TPB

Theoretical Population Biology

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Quasi equilibrium approximations of the fixation index under neutrality: The finite and infinite island models

Ola Hössjer [a,*], Per Erik Jorde [b,1], Nils Ryman [c]

[a] *Department of Mathematics, Div. of Mathematical Statistics, Stockholm University, SE 106 91 Stockholm, Sweden*
[b] *Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, P.O. Box 1066 Blindern, N-0371 Oslo, Norway*
[c] *Department of Zoology, Div. of Population Genetics, Stockholm University, SE 106 91 Stockholm, Sweden*

ARTICLE INFO

ABSTRACT

The fixation index F_{ST} and the coefficient of gene differentiation G_{ST} are analyzed for the finite island model under short time spans, ignoring mutations. Dividing the reproduction cycle into the three steps – gamete formation, fertilization, and migration – we develop a new approach for computing quasi equilibrium formulas for F_{ST} (and G_{ST}). Our formulas generalize earlier ones and reveal that the equilibrium value of F_{ST} is influenced not only by the migration rate and local effective population size, N_e, but also by the local census size N, particularly so when the migration rate is high. The order of migration and fertilization is found to have a smaller effect on F_{ST}. A major advantage compared to previous approaches is that stochastic allele frequency of migrants is easily accommodated, thereby avoiding underestimation of F_{ST} for large migration rates.

1. Introduction

The fixation index F_{ST} was introduced by Wright (1921, 1931, 1951). It is the most frequently used measure of genetic differentiation between subpopulations for models exhibiting spatial structure. It also quantifies the amount of inbreeding within subpopulations (S) relative to that of the total population (T). This makes it a highly relevant quantity to study for short term evolution in conservation genetics. During shorter time spans, the effect of mutations can be ignored, and the value of the fixation index is a delicate balance between genetic drift, which tends to increase F_{ST}, and migration, which tends to decrease F_{ST}. In this paper, we derive new methods for computing the equilibrium value of the fixation index resulting from this balance.

We derive formulas for the equilibrium F_{ST} of the finite island model under neutrality, using an approach which involves two main novelties: First, we work directly with the mutation free island model and derive a quasi equilibrium approximation of the fixation index. Second, we divide the reproduction cycle into three steps; gamete formation, fertilization, and migration and study several scenarios. In particular, we allow the local population

actual size N to differ from the local effective population size N_e in the gamete formation step. This is exemplified using several monoecious and dioecious models. We further allow the order of fertilization and migration to vary and consider fixed as well as stochastic allele frequencies of the migrants that enter an island.

Although much has been written about the island model, to the best of our knowledge no formulas for F_{ST} have yet been presented that distinguish N_e from N. This has impact also for estimating the variance effective size $N_{e,\text{tot}}$ of the whole population (Ryman et al., submitted for publication).

2. Theoretical background

2.1. The fixation index

We restrict ourselves to the island model (Wright, 1943a; Maruyama, 1970; Latter, 1973), the simplest possible way of describing a subdivided population, with the most important notation summarized in Table 1.

We thus assume that the diploid population evolves in non-overlapping generations, with the total population consisting of s islands of equal census and local size N and N_e respectively. We let P_{ti} be the frequency of a particular allele in island $i = 1, \ldots, s$ and generation t and m the expected fraction of newborns in each generation with parents originating from the total population. Since a fraction $1/s$ of these "immigrants" are actually from the focal island itself, the true immigration rate is

$$m' = m(s-1)/s.$$

* Corresponding author.
E-mail addresses: ola@math.su.se (O. Hössjer), p.e.jorde@bio.uio.no (P.E. Jorde), Nils.Ryman@popgen.su.se (N. Ryman).
[1] Present address: Institute of Marine Research, Flødevigen Research Station, N-4817 His, Norway.

Table 1
Mathematical notation.

Symbol	Definition
s	Number of islands.
N	Local census size.
N_e	Local (variance) effective size.
$N_{e,\text{tot}}$	Global (variance) effective size.
m	Fraction of newborns arriving from the total population.
m'	Migration rate ($=(s-1)m/s$).
m_{ij}	Migration rate from island i to j.
P_t	Allele frequency of the whole population, generation t.
P_{ti}	Allele freq. of individuals of island i and gen. t before reproduction cycle.
$\tilde{P}_{ti}$	Allele freq. of gamete pool of island i and gen. t, before migration.
P_{ti}^*	Allele freq. of individuals of island i and generation t, after fertilization and before migration (FM,I).
$\check{P}_{ti}$	Allele freq. of gamete pool of island i in generation t, after migration, before fertilization (MF).
$\tilde{P}_t$	Allele freq. of combined merged gamete pool, generation t (I).
P_{ti}'	Allele freq. of individuals of island i and generation t after fertilization from combined gamete pool (I).
$F_{ST,t}$	Fixation index, generation t.
$F_{ST,t}^{\text{adj}}$	Adjusted fixation index, generation t.
F_{ST}^{eq}	(Quasi) equilibrium limit of $F_{ST,t}$.
F_{ST}^{appr}	Approximation of F_{ST}^{eq}.
$G_{ST,t}$	Coefficient of gene differentiation, generation t.
G_{ST}^{eq}	(Quasi) equilibrium limit of $G_{ST,t}$.
G_{ST}^{appr}	Approximation of G_{ST}^{eq}.

For allele frequencies, the reproduction scenario(s) for which the notation appear(s), is indicated. Either FM (fertilization precedes migration), MF (migration precedes fertilization) or I (intermediate model).

In this paper we work with the parameter m and refer to it as the migration rate. The local (variance) effective size of each island i is

$$N_e = \frac{P_{ti}(1 - P_{ti})}{2\text{Var}(P_{t+1,i} - P_{ti}|P_{ti})} \quad \text{when } m = 0,$$

for a selectively neutral allele.

We define the fixation index of generation t as

$$F_{ST,t} = \frac{\sum_{i=1}^{s}(P_{ti} - P_t)^2}{sP_t(1 - P_t)}, \tag{1}$$

where $P_t = \sum_{i=1}^{s} P_{ti}/s$ is the generation t frequency of the allele in the whole population, see for instance Eq. (12.13) in Nei and Kumar (2000). It follows from Nei (1975, p. 123) and Cockerham and Weir (1987) that an equivalent formulation is

$$F_{ST,t} = \frac{f_S - f_T}{1 - f_T} = \frac{f_S - \left(\frac{1}{s}f_S + \frac{s-1}{s}f_D\right)}{1 - \left(\frac{1}{s}f_S + \frac{s-1}{s}f_D\right)}, \tag{2}$$

where f_S, f_T, and f_D are the probabilities that two randomly chosen genes of generation t are identical-by-state when drawn from the same subpopulation, the total population, or different subpopulations, respectively.

Other closely related definitions of $F_{ST,t}$ have slightly different denominators compared to (1)–(2), as reviewed by Charlesworth (1998). One of these, due to Weir and Cockerham (1984), is defined as the correlation between two alleles drawn from the same island, see also Cockerham (1969, 1973). Another adjusted version

$$F_{ST,t}^{\text{adj}} = \frac{\sum_{i=1}^{s}(P_{ti} - P_t)^2}{(s - 1)P_t(1 - P_t)} \tag{3}$$

of $F_{ST,t}$ has $s - 1$ rather than s in the denominator of (1). We will find below that in contrast to $F_{ST,t}$ the equilibrium value of $F_{ST,t}^{\text{adj}}$ is virtually independent of s.

Suppose samples of sizes n_i are taken for a subset $i \in I \subset \{1, \ldots, s\}$ of $k \le s$ islands, and let $\hat{P}_{ti}$ for each $i \in I$ be the sampled fraction of the given allele. Based on this data, we propose an estimator

$$\hat{F}_{ST,t} = \frac{\hat{S}^2 - \alpha\hat{P}_t(1 - \hat{P}_t)}{(1 - c\alpha)\hat{P}_t(1 - \hat{P}_t) + \hat{S}^2/k} \tag{4}$$

of the fixation index, where $\hat{P}_t = \sum_{i\in I}\hat{P}_{ti}/k$ and $\hat{S}^2 = \sum_{i\in I}(\hat{P}_{ti} - \hat{P}_t)^2/(k - 1)$ are the average and sample variance of the allele frequencies in the islands from which samples are taken. The constant α adjusts for sampling error within each island. It vanishes when $\hat{P}_{ti} = P_{ti}$ for all $i \in I$, and then $\hat{F}_{ST,t}$ equals an estimator proposed by Weir and Cockerham (1984). Finally, c is a fixed constant. Its optimal value will depend on the unknown fixation index and the sample size, although it should typically be chosen between 0 and 1. We verify in Appendix A that $\hat{F}_{ST,t}$ is an approximately unbiased estimator of $F_{ST,t}^{\text{adj}}$ rather than $F_{ST,t}$ when the fixation index is small.

Often the $s - k$ unsampled islands are unknown "ghost populations" (Wright, 1943b; Beerli, 2004; Slatkin, 2005). Then s is unknown and we cannot multiply $\hat{F}_{ST,t}$ by $(s-1)/s$ in order to get an approximately unbiased estimator of $F_{ST,t}$. This indicates that $F_{ST,t}^{\text{adj}}$ is a more stable function of s, and indeed, we will find below that the equilibrium approximation of $F_{ST,t}^{\text{adj}}$ is virtually independent of s.

The coefficient of gene differentiation (Nei, 1973; Nei and Chakravarti, 1977)

$$G_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T} \tag{5}$$

is a multiallelic and multilocus extension of F_{ST} rather than F_{ST}^{adj}. It is defined as the relative excess of the expected proportion of heterozygotes in the whole population, H_T, compared to that of the subpopulations, H_S. An adjusted version of G_{ST}, less dependent on s, is defined by Nei and Kumar (2000, Eq. (12.23)). However, it lacks the intuitive excess of heterozygosity interpretation.

Therefore, the unadjusted and adjusted fixation indexes both have their advantages. Since they only differ by a term $(s - 1)/s$, equilibrium results for the unadjusted fixation index can easily be translated to the adjusted version in theoretical studies when s is assumed to be known. In particular, both versions agree for the infinite ($s = \infty$) island model. In the sequel, we consider the unadjusted fixation index as default.

Other measures of genetic diversity have been and continue to be developed for special purposes, such as that of Chakraborty and Nei (1982) and Slatkin (1995), for microsatellite markers. Much work has also been devoted to evaluating effects of mutations on F_{ST} and other measures of genetic divergence. Instead, we focus on F_{ST} (and G_{ST}) and explore the role of demographic processes, including different migration and fertilization scenarios and departures from the ideal conditions (such as $N_e = N$) for this measure.

