113}
$$

where the terms in $\mathcal{N}_{\tau\alpha}$ correspond to all possible ways of picking $\alpha$ terms $\boldsymbol{\Pi U}$ and $\tau - \alpha$ terms $\boldsymbol{G}^0$. Taking the matrix norm of (113) and multiplying by $\boldsymbol{U}$ from the left and $\boldsymbol{\Pi}$ from the right, it follows from matrix norm inequalities that

$$
\begin{aligned}
&\|\boldsymbol{U}\| \|(\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau\| \|\boldsymbol{\Pi}\| \\
&\leq \sum_{\alpha=0}^{\tau} \sum_{n \in \mathcal{N}_{\tau\alpha}} \|\boldsymbol{U}\| \left\| \prod_{i=1}^{\alpha+1} (\boldsymbol{U}^{\{i>1\}} (\boldsymbol{G}^0)^{n_i - n_{i-1} - 1} \boldsymbol{\Pi}^{\{i<\alpha+1\}}) \right\| \|\boldsymbol{\Pi}\| \\
&\leq \sum_{\alpha=0}^{\tau} \sum_{n \in \mathcal{N}_{\tau\alpha}} \|\boldsymbol{U}\| \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}^{\{i>1\}} (\boldsymbol{G}^0)^{n_i - n_{i-1} - 1} \boldsymbol{\Pi}^{\{i<\alpha+1\}}\| \right) \|\boldsymbol{\Pi}\| \\
&\leq \sum_{\alpha=0}^{\tau} \sum_{n \in \mathcal{N}_{\tau\alpha}} \|\boldsymbol{U}\| \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}\|^{\{i>1\}} \|(\boldsymbol{G}^0)^{n_i - n_{i-1} - 1}\| \|\boldsymbol{\Pi}^{\{i<\alpha+1\}}\| \right) \|\boldsymbol{\Pi}\| \\
&= \sum_{\alpha=0}^{\tau} \sum_{n \in \mathcal{N}_{\tau\alpha}} \prod_{i=1}^{\alpha+1} (\|\boldsymbol{U}\| \|(\boldsymbol{G}^0)^{n_i - n_{i-1} - 1}\| \|\boldsymbol{\Pi}\|).
\end{aligned} \tag{114}
$$

Summing (114) over $\tau$, then changing the order of summation between $\alpha$ and $\tau$, and finally substituting $m_i = n_i - n_{i-1} - 1$, we find that

$$\begin{aligned}
\|\boldsymbol{\Pi}\|\|\boldsymbol{U}\|\text{Mixtime} &= \|\boldsymbol{U}\| \sum_{\tau=0}^{\infty} \|(\boldsymbol{G}^0 - \boldsymbol{\Pi}\boldsymbol{U})^\tau\| \|\boldsymbol{\Pi}\| \\
&\le \sum_{\tau=0}^{\infty} \sum_{\alpha=0}^{\tau} \sum_{n \in \mathcal{N}_{\tau\alpha}} \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}\| \|(\boldsymbol{G}^0)^{n_i - n_{i-1} - 1}\| \|\boldsymbol{\Pi}\| \right) \\
&= \sum_{\alpha=0}^{\infty} \sum_{\tau=\alpha}^{\infty} \sum_{n \in \mathcal{N}_{\tau\alpha}} \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}\| \|(\boldsymbol{G}^0)^{n_i - n_{i-1} - 1}\| \|\boldsymbol{\Pi}\| \right) \\
&= \sum_{\alpha=0}^{\infty} \sum_{m_1=0}^{\infty} \cdots \sum_{m_{\alpha+1}=0}^{\infty} \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}\| \|(\boldsymbol{G}^0)^{m_i}\| \|\boldsymbol{\Pi}\| \right) \\
&= \sum_{\alpha=0}^{\infty} \prod_{i=1}^{\alpha+1} \left( \|\boldsymbol{U}\| \sum_{m=0}^{\infty} \|(\boldsymbol{G}^0)^m\| \|\boldsymbol{\Pi}\| \right) \\
&= \sum_{\alpha=0}^{\infty} \left( \|\boldsymbol{U}\| \sum_{m=0}^{\infty} \|(\boldsymbol{G}^0)^m\| \|\boldsymbol{\Pi}\| \right)^{\alpha+1} \\
&= \|\boldsymbol{\Pi}\| \|\boldsymbol{U}\| \sum_{m=0}^{\infty} \|(\boldsymbol{G}^0)^m\| / (1 - \|\boldsymbol{\Pi}\| \|\boldsymbol{U}\| \sum_{m=0}^{\infty} \|(\boldsymbol{G}^0)^m\|). \qquad (115)
\end{aligned}$$

It can be seen that $(\boldsymbol{G}^0)^\tau \text{vec}(\boldsymbol{V}) = \text{vec}((\boldsymbol{B}^0)^\tau \boldsymbol{V} ((\boldsymbol{B}^0)^T)^\tau)$, by induction with respect to $\tau$. Writing $(\boldsymbol{G}^0)^\tau = (G^{0(\tau)}_{ij,kl})$ and $(\boldsymbol{B}^0)^\tau = (b^{0(\tau)}_{ik})$, this yields

$$G^{0(\tau)}_{ij,kl} = b^{0(\tau)}_{ik} b^{0(\tau)}_{jl},$$

and

$$\begin{aligned}
\|(\boldsymbol{G}^0)^\tau\| &= \max_{i,j} \sum_{k,l} |G^{0(\tau)}_{ij,kl}| \\
&= \max_{i,j} \sum_{k,l=1}^{s} |b^{0(\tau)}_{ik}| |b^{0(\tau)}_{jl}| \\
&= \max_{i} \sum_{k} |b^{0(\tau)}_{ik}| \cdot \max_{j} \sum_{l} |b^{0(\tau)}_{jl}| \\
&= \|(\boldsymbol{B}^0)^\tau\|^2. \qquad (116)
\end{aligned}$$

Formula (41) then follows from (115) to (116). In order to verify (42), we use (106) and the Jordan decomposition (104) to deduce

$$(\boldsymbol{B}^0)^\tau = \boldsymbol{Q} \text{diag}(0, \boldsymbol{\Lambda}_2^\tau, \ldots, \boldsymbol{\Lambda}_r^\tau) \boldsymbol{Q}^{-1},$$

where the middle matrix on the right hand side is block diagonal, with $\|\mathbf{\Lambda}_l^\tau\| = O(\tau^{j_l - j_{l-1}-1}|\lambda_l|^\tau)$ as $\tau \to \infty$, and $j_l - j_{l-1}$ the order of the square matrix $\mathbf{\Lambda}_l$, see Cox and Miller (1965) for details. In particular, this implies that $\|\mathbf{\Lambda}_l^\tau\|$ converges to zero at a faster rate than $(|\lambda_2| + \epsilon)^\tau$ as $\tau \to \infty$ for any $0 < \epsilon < 1 - |\lambda_2|$. Then (42) follows, since

$$\|(\boldsymbol{B}^0)^\tau\| \le \|\boldsymbol{Q}\| \left( \max_{2 \le l \le r} \|\mathbf{\Lambda}_l^\tau\| \right) \|\boldsymbol{Q}^{-1}\|.$$

Finally, (43) is a simple consequence of (41) and (42), since

$$\begin{aligned} \sum_{\tau=0}^{\infty} \|(\boldsymbol{B}^0)^\tau\|^2 &\le C^2 \sum_{\tau=0}^{\infty} (|\lambda_2| + \epsilon)^{2\tau} \\ &= \frac{C^2}{1 - (|\lambda_2| + \epsilon)^2}. \end{aligned}$$

□

## Appendix C: Proof of Theorem 1

We start by showing that $\mathrm{vec}(\boldsymbol{V}_t)$ and $\mathrm{vec}(\mathbf{\Sigma}_t)$ satisfy a similar system of equations as (33). To this end, since $\boldsymbol{\varepsilon}_t^0 = (\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})\boldsymbol{\varepsilon}_t$, the lower part of (22) implies a recursion

$$\begin{aligned} \boldsymbol{V}_{t+1} &= \frac{E_c\left(\boldsymbol{\varepsilon}_{t+1}^0 (\boldsymbol{\varepsilon}_{t+1}^0)^T | P_t\right)}{P_t(1-P_t)} + \frac{E_c\left(\boldsymbol{B}^0 \boldsymbol{P}_t^0 \left(\boldsymbol{B}^0 \boldsymbol{P}_t^0\right)^T | P_t\right)}{P_t(1-P_t)} + \boldsymbol{\xi}_{t+1} \\ &= (\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})\mathbf{\Sigma}_t(\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})^T + \boldsymbol{B}^0 \boldsymbol{V}_t (\boldsymbol{B}^0)^T + \boldsymbol{\xi}_{t+1}, \end{aligned} \tag{117}$$

where $\boldsymbol{\xi}_{t+1}$ is a remainder term that is nonzero since we conditioned on $P_t$ rather than $P_{t+1}$ and divided by $P_t(1-P_t)$ rather than $P_{t+1}(1-P_{t+1})$ on the right hand side of (117). Any departure of $E_c(\boldsymbol{\varepsilon}_{t+1}^0 | P_t)$ from $E(\boldsymbol{\varepsilon}_{t+1}^0 | P_t) = \boldsymbol{0}$ implies, in addition, that a cross covariance term is added to $\boldsymbol{\xi}_{t+1}$.

In vec format we may rewrite (117) as

$$\mathrm{vec}(\boldsymbol{V}_{t+1}) = \mathbf{\Pi}\mathrm{vec}(\mathbf{\Sigma}_t) + \boldsymbol{G}^0 \mathrm{vec}(\boldsymbol{V}_t) + \mathrm{vec}(\boldsymbol{\xi}_{t+1}), \tag{118}$$

with $\mathbf{\Pi}$ and $\boldsymbol{G}^0$ matrices defined by $\mathbf{\Pi}\mathrm{vec}(\mathbf{\Sigma}_t) = \mathrm{vec}((\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})\mathbf{\Sigma}_t(\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})^T))$ and $\boldsymbol{G}^0\mathrm{vec}(\boldsymbol{V}_t) = \mathrm{vec}(\boldsymbol{B}^0 \boldsymbol{V}_t (\boldsymbol{B}^0)^T)$ respectively. Hence their entries are as in (34) and (35).

For the standardized genetic drift covariance matrix, we first expand (31) as

$$\begin{aligned}\boldsymbol{\Sigma}_t &= \frac{E_c\left(E_c(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|P_t)|P_t\right)}{P_t(1-P_t)}\\ &= \frac{E_c\left(E(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|P_t)|P_t\right)}{P_t(1-P_t)} + \boldsymbol{\zeta}_t\\ &= \frac{E_c(\boldsymbol{\Omega}(P_t\mathbf{1}+\boldsymbol{P}_t^0)|P_t)}{P_t(1-P_t)} + \boldsymbol{\zeta}_t, \end{aligned} \tag{119}$$

where the remainder term $\boldsymbol{\zeta}_t$ occurs when replacing the inner expectation $E$ by $E_c$. Then we expand $\boldsymbol{\Omega}(P_t\mathbf{1}+\boldsymbol{P}_t^0)$ as in (30) and take expectation conditionally on $P_t$, and switch index from $t$ to $t+1$, to deduce that

$$\begin{aligned}\mathrm{vec}(\boldsymbol{\Sigma}_{t+1}) &= \boldsymbol{U}\underline{\mathbf{1}} - \boldsymbol{U}\mathrm{vec}(\boldsymbol{V}_{t+1}) + \boldsymbol{U}_{t+1}\boldsymbol{\mu}_{t+1} + \mathrm{vec}(\boldsymbol{\zeta}_{t+1})\\ &= \boldsymbol{U}\underline{\mathbf{1}} - \boldsymbol{U}\mathrm{vec}(\boldsymbol{V}_{t+1}) + \boldsymbol{\eta}_{t+1}. \end{aligned} \tag{120}$$

where $\boldsymbol{U}_t = (U_{tij,k})$ is an $s^2 \times s$ matrix, whose elements are defined as $U_{tij,k} = (1-2P_t)\sum_l U_{ij,kl}$, so that the last term on the right hand side of (30) can be written as $\boldsymbol{U}_t\boldsymbol{P}_t^0$. The last term on the right hand side of (120) is defined by

$$\boldsymbol{\eta}_t = \boldsymbol{U}_t\boldsymbol{\mu}_t + \mathrm{vec}(\boldsymbol{\zeta}_t)$$

with $\boldsymbol{\mu}_t$ as in (44).