2.2. Equilibrium of F_{ST}

Wright (1943a) studied the long term behavior of F_{ST} for the infinite island model ($s = \infty$) and derived the exact equilibrium value

$$F_{ST}^{\text{eq}} = \frac{(1 - m)^2}{2N(1 - (1 - m)^2) + (1 - m)^2} \tag{6}$$

when reproduction follows a Wright–Fisher model, where the local effective population size N_e equals N. This result was obtained under the assumption of *fixed immigrant allele frequencies*, i.e.

the allele frequency of all immigrants into a particular island is identical to the allele frequency of the total population. Moreover, the number of immigrating individuals to each island was fixed to Nm without random variation.

Sved and Latter (1977) considered a slightly more realistic scenario of *stochastic immigrant allele frequencies*, where the allele frequency of the immigrants into an island is obtained by means of binomial sampling from the total population, thus exhibiting some random variation. They derived recurrence relations for the variance of P_{ti}, i.e. the numerator of (1), in the limit $s = \infty$. Equilibrium values of the variance are obtained as the steady state solutions of these equations, and they can be divided by $P_t(1 - P_t)$ to provide the corresponding equilibrium values of the fixation index. In this way (6) is modified to

$$F_{ST}^{eq} = \frac{1}{2N(1 - (1 - m)^2) + 1 - m} \tag{7}$$

(derived from Eq. (8) in Sved and Latter (1977)) when $N_e = N$ and the number of diploid immigrants from the whole population is exactly Nm, and

$$F_{ST}^{eq} = \frac{1}{2N(1 - (1 - m)^2) + (1 - m)^2} \tag{8}$$

(derived from Eq. (11) of Sved and Latter (1977)) when $N_e = N$ and the number of immigrants $N\hat{m}$ from the whole population is stochastic, exhibiting binomial variation around $NE(\hat{m}) = Nm$. This corresponds to a scenario when gametes "select" parental island independently of each other. We remark that Sved and Latter (1977) use a slightly different terminology. Fixed and stochastic immigrant allele frequencies are in their paper referred to as fixed and stochastic migration respectively, and a fixed and stochastic number of immigrants are by them denoted fixed and stochastic migration rate.

When N is large and m small, we notice that (6)–(8) all reduce to the well known approximation

$$F_{ST}^{eq} \approx \frac{1}{4Nm + 1} \tag{9}$$

due to Wright (1943a), which is particularly appealing since it only depends on the expected number of immigrants Nm.

Finding and even defining F_{ST}^{eq} for the finite island model ($s < \infty$) involves some difficulties, since there is no equilibrium between genetic drift and migration in the absence of mutations. When $m > 0$, one allele will eventually become fixed in all subpopulations, although the time for this to happen is usually very long. The standard approach is to introduce a small mutation probability u per gamete and generation and then obtain recurrence relations for f_S and f_D, see Nei (1975) and Li (1976) for the island model. By finding the steady state solutions of f_S and f_D and inserting them into (2), an equilibrium value

$$F_{ST}^{eq} = \frac{(1 - m)^2(1 - u)^2}{\frac{s}{s-1}2N(1 - (1 - m)^2(1 - u)^2) + (1 - m)^2(1 - u)^2}$$

of the fixation index (and of G_{ST} as well) can be derived, as discussed for instance by Nei (1975), Takahata (1983), Crow and Aoki (1984), Takahata and Nei (1984) and Ryman and Leimar (2008). Taking the $u \to 0$ limit, we find that F_{ST}^{eq} simplifies to

$$F_{ST}^{eq} = \frac{(1 - m)^2}{\frac{s}{s-1}2N(1 - (1 - m)^2) + (1 - m)^2}, \tag{10}$$

which can be viewed as a generalization of (6) to the finite island model. Slatkin and Voelm (1991) have shown that (10) can be expressed in terms of mean coalescence times, and for small migration rates, it reduces to

$$F_{ST}^{eq} \approx \frac{1}{1 + \frac{s}{s-1}4Nm}, $$

which is an extension of (9) for the finite island model.

One may also derive F_{ST}^{eq} in other ways, utilizing recurrence relations of identical by state probabilities for more general migration structures (Malécot, 1951) or dioecious populations (Nagylaki, 1995), recurrence relations for characteristic functions (Rousset, 1996) or joint recurrence relations of the inbreeding coefficient and coancestry of individuals within and between islands (Chesser et al., 1993; Wang, 1997a,b).

3. Quasi equilibrium of F_{ST}

We will develop an approach for the finite island model that differs from (10) in three ways. First, we work with the neutral model without mutations, $u = 0$, directly, thereby avoiding to introduce mutations and then taking the $u \to 0$ limit.

Secondly, we allow for the allele frequencies of migrants from island i to j to vary not only with i, but also with j. This simple assumption turns out to be crucial in order to have 1 rather than $(1 - m)^2$ in the numerator of F_{ST}^{eq}. While this has minor effect for small migration rates, the effect can be quite substantial, at least in relative terms, when m is close to 1. Indeed, having $(1 - m)^2$ in the numerator suggests that all subpopulations have identical allele frequencies when $m = 1$. However, this is not possible since $s < \infty$, and either one allele has been fixed in all subpopulations, so that F_{ST} is not well defined, or, if this has not yet happened, some (small) random variability of allele frequency between islands makes F_{ST} positive.

Hössjer (submitted for publication) developed an idea of Latter and Sved (1981) and showed that $(1 - m)^2$ in the numerator of (10) also disappears when equilibrium equations for f_S and f_D are solved by first incorporating mutations and then letting the mutation probability tend to zero. The main idea is to redefine f_S as the probability of two genes drawn *with replacement* to be identical by state. This suggests that even though (10) is correct, it relies on a questionable definition of f_S.

Thirdly, since $s < \infty$, we don't treat $F_{ST,t}$ as fixed quantity but as a stochastic process. This is a well known approach, see for instance Nagylaki (1998) and Nei et al. (1977). In the latter paper, the authors derive approximate recursion formulas for the mean and variance of $F_{ST,t}$ when fixation of alleles are allowed and migrants from island i to other islands have identical allele frequencies. We will instead concentrate on the behavior *before* fixation. In doing so, we define the quasi equilibrium value as the limit

$$F_{ST}^{eq} = \lim_{t \to \infty} E(F_{ST,t}), \tag{11}$$

conditionally on that no allele becomes fixed in all islands. In this way, F_{ST}^{eq} becomes the mean value of the (typically small) fluctuations that the fixation index exhibits under quasi equilibrium. We argue that conditioning on non-fixation is appropriate when using neutral markers. Indeed, since they are *indicators* of populations substructure, they are only useful before fixation. This approach is also mentioned briefly by Nei et al. (1977) in simulations (see their Table 2), where they argue that (11) should be close to

$$E(F_{ST})^* = \frac{E\left(\sum_{i=1}^{s}(P_{ti} - P_t)^2\right)}{sE(P_t(1 - P_t))},$$

a quantity introduced by Nei (1975). We will use a related but slightly different approach, which also gives information about the nature of the oscillations of $F_{ST,t}$ around F_{ST}^{eq}. The starting point is to write the change of the fixation index from one generation to the next as a sum

$$F_{ST,t+1} = E(F_{ST,t+1}|t) + \epsilon_{t+1}, \tag{12}$$

of a deterministic and random component, where expectation is conditional on generation t and ϵ_{t+1} is an error term satisfying

$E(\epsilon_{t+1}|t) = 0$. For all versions of the island model considered in this paper, there exist constants A and B, which depend on various genetic model parameters, but not on $F_{ST,t}$, such that the expected value in (12) can be approximated as

$$E(F_{ST,t+1}|t)$$

$$\approx E\left(\sum_{i=1}^{s}(P_{t+1,i}-P_{t+1})^2\Big|t\right)\Big/\left(sE\left(P_{t+1}(1-P_{t+1})|t\right)\right)$$

$$= E\left(\sum_{i=1}^{s}(P_{t+1,i}-P_{t+1})^2\Big|t\right)\Big/(sP_t(1-P_t))(1-(2N_{e,\text{tot}})^{-1})$$

$$\approx AF_{ST,t}+B, \qquad (13)$$

with

$$N_{e,\text{tot}} = \frac{P_t(1-P_t)}{2\text{Var}(P_{t+1}-P_t|t)}, \qquad (14)$$

the (variance) effective size of the whole population, a quantity reviewed by Caballero (1994), Wang and Caballero (1999) and Waples (2002). We will find below that $(2N_{e,\text{tot}})^{-1}$ can be dropped in (13) with good accuracy, unless the migration rate m and the number of islands s are both small (see Appendix D for more details). Insertion of (13) into (12) gives an approximate characterization

$$F_{ST,t+1} \approx AF_{ST,t}+B+\epsilon_{t+1}, \qquad (15)$$

of the fixation index process, conditionally on that no allele gets fixed. This is an autoregressive process of order 1 (Brockwell and Davis, 1987), whose mean value

$$F_{ST}^{\text{appr}} = \frac{B}{1-A} \qquad (16)$$

is used as an approximation of F_{ST}^{eq} in (11). As mentioned in the Introduction, the quasi equilibrium approximation of the adjusted fixation index, $F_{ST}^{\text{appr,adj}}$, is obtained by multiplying (16) with $(s-1)/s$.

The accuracy of (16) is dependent on how good the approximations in (13) are. In the first line, we assumed that the expected value of the ratio is close to the the ratio of the expected values. As rule of thumb, the larger $N_{e,\text{tot}}$ is, the smaller the coefficient of variation of the denominator, and the more accurate this step. The last step of (13) is exact for some models and a large population approximation for others, see Appendix B for further details.

The quasi equilibrium fluctuations of $F_{ST,t}$ are caused by a genetic drift that varies between generations, as manifested by the error term ϵ_{t+1} in (12) and (15). We will find in the Numerical results section, that the magnitude of these oscillations decrease with the number of subpopulations when the size of each island is kept fixed. Somewhat analogous variations can be observed in the effective size of the total population for homogeneous ($s = 1$) populations. It has been shown analytically and through simulations (Waples and Faulkner, 2009) that the variance of these oscillations is inversely proportional to the size of the population.

In Appendix C, we will motivate that $G_{ST,t}$, the coefficient of gene differentiation at time t, is also approximated by an autoregressive process (15), with the same constants A and B, so that the right hand side of (16) also becomes an approximation

$$G_{ST}^{\text{appr}} = \frac{B}{1-A} \qquad (17)$$

of

$$G_{ST}^{\text{eq}} = \lim_{t\to\infty} E(G_{ST,t}).$$

Analogously to (11), the defining equation for G_{ST}^{eq} is conditional on that no allele gets fixed in all islands at any locus. Part of

the difference between G_{ST}^{appr} and G_{ST}^{eq} depends on the number of loci n, the so called inter locus error (Nei, 1975; Waples, 1998). We motivate in Appendix C that as n grows the interlocus error decreases and (17) becomes a more accurate approximation of G_{ST}^{eq}. Moreover, the magnitude of the quasi equilibrium oscillations decreases with n as well.