Now (118) and (120) define a system of equations which only differs from (33) in that the remainder terms $\mathrm{vec}(\boldsymbol{\xi}_{t+1})$ and $\boldsymbol{\eta}_{t+1}$ have been added. For simplicity of notation, we write $\tilde{\boldsymbol{\xi}}_t = \mathrm{vec}(\boldsymbol{\xi}_t) = \mathrm{vec}(\xi_{t,ij};\, 1 \le i, j \le s)$, a column vector of length $s^2$. Combining and (118) and (120), we get

$$\begin{pmatrix}\mathrm{vec}(\boldsymbol{\Sigma}_{t+1})\\ \mathrm{vec}(\boldsymbol{V}_{t+1})\end{pmatrix} = \boldsymbol{T}\begin{pmatrix}\mathrm{vec}(\boldsymbol{\Sigma}_t)\\ \mathrm{vec}(\boldsymbol{V}_t)\end{pmatrix} + \begin{pmatrix}\boldsymbol{U}\underline{\mathbf{1}}\\ \boldsymbol{0}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\eta}_{t+1} - \boldsymbol{U}\tilde{\boldsymbol{\xi}}_{t+1}\\ \tilde{\boldsymbol{\xi}}_{t+1}\end{pmatrix}, \tag{121}$$

where

$$\boldsymbol{T} = \begin{pmatrix}\boldsymbol{0} & -\boldsymbol{U}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{\Pi} & \boldsymbol{G}^0\end{pmatrix}.$$

On the other hand, it follows from (33) that

$$\begin{pmatrix}\mathrm{vec}(\boldsymbol{\Sigma})\\ \mathrm{vec}(\boldsymbol{V})\end{pmatrix} = \boldsymbol{T}\begin{pmatrix}\mathrm{vec}(\boldsymbol{\Sigma})\\ \mathrm{vec}(\boldsymbol{V})\end{pmatrix} + \begin{pmatrix}\boldsymbol{U}\underline{\mathbf{1}}\\ \boldsymbol{0}\end{pmatrix}. \tag{122}$$

Taking the difference of (121) and (122), we find that

$$\boldsymbol{\delta}_t = \begin{pmatrix}\mathrm{vec}(\Delta\boldsymbol{\Sigma}_t)\\ \mathrm{vec}(\Delta\boldsymbol{V}_t)\end{pmatrix},$$

satisfies

$$\delta_{t+1} = T\delta_t + \begin{pmatrix} \eta_{t+1} - U\tilde{\xi}_{t+1} \\ \tilde{\xi}_{t+1} \end{pmatrix} \Longrightarrow \delta_t = \sum_{\tau=0}^{\infty} T^{\tau} \begin{pmatrix} \eta_{t-\tau} - U\tilde{\xi}_{t-\tau} \\ \tilde{\xi}_{t-\tau} \end{pmatrix}, \quad (123)$$

provided that the series converges. It can be shown by induction with respect to $\tau$ that

$$T^{\tau} = \begin{pmatrix} -U(G^0 - \Pi U)^{\tau-1}\Pi & -U(G^0 - \Pi U)^{\tau-1}G^0 \\ (G^0 - \Pi U)^{\tau-1}\Pi & (G^0 - \Pi U)^{\tau-1}G^0 \end{pmatrix}$$

for all $\tau \geq 1$. Inserting this formula into (123), one obtains

$$\delta_t = \begin{pmatrix} -\eta_t \\ 0 \end{pmatrix} + \begin{pmatrix} -U \\ I \end{pmatrix} \sum_{\tau=0}^{\infty} (G^0 - \Pi U)^{\tau} \left( \Pi \eta_{t-\tau-1} + \tilde{\xi}_{t-\tau} \right). \quad (124)$$

Since $\tilde{\xi}_t$ contains the same elements as $\xi_t$, we have that $|\tilde{\xi}_t|_\infty = |\xi_t|_\infty$, and moreover, $|\eta_t|_\infty \leq |\zeta_t|_\infty + \|U_t\||\mu_t|_\infty$. Hence it follows, by taking the $|\cdot|_\infty$-norm of the upper and lower part of (124), that

$$\begin{aligned} |\Delta\Sigma_t|_\infty \leq\ & |\zeta_t|_\infty + \|U_t\||\mu_t|_\infty \\ & + \|U\| \sum_{\tau=0}^{\infty} \|(G^0 - \Pi U)^{\tau}\| \left( \|\Pi\|(|\zeta_t|_\infty + \|U_{t-\tau-1}\||\mu_{t-\tau-1}|_\infty) + |\xi_{t-\tau}|_\infty \right) \end{aligned} \quad (125)$$

and

$$|\Delta V_t|_\infty \leq \sum_{\tau=0}^{\infty} \|(G^0 - \Pi U)^{\tau}\| \left( \|\Pi\|(|\zeta_t|_\infty + \|U_{t-\tau-1}\||\mu_{t-\tau-1}|_\infty) + |\xi_{t-\tau}|_\infty \right). \quad (126)$$

Since

$$\sum_k |U_{tij,k}| \leq \sum_k \left| \sum_l U_{ij,kl} \right| \leq \sum_{kl} |U_{ij,kl}|,$$

if follows that $\|U_t\| \leq \|U\|$. Hence we may replace $\|U_t\|$ and $\|U_{t-\tau-1}\|$ in (125)–(126) by their upper bounds $\|U\|$, take conditional expectation $E_c$ on both sides of these two inequalities, and finally letting $t \to \infty$, thereby obtaining (48) and (49). □

## Appendix D: Verifying formulas for $\Omega(P_t)$ and $N_{eV}^{\text{appr}}$ for various reproduction and migration models.

We will start by verifying (30) (and hence also (120)) separately for reproduction scenarios 1, 2 and 3.

**Reproduction scenario 1.** For this reproduction scenario, we write

$$P^*_{tki} = P_{tk} + (\tilde{P}_{tk} - P_{tk}) + (P^*_{tki} - \tilde{P}_{tk}).$$

It follows from (10) and (53) that

$$\varepsilon_{t+1,i} = \sum_{k=1}^{s} b_{ik}(\tilde{P}_{tk} - P_{tk}) + \sum_{k=1}^{s} b_{ik}(P^*_{tki} - \tilde{P}_{tk}).$$

We further have that

$$\mathrm{Var}(\tilde{P}_{tk} - P_{tk}|\boldsymbol{P}_t) = \left(\frac{1}{2N_{ek}} - \frac{1}{2Nu_k}\right) P_{tk}(1 - P_{tk})(1 + o(1)) \tag{127}$$

and

$$\mathrm{Var}(P^*_{tki} - \tilde{P}_{tk}|\boldsymbol{P}_t) = \frac{P_{tk}(1 - P_{tk})}{2Nu_k m_{ki}}(1 + o(1)).$$

Combining the last three displayed expressions, we arrive at (54). □

**Reproduction scenario 2.** Write

$$\varepsilon_{t+1,i} = \sum_{k=1}^{s} b_{ik}(P^*_{tki} - P_{tk}), \tag{128}$$

introduce $C_{kij} = \mathrm{Cov}(\nu^l_{ki}, \nu^l_{kj})$ and $\tilde{C}_{kij} = \mathrm{Cov}(\nu^l_{ki}, \nu^{l'}_{kj})$ when $l \neq l'$. Because of the assumed exchangeability of $\{\boldsymbol{\nu}^l_k\}_{l=1}^{2Nu_k}$, $C_{kij}$ and $\tilde{C}_{kij}$ do not depend on $l$ and $(l, l')$ respectively. Since (2) holds exactly, with remainder term $o(1)$ equal to zero, the variance of the left hand side must be zero, and this implies $\tilde{C}_{kij} = -C_{kij}/(2Nu_k - 1)$. Therefore, it follows from (55) that

$$\begin{aligned}\mathrm{Cov}(P^*_{tki}, P^*_{tkj}|\boldsymbol{P}_t) &= \frac{2Nu_k P_{tk} C_{kij} + 2Nu_k P_{tk}(2Nu_k P_{tk} - 1)\tilde{C}_{kij}}{(2Nu_k)^2 m_{ki} m_{kj}} \\ &\sim \frac{C_{kij}}{m_{ki} m_{kj}} \frac{P_{tk}(1 - P_{tk})}{2Nu_k}.\end{aligned}$$

Combining this with (128), we arrive at

$$\boldsymbol{\Omega}(\boldsymbol{P}_t)_{ij} = \sum_{k=1}^{s} b_{ik} b_{jk} \frac{C_{kij}}{2Nu_k m_{ki} m_{kj}} P_{tk}(1 - P_{tk}),$$

which is equivalent to (56). □

**Reproduction scenario 3.** In order to verify (120), we first notice from (10) and (57) that

$$\varepsilon_{t+1,i} = (P_{t+1,i} - \check{P}_{ti}) + \sum_{k=1}^{s} b_{ik}(\tilde{P}_{tk} - P_{tk}) + \sum_{k=1}^{s}(B_{ik} - b_{ik})P_{tk} + \text{rem}, \qquad (129)$$

with $\text{rem} = \sum_{k=1}^{s}(B_{ik} - b_{ik})(\tilde{P}_{tk} - P_{tk})$ a remainder term that vanishes when $N_{ek} = Nu_k$ for all $k$ and which is otherwise asymptotically negligible when $\alpha_i \to \infty$ as $N \to \infty$. It follows from (57) and (59) that

$$\text{Var}(P_{t+1,i} - \check{P}_{ti}|\boldsymbol{P}_t) \sim \frac{(\boldsymbol{B}\boldsymbol{P}_t)_i(1 - (\boldsymbol{B}\boldsymbol{P}_t)_i)}{2Nu_i}(1 + o(1)),$$

and

$$\begin{aligned}\text{Var}\left(\sum_{k=1}^{s}(B_{ik} - b_{ik})P_{tk}|\boldsymbol{P}_t\right) &= \frac{1}{\alpha_i + 1}\sum_{k=1}^{s} P_{tk}^2 b_{ik} - \frac{1}{\alpha_i + 1}\sum_{k,l=1}^{s} P_{tk}P_{tl}b_{ik}b_{il} \\ &= \frac{1}{\alpha_i + 1}\sum_{k=1}^{s}(P_{tk} - (\boldsymbol{B}\boldsymbol{P}_t)_i)^2 b_{ik}.\end{aligned}$$

In conjunction with (127) and (129), this proves (60). □

**Verifying** (65). The reproduction scenario 3 expression for $\boldsymbol{\Sigma}$ is obtained by combining the upper equation of (33) with the relevant entries for $U_{ij,kl}$ in Table 3. When $N_{ek} = Nu_k$ for $k = 1, \ldots, s$, all non-diagonal ($i \neq j$) terms vanish and then the denominator of (62) can be written as

$$\begin{aligned}2\boldsymbol{u}\boldsymbol{\Sigma}\boldsymbol{u}^T &= 2\sum_{i,j,k,l} u_i u_j U_{ij,kl} - 2\sum_{i,j} u_i u_j \sum_{k,l} U_{ij,kl}V_{kl} \\ &= 2\sum_{i,k,l} u_i^2 U_{ii,kl} - 2\sum_{i} u_i^2 \sum_{k,l} U_{ii,kl}V_{kl} \\ &= \frac{1}{N}\left(1 - \sum_i u_i(\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T)_{ii}\right) + 2\sum_i \frac{u_i^2}{\alpha_i + 1}\left(\sum_k b_{ik}V_{kk} - (\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T)_{ii}\right),\end{aligned}$$

which yields (65). □

**Deriving explicit expressions of $N_{eV}^{\text{appr}}$ and $F_{ST}^{\text{appr}}$ for the island model.** Since $\boldsymbol{\gamma} = \boldsymbol{u}$ for the island model, we can apply (62) and (63), with $\boldsymbol{u} = \mathbf{1}^T/s$, to deduce

$$N_{eV}^{\text{appr}} = \frac{1}{2\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}/s^2} \qquad (130)$$

and

$$F_{ST}^{\text{appr}} = \frac{1}{s}\text{tr}(\boldsymbol{V}). \qquad (131)$$

We will start by giving a more explicit expression for $\boldsymbol{V}$. It follows from (66) that $\boldsymbol{Bq} = (1-m)\boldsymbol{q}$ for any vector $\boldsymbol{q}$ with $(\boldsymbol{q}, \mathbf{1}) = 0$. Hence $\lambda_2 = \cdots = \lambda_s = 1-m$. In this case it is particularly convenient to put $\lambda_1^0 = 1-m$ in the definition of $\boldsymbol{B}^0$, since then, according to (106), $\boldsymbol{B}^0 = (1-m)\boldsymbol{I}$. The lower part of (33) can be written as $\boldsymbol{V} = \boldsymbol{B}^0\boldsymbol{V}(\boldsymbol{B}^0)^T + \tilde{\boldsymbol{\Sigma}}$, where $\tilde{\boldsymbol{\Sigma}} = (\boldsymbol{I} - \mathbf{1}\boldsymbol{\gamma})\boldsymbol{\Sigma}(\boldsymbol{I} - \boldsymbol{\gamma})^T$. We can repeatedly apply this equation to deduce that

$$\boldsymbol{V} = \sum_{r=0}^{\infty}(1-m)^{2r}\tilde{\boldsymbol{\Sigma}} = \frac{\tilde{\boldsymbol{\Sigma}}}{1-(1-m)^2},$$

and hence (131) can be rewritten as

$$\left(1-(1-m)^2\right)F_{ST}^{\text{appr}} = \frac{1}{s}\text{tr}(\tilde{\boldsymbol{\Sigma}}) = \frac{1}{s}\left(\text{tr}(\boldsymbol{\Sigma}) - \frac{1}{s}\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}\right). \tag{132}$$

Therefore, in view of (130) and (132), it remains to find $\boldsymbol{\Sigma}$.

For reproduction scenario 1, it can be deduced from (120) that (54) simplifies to

$$\begin{aligned}\Sigma_{ij} = &\left(\frac{1}{2N_e} - \frac{1}{2N/s}\right)\left(\frac{2m-m^2}{s} + (1-m)^2 1_{\{i=j\}}\right) + \frac{1_{\{i=j\}}}{2N/s}\\ &- \left(\frac{1}{2N_e} - \frac{1}{2N/s}\right)\left(\frac{m^2}{s^2}\text{tr}(\boldsymbol{V}) + \frac{V_{ii}+V_{jj}}{2}\left(2\frac{m}{s}(1-m) + 1_{\{i=j\}}(1-m)^2\right)\right)\\ &- \frac{1_{\{i=j\}}}{2N/s}\left(\frac{m}{s}\text{tr}(\boldsymbol{V}) + (1-m)V_{ii}\right)\end{aligned}$$

for the island model, so that

$$\frac{2}{s^2}\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1} = \frac{1}{sN_e}\left(1 - \frac{1}{s}\text{tr}(\boldsymbol{V})\right) = \frac{1}{sN_e}(1 - F_{ST}^{\text{appr}}) \tag{133}$$

and

$$\frac{1}{s}\text{tr}(\tilde{\boldsymbol{\Sigma}}) = \frac{s-1}{s}\frac{1}{2\tilde{N}}(1 - F_{ST}^{\text{appr}}). \tag{134}$$

Combining (130) and (133) we arrive at (67), and inserting (134) into (132) and solving for $F_{ST}^{\text{appr}}$ we arrive at (68).

For reproduction scenario 3, a similar simplification of (60) leads to

$$\begin{aligned}\frac{2}{s^2}\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1} &= \frac{1}{sN_e} - \left(\frac{1}{N_e} - \frac{1-(1-m)^2}{N/s}\right)\frac{1}{s^2}\text{tr}(\boldsymbol{V}) + \frac{2\left(1-(1-m)^2\right)}{\alpha+1}\frac{1}{s^2}\text{tr}(\boldsymbol{V})\\ &= \frac{1}{sN_e} - \left(\frac{1}{N_e} - \frac{1-(1-m)^2}{N/s}\right)\frac{1}{s}F_{ST}^{\text{appr}} + \frac{2\left(1-(1-m)^2\right)}{\alpha+1}\frac{1}{s}F_{ST}^{\text{appr}}.\end{aligned} \tag{135}$$

and

$$\frac{1}{s}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}) = \frac{s-1}{s}\left(\frac{1}{2\tilde{N}} - \frac{(1-m)^2}{2N_e}\frac{1}{s}\mathrm{tr}(\boldsymbol{V}) + \frac{1-(1-m)^2}{\alpha+1}\frac{1}{s}\mathrm{tr}(\boldsymbol{V})\right)$$
$$= \frac{s-1}{s}\left(\frac{1}{2\tilde{N}} - \frac{(1-m)^2}{2N_e}F_{ST}^{\mathrm{appr}} + \frac{1-(1-m)^2}{\alpha+1}F_{ST}^{\mathrm{appr}}\right). \quad (136)$$

Inserting (135) into (130) we arrive at (69), and plugging (136) into (132) and solving for $F_{ST}^{\mathrm{appr}}$ we arrive at (70). □

## Appendix E: Proof of Theorem 2

In order to prove Theorem 2, we first need two lemmas, which we state for a single biallelic locus:

**Lemma 1** *In the one locus biallelic definitions (12) and (13) of* $N_{eV,t}^w = Y/X$ *and* $F_{ST,t}^w = Z/Y$*, the conditional expected values of the numerators and denominators equal*

$$E_c(Y|P_t) = E_c\left(P_t^w(1-P_t^w)|P_t\right)$$
$$= \left(1-(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}_t(\boldsymbol{w}-\boldsymbol{\gamma})^T + (1-2P_t)(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{\mu}_t\right)P_t(1-P_t)$$
$$= (1-tr(\boldsymbol{C}_Y\boldsymbol{V}_t) + \boldsymbol{c}_Y\boldsymbol{\mu}_t)P_t(1-P_t), \quad (137)$$
$$E_c(Z|P_t) = E_c\left(\sum_{i=1}^s w_i(P_{ti}-P_t^w)^2|P_t\right)$$
$$= \left(\sum_{i=1}^s w_i\boldsymbol{V}_{tii} - (\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}_t(\boldsymbol{w}-\boldsymbol{\gamma})^T\right)P_t(1-P_t)$$
$$= tr(\boldsymbol{C}_Z\boldsymbol{V}_t)P_t(1-P_t) \quad (138)$$

*and*

$$E_c(X|P_t) = 2E_c\left(E((P_{t+1}^w - P_t^w)^2|P_t^w)|P_t\right)$$
$$= 2\left(\boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})(\boldsymbol{V}_t-\boldsymbol{\varsigma}_t)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T + \boldsymbol{w}(\boldsymbol{\Sigma}_t-\boldsymbol{\zeta}_t)\boldsymbol{w}^T\right)P_t(1-P_t)$$
$$= \left(tr\left(\boldsymbol{C}_X(\boldsymbol{V}_t-\boldsymbol{\varsigma}_t)\right) + tr\left(\boldsymbol{C}'_X(\boldsymbol{\Sigma}_t-\boldsymbol{\zeta}_t)\right)\right)P_t(1-P_t) \quad (139)$$

*respectively, where* $\boldsymbol{C}_Y = (\boldsymbol{w}-\boldsymbol{\gamma})^T(\boldsymbol{w}-\boldsymbol{\gamma})$, $\boldsymbol{c}_Y = (1-2P_t)(\boldsymbol{w}-\boldsymbol{\gamma})$, $\boldsymbol{C}_Z = diag(\boldsymbol{w}) - \boldsymbol{w}^T\boldsymbol{w}$, $\boldsymbol{C}_X = 2(\boldsymbol{B}-\boldsymbol{I})^T(\boldsymbol{w}-\boldsymbol{\gamma})^T(\boldsymbol{w}-\boldsymbol{\gamma})(\boldsymbol{B}-\boldsymbol{I})$, $\boldsymbol{C}'_X = 2\boldsymbol{w}^T\boldsymbol{w}$, $\boldsymbol{\mu}_t$ *and* $\boldsymbol{\zeta}_t$ *are the remainder terms defined in (44) and (119), and* $\boldsymbol{\varsigma}_t$ *another remainder term defined below, in (140).*

*Proof* We only prove the first parts of (137)–(139), and leave the second part to the reader. Starting with (137), we find that

$$\begin{aligned}
&E_c\left(P_t^w(1-P_t^w)|P_t\right)\\
&\quad= P_t(1-P_t) - E_c\left((P_t^w-P_t)^2|P_t\right) + (1-2P_t)E_c\left(P_t^w-P_t|P_t\right)\\
&\quad= P_t(1-P_t) - E_c\left((\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{P}_t^0)^2|P_t\right) + (1-2P_t)E_c\left((\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{P}_t^0|P_t\right)\\
&\quad= P_t(1-P_t)\left(1-(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}_t(\boldsymbol{w}-\boldsymbol{\gamma})^T + (1-2P_t)(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{\mu}_t\right),
\end{aligned}$$

where in the second equality we used $P_t^w - P_t = (\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{P}_t = (\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{P}_t^0$. For (139) we use (21) and $(\boldsymbol{B}-\boldsymbol{I})\mathbf{1} = \boldsymbol{0}$ to deduce

$$\begin{aligned}
P_{t+1}^w &= \boldsymbol{w}\boldsymbol{P}_{t+1}\\
&= \boldsymbol{w}\boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{w}\boldsymbol{\varepsilon}_{t+1}\\
&= P_t^w + \boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{P}_t + \boldsymbol{w}\boldsymbol{\varepsilon}_{t+1}\\
&= P_t^w + \boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{P}_t^0 + \boldsymbol{w}\boldsymbol{\varepsilon}_{t+1}.
\end{aligned}$$

We introduce the ascertainment bias term

$$\boldsymbol{\varsigma}_t = \frac{E_c\left(E_c\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|P_t^w, P_t\right)|P_t\right) - E_c\left(E\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|P_t^w, P_t\right)|P_t\right)}{P_t(1-P_t)}, \tag{140}$$

which quantifies the effect of replacing the inner expectation $E$ of $\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T$ by $E_c$. Then we can write

$$\begin{aligned}
&E_c\left(E\left((P_{t+1}^w - P_t^w)^2|P_t^w\right)|P_t\right)\\
&= E_c\left(E\left((P_{t+1}^w - P_t^w)^2|P_t^w, P_t\right)|P_t\right)\\
&= \boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})E_c\left(E\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|P_t^w, P_t\right)|P_t\right)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T\\
&\quad+\boldsymbol{w}E_c\left(E\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|P_t^w, P_t\right)|P_t\right)\boldsymbol{w}^T\\
&= \boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})E_c\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|P_t\right)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T - \boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\varsigma}_t(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T P_t(1-P_t)\\
&\quad+\boldsymbol{w}E_c\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|P_t\right)\boldsymbol{w}^T - \boldsymbol{w}\boldsymbol{\zeta}_t\boldsymbol{w}^T P_t(1-P_t)\\
&= P_t(1-P_t)\left(\boldsymbol{w}(\boldsymbol{B}-\boldsymbol{I})(\boldsymbol{V}_t-\boldsymbol{\varsigma}_t)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{w}^T + \boldsymbol{w}(\boldsymbol{\Sigma}_t-\boldsymbol{\zeta}_t)\boldsymbol{w}^T\right).
\end{aligned}$$

In order to verify (138), we first write

$$\boldsymbol{P}_t - P_t^w\mathbf{1} = (\boldsymbol{I}-\mathbf{1}\boldsymbol{w})\boldsymbol{P}_t = (\boldsymbol{I}-\mathbf{1}\boldsymbol{w})\boldsymbol{P}_t^0,$$

which leads to

$$E_c\left((P_{ti}-P_t^w)^2|P_t\right)=P_t(1-P_t)\left((\boldsymbol{I}-\mathbf{1}\boldsymbol{w})\boldsymbol{V}_t(\boldsymbol{I}-\mathbf{1}\boldsymbol{w})^T\right)_{ii},$$

and then (138) follows since

$$\sum_{i=1}^{s} w_i\left((\boldsymbol{I}-\mathbf{1}\boldsymbol{w})\boldsymbol{V}_t(\boldsymbol{I}-\mathbf{1}\boldsymbol{w})^T\right)_{ii}=\sum_{i=1}^{s} w_i\boldsymbol{V}_{tii}-(\boldsymbol{w}-\boldsymbol{\gamma})\boldsymbol{V}_t(\boldsymbol{w}-\boldsymbol{\gamma})^T.$$

□

**Lemma 2** *Let* $\boldsymbol{c}$ *be a* $1\times s$ *vector,* $\boldsymbol{C}$ *an* $s\times s$ *matrix, and define*

$$\epsilon=\boldsymbol{c}(\boldsymbol{P}_t^0-\boldsymbol{\mu}_t P_t(1-P_t))+tr\left(\boldsymbol{C}(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T-\boldsymbol{V}_t P_t(1-P_t))\right).$$

*Then*

$$E_c(\epsilon^2|P_t)\le 2|\boldsymbol{c}|_1^2|\boldsymbol{V}_t|_\infty P_t(1-P_t)+2|\boldsymbol{C}|_1^2\kappa_t, \tag{141}$$