As mentioned above, our method differs from (10) and is rather a generalization of the infinite island model approach of Sved and Latter (1977) to the finite island model.

4. Models for the reproduction cycle

The reproduction cycle from generation t to $t + 1$ is divided into three steps, i.e. gamete formation, fertilization, and migration. Similar reproduction steps have been considered before, see for instance Nagylaki (1983) and Sampson (2006), but the novelty of our approach is that a large number of gamete formation, migration, and fertilization scenarios can be treated within a unified framework. For the examples below, it is sufficient with four parameters to summarize them all; the migration rate m, the number of islands s, and the local effective and local census population sizes N_e and N.

We first consider gamete formation, using the notation for allele frequencies of various groups summarized in Table 1.

4.1. Gamete formation

An infinitely large gamete pool is constructed from the individuals of island i in generation t, with allele frequency $\tilde{P}_{ti}$. To account for varying reproductive rate among the individuals of island i, $\tilde{P}_{ti}$ may differ from P_{ti} in a way quantified as

$$\frac{E\left((\tilde{P}_{ti}-P_{ti})^2|P_{ti}\right)}{P_{ti}(1-P_{ti})} \sim \frac{1}{2N_e} - \frac{1}{2N}, \qquad (18)$$

where $x \sim y$ means that x/y tends to one as N and N_e get large.

Formula (18) is crucial for this paper. It implies that in the absence of migration, the total amount of genetic drift within each island (size proportional to $(2N_e)^{-1}$) can be divided into two variance components. One is due to fertilization (size proportional to $(2N)^{-1}$) and the other (the remainder) is due to gamete formation (size proportional to $(2N_e)^{-1} - (2N)^{-1}$). Although (18) seems new for the island model, a similar decomposition into variance components has implicitly been noted for homogeneous ($s = 1$) populations, in the context of estimating genetic drift by means of the temporal method (Nei and Tajima, 1983; Waples, 1989a).

For the Wright–Fisher model, when $N_e = N$, the gamete pool is an exact deterministic copy of island i in terms of allele frequency ($\tilde{P}_{ti} = P_{ti}$), so that the gamete formation variance component vanishes. At the other extreme, when $N_e \ll N$, the gamete formation variance component dominates, i.e. most of the random variation of the local allele frequency takes place when the gamete pool is formed rather than during fertilization.

There are several ways in which N_e can be smaller than N. In this paper we consider three (monoecious and dioecious) examples, all of which satisfy (18), and thus having the same (quasi) equilibrium expressions for F_{ST}:

4.1.1. Subgroup of breeders with same expected amount of gamete formation

As a first example of a situation with $N_e < N$ we consider the case where a subset of $2N_e$ genes are selected for replication

Table 2

Approximate (quasi) equilibrium expressions of expected F_{ST} ($= F_{ST}^{\mathrm{appr}}$) for the neutral island model.

s	Reproduction	$N_e = N$	$N_e \neq N$
∞	FM/FF	$\dfrac{(1-m)^2}{2N(1-(1-m)^2)+(1-m)^2}$ (6)	$\dfrac{(1-m)^2}{2N_e(1-(1-m)^2)+(1-m)^2}$
∞	FM/FS	$\dfrac{1}{2N(1-(1-m)^2)+1}$	$\dfrac{1}{2\tilde{N}(1-(1-m)^2)+1}$
∞	MF/FF	$\dfrac{1}{2N(1-(1-m)^2)+(1-m)^2}$ (8)	$\dfrac{1}{2\tilde{N}(1-(1-m)^2)+\frac{\tilde{N}}{N_e}(1-m)^2}$
∞	I/F	$\dfrac{1}{2N(1-(1-m)^2)+1-m}$ (7)	$\dfrac{1}{2\tilde{N}(1-(1-m)^2)+1-\frac{\tilde{N}}{N}m}$
$<\infty$	FM/FF	$\dfrac{(1-m)^2}{\frac{s}{s-1}2N\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+(1-m)^2}$	$\dfrac{(1-m)^2}{\frac{s}{s-1}2N_e\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+(1-m)^2}$ (25)
$<\infty$	FM/FS	$\dfrac{1}{\frac{s}{s-1}2N\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+1}$	$\dfrac{1}{\frac{s}{s-1}2\tilde{N}\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+1}$ (27)
$<\infty$	MF/FF	$\dfrac{1}{\frac{s}{s-1}2N\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+(1-m)^2}$	$\dfrac{1}{\frac{s}{s-1}2\tilde{N}\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+\frac{\tilde{N}}{N_e}(1-m)^2}$ (32)
$<\infty$	I/F	$\dfrac{1}{\frac{s}{s-1}2N\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+1-m}$	$\dfrac{1}{\frac{s}{s-1}2\tilde{N}\left(1-(1-m)^2-\frac{1}{2N_{e,tot}}\right)+1-\frac{\tilde{N}}{N}m}$ (37)

Equations for the infinite ($s = \infty$) and finite ($s < \infty$) island models are shown, and the local effective population size N_e either equals or differs from the actual one, N. The four reproduction scenarios are FM/FF, FM/FS (fertilization precedes migration with fixed migrant proportions and fixed or stochastic migrant allele frequencies), MF/FF (migration precedes fertilization with fixed migrant proportions and migrant allele frequencies) and I/F (intermediate model, fixed migrant proportions). Equation numbers refer to those in the text; (5)–(7) represent parameter combinations giving results of Wright (1943a), Sved and Latter (1977) and Nei (1975); and the others are those derived in this paper. $N_{e,\mathrm{tot}}$ in (14) is either chosen as ∞ and hence dropped, or (as in all the tables with numerical results) a function (D.1) of s, m, N_e and N that slightly depends on the reproduction model. The upper part ($s = \infty$) is obtained from the lower part ($s < \infty$) by replacing $s/(s-1)$ with 1. $\tilde{N}$ is an harmonic average of N_e and N, cf. (28), and column $N_e = N$ is obtained from column $N_e \neq N$ by putting $\tilde{N} = N_e$ everywhere. The discrepancy between $\tilde{N}$ and N_e is larger the higher m is. Hence, for models FM/FS, MF/FF and I/F, the effect of $N_e \neq N$ on F_{ST}^{appr} is most pronounced for large migration rates, whereas F_{ST}^{appr} only depends on N_e for FM/FF. Model FM/FF has systematically lower values of F_{ST}^{appr} due to the $(1-m)^2$ term of the numerator.

during gamete formation (for a diploid organism this corresponds to N_e breeders). The quantity $\tilde{P}_{ti}$ is then determined by drawing $2N_e$ genes of breeders randomly *without replacement* from the $2N$ genes of island i in generation t. If the breeders' genes have equal opportunities to contribute to the infinite gamete pool, a hypergeometric distribution of the allele frequency

$$\tilde{P}_{ti} = \mathrm{Hyp}(2N, 2N_e, P_{ti})/(2N_e)$$

of the gamete pool is obtained. By second moment properties of the hypergeometric distribution, it follows that (18) holds.

4.1.2. Variable amount of gamete formation

The breeders' $2N_e$ genes of the previous example contributed with the same fraction $1/(2N_e)$ to gamete pool i. More generally, we can number the island i genes of generation t as $1, \ldots, 2N$ and let $w = (w_1, \ldots, w_{2N})$ denote the vector of relative contributions of all these genes to gamete pool i, so that $\sum_{j=1}^{2N} w_j = 1$ and

$$\tilde{P}_{ti} = \sum_{j:j \text{ has allele}1} w_j.$$

Suppose $w = \mathrm{Dir}(\alpha/(2N), \ldots, \alpha/(2N))$ has a Dirichlet distribution. The larger α is, the less variability there is in breeding intensity. It is easy to see, using the marginalization property of the Dirichlet distribution, that

$$\tilde{P}_{ti} = \mathrm{Beta}\left(\alpha(1 - P_{ti}), \alpha P_{ti}\right).$$

Since $\mathrm{Var}(\tilde{P}_{ti}) = P_{ti}(1 - P_{ti})/(\alpha + 1)$, it can be shown that α should be selected as

$$\alpha = \frac{2N_e N}{N - N_e} - 1$$

in order to satisfy (18). The degenerate case $N = N_e$ corresponds to $\alpha = \infty$, $w_j \equiv 1/(2N)$ and $\tilde{P}_{ti} = P_{ti}$.

4.1.3. Dioecious model with fixed sex ratio

Assume that the N individuals of each island in any generation of a diploid, dioecious population consists of N_m males and N_f females, with $N = N_m + N_f$. Write

$$P_{ti} = \frac{1}{2}P_{tim} + \frac{1}{2}P_{tif},$$

where P_{tim} and P_{tif} are the allele frequency of the N genes in island i and generation t inherited from males and females in the previous generation respectively. Gamete pool i of generation t is divided into a male and female part with allele frequencies $\tilde{P}_{tim}$ and $\tilde{P}_{tif}$. Because of the two-sex reproduction, the male and female genes will contribute in *equal proportions* to the gamete pool, regardless of the sex ratio N_m/N, so that the overall allele frequency of the combined gamete pool is

$$\tilde{P}_{ti} = \frac{1}{2}\tilde{P}_{tim} + \frac{1}{2}\tilde{P}_{tif}.$$

The male gene pool is obtained by drawing randomly *without replacement* N_m genes from the set of all N paternally inherited genes in island i, of which $X_{ti} = \mathrm{Hyp}(N, N_m, P_{tim})$ have the specified allele, and in addition drawing N_m genes without replacement from the set of all N maternally inherited genes, of which $Y_{ti} = \mathrm{Hyp}(N, N_m, P_{tif})$ have the specified allele. Then $\tilde{P}_{tim} = (X_{ti} + Y_{ti})/(2N_m)$ and $\tilde{P}_{tif} = (2NP_{ti} - X_{ti} - Y_{ti})/(2N_f)$. Using moment properties of the hypergeometric distribution and independence of X_{ti} and Y_{ti}, it can be shown that (18) holds with

$$N_e = \frac{4N_m N_f}{N},$$

in agreement with, for instance, Section 3.11 of Crow and Kimura (1970).

4.2. Fertilization precedes migration

In this case the gamete formation step of the reproduction cycle is first succeeded by fertilization, followed by migration. This order is of biological relevance when diploid individuals of a monoecious or dioecious organism migrate. One generation cycle of this model is illustrated in Fig. 1, and described in more detail as follows:

For each island i, a pre-migration population of $2N$ genes is formed by drawing $2N$ genes from gamete pool i after gamete formation. The resulting allele frequency is P_{ti}^*, with

$$\frac{E\left((P_{ti}^* - \tilde{P}_{ti})^2|\tilde{P}_{ti}\right)}{\tilde{P}_{ti}(1 - \tilde{P}_{ti})} \sim \frac{1}{2N}. \tag{19}$$
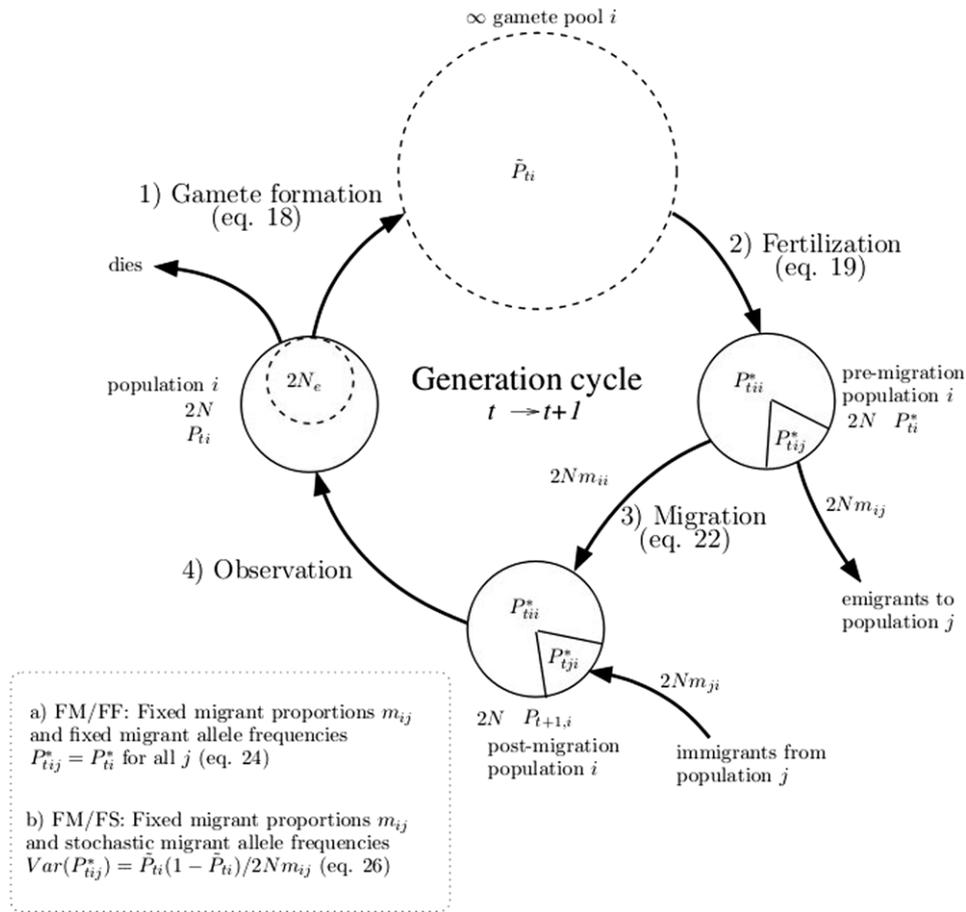
Fig. 1. One generation cycle when fertilization precedes migration, from the perspective of island i, with diploid migrants. For simplicity, only migration to one other island j is visualized. When migration to all islands $j' \neq i$ is taken into account, $m_{ii} = 1 - \sum_{j' \neq i} m_{ij'}$.

For the monoecious models, we achieve (19) by means of binomial sampling,

$$P_{ti}^* = \frac{\text{Bin}(2N, \tilde{P}_{ti})}{2N}, \tag{20}$$

and for the dioecious model paternally and maternally inherited genes are sampled separately to retain equal proportions from the gamete pool, i.e.

$$P_{ti}^* = \frac{1}{2}P_{tim}^* + \frac{1}{2}P_{tif}^*, \tag{21}$$

where $P_{tim}^* = \text{Bin}(N, \tilde{P}_{tim})/N$ and $P_{tif}^* = \text{Bin}(N, \tilde{P}_{tif})/N$ refer to the allele frequencies of the genes sampled from the male and female parts of the gamete pool.

In the next step migration among the s subpopulations takes place. Let m_{ij} denote the proportion of the $2N$ genes of island i that migrate to island j. In particular, m_{ii} is the proportion of genes of island i that do not migrate. Let P_{tij}^* be the *migrant allele frequency* of the genes migrating from i to j, so that the pre-migration allele frequency of island i is a mixture

$$P_{ti}^* = \sum_{j=1}^{s} m_{ij} P_{tij}^*.$$

After migration, the allele frequency of island i and generation $t+1$ is a (different) mixture of P_{tii}^* and the migrant allele frequencies P_{tji}^* of individuals migrating to i from various other islands j;

$$P_{t+1,i} = \sum_{j=1}^{s} m_{ji} P_{tji}^*. \tag{22}$$

For the dioecious model, we can split (22) into male and female parts, as

$$P_{t+1,im} = \sum_{j=1}^{s} m_{ji} P_{tjim}^*$$

$$P_{t+1,if} = \sum_{j=1}^{s} m_{ji} P_{tjif}^*,$$

where P_{tijm}^* and P_{tijf}^* are the migrant allele frequencies of the genes from the male and female gamete pools that migrate from i to j. Since individuals, not gametes, migrate, we obviously have $P_{tij}^* = (P_{tijm}^* + P_{tijf}^*)/2$.

The mixture (22) could either be in fixed or stochastic proportions, with a fixed or stochastic allele frequency of migrants. We now consider some of these possibilities:

4.2.1. Fixed migrant proportions and fixed migrant allele frequencies

We assume that

$$m_{ij} = \begin{cases} 1 - m + m/s, & i = j, \\ m/s, & i \neq j, \end{cases} \tag{23}$$

are constant, referred to as *fixed migrant proportions*. Then, the number of diploid immigrants to each island is exactly Nm', so that fixed migrant proportions implies a fixed number of immigrants. However, in the model of the next section, fixed migrant proportions refers to gametes, not individuals, and does *not* imply a fixed number of immigrants. This shows that in general the two concepts are different.

We further assume *fixed migrant allele frequencies*

$$P^*_{tij} = P^*_{ti} \quad \text{for all } i, j, \tag{24}$$

i.e. the allele frequencies are the same for all subpopulations j of individuals in island i that migrate to various islands j (including those that remain in island i, i.e. $j = i$). It is verified in Appendix B that fixed migrant proportions and allele frequencies imply fixed immigrant allele frequencies when $s \to \infty$, regardless of whether N_e equals N or not. It is also shown that (23)–(24) lead to

$$F^{\text{appr}}_{ST} = \frac{(1-m)^2}{\frac{s}{s-1} 2N_e \left(1 - (1-m)^2 - \frac{1}{2N_{e,tot}}\right) + (1-m)^2}, \tag{25}$$

a generalization of (10) to $N_e \neq N$, or, of (6) to $s < \infty$ and $N_e \neq N$. Nei et al. (1977) derive an expression for $E(F_{ST})^*$ somewhat related to (11) under the assumptions that $N_e = N$, fertilization precedes migration and fixed migrant allele frequencies (see (28) and the formula three lines below this equation in their paper).

4.2.2. Fixed migrant proportions and stochastic migrant allele frequencies

We retain the fixed migrant proportions (23) but assume that the $2N$ genes of the pre-migration population j are randomly divided into s subpopulations of relative sizes $m_{j1}, \ldots, m_{js}$. This is equivalent to drawing these subpopulations independently from the gamete pool of island j in the gamete formation step, so that there will be variation among the migrant allele frequencies P^*_{tji}. In particular,

$$\text{Var}(P^*_{tji}|\tilde{P}_{tj}) = \frac{\tilde{P}_{tj}(1-\tilde{P}_{tj})}{2Nm_{ji}} \sim \frac{P_{tj}(1-P_{tj})}{2Nm_{ji}}, \tag{26}$$

which is achieved for the monoecious and dioecious models analogously to (20) and (21). It is shown in Appendix B that

$$F^{\text{appr}}_{ST} = \frac{1}{\frac{s}{s-1} 2\tilde{N} \left(1 - (1-m)^2 - \frac{1}{2N_{e,tot}}\right) + 1}, \tag{27}$$

where

$$\frac{1}{\tilde{N}} = \frac{(1-m)^2}{N_e} + \frac{2m - m^2}{N} \tag{28}$$

is a weighted harmonic average of N_e and N, with $\tilde{N} = N$ under panmixia ($m = 1$) and $\tilde{N} = N_e$ under complete isolation ($m = 0$).

When $s = \infty$ and $N_e = N$, (27) reduces to none of (6)–(8), although it is much closer to (7)–(8), both of which assume stochastic immigrant allele frequencies, than to (6), which relies on fixed immigrant allele frequencies.

4.3. Migration precedes fertilization

We now reverse the migration and fertilization steps, so that parts of gamete pools rather than individuals migrate (or mix). This could be of biological relevance for, e.g., pollination in plants. Fig. 2 illustrates one generation cycle of this model, a detailed description of which is as follows:

The gamete pools mix to produce s new gamete pools of infinite size. Let m_{ji} denote the proportion of the gamete pool of island j before migration that ends up in the gamete pool of island i after migration. The post-migration allele frequency of island i is then

$$\check{P}_{ti} = \sum_{j=1}^{s} m_{ji}\tilde{P}_{tj}. \tag{29}$$

In (29) we assume that the subpopulation of gamete pool j that migrates to island i consists of so many gametes that its allele

frequency $\tilde{P}_{tj}$ is the same regardless of i. We refer to this as fixed migrant allele frequencies of gametes.

For the dioecious gamete formation model, we achieve (29) (somewhat unrealistically) by assuming equal migration proportions for the male and female alleles, so that $\check{P}_{ti} = (\check{P}_{tim} + \check{P}_{tif})/2$, where $\check{P}_{tim} = \sum_{j=1}^{s} m_{ji}\tilde{P}_{tjm}$ and $\check{P}_{tif} = \sum_{j=1}^{s} m_{ji}\tilde{P}_{tjf}$.

The generation $t+1$ population of island i is defined by drawing $2N$ genes from the post-migration gamete pools of step 2. The resulting allele frequency $P_{t+1,i}$ satisfies

$$\frac{E\left((P_{t+1,i} - \check{P}_{ti})^2|\check{P}_{ti}\right)}{\check{P}_{ti}(1-\check{P}_{ti})} \sim \frac{1}{2N}. \tag{30}$$

For the monoecious models, (30) is achieved by means of binomial sampling, and for the dioecious model through

$$\begin{aligned} P_{t+1,im} &= \text{Bin}(N, \check{P}_{tim})/N, \\ P_{t+1,if} &= \text{Bin}(N, \check{P}_{tif})/N. \end{aligned} \tag{31}$$

4.3.1. Fixed migrant proportions

Assume that the migrant proportions (23) are fixed, and let $\widehat{m'}_{i,t+1}$ refer to the proportion of alleles of island i and generation $t+1$ that originate from other islands. Since alleles are drawn randomly in the fertilization step, after migration, the number of immigrants $N\widehat{m'}_{i,t+1}$ in the diploid case will exhibit random variation around its mean value Nm', even though the migrant proportions are fixed.