*with*

$$\kappa_t=\max_{1\le i\le s} E_c((P_{ti}^0)^4|P_t).$$

*Proof* Put $\boldsymbol{c}=(c_1,\ldots,c_s)$ and $\boldsymbol{C}=(C_{ij})_{i,j=1}^s$. For simplicity, we omit conditioning on $P_t$ in the notation, writing $E_c(\cdot)=E_c(\cdot|P_t)$. Then

$$\begin{aligned}
E_c(\epsilon^2)&\le 2E_c\left(\left(\sum_i c_i(P_{ti}^0-\mu_{ti}P_t(1-P_t))\right)^2\right)+2E_c\left(\left(\sum_{ij}C_{ji}P_{ti}^0P_{tj}^0\right)^2\right)\\
&\le 2E_c\left(\left(\sum_i c_iP_{ti}^0\right)^2\right)+2E_c\left(\left(\sum_{ij}C_{ji}P_{ti}^0P_{tj}^0\right)^2\right)\\
&\le 2\sum_{i,j}|c_i||c_j|E_c(|P_{ti}^0P_{tj}^0|)+2\sum_{ijkl}|C_{ji}||C_{lk}|E_c(\left|P_{ti}^0P_{tj}^0P_{tk}^0P_{tl}^0\right|)\\
&\le \sum_{i,j}|c_i||c_j|(E_c((P_{ti}^0)^2)+E_c((P_{tj}^0)^2)\\
&\quad+0.5\sum_{ijkl}|C_{ji}||C_{lk}|(E_c(P_{ti}^0)^4+E_c(P_{tj}^0)^4+E_c(P_{tk}^0)^4+E_c(P_{tl}^0)^4)\\
&\le 2\sum_{i,j}|c_i||c_j||\boldsymbol{V}_t|_\infty P_t(1-P_t)+2\sum_{ijkl}|C_{ji}||C_{lk}|\kappa_t,
\end{aligned}$$

using the Cauchy Schwarz Inequality in the fourth step. The last term is identical to the right hand side of (141). □

*Proof of Theorem 2* When all loci are biallelic ($n(x) \equiv 2$), formulas (75) and (78) simplify to

$$G^w_{ST,t} = Z/Y,$$
$$N^w_{eV,t} = Y/X,$$

respectively, where

$$X = 2\sum_{x=1}^{n} E\left(\left(P^w_{t+1}(x) - P^w_t(x)\right)^2 | P^w_t(x)\right),$$
$$Y = \sum_{x=1}^{n} P^w_t(x)(1 - P^w_t(x)),$$
$$Z = \sum_{x=1}^{n} \mathrm{tr}\left(\boldsymbol{C}_Z \boldsymbol{P}^0_t(x)\boldsymbol{P}^0_t(x)^T\right)$$

are multilocus extensions of the corresponding numerators and denominators $X, Y, Z$ of Lemma 1, where also $\boldsymbol{C}_Z$ is defined. We assume that $P_{ti}(x)$ is the value of the overall allele frequency $P_{ti}(x, a)$ of some (arbitrary) of the two alleles $a = 1, 2$ at locus $x$ and subpopulation $i = 1, \ldots, s$, $P_t(x) = \sum_{i=1}^{s} \gamma_i P_{ti}(x)$, $P^0_{ti}(x) = P_{ti}(x) - P_t(x)$ and $\boldsymbol{P}^0_t(x) = (P^0_{ti}(x); i = 1, \ldots, s)^T$.

It will be convenient to condition on the allele frequency spectrum $\mathcal{P}_t = \{P_t(x);\ x = 1, \ldots, n\}$, writing

$$\begin{aligned} G^w_{ST,t} &= (\bar{Z} + \epsilon_Z)/(\bar{Y} + \epsilon_Y), \\ N^w_{eV,t} &= (\bar{Y} + \epsilon_Y)/(\bar{X} + \epsilon_X), \end{aligned} \tag{142}$$

where

$$\begin{aligned} \bar{X} &= E_c(X|\mathcal{P}_t) \\ &= \sum_x P_t(x)(1 - P_t(x))\left(\mathrm{tr}\left(\boldsymbol{C}_X(\boldsymbol{V}_t(x) - \boldsymbol{\varsigma}_t(x))\right) + \mathrm{tr}\left(\boldsymbol{C}'_X(\boldsymbol{\Sigma}_t(x) - \boldsymbol{\zeta}_t(x))\right)\right), \\ \bar{Y} &= E_c(Y|\mathcal{P}_t) = \sum_x P_t(x)(1 - P_t(x))\left(1 - \mathrm{tr}\left(\boldsymbol{C}_Y \boldsymbol{V}_t(x)\right) + \boldsymbol{c}_Y(x)\boldsymbol{\mu}_t\right), \\ \bar{Z} &= E_c(Z|\mathcal{P}_t) = \sum_x P_t(x)(1 - P_t(x))\mathrm{tr}(\boldsymbol{C}_Z \boldsymbol{V}_t(x)), \end{aligned}$$

can be deduced from Lemma 1, using the same definitions of $\boldsymbol{C}_X, \boldsymbol{C}'_X$ and $\boldsymbol{C}_Y$ as there. Moreover, $\boldsymbol{V}_t(x)$, $\boldsymbol{\Sigma}_t(x)$, $\boldsymbol{c}_Y(x) = (1 - 2P_t(x))(\boldsymbol{w} - \boldsymbol{\gamma})$, $\boldsymbol{\mu}_t(x)$, $\boldsymbol{\zeta}_t(x)$ and $\boldsymbol{\varsigma}_t(x)$ are the values of $\boldsymbol{V}_t$, $\boldsymbol{\Sigma}_t$, $\boldsymbol{c}_Y$, $\boldsymbol{\mu}_t$, $\boldsymbol{\zeta}_t$ and $\boldsymbol{\varsigma}_t$ at locus $x$. The remaining three quantities of (142) are the residual terms

$$\begin{aligned} \epsilon_X = 2\sum_x E\left(\left(P^w_{t+1}(x) - P^w_t(x)\right)^2 | P^w_t(x)\right) \\ -2\sum_x E_c\left(E\left(\left(P^w_{t+1}(x) - P^w_t(x)\right)^2 | P^w_t(x)\right) | P_t(x)\right), \end{aligned}$$

$$\begin{aligned}\epsilon_Y &= \sum_x \boldsymbol{c}_Y(x)(\boldsymbol{P}_t^0(x,a) - \boldsymbol{\mu}_t P_t(x)(1-P_t(x))), \\ &\quad - \sum_x \mathrm{tr}\left(\boldsymbol{C}_Y(\boldsymbol{P}_t^0(x)(\boldsymbol{P}_t^0(x))^T - \boldsymbol{V}_t(x)P_t(x)(1-P_t(x)))\right), \\ \epsilon_Z &= \sum_x \mathrm{tr}\left(\boldsymbol{C}_Z(\boldsymbol{P}_t^0(x)(\boldsymbol{P}_t^0(x))^T - \boldsymbol{V}_t(x)P_t(x)(1-P_t(x)))\right).\end{aligned} \tag{143}$$

It follows from the definitions of $G_{ST}^{\mathrm{appr},w}$ and $N_{eV}^{\mathrm{appr},w}$ in (51) and (50) that we can write

$$\begin{aligned}G_{ST}^{\mathrm{appr},w} &= Z^{\mathrm{appr}}/Y^{\mathrm{appr}}, \\ N_{eV}^{\mathrm{appr},w} &= Y^{\mathrm{appr}}/X^{\mathrm{appr}}\end{aligned} \tag{144}$$

with

$$\begin{aligned}X^{\mathrm{appr}} &= \sum_x P_t(x)(1-P_t(x))\left(\mathrm{tr}\left(\boldsymbol{C}_X\boldsymbol{V}\right) + \mathrm{tr}\left(\boldsymbol{C}_X'\boldsymbol{\Sigma}\right)\right), \\ Y^{\mathrm{appr}} &= \sum_x P_t(x)(1-P_t(x))\left(1-\mathrm{tr}\left(\boldsymbol{C}_Y\boldsymbol{V}\right)\right), \\ Z^{\mathrm{appr}} &= \sum_x P_t(x)(1-P_t(x))\mathrm{tr}(\boldsymbol{C}_Z\boldsymbol{V}).\end{aligned}$$

Taking the difference of (142) and (144), we find that

$$\begin{aligned}G_{ST,t}^{w} - G_{ST}^{\mathrm{appr},w} &= \frac{\bar{Z}}{\bar{Y}} - \frac{Z^{\mathrm{appr}}}{Y^{\mathrm{appr}}} + \frac{\bar{Z}+\epsilon_Z}{\bar{Y}+\epsilon_Y} - \frac{\bar{Z}}{\bar{Y}} \\ &\approx \frac{\bar{Z}}{\bar{Y}} - \frac{Z^{\mathrm{appr}}}{Y^{\mathrm{appr}}} + \frac{1}{\bar{Y}}\epsilon_Z - \frac{\bar{Z}}{\bar{Y}^2}\epsilon_Y - \frac{1}{\bar{Y}^2}\epsilon_Y\epsilon_Z + \frac{\bar{Z}}{\bar{Y}^3}\epsilon_Y^2,\end{aligned} \tag{145}$$

where in the last step we made a second order Taylor expansion. The first term on the right hand side of (145) can be further approximated as

$$\begin{aligned}\frac{\bar{Z}}{\bar{Y}} - \frac{Z^{\mathrm{appr}}}{Y^{\mathrm{appr}}} &= \frac{1}{\bar{Y}}(\bar{Z} - Z^{\mathrm{appr}}) - \frac{Z^{\mathrm{appr}}}{\bar{Y}Y^{\mathrm{appr}}}(\bar{Y} - Y^{\mathrm{appr}}) \\ &\approx \frac{1}{Y^{\mathrm{appr}}}(\bar{Z} - Z^{\mathrm{appr}}) - \frac{Z^{\mathrm{appr}}}{(Y^{\mathrm{appr}})^2}(\bar{Y} - Y^{\mathrm{appr}}) \\ &\approx \frac{2}{H_T^{\mathrm{eq}}(1 - tr\,(\boldsymbol{C}_Y\boldsymbol{V}))} \cdot \frac{1}{n}(\bar{Z} - Z^{\mathrm{appr}}) \\ &\quad - \frac{2tr(\boldsymbol{C}_Z V)}{H_T^{\mathrm{eq}}(1 - tr\,(\boldsymbol{C}_Y V))^2} \cdot \frac{1}{n}(\bar{Y} - Y^{\mathrm{appr}}) \\ &:= \frac{C_1}{n}(\bar{Z} - Z^{\mathrm{appr}}) - \frac{C_2}{n}(\bar{Y} - Y^{\mathrm{appr}}),\end{aligned} \tag{146}$$

where in the second step we replaced $\bar{Y}$ and $\bar{Z}$ by $Y^{\text{appr}}$ and $Z^{\text{appr}}$, in the third step we approximated the gene diversity

$$H_{Tt} = H_{T_t}^{\boldsymbol{\gamma}} = \frac{2}{n}\sum_x P_t(x)(1 - P_t(x)) \approx H_T^{\text{eq}},$$

in (76) by its quasi equilibrium limit (20), which is accurate, by a Law of Large Numbers argument, for large $n$. In the last step of (147) we introduced the constants $C_1$ and $C_2$ in order to simplify notation.

By the definition of $\epsilon_Y$ and $\epsilon_Z$ we have $E_c(\epsilon_Y|\mathcal{P}_t) = E_c(\epsilon_Z|\mathcal{P}_t) = 0$, and, since $\bar{Y}$ and $\bar{Z}$ are both functions of $\mathcal{P}_t$, it follows that the first two terms of the last line of (145) have zero mean. Since all loci are in linkage equilibrium, the terms on the right hand sides of all three equations in (143) are independent for different $x$. By Lemma 2 it then follows, after some computations, that

$$|E_c\left(-\frac{1}{\bar{Y}^2}\epsilon_Y\epsilon_Z + \frac{\bar{Z}}{\bar{Y}^3}\epsilon_Y^2\right)| \le \frac{C_3'}{n} \tag{147}$$

for some constant $C_3'$, independently of $n$. Combining (145), (147) and (147), using $P_t(x)(1 - P_t(x)) \le 1/4$, $|\text{tr}(\boldsymbol{C}_Y(\boldsymbol{V}_t(x) - \boldsymbol{V})| \le |\boldsymbol{C}_Y|_1|\boldsymbol{V}_t(x) - \boldsymbol{V}|_\infty$ and analogous estimates for all $x = 1, \ldots, n$, we find that

$$\begin{aligned}|E_c(G_{ST,t}^w) - G_{ST}^{\text{appr},w}| &\le \frac{C_1}{n}E_c|\bar{Z} - Z^{\text{appr}}| + \frac{C_2}{n}E_c|\bar{Y} - Y^{\text{appr}}| + \frac{C_3'}{n}\\ &\le \frac{C_1|C_Z|_1 + C_2|C_Y|_1}{4n}\sum_x E_c(|\boldsymbol{V}_t(x) - \boldsymbol{V}|_\infty)\\ &\quad + \frac{C_2|\boldsymbol{w} - \boldsymbol{\gamma}|_1}{4n}\sum_x E_c(|\boldsymbol{\mu}_t(x)|_\infty) + \frac{C_3'}{n}.\end{aligned}$$