It is shown in Appendix B that

$$F^{\text{appr}}_{ST} = \frac{1}{\frac{s}{s-1} 2\tilde{N} \left(1 - (1-m)^2 - \frac{1}{2N_{e,tot}}\right) + \frac{\tilde{N}}{N_e}(1-m)^2}, \tag{32}$$

thus generalizing (8) to $s < \infty$ and $N_e \neq N$. Hence, a stochastic number of immigrants with a stochastic immigrant allele frequency can be interpreted as migration with fixed migrant proportions preceding fertilization.

4.4. Intermediate model

It turns out that a fixed number of immigrants with a stochastic migrant allele frequency (Maruyama, 1970; Sved and Latter, 1977) can be described by means of an intermediate model. After the initial gamete formation step, fertilization takes place within each island to produce a fraction $1 - m$ of individuals, without any migration. In addition, a combined gamete pool is formed, with contributions from all islands, and then, the remaining fraction m of offspring of each island are drawn from the combined gamete pool. In more detail, the steps of the reproduction cycle after gamete formation, can be described as follows:

Fertilization takes place within island i by drawing an exact number $2N(1 - m)$ of genes from the gamete pool with allele frequency $\tilde{P}_{ti}$. Denote the allele frequency of these genes by P^*_{ti}, where

$$\frac{E\left((P^*_{ti} - \tilde{P}_{ti})^2|\tilde{P}_{ti}\right)}{\tilde{P}_{ti}(1-\tilde{P}_{ti})} \sim \frac{1}{2N(1-m)}. \tag{33}$$

The gamete pools from gamete formation within each island are merged, with proportions $w_1, \ldots, w_s$, to a combined gamete pool, with allele frequency

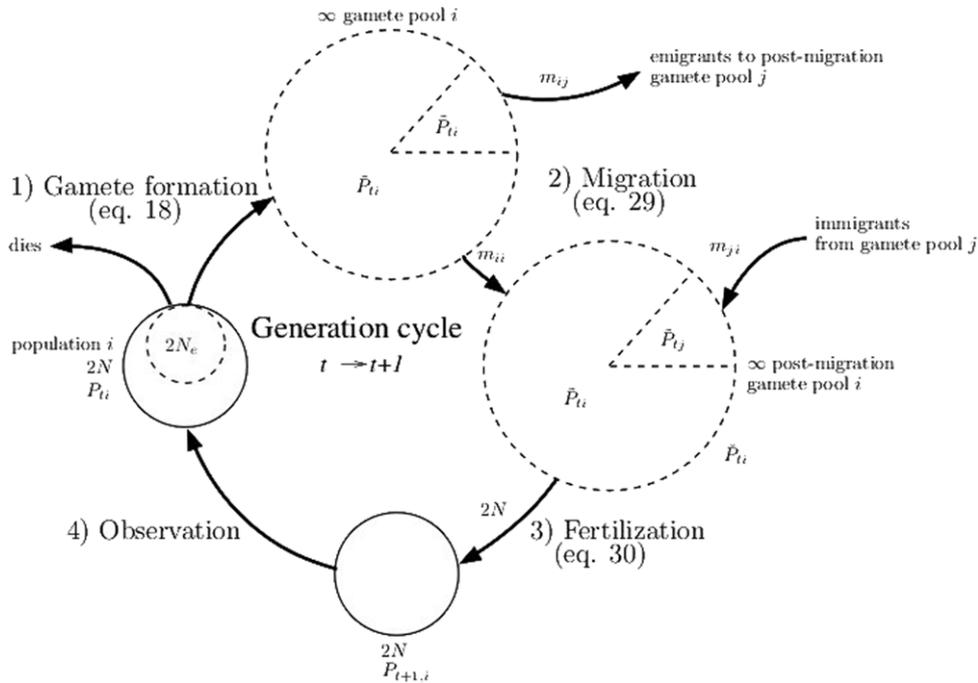$$\tilde{P}_t = \sum_{i=1}^{s} w_i\tilde{P}_{ti}. \tag{34}$$

Fig. 2. One generation cycle when migration precedes fertilization, from the perspective of island i, with haploid migrants. For simplicity, only migration to one other island j is visualized. Fixed migrant proportions m_{ij} and fixed migrant allele frequencies $\tilde{P}_{ti}$ (MF/FF).

For each $i = 1, \ldots, s$, an exact number $2Nm$ of genes are drawn from the combined gamete pool and then migrate to island i. Denote the allele frequency of the genes that end up in island i in this way by P'_{ti}, where

$$\frac{E\left((P'_{ti} - \tilde{P}_t)^2 | \tilde{P}_t\right)}{\tilde{P}_t(1 - \tilde{P}_t)} \sim \frac{1}{2Nm}. \tag{35}$$

Combining the $2N(1 - m)$ genes that are drawn from the gamete pool of island i with the $2Nm$ genes that are drawn from the combined gamete pool and then migrate to island i, we finally obtain the allele frequency

$$P_{t+1,i} = (1 - m)P^*_{ti} + m\hat{P}_{ti} \tag{36}$$

of island i and generation $t + 1$.

Notice that (33) and (35) can be achieved for the monoecious gamete formation model by means of binomial sampling. For the dioecious model, the female and male subpopulations are each sampled binomially, as in (21) and (31).

It is shown in Appendix B that when all islands contribute equally to the combined gamete pool (so called *fixed migrant proportions*, $w_i = 1/s$), (7) generalizes to

$$F^{\text{appr}}_{ST} = \frac{1}{\frac{s}{s-1}2\tilde{N}\left(1 - (1-m)^2 - \frac{1}{2N_{e,tot}}\right) + 1 - \frac{\tilde{N}}{N}m}. \tag{37}$$

This is the natural counterpart of a fixed number of immigrants with stochastic immigrant allele frequency (cf. (7)) when $s < \infty$ and $N_e \neq N$, since exact proportions $1 - m$ and m of all genes are drawn from the same island and the combined gamete pool respectively, but with randomly varying allele frequencies.

5. Numerical results

Expressions for F^{appr}_{ST}, either derived in this paper or previously known, are summarized in Table 2 for the following reproduction models: Fertilization precedes migration with fixed migrant proportions and fixed (FM/FF) or stochastic (FM/FS) migrant allele frequencies, migration precedes fertilization with fixed migrant proportions and migrant allele frequencies (MF/FF) and the intermediate model with fixed migrant proportions (I/F). The X/YZ is such that X specifies the order of fertilization and migration, Y the type of migrant proportions and Z the type of migrant allele frequencies. For the intermediate model Z is not well defined since both gametes and individuals migrate. For none of the models did we stipulate whether the immigrant allele frequencies are stochastic or not, since this is not part of the model specification, but rather a consequence of it.

In order to illustrate these expressions, we evaluated F^{appr}_{ST} numerically as function of m (Table 3), N (Table 4) or s (Table 5). All F^{appr}_{ST} formulas for the finite island model need $N_{e,tot}$ in (14) to be specified. Often this term can be dropped with good accuracy, but more refined choices of $N_{e,tot}$ are described in Appendix D, with (D.1) the default choice unless stated otherwise.

We also performed computer simulations to check on our analytical expressions. Simulations were performed on sets of populations connected by migration as follows. Each simulation run represented a single locus with two alleles, with initial allele frequencies $P_{0i} = 0.5$, $i = 1, \ldots, s$, for the starting generation 0. For scenario FM/FS, a fixed number $2Nm_{ij}$ of migrating genes were drawn (with replacement) from gamete pool i and targeted for population j. When migrants between all $s \times (s - 1)$ population pairs were allotted they were placed in their respective target population, and taken to represent the new generation. Within each generation, non-migrants were similarly drawn with replacement from the parental gamete pool. For large s this approach is not feasible, since $2Nm_{ij}$ is too small (a fraction of a gene). For large s we instead performed simulations from model I/F, sampling immigrants (with replacement) from a conceptually infinite pool with allele frequency equal to the average (34) over populations in the paternal generation. Simulations were carried out for a sufficiently large number of generations (=50) to stabilize $F_{ST,t}$, and repeated in 10 000 replicates.

Table 3
Values of F_{ST}^{appr} for varying m.

m	$N = N_e = 100, s = \infty$					$N = 1000, N_e = 10, s = 5$				
	FM/FF	FM/FS	MF/FF	I/F	Sim	FM/FF	FM/FS	MF/FF	I/F	Sim
0	1.0000	1.0000	1.0000	1.0000	–	1.0000	1.0000	1.0000	1.0000	–
0.1	0.0209	0.0256	0.0258	0.0257	0.0257	0.1515	0.1518	0.1518	0.1518	0.1444
0.2	0.0088	0.0137	0.0138	0.0137	0.0138	0.0680	0.0684	0.0684	0.0684	0.0651
0.3	0.0048	0.0097	0.0098	0.0097	0.0098	0.0377	0.0381	0.0381	0.0381	0.0369
0.4	0.0028	0.0078	0.0078	0.0078	0.0078	0.0223	0.0227	0.0227	0.0227	0.0221
0.5	0.0017	0.0066	0.0067	0.0066	0.0066	0.0133	0.0137	0.0137	0.0137	0.0136
0.6	0.0010	0.0059	0.0059	0.0059	0.0059	0.0077	0.0081	0.0081	0.0081	0.0080
0.7	0.0005	0.0055	0.0055	0.0055	0.0055	0.0040	0.0044	0.0044	0.0044	0.0044
0.8	0.0002	0.0052	0.0052	0.0052	0.0052	0.0017	0.0021	0.0021	0.0021	0.0020
0.9	0.0001	0.0050	0.0051	0.0050	0.0050	0.0004	0.0008	0.0008	0.0008	0.0008
1.0	0.0000	0.0050	0.0050	0.0050	0.0050	0.0000	0.0004	0.0004	0.0004	0.0004

The reproduction models FM/FF, FM/FS, MF/FF and I/F are defined in Table 2 and the simulations (Sim) in the main text.

Table 4
Values of F_{ST}^{appr} for varying N when $N_e = 10$.

N	$m = 0.1, s = \infty$					$m = 1, s = \infty$				
	FM/FF	FM/FS	MF/FF	I/F	Sim	FM/FF	FM/FS	MF/FF	I/F	Sim
10	0.1757	0.2083	0.2169	0.2128	0.2141	0	0.0476	0.0500	0.0500	0.0501
30	0.1757	0.1869	0.1894	0.1882	0.1901	0	0.0164	0.0167	0.0167	0.0166
100	0.1757	0.1791	0.1798	0.1795	0.1804	0	0.0050	0.0050	0.0050	0.0050
300	0.1757	0.1768	0.1771	0.1770	0.1779	0	0.0017	0.0017	0.0017	0.0017
1000	0.1757	0.1760	0.1761	0.1761	0.1777	0	0.0005	0.0005	0.0005	0.0005

The reproduction models FM/FF, FM/FS, MF/FF and I/F are defined in Table 2 and the simulations (Sim) in the main text.