Then we use $|\boldsymbol{C}_Z|_1 \le 2$ and $|\boldsymbol{C}_Y|_1 \le |\boldsymbol{w} - \boldsymbol{\gamma}|_1^2$ and let $t \to \infty$, in order to deduce that

$$\begin{aligned}\lim_{t\to\infty}|E_c(G_{ST,t}^w) - G_{ST}^{\text{appr},w}| &\le \frac{2C_1 + C_2|\boldsymbol{w} - \boldsymbol{\gamma}|_1^2}{4}|\Delta\boldsymbol{V}|^{\text{eq}} + \frac{C_2|\boldsymbol{w} - \boldsymbol{\gamma}|_1}{4}|\boldsymbol{\mu}|^{\text{eq}} + \frac{C_3'}{n}\\ &=: C_1'|\Delta\boldsymbol{V}|^{\text{eq}} + C_2'|\boldsymbol{\mu}|^{\text{eq}} + \frac{C_3'}{n},\end{aligned}$$

since, for instance, the limit $\lim_{t\to\infty} E_c(|\boldsymbol{V}_t(x) - \boldsymbol{V}|_\infty) = |\Delta\boldsymbol{V}|^{\text{eq}}$ in (48) exists for all $x$. As similar analysis shows that

$$\begin{aligned}&\lim_{t\to\infty}|E_c(N_{eV,t}^w) - N_{eV}^{\text{appr},w}| \le \frac{C_3}{n}\lim_{t\to\infty}E_c|\bar{Y} - Y^{\text{appr}}|\\ &\quad + \frac{C_4}{n}\lim_{t\to\infty}E_c|\bar{X} - X^{\text{appr}}| + \frac{C_9'}{n}\\ &\quad \le \frac{C_3|\boldsymbol{C}_Y|_1 + C_4|\boldsymbol{C}_X|_1}{4}|\Delta\boldsymbol{V}|^{\text{eq}} + \frac{C_4|\boldsymbol{C}_X'|_1}{4}|\Delta\boldsymbol{\Sigma}|^{\text{eq}} + \frac{C_3|\boldsymbol{w} - \boldsymbol{\gamma}|_1}{4}|\boldsymbol{\mu}|^{\text{eq}}\end{aligned}$$

$$
\begin{aligned}
&+\frac{C_4|\boldsymbol{C}_X|_1}{4}|\boldsymbol{\varsigma}|^{\mathrm{eq}}+\frac{C_4|\boldsymbol{C}'_X|_1}{4}|\boldsymbol{\zeta}|^{\mathrm{eq}}+\frac{C'_9}{n}\\
&\le \frac{C_3|\boldsymbol{w}-\boldsymbol{\gamma}|_1^2+4C_4|\boldsymbol{w}-\boldsymbol{\gamma}|_1^2}{4}|\Delta\boldsymbol{V}|^{\mathrm{eq}}+\frac{2C_4}{4}|\Delta\boldsymbol{\Sigma}|^{\mathrm{eq}}+\frac{C_3|\boldsymbol{w}-\boldsymbol{\gamma}|_1}{4}|\boldsymbol{\mu}|^{\mathrm{eq}}\\
&+\frac{4C_4|\boldsymbol{w}-\boldsymbol{\gamma}|_1^2}{4}|\boldsymbol{\varsigma}|^{\mathrm{eq}}+\frac{2C_4}{4}|\boldsymbol{\zeta}|^{\mathrm{eq}}+\frac{C'_9}{n}\\
&=: C'_4|\Delta\boldsymbol{V}|^{\mathrm{eq}}+C'_5|\Delta\boldsymbol{\Sigma}|^{\mathrm{eq}}+C'_6|\boldsymbol{\mu}|^{\mathrm{eq}}+C'_7|\boldsymbol{\varsigma}|^{\mathrm{eq}}+C'_8|\boldsymbol{\zeta}|^{\mathrm{eq}}+\frac{C'_9}{n},
\end{aligned}
$$

where

$$
|\boldsymbol{\varsigma}|^{\mathrm{eq}}=\lim_{t\to\infty}E_c(|\boldsymbol{\varsigma}_t|_\infty) \tag{148}
$$

is an asymptotic upper bound for the remainder terms $\boldsymbol{\varsigma}_t(x)$, defined in the same way as (45)–(47), and

$$
\begin{aligned}
C_3&=2/\left(H_T^{\mathrm{eq}}(\mathrm{tr}(\boldsymbol{C}_X\boldsymbol{V})+\mathrm{tr}(\boldsymbol{C}'_X\boldsymbol{\Sigma}))\right),\\
C_4&=2\left(1-\mathrm{tr}(\boldsymbol{C}_Y\boldsymbol{V})\right)/\left(H_T^{\mathrm{eq}}(\mathrm{tr}(\boldsymbol{C}_X\boldsymbol{V})+\mathrm{tr}(\boldsymbol{C}'_X\boldsymbol{\Sigma}))^2\right).
\end{aligned}
$$

□

## Appendix F: Details from Sect. 11

*Proof of Proposition 4* Let $Q_{ij,kl}$ denote the probability that two different genes from subpopulations $i$ and $j$ have their parents in subpopulations $k$ and $l$ respectively, and let $p_{ijk}$ be the coalescence probability defined in (84).

It is possible to compute $q_{t+1,ij}$ by conditioning on the parental subpopulation $k$ and $l$ one generation back in time, and then look at the ancestry of the parents $t$ generations back in time. Since coalescence can only appear when $k=l$, we find that

$$
q_{t+1,ij}=\sum_{k,l}Q_{ij,kl}(1-p_{ijk})^{\{k=l\}}q_{t,kl}.
$$

This equals the recursion in (82), with

$$
D_{ij,kl}=Q_{ij,kl}(1-p_{ijk})^{\{k=l\}}. \tag{149}
$$

On the other hand, we can rewrite the gene diversity recursion (26) as

$$
E(H_{t+1,ij}|\boldsymbol{P}_t)=\left(1-\frac{1}{2Nu_i}\right)^{\{i=j\}}\sum_{k,l}Q_{ij,kl}(1-p_{ijk})^{\{k=l\}}\frac{H_{tkl}}{\left(1-\frac{1}{2Nu_k}\right)^{\{k=l\}}},
$$

since $(1-1/(2Nu_i))^{\{i=j\}}$ is the probability that two genes, drawn with replacement from subpopulations $i$ and $j$ in generation $t+1$ are different, and $H_{tkl}/(1-$

$1/(2Nu_k))^{\{k=l\}}$ is the probability that two different genes from subpopulations $k$ and $l$ in generation $t$ have different alleles. Hence we see from (26) that

$$A_{ij,kl} = \frac{\left(1 - \frac{1}{2Nu_i}\right)^{\{i=j\}}}{\left(1 - \frac{1}{2Nu_k}\right)^{\{k=l\}}} Q_{ij,kl}(1 - p_{ijk})^{\{k=l\}},$$

from which (83) follows. □

We will derive explicit expressions of the matrix elements $D_{ij,kl}$ in Proposition 4. To this end, one could either calculate the coefficients $U_{ij,kl}$ of the covariance matrix expansion (25), and then use Propositions 1 and 4 in order to find $D_{ij,kl}$. Alternatively, one may employ coalescence probabilities and obtain the elements of $\boldsymbol{D}$ directly from (82). We use this latter approach in order to prove the following:

**Proposition 7** *Asymptotically, for large populations and reproduction scenario 2, the elements of* $\boldsymbol{D}$ *have the form*

$$D_{ij,kl} = b_{ik}\left(\frac{b_{il} - \frac{1_{\{k=l\}}}{2Nu_i}}{1 - \frac{1}{2Nu_i}}\right)^{\{i=j\}} b_{jl}^{\{i\neq j\}}\left(1 - p_{ijk}\right)^{\{k=l\}} + o(N^{-1}), \qquad (150)$$

*where* $p_{ijk}$ *is the coalescence probability (84) that two genes from subpopulations* $i$ *and* $j$*, that have their parents in* $k$*, have the same parent, and*

$$\sigma_{ijk}(N) = \frac{1}{m_{ki}m_{kj}} \cdot \begin{cases} E\left(\nu_{ki}^l(\nu_{ki}^l - 1)\right), & i = j, \\ E\left(\nu_{ki}^l \nu_{kj}^l\right), & i \neq j. \end{cases} \qquad (151)$$

*For reproduction scenario 3 with* $\alpha_i \equiv \infty$*, it holds that*

$$D_{ij,kl} = b_{ik}b_{jl}\left(1 - p_{ijk}\right)^{\{k=l\}} + o(N^{-1}), \qquad (152)$$

*with coalescence probability* $p_{ijk} = 1/(2N_{ek})$*, so that* $\sigma_{ijk}(N)$ *in (84) equals*

$$\sigma_{ijk}(N) = \frac{Nu_k}{N_{ek}}. \qquad (153)$$

Nagylaki (2000) has derived a recursion that generalizes (152) when $N_{ek} = Nu_k$ for probabilities that concern not only the time when but also the subpopulation where coalescence of two genes from subpopulations $i$ and $j$ occurs. The constant $\sigma_{ijk}(N)$ was defined in Hössjer (2011). As mentioned in Sect 11.1, it can be interpreted as the coalescence rate of a pair of lines from subpopulations $i$ and $j$, when both of these migrate backwards to $k$.

*Proof of Proposition 7* In order to establish (150) and (152), we will use (149), and hence we need to find expressions for $Q_{ij,kl}$ and $p_{ijk}$. Starting with reproduction scenario 2, we have

$$Q_{ij,kl} = \begin{cases} b_{ik}b_{jl}, & i \neq j, \\ 2Nu_i b_{ik}(2Nu_i b_{ik} - 1)/(2Nu_i(2Nu_i - 1)), & i = j, k = l, \\ 2Nu_i b_{ik} \cdot 2Nu_i b_{il}/(2Nu_i(2Nu_i - 1)), & i = j, k \neq l, \end{cases} \tag{154}$$

since the two genes are drawn without replacement, and an exact fraction $b_{ik}$ of the parents of the offspring genes of subpopulation $i$ originate from subpopulation $k$, and similarly, an exact fraction $b_{jl}$ of the genes in $j$ to have their parent in $l$. We can rewrite (154) more compactly as

$$Q_{ij,kl} = b_{ik}\left(\frac{b_{il} - \frac{1_{\{k=l\}}}{2Nu_i}}{1 - \frac{1}{2Nu_i}}\right)^{\{i=j\}} b_{jl}^{\{i\neq j\}}.$$

It follows for instance from Hössjer (2011) that the coalescence probability $p_{ijk}$ has the form (84), and this completes the proof of (150).

For reproduction scenario 3 with $\alpha_i \equiv \infty$, we simply have

$$Q_{ij,kl} = b_{ik}b_{jl},$$

since the parental subpopulations are drawn independently for two genes of subpopulations $i$ and $j$, from the probability distributions corresponding to rows $i$ and $j$ of $\boldsymbol{B}$. Moreover, the coalescence probability is $1/(2N_{ek})$, since this is the probability that the two parents in $k$ originate from the same gene of a breeder, and this completes the proof of (152). □

*Proof of Theorem 3* We will use (87) in order to prove (88). By Perron–Frobenius' Theorem, there exists a unique largest eigenvalue $\lambda$ of $\boldsymbol{D}$, with corresponding left and right eigenvectors $\boldsymbol{l} = (l_{ij})$ and $\boldsymbol{r} = (r_{ij})$, which can be normalized so that

$$\sum_{ij} l_{ij} = 1,$$
$$\sum_{ij} l_{ij}r_{ij} = 1.$$

By a Jordan decomposition of $\boldsymbol{D}$, it follows that

$$\boldsymbol{D}^\tau = \lambda^\tau \boldsymbol{rl} + o(\lambda^\tau) \text{ as } \tau \to \infty.$$

Our asymptotic analysis $N \to \infty$ is equivalent to letting the perturbation parameter

$$\varepsilon = \frac{1}{2N}$$

tend to zero. In order to highlight the dependence of $\boldsymbol{D} = \boldsymbol{D}(\varepsilon)$ on $\varepsilon$, we Taylor expand its elements around $\varepsilon = 0$, as

$$D_{ij,kl} = D_{ij,kl}(\varepsilon) = b_{ik}b_{jl} + \dot{D}_{ij,kl}\varepsilon + o(\varepsilon).$$

It follows from (150) that $\dot{\boldsymbol{D}} = (\dot{D}_{ij,kl})$ has elements

$$\dot{D}_{ij,kl} = -1_{\{k=l\}} u_k^{-1} b_{ik} b_{jl} \sigma_{ijk} + 1_{\{i=j\}} u_i^{-1} b_{ik} (b_{il} - 1_{\{k=l\}})$$

for reproduction scenario 2 and

$$\dot{D}_{ij,kl} = -1_{\{k=l\}} u_k^{-1} b_{ik} b_{jl} \sigma_{ijk}$$

for reproduction scenario 3 with $\alpha_i \equiv \infty$. Clearly, $\boldsymbol{D}(0) = \boldsymbol{B} \otimes \boldsymbol{B}$ is the Kronecker product of $\boldsymbol{B}$ with itself for either reproduction scenario. It has largest eigenvalue $\lambda(0) = 1$, since $\boldsymbol{B}$ is the transition matrix of an irreducible Markov chain, with a unique largest eigenvalue 1. Moreover, the form of the left and right eigenvectors $\boldsymbol{l} = \boldsymbol{l}(\varepsilon)$ and $\boldsymbol{r} = \boldsymbol{r}(\varepsilon)$ can be deduced from the left and right eigenvectors of $\boldsymbol{B}$ when $\varepsilon = 0$, as

$$\begin{aligned} l_{ij}(0) &= \gamma_i \gamma_j, \\ r_{ij}(0) &= 1. \end{aligned}$$

It follows from perturbation theory of matrices (see for instance Nagylaki (1980) and Van der AA et al. 2007), that

$$\lambda(\varepsilon) = 1 + \dot{\lambda}\varepsilon + o(\varepsilon) \text{ as } \varepsilon \to 0,$$

where

$$\begin{aligned} \dot{\lambda} &= \boldsymbol{l}(0)\dot{\boldsymbol{D}}\boldsymbol{r}(0) \\ &= -\sum_{ijk} \gamma_i \gamma_j u_k^{-1} b_{ik} b_{jl} \sigma_{ijk} + \sum_{ikl} \gamma_i^2 u_i^{-1} b_{ik} (b_{il} - 1_{\{k=l\}}) \\ &= -C + \sum_i \gamma_i^2 u_i^{-1} (1 - 1) \\ &= -C, \end{aligned}$$

for reproduction scenario 2, with $C$ as defined in (89). A similar (but simpler) analysis shows that $\dot{\lambda} = -C$ for reproduction scenario 3 with $\alpha_i \equiv \infty$. In view of (87), this implies

$$\begin{aligned} N_{e\pi} &= \frac{1}{2} \boldsymbol{W}_T (\boldsymbol{I} - \boldsymbol{D})^{-1} \underline{\mathbf{1}} \\ &= \frac{1}{2} \boldsymbol{W}_T \left( \sum_{\tau=0}^{\infty} \boldsymbol{D}^{\tau} \right) \underline{\mathbf{1}} \\ &= \frac{1}{2} \boldsymbol{W}_T \left( \sum_{\tau=0}^{\infty} \left( \lambda^{\tau} \boldsymbol{r}\boldsymbol{l} + o(\lambda^{\tau}) \right) \right) \underline{\mathbf{1}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{\tau=0}^{\infty}\left((\boldsymbol{W}_T\boldsymbol{r}\boldsymbol{l}\underline{\mathbf{1}})\lambda^\tau + o(\lambda^\tau)\right)\\
&= \frac{1}{2}\sum_{\tau=0}^{\infty}\left(\boldsymbol{W}_T(\underline{\mathbf{1}} + o(1))\lambda^\tau + o(\lambda^\tau)\right)\\
&= \frac{1}{2}\sum_{\tau=0}^{\infty}\left(\lambda^\tau + o(\lambda^\tau)\right)\\
&= \frac{1}{2(1-\lambda)}(1+o(1))\\
&= \frac{1}{2C\varepsilon}(1+o(1))\\
&= \frac{N}{C}(1+o(1))
\end{aligned}
\tag{155}
$$

as $\varepsilon \to 0$, or equivalently, as $N \to \infty$, thereby proving (88). In the fifth equality of (156) we used that $\boldsymbol{r} = \boldsymbol{r}(\varepsilon) = \underline{\mathbf{1}} + o(1)$ as $\varepsilon \to 0$, and in the sixth equality $\boldsymbol{W}_T\underline{\mathbf{1}} = \sum_{i,j} w_i w_j = 1$, regardless of the choice of weight vector $\boldsymbol{w}$.

We now turn to the proof of (90). It follows from Table 3 that $\|\boldsymbol{U}\| = O(N^{-1})$ for both reproduction scenarios 2 and 3 (with $\alpha_i \equiv \infty$). Invoking the upper part of (33) and (38), we deduce that

$$
\mathrm{vec}(\boldsymbol{\Sigma}) = \boldsymbol{U}\underline{\mathbf{1}}\left(1 + O(N^{-1}\mathrm{Mixtime})\right) = \boldsymbol{U}\underline{\mathbf{1}}\left(1 + O(N^{-1}\right),
$$

where the last step follows from Proposition 3 and the fact that the migration rates are kept fixed. Inserting the last expression into (50), we find that

$$
N_{eV}^{\mathrm{appr}} = \frac{N}{C'} + o(N), \tag{156}
$$

where

$$
C' = 2N\sum_{i,j=1}^{s}\gamma_i\gamma_j(\boldsymbol{U}\underline{\mathbf{1}})_{ij}. \tag{157}
$$

It thus remains to verify, for both reproduction scenarios, that $C' = C$. Starting with reproduction scenario 2, we find from Table 3 that

$$
(\boldsymbol{U}\underline{\mathbf{1}})_{ij} = \sum_{k,l=1}^{s} U_{ij,kl} = \sum_{k=1}^{s}\frac{C_{kij}u_k}{2Nu_iu_j}, \tag{158}
$$

with $C_{kij} = \mathrm{Cov}(\nu_{ki}^l, \nu_{kj}^l)$. By the assumptions of the theorem, the quantities $\sigma_{ijk}(N)$ in (151) will converge as $N \to \infty$. Since the migration rates in $\boldsymbol{M}$ are fixed, it follows that the covariances $C_{kij} = C_{kij}(N)$ will converge as well. With a slight abuse of

notation, we write $C_{kij}$ also for the asymptotic $N \to \infty$ limits. Inserting (158) into (157), we find that

$$C' = \sum_{i,j,k=1}^{s} \gamma_i \gamma_j \frac{C_{kij} u_k}{u_i u_j}.$$

On the other hand, it follows from the definition of $\sigma_{ijk}$ in (151), that each covariance term $C_{kij}$ can be rewritten as

$$C_{kij} = \sigma_{ijk} m_{ki} m_{kj} - m_{ki} m_{kj} + m_{ki} 1_{\{i=j\}}. \tag{159}$$

Inserting (159) into (157), it follows, after some computations, that

$$\begin{aligned} C' &= \sum_{ijk} \gamma_i \gamma_j u_k^{-1} b_{ik} b_{jk} \left( \sigma_{kij} - 1 + m_{ki}^{-1} 1_{\{i=j\}} \right) \\ &= \sum_{ijk} \gamma_i \gamma_j u_k^{-1} b_{ik} b_{jk} \sigma_{kij} - \sum_{ijk} \gamma_i \gamma_j u_k^{-1} b_{ik} b_{jk} + \sum_{ik} \gamma_i^2 u_k^{-1} m_{ki}^{-1} b_{ik}^2 \\ &= C - \sum_k u_k^{-1} \gamma_k^2 + \sum_i u_i^{-1} \gamma_i^2, \\ &= C, \end{aligned}$$

and in view of (156), this proves (90).

For reproduction scenario 3 with $\alpha_i \equiv \infty$, it follows from Table 3 that

$$(\boldsymbol{U}\underline{\mathbf{1}})_{ij} = \sum_{k,l=1}^{s} U_{ij,kl} = \sum_{k=1}^{s} b_{ik} b_{jk} \left( \frac{1}{2N_{ek}} - \frac{1}{2Nu_k} \right) + \frac{1_{\{i=j\}}}{2Nu_i}.$$

Insertion of this expression into (157) leads to

$$\begin{aligned} C' &= 2N \sum_{i,j,k=1}^{s} \gamma_i \gamma_j b_{ik} b_{jk} \left( \frac{1}{2N_{ek}} - \frac{1}{2Nu_k} \right) + 2N \sum_{i=1}^{s} \frac{\gamma_i^2}{2Nu_i} \\ &= 2N \sum_{k=1}^{s} \gamma_k^2 \left( \frac{1}{2N_{ek}} - \frac{1}{2Nu_k} \right) + 2N \sum_{i=1}^{s} \frac{\gamma_i^2}{2Nu_i} \\ &= \sum_{k=1}^{s} u_k^{-1} \gamma_k^2 \cdot \frac{2Nu_k}{2N_{ek}} \\ &= \sum_{k=1}^{s} u_k^{-1} \gamma_k^2 \sigma_k \\ &= C, \end{aligned} \tag{160}$$

where $\sigma_k = \sigma_{ijk}$ is defined in (153). The last step of (160) follows easily by adding a term $\sigma_k$ on both sides of Eq. (91). □

Given two random variables $X$ and $Y$, we put $E_0(Y/X)^* = E_0(Y)/E_0(X)$, where $E_0(X) = E(X|\boldsymbol{P}_0 = P_0\underline{\mathbf{1}})$, a prediction of $Y/X$ given that the allele frequencies of the founder generation are the same in all subpopulations. The following proposition shows that $\bar{f}_{ST}^w$ and $f_{ST}^w$ are weighted averages over $t$ of $E_0(\bar{F}_{ST,t}^w)^*$ and $E_0(F_{ST,t}^w)^*$ respectively:

**Proposition 8** *The matrix $\bar{\boldsymbol{H}}_t = (\bar{H}_{tij})_{i,j=1}^s$ of gene diversities, defined for a pair of distinct genes, satisfies*

$$E_0\left(vec(\bar{\boldsymbol{H}}_t)\right) = 2P_0(1-P_0)\boldsymbol{D}^t\left(\underline{\mathbf{1}} + O(N^{-1})\right), \tag{161}$$

*and the fixation index in (98) is a weighted average*

$$\bar{f}_{ST}^w = \sum_{t=0}^{\infty}\bar{\omega}_t E_0\left(\bar{F}_{ST,t}^w\right)^* + O(N^{-1}) = \sum_{t=0}^{\infty}\bar{\omega}_t\frac{E_0\left(\bar{H}_{Tt}^w - \bar{H}_{St}^w\right)}{E_0\left(\bar{H}_{Tt}^w\right)} + O(N^{-1}), \tag{162}$$

*of predictions $E_0\left(\bar{F}_{ST,t}^w\right)^*$ of the fixation index (95) over different time horizons $t$, with weights*

$$\bar{\omega}_t = \frac{\boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}}{\sum_{\tau=0}^{\infty}\boldsymbol{W}_T\boldsymbol{D}^\tau\underline{\mathbf{1}}}.$$

*Analogously, the matrix $\boldsymbol{H}_t = (H_{tij})_{i,j=1}^s$ of gene diversities, when the pair of genes is drawn with replacement, satisfies*

$$E_0\left(vec(\boldsymbol{H}_t)\right) = 2P_0(1-P_0)\boldsymbol{A}^t\underline{\mathbf{1}}, \tag{163}$$

*and the fixation index (99) is a weighted average*

$$f_{ST}^w = \sum_{t=0}^{\infty}\omega_t E_0\left(F_{ST,t}^w\right)^* = \sum_{t=0}^{\infty}\omega_t\frac{E_0\left(H_{Tt}^w - H_{St}^w\right)}{E_0\left(H_{Tt}^w\right)}, \tag{164}$$

*with weights*

$$\omega_t = \frac{\boldsymbol{W}_T\boldsymbol{A}^t\underline{\mathbf{1}}}{\sum_{\tau=0}^{\infty}\boldsymbol{W}_T\boldsymbol{A}^\tau\underline{\mathbf{1}}}. \tag{165}$$

It is implicit from the proof of Theorem 3 that the weights (165) correspond to a probability distribution with mean $O(N)$, as discussed in Subsection 11.2.