Table 5
Values of F_{ST}^{appr} for varying s.

s	Unadjusted values				Adjusted values based on (3)			
	FM/FF	FM/FS	MF/FF	I/F	FM/FF	FM/FS	MF/FF	I/FF
2	0.0107	0.0109	0.0109	0.0109	0.0214	0.0219	0.0219	0.0219
4	0.0158	0.0162	0.0162	0.0162	0.0211	0.0216	0.0216	0.0216
10	0.0189	0.0193	0.0193	0.0193	0.0210	0.0215	0.0215	0.0215
30	0.0202	0.0207	0.0207	0.0207	0.0209	0.0214	0.0214	0.0214
100	0.0207	0.0211	0.0212	0.0212	0.0209	0.0214	0.0214	0.0214
300	0.0208	0.0213	0.0213	0.0213	0.0209	0.0214	0.0214	0.0214
∞	0.0209	0.0214	0.0214	0.0214	0.0209	0.0214	0.0214	0.0214

$N = 1000, N_e = 100$ and $m = 0.1$. The reproduction models FM/FF, FM/FS, MF/FF and I/F are defined in Table 2.

Table 6
Values of G_{ST}^{appr} and $G_{ST,t}$ for varying m.

m	FM/FF	FM/FS	MF/FF	I/F	Sim1	Sim2	Sim3	t
0.1	0.0371	0.0454	0.0458	0.0457	0.0455	0.0469	0.0499	50
0.2	0.0158	0.0245	0.0247	0.0246	0.0244	0.0210	0.0252	20
0.3	0.0086	0.0174	0.0175	0.0175	0.0160	0.0168	0.0175	20
0.4	0.0050	0.0139	0.0140	0.0140	0.0145	0.0135	0.0131	20
0.5	0.0030	0.0119	0.0120	0.0119	0.0113	0.0119	0.0120	20
0.6	0.0017	0.0106	0.0107	0.0107	0.0102	0.0107	0.0105	10
0.7	0.0009	0.0098	0.0099	0.0099	0.0096	0.0101	0.0097	10
0.8	0.0004	0.0093	0.0094	0.0094	0.0094	0.0094	0.0093	10
0.9	0.0001	0.0090	0.0091	0.0091	0.0092	0.0090	0.0091	10
1.0	0.0000	0.0089	0.0090	0.0090	0.0089	0.0089	0.0090	10

$N = N_e = 50$ and $s = 10$. The reproduction models FM/FF, FM/FS, MF/FF and I/F are defined in Table 2. For each parameter combination, simulated values of $G_{ST,t}$ are displayed based on three runs Sim1–Sim3 with the EASYPOP computer program (Balloux, 2001) with 100 loci having 99 alleles each. For each of the three replicates, t generations were generated in order for $G_{ST,t}$ to attain some stability. Easypop assumes real data allele frequencies $\hat{P}_{ti}$, estimated from finite samples, and therefore corrects the resulting estimates $\hat{G}_{ST,t}$ by removing the sampling bias. Since our analysis is based on the true allele frequencies P_{ti} we have removed this finite sample correction.

The three stochastic immigrant allele frequency models FM/FS, MF/FF, and I/F have almost identical values of F_{ST}^{appr}, whereas those of the fixed immigrant allele frequency model FM/FF are sometimes substantially smaller, particularly for large m and small N; see the left part of Table 3 and right of Table 4. When the immigrant allele frequency is stochastic and the migration rate high, it turns out that F_{ST}^{appr} varies quite a lot with N, at least in relative terms, even though N_e is kept fixed, as shown in the right part of Table 4.

By increasing s from 2 to ∞ we essentially double F_{ST}^{appr}, see the left part of Table 5. In order to decrease the dependence on s, we have included the corresponding equilibrium values for the adjusted fixation index in the right part of Table 5, obtained by multiplying the unadjusted values with $s/(s-1)$. Indeed, one notices that the adjusted equilibrium values are virtually independent of s, as discussed in the Introduction.

Fig. 3 shows the dynamics of the fixation index for $N_e = 50$, $N = 100$ and four different combinations of s and m. For each model, one time series $F_{ST,t}^{\text{adj}}$ is plotted (solid lines) as function of t, together with the average $\bar{F}_{ST,t}^{\text{adj}}$ (dashed line), estimated from 10 000 simulations, and the approximate quasi equilibrium value $F_{ST}^{\text{appr,adj}}$ (horizontal dotted line). It is seen that $F_{ST,t}^{\text{adj}}$ converges much more rapidly to its quasi equilibrium limit when m is large and the magnitude of the oscillations around this limit decreases with increasing s and m.

On the other hand, for the multilocus extension $G_{ST,t}$ of the fixation index, the oscillations around the equilibrium limit are much smaller within as well as between replicates, see Table 6.

All treatments up till now refer to population allele frequencies. In practical applications only sample allele frequencies are available. In order to study the performance of the estimator $\hat{F}_{ST,t}$ in (4) (with $c = 1$), we compared it in a simulation study with three well known estimators due to Weir and Cockerham (1984), Nei (1973), and Nei and Chesser (1983). The simulations refer to a diploid, dioecious organism in which N_e out of N individuals in each population ($s = 2$ or 5 populations) participate in breeding each generation (population model FM/FS). Both equal ($P_{0i} = 0.5$) and skewed ($P_{0i} = 0.9$) initial allele frequencies are simulated. Samples of n individuals for genetic analyzes (10 loci) are drawn from 2 subpopulations without replacement after 25 generations. The results are reported in Table 7. It is seen that $\hat{F}_{ST,t}$ has the overall smallest bias. The same good performance of $\hat{F}_{ST,t}$ (results not shown) remains when N differs from N_e.
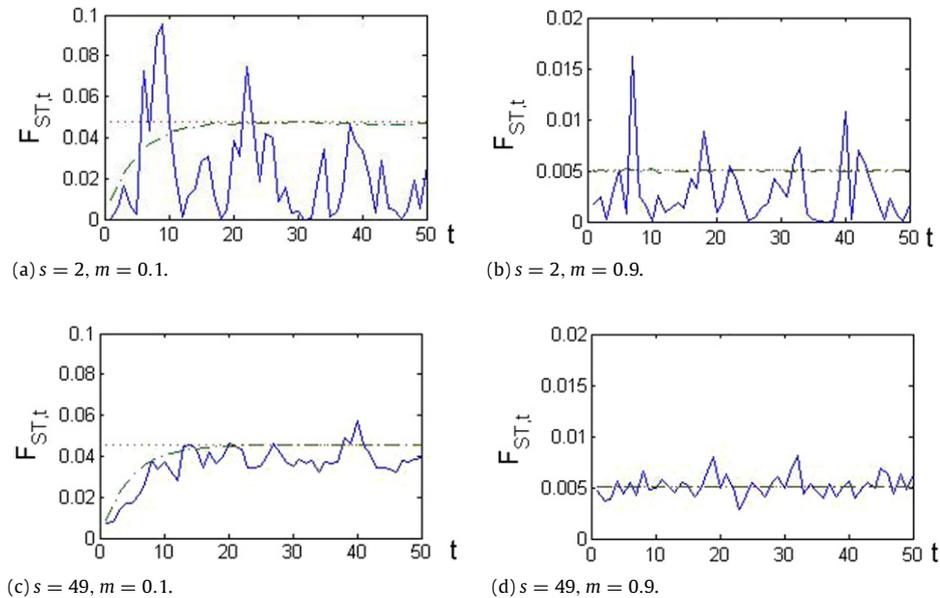
(a) $s = 2$, $m = 0.1$. (b) $s = 2$, $m = 0.9$.

(c) $s = 49$, $m = 0.1$. (d) $s = 49$, $m = 0.9$.

Fig. 3. One locus plots of $F_{ST,t}^{\text{adj}}$ as function of t when $N_e = 50$, $N = 100$ and (a) $s = 2$, $m = 0.1$, (b) $s = 2$, $m = 0.9$, (c) $s = 49$, $m = 0.1$ and (d) $s = 49$, $m = 0.9$. The solid lines show $F_{ST,t}^{\text{adj}}$ from one simulation, the dash-dotted lines are averages of $F_{ST,t}^{\text{adj}}$ from 10 000 simulations and the dotted horizontal line corresponds to the approximate limit as t increases, that is, $F_{ST}^{\text{appr,adj}}$. We have chosen the adjusted values rather than the unadjusted ones in order to facilitate the effect of varying s (the subplots within each column). Notice that the average $F_{ST,t}^{\text{adj}}$ converges much more quickly to the quasi equilibrium limit when $m = 0.9$, and that the magnitude of the oscillations decreases with increasing s and m. See the numerical results section for details on the simulations.

Table 7

Expected (quasi) equilibrium values of $F_{ST,t}^{\text{adj}}$ and estimates thereof for $k = 2$ sampled islands.

Parameters						Simulated expected values				
s	N_e	n	m	P_{0i}	F_{ST}^{appr}	$F_{ST,t}^{\text{adj}}$	$\hat{F}_{ST,t}^{\text{WC}}$	$\hat{G}_{ST,t}$	$\hat{F}_{ST,t}^{\text{NC}}$	$\hat{F}_{ST,t}$
2	100	50	0.1	0.5	0.0263	0.0260	0.0207	0.0310	0.0210	0.0257
2	100	50	0.5	0.5	0.0066	0.0066	0.0016	0.0116	0.0016	0.0066
2	100	50	1	0.5	0.0050	0.0050	0.0001	0.0100	0.0000	0.0050
2	100	100	0.1	0.5	0.0262	0.0260	0.0210	0.0262	0.0212	0.0258
2	100	100	0.5	0.5	0.0066	0.0066	0.0015	0.0066	0.0016	0.0065
2	100	100	1	0.5	0.0050	0.0050	0.0000	0.0050	0.0000	0.0050
2	100	50	0.1	0.9	0.0264	0.0260	0.0214	0.0318	0.0218	0.0263
2	100	50	0.5	0.9	0.0066	0.0066	0.0016	0.0116	0.0016	0.0066
2	100	50	1	0.9	0.0050	0.0050	−0.0001	0.0100	0.0000	0.0049
2	100	100	0.1	0.9	0.0257	0.0260	0.0204	0.0258	0.0208	0.0253
2	100	100	0.5	0.9	0.0067	0.0066	0.0017	0.0066	0.0016	0.0066
2	100	100	1	0.9	0.0049	0.0050	−0.0001	0.0050	0.0000	0.0049
5	100	50	0.1	0.5	0.0260	0.0260	0.0209	0.0314	0.0212	0.0259
5	100	50	0.5	0.5	0.0066	0.0066	0.0016	0.0116	0.0016	0.0066
5	100	50	1	0.5	0.0049	0.0050	−0.0001	0.0100	−0.0002	0.0049
5	100	100	0.1	0.5	0.0258	0.0260	0.0208	0.0260	0.0210	0.0257
5	100	100	0.5	0.5	0.0067	0.0066	0.0017	0.0068	0.0018	0.0067
5	100	100	1	0.5	0.0050	0.0050	0.0000	0.0050	0.0000	0.0049
5	100	50	0.1	0.9	0.0258	0.0260	0.0208	0.0312	0.0212	0.0257
5	100	50	0.5	0.9	0.0066	0.0066	0.0016	0.0116	0.0016	0.0066
5	100	50	1	0.9	0.0050	0.0050	0.0000	0.0100	0.0000	0.0050
5	100	100	0.1	0.9	0.0258	0.0260	0.0207	0.0260	0.0210	0.0256
5	100	100	0.5	0.9	0.0066	0.0066	0.0016	0.0066	0.0016	0.0066
5	100	100	1	0.9	0.0050	0.0050	0.0000	0.0050	0.0000	0.0049