*Proof of Proposition 8* By means of an expansion $(\boldsymbol{I}-\boldsymbol{D})^{-1} = \sum_{t=0}^{\infty}\boldsymbol{D}^t$, it is clear that (98) can be rewritten as

$$\bar{f}_{ST}^w = \sum_{t=0}^{\infty}\bar{\omega}_t\frac{(\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{D}^t\underline{\mathbf{1}}}{\boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}}, \tag{166}$$

given an assumption that the $\mu \to 0$ approximation in (98) is exact. On the other hand, as in the proof of (82), it follows that we get a gene diversity recursion

$$E\left(\mathrm{vec}(\bar{\boldsymbol{H}}_{t+1})|\boldsymbol{P}_t\right) = \boldsymbol{D}\mathrm{vec}(\bar{\boldsymbol{H}}_t) \tag{167}$$

instead of (29) when two genes are drawn without replacement. We prove (161) by repeated use of (167). This yields

$$\begin{aligned} E_0(\mathrm{vec}(\bar{\boldsymbol{H}}_t)) &= E(\mathrm{vec}(\bar{\boldsymbol{H}}_t)|\boldsymbol{P}_0 = P_0\mathbf{1}) \\ &= \boldsymbol{D}^t\,\mathrm{vec}(\bar{\boldsymbol{H}}_0) \\ &= 2P_0(1-P_0)\boldsymbol{D}^t\left(\underline{\mathbf{1}} + O(N^{-1})\right), \end{aligned}$$

applying (94) with $t = 0$ in the last step. Invoking the definitions of $\bar{H}^w_{Tt}$ and $\bar{H}^w_{St}$ into (161), this yields

$$\begin{aligned} E_0\left(\bar{H}^w_{Tt} - \bar{H}^w_{Tt}\right) &= 2P_0(1-P_0)(\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{D}^t\underline{\mathbf{1}} + O(N^{-1}), \\ E_0\left(\bar{H}^w_{Tt}\right) &= 2P_0(1-P_0)\boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}\left(1 + O(N^{-1})\right), \end{aligned}$$

where the last step follows as in the proof of Theorem 3 (see in particular (156)), since

$$\begin{aligned} \boldsymbol{W}_T\boldsymbol{D}^t\left(\underline{\mathbf{1}} + O(N^{-1})\right) &= \lambda^t\boldsymbol{W}_T\boldsymbol{r}\boldsymbol{l}(\underline{\mathbf{1}} + O(N^{-1})) + o(\lambda^t) \\ &= \lambda^t\boldsymbol{W}_T\boldsymbol{r}(1 + O(N^{-1})) + o(\lambda^t) \\ &= \lambda^t(1 + O(N^{-1})) + o(\lambda^t) \\ &= \boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}\left(1 + O(N^{-1})\right). \end{aligned}$$

Hence it follows that

$$E_0\left(\bar{F}^w_{ST,t}\right) = \frac{(\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{D}^t\underline{\mathbf{1}} + O(N^{-1})}{\boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}\left(1 + O(N^{-1})\right)} = \frac{(\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{D}^t\underline{\mathbf{1}}}{\boldsymbol{W}_T\boldsymbol{D}^t\underline{\mathbf{1}}} + O(N^{-1}).$$

By inserting the last equation into (166) we arrive at (162).

Equations (163) and (164) are derived analogously, although the proof is simpler. The reason is that the $O(N^{-1})$ remainder terms vanish, since $\mathrm{vec}(\boldsymbol{H}_0) = 2P_0(1-P_0)\underline{\mathbf{1}}$ holds exactly when $\boldsymbol{P}_0 = P_0\mathbf{1}$. □

*Proof of Theorem 4* It will be convenient to rewrite (27) as

$$\boldsymbol{A} = \boldsymbol{B}\otimes\boldsymbol{B} - \boldsymbol{U} = \boldsymbol{G} - \boldsymbol{U}, \tag{168}$$

where $\boldsymbol{G} = (G_{ij,kl})$ has elements $G_{ij,kl} = b_{ik}b_{jl}$. The Jordan decomposition of $\boldsymbol{B}$ in Appendix A implies that $\boldsymbol{B}^0\boldsymbol{B}^{t-1} = \boldsymbol{B}^{t-1}\boldsymbol{B}^0 = (\boldsymbol{B}^0)^t$ for any non-negative integer $t$. Since $\boldsymbol{G}^0 = \boldsymbol{B}^0\otimes\boldsymbol{B}^0$ and $\boldsymbol{G} = \boldsymbol{B}\otimes\boldsymbol{B}$, it is easy to see that this implies

$$\boldsymbol{G}^0\boldsymbol{G}^{t-1} = \boldsymbol{G}^{t-1}\boldsymbol{G}^0 = (\boldsymbol{G}^0)^t. \tag{169}$$

A similar calculation as in the proof of Theorem 3 (see in particular (156)) yields

$$\sum_{t=0}^{\infty}(\boldsymbol{G}-\boldsymbol{U})^t\underline{\mathbf{1}} = (1-\lambda)^{-1}\underline{\mathbf{1}} + o\left((1-\lambda)^{-1}\right), \tag{170}$$

where $\lambda$ is the unique largest eigenvalue of $\boldsymbol{G}-\boldsymbol{U}$. We will also make use of the fact that

$$(\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{G} = (\boldsymbol{W}_T - \boldsymbol{W}_S)\boldsymbol{G}^0 = -\boldsymbol{W}_S\boldsymbol{G}^0, \tag{171}$$

which follows since $\boldsymbol{w} = \boldsymbol{\gamma}$ and

$$\begin{aligned}
\boldsymbol{W}_T\boldsymbol{G} &= \text{vec}\left((\boldsymbol{\gamma}\boldsymbol{B})\otimes(\boldsymbol{\gamma}\boldsymbol{B})\right)^T \\
&= \text{vec}(\boldsymbol{\gamma}\otimes\boldsymbol{\gamma})^T \\
&= \boldsymbol{W}_T, \\
\boldsymbol{W}_S\boldsymbol{G} &= \boldsymbol{W}_S\left((\mathbf{1}\boldsymbol{\gamma})\otimes(\mathbf{1}\boldsymbol{\gamma})\right) + \boldsymbol{W}_S\left((\mathbf{1}\boldsymbol{\gamma})\otimes\boldsymbol{B}^0\right) + \boldsymbol{W}_S(\boldsymbol{B}^0\otimes(\mathbf{1}\boldsymbol{\gamma})) + \boldsymbol{W}_S\boldsymbol{G}^0 \\
&= \boldsymbol{W}_T + 0 + 0 + \boldsymbol{W}_S\boldsymbol{G}^0 \\
&= \boldsymbol{W}_T + \boldsymbol{W}_S\boldsymbol{G}^0,
\end{aligned}$$

with $\mathbf{1}$ a column vector of length $s$, and

$$\begin{aligned}
\boldsymbol{W}_T\boldsymbol{G}^0 &= \text{vec}\left((\boldsymbol{\gamma}\boldsymbol{B}^0)\otimes(\boldsymbol{\gamma}\boldsymbol{B}^0)\right)^T \\
&= \text{vec}\,(\mathbf{0}\otimes\mathbf{0})^T \\
&= \mathbf{0}.
\end{aligned}$$

Based on these preliminaries, we can rewrite the numerator of (99) as

$$\begin{aligned}
&(\boldsymbol{W}_T - \boldsymbol{W}_S)\left(\boldsymbol{I} - (\boldsymbol{G}-\boldsymbol{U})\right)^{-1}\underline{\mathbf{1}} = (\boldsymbol{W}_T - \boldsymbol{W}_S)\sum_{t=0}^{\infty}(\boldsymbol{G}-\boldsymbol{U})^t\underline{\mathbf{1}} \\
&= (\boldsymbol{W}_S - \boldsymbol{W}_T)\sum_{t=0}^{\infty}\left(-\boldsymbol{G}^t + \sum_{\tau=0}^{t-1}\boldsymbol{G}^\tau\boldsymbol{U}(\boldsymbol{G}-\boldsymbol{U})^{t-\tau-1}\right)\underline{\mathbf{1}} \\
&= (\boldsymbol{W}_S - \boldsymbol{W}_T)\sum_{t=0}^{\infty}\sum_{\tau=0}^{t-1}(\boldsymbol{G}^0)^\tau\boldsymbol{U}(\boldsymbol{G}-\boldsymbol{U})^{t-\tau-1}\underline{\mathbf{1}} \\
&= (\boldsymbol{W}_S - \boldsymbol{W}_T)\sum_{\tau=0}^{\infty}(\boldsymbol{G}^0)^\tau\boldsymbol{U}\sum_{\alpha=0}^{\infty}(\boldsymbol{G}-\boldsymbol{U})^\alpha\underline{\mathbf{1}} \\
&= (1-\lambda)^{-1}(\boldsymbol{W}_S - \boldsymbol{W}_T)(\boldsymbol{I}-\boldsymbol{G}^0)^{-1}\boldsymbol{U}\underline{\mathbf{1}} + o\left(\|\boldsymbol{U}\|(1-\lambda)^{-1}\right),
\end{aligned}$$

using (169), (171) and the fact that $(\boldsymbol{W}_S - \boldsymbol{W}_T)\boldsymbol{G}^t\underline{\mathbf{1}} = (\boldsymbol{W}_S - \boldsymbol{W}_T)\underline{\mathbf{1}} = 0$ in the third step, a change of variables $\alpha = t - \tau - 1$ in the fourth step and (170) in the fifth step. Formula (170) also implies that the denominator of (99) equals

$$\boldsymbol{W}_T\left(\boldsymbol{I} - (\boldsymbol{G} - \boldsymbol{U})\right)^{-1}\underline{\mathbf{1}} = (1-\lambda)^{-1} + o\left((1-\lambda)^{-1}\right).$$

In view of (168), we obtain formula (100) by taking the ratio of the last two displayed equations. In order to prove that $F_{ST}^{\text{appr}}$ equals the right hand side of (100) as well, it follows, by the definition of $\boldsymbol{\Pi}$ in (34), that

$$\begin{aligned}
(\boldsymbol{W}_S\boldsymbol{\Pi})_{kl} &= \sum_{i,j=1}^{s} \gamma_i 1_{\{i=j\}}\Pi_{ij,kl} \\
&= \sum_{i=1}^{s} \gamma_i \Pi_{ii,kl} \\
&= \sum_{i=1}^{s} \gamma_i \left(1_{\{(i,i)=(k,l)\}} - \gamma_k 1_{\{i=l\}} - \gamma_l 1_{\{i=k\}} + \gamma_k\gamma_l\right) \\
&= \gamma_k 1_{\{k=l\}} - \gamma_k\gamma_l \\
&= (\boldsymbol{W}_S - \boldsymbol{W}_T)_{kl},
\end{aligned}$$

which we can rewrite in vector format, as

$$\boldsymbol{W}_S\boldsymbol{\Pi} = \boldsymbol{W}_S - \boldsymbol{W}_T. \tag{172}$$

A similar calculation shows that

$$\begin{aligned}
(\boldsymbol{G}^0\boldsymbol{\Pi})_{ij,kl} &= \sum_{m,n=1}^{s} (G^0)_{ij,mn}\Pi_{mn,kl} \\
&= \sum_{m,n=1}^{s} b^0_{im}b^0_{jn}\left(1_{\{(m,n)=(k,l)\}} - \gamma_k 1_{\{m=l\}} - \gamma_l 1_{\{n=k\}} + \gamma_k\gamma_l\right) \\
&= b^0_{ik}b^0_{jl} - \gamma_k b^0_{il}\sum_{n=1}^{s} b^0_{jn} - \gamma_l b^0_{jk}\sum_{m=1}^{s} b^0_{im} + \gamma_k\gamma_l\sum_{m=1}^{s} b^0_{im}\sum_{n=1}^{s} b^0_{jn} \\
&= b^0_{ik}b^0_{jl} \\
&= G^0_{ij,kl},
\end{aligned}$$

which we rewrite as

$$\boldsymbol{G}^0\boldsymbol{\Pi} = \boldsymbol{G}^0. \tag{173}$$

This yields

$$\begin{aligned} F_{ST}^{\text{appr},\boldsymbol{\gamma}} &= \sum_{i=1}^{s} \gamma_i V_{ii} \\ &= \boldsymbol{W}_S \text{vec}(\boldsymbol{V}) \\ &= \boldsymbol{W}_S \sum_{\tau=0}^{\infty} (\boldsymbol{G}^0 - \boldsymbol{\Pi U})^\tau \boldsymbol{\Pi U} \underline{\mathbf{1}} \\ &= \boldsymbol{W}_S \sum_{\tau=0}^{\infty} (\boldsymbol{G}^0)^\tau \boldsymbol{\Pi U} \underline{\mathbf{1}} + O(\|\boldsymbol{U}\|^2) \\ &= (\boldsymbol{W}_S - \boldsymbol{W}_T) \sum_{\tau=0}^{\infty} (\boldsymbol{G}^0)^\tau \boldsymbol{U} \underline{\mathbf{1}} + O(N^{-2}) \\ &= (\boldsymbol{W}_S - \boldsymbol{W}_T) \left( \boldsymbol{I} - \boldsymbol{G}^0 \right)^{-1} \boldsymbol{U} \underline{\mathbf{1}} + O(N^{-2}), \end{aligned}$$

where in the first step we used the definition (51) of $F_{ST}^{\text{appr},\boldsymbol{\gamma}}$, in the third step the expansion (36) of $\text{vec}(\boldsymbol{V})$ and in the fifth step the assumption $\|\boldsymbol{U}\| = O(N^{-1})$, (172), (173) and the second part of (171).