The reproduction scheme is FM/FS with $N = N_e$. Expected values of $F_{ST,t}^{\text{adj}}$ are compared with those of four different estimators based on 5000 simulations. Sampling of $n = n_1 = n_2$ individuals is with replacement from $k = 2$ islands over 10 loci after $t = 25$ generations. Simulations start at generation 0, with all the s islands having identical allele frequencies P_{0i}. The four estimators are (I) $\hat{F}_{ST,t}^{\text{WC}}$ (Weir and Cockerham, 1984, Formula 6), (II) $\hat{G}_{ST,t}$ (Nei, 1973), (III) $\hat{F}_{ST,t}^{\text{NC}}$ (Nei and Chesser, 1983) and (IV) $\hat{F}_{ST,t}$ in (4), with $c = 1$. Estimators II–III are multiplied by a factor 2 to compensate for $k = 2$. For comparison, we also included F_{ST}^{appr} in (27), with $N_{e,\text{tot}} = sN_e$. In contrast to I–IV, F_{ST}^{appr} is computed from population allele frequencies (P_{ti}) rather than sampled ones ($\hat{P}_{ti}$).

6. Discussion

6.1. Summary and extensions

In this paper, we have introduced a novel approach for computing approximations F_{ST}^{appr} of the quasi equilibrium limit F_{ST}^{eq} of the fixation index, for spatial mutation free models that exhibit migra-tion between subpopulations. This approach is suitable for the rel-atively short time spans encountered, e.g. in conservation genetics and requires that the mutation rate is of smaller order than the mi-gration rate.

We applied our methodology to the finite island model and our findings can be summarized as follows:

1. By highlighting gamete formation as one part of the reproduction cycle, we are able to distinguish the local census size N from the effective local population size N_e.

2. A number of different reproduction models can be defined with different order of fertilization and migration, fixed or stochastic migrant proportions, and fixed or stochastic migrant allele frequencies. They are all based on a number of simplifying assumptions concerning the reproduction cycle.

3. We evaluated F_{ST}^{appr} for four models; FM/FF, FM/FS, MF/FF and I/F, analytically and some of them by simulation, as function of s, m, N_e and N.

4. Three of the four models; FM/FS, MF/FF and I/F, have almost identical values of F_{ST}^{appr} for all parameter values. Their common feature is a stochastic allele frequency of the immigrants to each island. This indicates that the order of fertilization and migration is less important, a similar conclusion as obtained by Nagylaki (1983) in a slightly different framework.

5. The fourth model, FM/FF, has consistently lower values of F_{ST}^{appr}. This is caused by the rather unrealistic assumption of a fixed allele frequency of immigrants, as manifested by an extra term $(1 - m)^2$ in the numerator of the defining equation of F_{ST}^{appr}. Similar values of F_{ST}^{appr} as for FM/FF are obtained with the traditional identical-by-state probabilities approach. The difference in F_{ST}^{appr} between FM/FF and the other three models is minor when m is small, but quite substantial for large m.

6. A interesting finding is that F_{ST}^{appr} depends not only on N_e, but also on N, for all parameter combinations. However, the dependence of F_{ST}^{appr} on N is substantial only when m is large and N_e small.

7. For most practical purposes, it is enough to distinguish FM/FF from the group FM/FS, MF/FF and I/F as far as computation of F_{ST}^{appr} is concerned. The same can be said for the total variance effective population size $N_{e,\mathrm{tot}}$.

In order to highlight the novelties of our approach, we limited our study to relatively simple models. However, several extensions of our work are possible. First, it is of interest to investigate the effect of having $N \neq N_e$ for other quantities than F_{ST}. It turns out that the value of $N \neq N_e$ can impact the estimate of $N_{e,\mathrm{tot}}$ by the temporal method when subpopulation structure is ignored (Ryman et al., submitted for publication). Second, one may introduce an extra source of variability by allowing for randomly varying migrant proportions, thereby increasing F_{ST}. Third, it would be of interest to consider unequal male and female migration rates when gametes, not individuals, migrate, as in Section 4.3. Fourth, more general spatial structures than the island model can be treated, including, for instance, the hierarchical island model, the one- and two-dimensional stepping stone models and the circular stepping stone model. This has been done by Sawyer (1976) and Nagylaki (1980, 1983) using the Malécot (1951) recursions for identical by state probabilities. In Hössjer and Ryman (submitted for publication) we do the same, using instead the quasi equilibrium approach. Fifth, in real populations the sizes of the subpopulations and the sex-ratios of dioecious models change with time. Waples and Do (1994) show that even small random fluctuations of the sex ratios decrease the effective population size and Olsson et al. (submitted for publication) find explicit expressions for the variance effective size of age-structured models with variable sizes of the age classes. A natural extension of the present work is to investigate how the fixation index is affected by subpopulation fluctuations.

6.2. Interpretation of results

One of the major findings of this paper is that the equilibrium fixation index depends not only on N_e and m, but also on N, whenever migration occurs, for all reproduction models except FM/FF, i.e. the original island model of Wright. Since the equilibrium fixation index results from a balance between genetic drift and migration, it is natural to look at this balance more closely, in order to understand why $\tilde{N}$ rather than N_e appears in the formulas for F_{ST}^{appr} for all reproduction scenarios except FM/FF. It follows from the calculations in Appendix B that (15) is well approximated by

$$F_{ST,t+1} \approx (1 - m)^2 F_{ST,t} + B + \epsilon_{t+1}, \tag{38}$$

for large populations. The right hand side of (38) illustrates how the balance between migration and genetic drift affects the fixation index from one generation to the next. The term $(1 - m)^2$ gets smaller the larger the migration rate is, and it tends to decrease the fixation index. The genetic drift, on the other hand, is involved in the non-negative term B and hence tends to increase the fixation index. The genetic drift can be decomposed into two parts, and only the first, which concerns differentiation between islands, affects the fixation index, so that B is the average effect of random differentiation between islands. The second part, which concerns random drift of the allele frequency of the total population, P_t, has no effect on the fixation index.

It also follows from Appendix B that

$$B \approx \frac{s-1}{s} \left\{ (1 - m)^2 \left(\frac{1}{2N_e} - \frac{1}{2N} \right) + 1 \cdot \frac{1}{2N} \right\}$$

$$= \frac{s-1}{s} \frac{1}{2\tilde{N}} \tag{39}$$

is sufficiently accurate, for reproduction scenarios FM/FS, MF/FF and I/F, in order to explain the main features of the recursion (38). The $(s-1)/s$ term is only present in the recursion of the unadjusted fixation index. It corresponds to the relative size of that part of the genetic drift that affects differentiation between subpopulations. The effect of migration is to reduce gamete formation variance by a term $(1 - m)^2$, whereas the fertilization variance remains unchanged, as illustrated by a term 1. The removed part of the gamete formation variance only affects the total population as a whole; not differentiation between islands.

On the other hand, for reproduction scenario FM/FF,

$$B \approx \frac{s-1}{s} \left\{ (1 - m)^2 \left(\frac{1}{2N_e} - \frac{1}{2N} \right) + (1 - m)^2 \cdot \frac{1}{2N} \right\}$$

$$= \frac{s-1}{s} \frac{(1 - m)^2}{2N_e}. \tag{40}$$

The only difference compared to (39) is that migration now also affects the fertilization variance by a term $(1 - m)^2$, since the immigrant allele frequencies of all islands are the same and hence will not contribute to any change of $F_{ST,t}$.

The gamete formation term $(1 - m)^2 \left((2N_e)^{-1} - (2N)^{-1} \right)$ that appears in (39)–(40) can be interpreted as an Allendorf–Phelps effect (Allendorf and Phelps, 1981; Waples, 1998), by which an extra source of variation is added when offspring are not drawn from the whole parental generation, but rather from a subset of breeders.

By inserting $A = (1 - m)^2$ and either (39) or (40) into (13), we find that the overall effect on F_{ST}^{appr} of having a fixed immigrant allele frequency is that N_e replaces $\tilde{N}$ in the denominator of F_{ST}^{appr} and a multiplicative term $(1 - m)^2$ appears in the numerator of F_{ST}^{appr}. Hence, the appearance of N_e in the well-known expression (9) for the standard infinite island model of Wright, is seen to be a consequence of fixed migrant allele frequencies.

6.3. Potential applications

Suppose that weak differentiation has been observed between the subpopulations of a population. In order to determine whether the observed allele frequency differences are significant, one may set up a test and reject the null hypothesis H_0 of panmixia ($m = 1$) when an estimate $\hat{F}_{ST}$ (or $\hat{G}_{ST}$) gets too large. In order to compute p-values, the sampling distribution of $\hat{F}_{ST}$ under H_0 has to be known. However, it will not only depend on the sampling scheme and sample size, but also on the quasi equilibrium distribution of F_{ST} under panmixia. See also Waples (1989b), where similar issues are discussed when temporal rather than spatial variation of allele frequencies are tested.

First, formula (10) implies $F_{ST}^{eq} = 0$ and hence $F_{ST} \equiv 0$ under panmixia, since F_{ST}^{eq} is defined as the mean of the quasi equilibrium distribution. This suggests that any significant departure of $\hat{F}_{ST}$ from zero should be interpreted as genetic differentiation. Indeed, Hauser and Carvalho (2008) report a number of significant findings of subpopulations structure for marine fishes. However, our results can be summarized as

$$F_{ST}^{appr,adj} \approx \begin{cases} 1/(4N_e m + 1), \\ \quad m \to 0, \text{ FM/FS, MF/FF, I/F, FM/FF,} \\ 1/(2N_e), \quad m = 1, \text{ FM/FF,} \\ 1/(2N), \quad m = 1, \text{ FM/FS, MF/FF and I/F,} \end{cases} \quad (41)$$

for small and large migration rates. In particular, $F_{ST}^{appr,adj}$ is close to $1/(2N)$ under panmixia, regardless of the value of N_e, for all reproduction scenarios except the least realistic one, FM/FF. It may be the case, for a well powered test based on large sample sizes, that the overly optimistic assumption $F_{ST}^{eq} = 0$ when $m = 1$ leads to *too many* reported cases of subpopulation substructure. We therefore suggest more research in order to work out a more realistic null distribution of $\hat{F}_{ST}$. Such an analysis should take overlapping generations as well as the number of investigated loci into account.