Finally, formula (101) is proved in the same way as (100), replacing $\boldsymbol{U}$ by $\bar{\boldsymbol{U}} = \boldsymbol{B} \otimes \boldsymbol{B} - \boldsymbol{D}$ everywhere. □

In order to compare the sizes of the fixation indeces when genes are drawn with and without replacement, we formulate the following result:

**Proposition 9** *The fixation index in (99) can be written as*

$$f_{ST}^w = \frac{\bar{h}_T^w - \bar{h}_S^w + \sum_{i=1}^s \frac{w_i - w_i^2}{2Nu_i} \bar{h}_{ii}}{\bar{h}_T^w - \sum_{i=1}^s \frac{w_i^2}{2Nu_i} \bar{h}_{ii}}. \tag{174}$$

*In particular, for a strong migration limit where $N \to \infty$ while the migration rates in $\boldsymbol{M}$ are kept fixed, it holds that*

$$\begin{aligned} f_{ST}^w &= \bar{f}_{ST}^w + \sum_{i=1}^{s} \frac{w_i - w_i^2}{2Nu_i} + o(N^{-1}) \\ &\overset{w_i = u_i = 1/s}{=} \bar{f}_{ST}^w + \frac{s-1}{2N} + o(N^{-1}). \end{aligned} \tag{175}$$

In order to illustrate this result, consider the island model under panmixia ($m = 1$), for which it is well known that $\bar{f}_{ST} = 0$ for the canonical and uniform weighting scheme $w_i = 1/s$, reflecting the fact that subpopulations on the average are identical. However, even under panmixia, there will still be small differences between subpopulations. It is shown in Hössjer (2013) (see also Latter and Sved 1981) that

the replacement version $f_{ST}$ of the fixation index captures this, in terms of a nonzero value $f_{ST} = (s-1)/(2N) + o(N^{-1})$. It also follows from Hössjer et al. (2013) or (68) that the replacement version of the quasi equilibrium approximation of the fixation index satisfies $F_{ST}^{\text{appr}} = (s-1)/(2N)$ under panmixia.

*Proof of Proposition 9* We have that

$$h_{ij}^w = \left(1 - \frac{1}{2Nu_i}\right)^{\{i=j\}} \bar{h}_{ij}^w,$$

since the probability is $(1 - 1/(2Nu_i))^{\{i=j\}}$ that two genes are not the same when drawn with replacement, and given this, they are different by state with probability $\bar{h}_{ij}^w$, as defined in (96). It then follows from (97), and the analogous definitions of $h_S^w$ and $h_T^w$ in terms of $h_{ij}^w$, that

$$h_S^w = \bar{h}_S^w - \sum_{i=1}^s \frac{w_i}{2Nu_i} \bar{h}_{ii},$$
$$h_T^w = \bar{h}_T^w - \sum_{i=1}^s \frac{w_i^2}{2Nu_i} \bar{h}_{ii}.$$

By inserting these two equations into (99), we arrive at (174).

When migration rates are fixed and $N \to \infty$, we have $\bar{h}_{ij} = \bar{h}_T^w(1 + O(N^{-1}))$ for all $i$, $j$, and hence (174) implies

$$f_{ST}^w = \bar{f}_{ST}^w + \frac{\bar{h}_T^w \sum_{i=1}^s \frac{w_i - w_i^2}{2Nu_i} + O(N^{-2})}{\bar{h}_T^{\boldsymbol{w}} \left(1 + O(N^{-1})\right)},$$

which can be simplified to (175). □

## References


Allendorf F, Ryman N (2002) The role of genetics in population viability analysis. In: Bessinger SR, McCullogh DR (eds) Population viability analysis. The University of Chicago Press, Chicago

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell, Malden

Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity 56:409–415

Brockwell PJ, Davis RA (1987) Time series: theory and methods. Springer, New York

Caballero A (1994) Developments in the prediction of effective population size. Heredity 73:657–679

Cannings C (1974) The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv Appl Prob 6:260–290

Caswell H (2001) Matrix population models, 2nd edn. Sinauer, Sunderland

Cattiaux P, Collet P, Lambert A, Martínez SM, Martín JS (2009) Quasi-stationary distributions and diffusion models in population dynamics. Ann Probab 37(5):1926–1969

Chakraborty R, Leimar O (1987) Genetic variation within a subdivided population. In: Ryman N, Utter R (eds) Population genetics and fishery management. Washington Sea Grant Program, Seattle, WA. Reprinted 2009 by The Blackburn Press, Caldwell

Collet P, Martinez S (2013) Quasi stationary distributions, Markov chains, diffusions and dynamical systems. Springer, Berlin
Cox DR, Miller HD (1965) The theory of stochastic processes. Methuen & Co Ltd, London
Crow JF (2004) Assessing population subdivision. In: Wasser SP (ed) Evolutionary theory and processes: modern horizons. Papers in Honour of Eviator Nevo. Springer Science+Business Media Dordrecht, Berlin, pp 35–42
Crow JF, Aoki K (1982) Group selection for a polygenic behavioral trait: a differential proliferation model. Proc Natl Acad Sci 79:2628–2631
Crow JF, Aoki K (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc Natl Acad Sci 81:6073–6077
Crow JF, Kimura M (1970) An introduction to population genetics theory. The Blackburn Press, Caldwell
Durrett R (2008) Probability models for DNA sequence evolution, 2nd edn. Springer, New York
Engen S, Lande R, Saether B-E (2005a) Effective size of a fluctuating age-structured population. Genetics 170:941–954
Engen S, Lande R, Saether B-E, Weimerskirch H (2005b) Extinction in relation to demographic and environmental stochasticity in age-structured models. Math Biosci 195:210–227
Engle RF, Granger CWJ (1987) Co-integration and error correction: Representation, estimation and testing. Econometrica 55:251–276
Ethier SN, Nagylaki T (1980) Diffusion approximation of Markov chains with two time scales and applications to genetics. Adv Appl Prob 12:14–49
Ewens WJ (1982) On the concept of effective population size. Theoret Popul Biol 21:373–378
Ewens WJ (2004) Mathematical Population Genetics. I. Theoretical introduction, 2nd edn. Springer, New York
Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. Genetics 68:581–597
Fisher RA (1958) The genetical theory of natural selection, 2nd edn. Dover, New York
Granger CWJ (1981) Some properties of time series data and their use in econometric model specification. J Econom 16:121–130
Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83:145–154
Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population model. Mol Ecol Notes 2:618–620
Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples R, Ruegg K, Palstra F (2011) Understanding and estimating effective population size for practical applications in marine species management. Conserv Biol 25(3):438–449
Hössjer O (2011) Coalescence theory for a general class of structured populations with fast migration. Adv Appl Probab 43(4):1027–1047
Hössjer O (2013) Spatial autocorrelation for subdivided populations with invariant migration schemes. Methodol Comput Appl Probab. doi:10.1007/s11009-013-9321-3
Hössjer O, Jorde PE, Ryman N (2013) Quasi equilibrium approximations of the fixation index of the island model under neutrality. Theoret Popul Biol 84:9–24
Jamieson IG, Allendorf FW (2012) How does the 50/500 rule apply to MVPs? Trends Ecol Evol 27(10): 578–584
Jorde P-E, Ryman N (2007) Unbiased estimator of genetic drift and effective population size. Genetics 177:927–935
Karlin S (1966) A first course in stochastic processes. Academic Press, New York
Kimura M (1953) 'Stepping stone' model of population. Ann Rep Natl Inst Genet Japan 3:62–63
Kimura M (1955) Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci USA 41:141–150
Kimura M (1964) Diffusion models in population genetics. J Appl Prob 1:177–232
Kimura M (1971) Theoretical foundations of population genetics at the molecular level. Theor Popul Biol 2:174–208
Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 61:763–771
Kingman JFC (1982) The coalescent. Stoch Proc Appl 13:235–248
Latter BDH, Sved JA (1981) Migration and mutation in stochastic models of gene frequency change. II. Stochastic migration with a finite number of islands. J Math Biol 13:95–104

Leviyang S (2011a) The distribution of $F_{ST}$ for the island model in the large population, weak mutation limit. Stoch Anal Appl 28:577–601
Leviyang S (2011b) The distribution of $F_{ST}$ and other genetic statistics for a class of population structure models. J Math Biol 62:203–289
Leviyang S, Hamilton MB (2011) Properties of Weir and Cockerham's $F_{ST}$ estimator and associated bootstrap confidence intervals. Theoret Populat Biol 79:39–52
Malécot G (1946) La consanguinité dans une population limitée. C R Acad Sci (Paris) 222:841–843
Maruyama T (1970a) On the rate of decrease of heterozygosity in circular stepping stone models of populations. Theor Popul Biol 1:101–119
Maruyama T (1970b) Effective number of alleles in subdivided populations. Theor Popul Biol 1:273–306
Möhle M (2010) Looking forwards and backwards in the multi-allelic neutral Cannings population model. J Appl Prob 47:713–731
Nagylaki T (1980) The strong migration limit in geographically structured populations. J Math Biol 9: 101–114
Nagylaki T (1982) Geographical invariance in population genetics. J Theor Biol 99:159–172
Nagylaki T (1998) The expected number of heterozygous sites in a subdivided population. Genetics 149:1599–1604
Nagylaki T (2000) Geographical invariance and the strong-migration limit in subdivided populations. J Math Biol 41:123–142
Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323
Nei M (1975) Molecular evolution and population genetics. North-Holland, Amsterdam
Nei M (1977) $F$-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41: 225–233
Nei M, Chakravarti A, Tateno Y (1977) Mean and variance of $F_{ST}$ in a finite number of incompletely isolated populations. Theoret Popul Biol 11:291–306
Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford
Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. Genetics 98:625–640
Nordborg M, Krone S (2002) Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin M, Veuille M (eds) Modern development in theoretical population genetics. Oxford Univ Press, Oxford, pp 194–232
Nunney L (1999) The effective size of a hierarchically-structured population. Evolution 53:1–10
Olsson F, Hössjer O, Laikre L, Ryman N (2013) Variance effective population size of populations in which size and age composition fluctuate. Theoret, Popul Biol (to appear)
Orive ME (1993) Effective population size in organisms with complex life-histories. Theoret Popul Biol 44:316–340
Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild populations persistence? Mol Ecol 17:3428–3447
Rottenstreich S, Miller JR, Hamilton MB (2007) Steady state of homozygosity and $G_{ST}$ for the island model. Theoret Popul Biol 72:231–244
Ryman N, Allendorf FW, Jorde PE, Laikre L, Hössjer O (2013) Samples from structured populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. Mol Ecol Resour (to appear)
Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations. Evolution 62(9):2250–2259
Sagitov S, Jagers P (2005) The coalescent effective size of age-structured populations. Ann Appl Probab 15(3):1778–1797
Sampson KY (2006) Structured coalescent with nonconservative migration. J Appl Prob 43:351–362
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
Slatkin M (1981) Estimating levels of gene flow in natural populations. Genetics 99:323–335
Slatkin M (1985) Rare alleles as indicators of gene flow. Evolution 39:53–65
Slatkin M (1991) Inbreeding coefficients and coalescence times. Genet Res 58:167–175
Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population genetics. Am Nat 138(2):499–517
Sved JA, Latter BDH (1977) Migration and mutation in stochastic models of gene frequency change. J Math Biol 5:61–73

Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. Biol J Linnean Soc 60:73–93
Takahata N (1983) gene identity and genetic differentiation of populations in the finite island model. Genetics 104 (3): 497–512
Takahata N, Nei M (1984) $F_{ST}$ and $G_{ST}$ statistics in the Finite island model. Genetics 107 (3): 501–504
Van der AA NP, Ter Morsche HG, Mattheij RRM (2007) Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem. Electron J Linear Algebra 16:300–314
Wakeley J (1999) Nonequilibrium migration in human history. Genetics 153:1863–1871
Wakeley J, Takahashi T (2004) The many-demes limit for selection and drift in a subdivided population. Theoret Popul Biol 66:83–91
Wang J, Caballero A (1999) Developments in predicting the effective size of subdivided populations. Heredity 82:212–226
Waples RS (1989) A generalized approach for estimating effective population size from temporal changes of allele frequency. Genetics 121:379–391
Waples RS (2002) Definition and estimation of effective population size in the conservation of endangered species. In: Beissinger SR, McCullogh DR (eds) Populations viability analysis. The University of Chicago Press, Chicago
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–1439
Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. Genetics 175:219–233
Ward RD, Woodward M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater and anadromous fishes. J Fish Biol 44:213–232
Weir BS, Cockerham CC (1984) Estimating $F$-statistics for the analysis of population structure. Evolution 38(6):1358
Weiss GH, Kimura M (1965) A mathematical analysis of the stepping stone model of genetic correlation. J Appl Probab 2:129–149
Whitlock MC, Barton NH (1997) The effective size of a subdivided population. Genetics 145:427–441
Wilkinson-Herbots HM (1998) Genealogy and subpopulation differentiation under various models of population structure. J Math Biol 37:535–585
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 87:430–431
Wright S (1943) Isolation by distance. Genetics 28:114–138
Wright S (1946) Isolation by distance under diverse systems of mating. Genetics 31:39–59
Wright S (1951) The general structure of populations. Ann Eugenics 15:323–354
Wright S (1978) Variability within and among genetic populations. Evolution and the genetics of populations, vol 4. University of Chicago Press, Chicago