Acknowledgments

Appendix A. Approximate unbiasedness of $\hat{F}_{ST,t}$ in (4)

Let $E(\cdot)$ denote expectation when first a random subset $I \subset \{1, \ldots, s\}$ of k islands are selected, and then $2n_i$ genes are sampled for each $i \in I$ from $2N$, with or without replacement. We use the approximation

$$E(\hat{F}_{ST,t}) \approx \frac{E(\hat{S}^2) - \alpha P_t(1 - P_t)}{(1 - c\alpha)E(\hat{P}_t(1 - \hat{P}_t)) + E(\hat{S}^2)/k}, \quad (A.1)$$

which is more accurate the larger $2 \leq k \leq s$ and all n_i are, since the coefficients of variation of the numerator and denominator of $\hat{F}_{ST,t}$ are both decreasing functions of k and $\{n_i; i \in I\}$. The constant α is defined as

$$\alpha = \begin{cases} \dfrac{1}{k} \displaystyle\sum_{i \in I} \dfrac{1}{2n_i}, & \text{sampling with replacement,} \\ \dfrac{1}{k} \displaystyle\sum_{i \in I} \dfrac{1}{2n_i} \dfrac{2N - 2n_i}{2N - 1}, & \text{sampling without replacement.} \end{cases}$$

In practice N should be replaced by some estimator $\hat{N}$ when sampling is without replacement.

Letting $\delta_{ti} = P_{ti} - P_t$, we find, after some computations, that

$$S^2 = \frac{1}{k - 1} \sum_{i \in I} (P_{ti} - P_t)^2$$

$$= \frac{1}{k - 1} \sum_{i \in I} \left(\delta_{ti} - \frac{1}{k} \sum_{j \in I} \delta_{tj} \right)^2$$

$$= \frac{1}{k} \sum_{i \in I} \delta_{ti}^2 - \frac{2}{k(k-1)} \sum_{(i,j) \in I; i < j} \delta_{ti}\delta_{tj}.$$

If $I = (I_1, \ldots, I_k)$, with the indexes numbered in random order, it follows that

$$E(S^2) = E(\delta_{tI_1}^2) - E(\delta_{tI_1}\delta_{tI_2})$$

$$= \frac{s}{s - 1} E(\delta_{tI_1}^2)$$

$$= \frac{s}{s - 1} P_t(1 - P_t) F_{ST,t}$$

$$= P_t(1 - P_t) F_{ST,t}^{adj}. \quad (A.2)$$

Since (A.1) is verified in the same way for both sampling schemes, we assume for simplicity that samples are taken with replacement. Then the sampling errors $\Delta_{ti} = \hat{P}_{ti} - P_{ti}$ satisfy

$$E(\Delta_{ti}|P_{ti}) = 0,$$

$$E(\Delta_{ti}^2|P_{ti}) = \frac{P_{ti}(1 - P_{ti})}{2n_i}.$$

Conditioning on $\mathcal{P} = (P_{ti}; i \in I)$ we thus obtain

$$E(\hat{S}^2) - E(S^2) = \frac{1}{k - 1} E\left(E\left(\sum_{i \in I} (\Delta_{ti} - \Delta_t)^2 \Big| \mathcal{P} \right) \right)$$

$$= \frac{1}{k} E\left(\sum_{i \in I} E(\Delta_{ti}^2 | P_{ti}) \right)$$

$$= \frac{1}{k} E\left(\sum_{i \in I} \frac{P_{ti}(1 - P_{ti})}{2n_i} \right)$$

$$= \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2n_i} \cdot \frac{1}{s} \sum_{j=1}^{s} P_{tj}(1 - P_{tj})$$

$$= \alpha P_t(1 - P_t)(1 - F_{ST,t}), \quad (A.3)$$

where $\Delta_t = \sum_{i \in I} \Delta_{ti}/k$, and the outer expectation was taken with respect to variation in I.

Putting $\tilde{P}_t = \sum_{i \in I} P_{ti}/k$, we find that

$$E(\tilde{P}_t(1 - \tilde{P}_t)) = P_t(1 - P_t) - E\left(\frac{1}{k} \sum_{i \in I} \delta_{ti} \right)^2$$

$$= P_t(1 - P_t) - \frac{1}{k} E(\delta_{tI_1}^2) - \frac{k - 1}{k} E(\delta_{tI_1}\delta_{tI_2})$$

$$= P_t(1 - P_t) + \frac{(k-1)/(s-1) - 1}{k} E(\delta_{tI_1}^2)$$

$$= P_t(1 - P_t) \left(1 - \frac{s - k}{k(s - 1)} F_{ST,t} \right), \quad (A.4)$$

and

$$E(\hat{P}_t(1 - \hat{P}_t)) - E(\tilde{P}_t(1 - \tilde{P}_t)) = -E(E(\Delta_t^2|\mathcal{P}))$$

$$= -\frac{1}{k^2} E\left(\sum_{i \in I} \frac{1}{2n_i} P_{ti}(1 - P_{ti}) \right)$$

$$= -\frac{\alpha}{k} P_t(1 - P_t)(1 - F_{ST,t}). \quad (A.5)$$

Inserting (A.2)–(A.5) into (A.1), we arrive, after some computations, at

$$E(\hat{F}_{ST,t}) \approx F_{ST,t}^{\text{adj}}$$

$$\times \frac{1 - \frac{k-1}{k}\alpha + \frac{s-1}{ks}\frac{1-F_{ST,t}}{F_{ST,t}}\alpha^2}{1 + \frac{F_{ST,t}}{s-1} - c\alpha + c\alpha\left(\frac{s-k}{k(s-1)}F_{ST,t} + \frac{1-F_{ST,t}}{k}\alpha\right)}. \quad (A.6)$$

The right hand side of (A.6) essentially equals $F_{ST,t}^{\text{adj}}$ when $\alpha^2 \ll F_{ST,t}$ and either $F_{ST,t}$ is small or s is large. The optimal c, in terms minimal bias, will depend on α, $F_{ST,t}$, k and s, although $c = 1$ seems to be a good choice for quite a wide range of parameter values. □

Appendix B. Explicit expressions for $F_{ST,t+1}^{\text{appr}}$

In order to compute F_{ST}^{appr}, we need to find expressions for A and B in (13) to insert into (16). To this end, we consider models for which the allele frequency of island i and generation $t + 1$ can be expressed recursively as

$$P_{t+1,i} = (1 - m)P_{ti} + mP_t + \varepsilon_{t+1,i}, \quad (B.1)$$

with fraction $1-m$ and m contributions from the allele frequencies of the previous generation t of the same island i and the total population respectively, and additionally a random error term $\varepsilon_{t+1,i}$. All expected values and covariances are conditional on generation t, but for simplicity of notation, we don't write this out. Assuming that $E(\varepsilon_{t+1,i}) = 0$ and putting $\text{Cov}(\varepsilon_{t+1,i}, \varepsilon_{t+1,j}) = \sigma_{ij}$, we can rewrite the recursion (13) as

$$E(F_{ST,t+1}) = \frac{(1-m)^2 F_{ST,t}}{1 - (2N_{e,\text{tot}})^{-1}}$$

$$+ \frac{\frac{s-1}{s^2}\text{tr}(\Sigma) - \frac{1}{s^2}\sum_{i\neq j}\sigma_{ij}}{P_t(1-P_t)\left(1 - (2N_{e,\text{tot}})^{-1}\right)}, \quad (B.2)$$

where $\text{tr}(\Sigma) = \sum_i \sigma_{ii}$ is the diagonal sum of the covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^s$. Indeed, it follows from (B.1) that

$$P_{t+1,i} - P_{t+1} = (1 - m)(P_{ti} - P_t) + \varepsilon_{t+1,i} - \varepsilon_{t+1}, \quad (B.3)$$

with $\varepsilon_{t+1} = \sum_{i=1}^s \varepsilon_{t+1,i}/s$, and hence

$$E(F_{ST,t+1}) = \frac{\sum_{i=1}^s E\left((1-m)(P_{ti} - P_t) + \varepsilon_{t+1,i} - \varepsilon_{t+1}\right)^2}{sP_t(1-P_t)\left(1 - (2N_{e,\text{tot}})^{-1}\right)}$$

$$= \frac{(1-m)^2 F_{ST,t}}{1 - (2N_{e,\text{tot}})^{-1}} + \frac{\sum_{i=1}^s E\left(\varepsilon_{t+1,i} - \varepsilon_{t+1}\right)^2}{sP_t(1-P_t)\left(1 - (2N_{e,\text{tot}})^{-1}\right)},$$

which after some calculations simplifies to (B.2).

If in addition the error term in (B.1) decomposes as

$$\varepsilon_{t+1,i} = \eta_{t+1,i} + \xi_{t+1}, \quad (B.4)$$

with ξ_{t+1} identical for all i and the covariance matrix $D = (\text{Cov}(\eta_{t+1,i}, \eta_{t+1,j}))_{i,j=1}^s = (d_{ij})_{i,j=1}^s$ diagonal, (B.2) simplifies to

$$E(F_{ST,t+1}) = \frac{(1-m)^2 F_{ST,t}}{1 - (2N_{e,\text{tot}})^{-1}}$$

$$+ \frac{s-1}{s} \cdot \frac{\frac{1}{s}\text{tr}(D)}{P_t(1-P_t)} \cdot \frac{1}{1 - (2N_{e,\text{tot}})^{-1}}. \quad (B.5)$$

Indeed, (B.4) implies that

$$P_{t+1,i} - P_{t+1} = (1 - m)(P_{ti} - P_t) + \eta_{t+1,i} - \eta_{t+1}, \quad (B.6)$$

with $\eta_{t+1} = \sum_{i=1}^s \eta_{t+1,i}/s$. Since (B.6) is analogous with (B.3), with $\eta_{t+1,i}$ instead of (B.4), it follows that (B.5) is proved in the same way as (B.2) with D instead of Σ, recalling that D is diagonal.

It turns out that (B.4) holds for all breeding, fertilization and migration scenarios of the island model considered in this paper. We will see below, separately for each model, how A and B can be derived from (B.5) in order to compute F_{ST}^{appr} from (16). □

Fixed migrant proportions and allele frequencies imply fixed immigrant allele frequencies. From (22)–(24) it follows that the allele frequency of island i after a completed reproduction cycle $t \to t + 1$ is

$$P_{t+1,i} = (1 - m)P_{tii}^* + \frac{m}{s}\sum_{j=1}^s P_{tji}^*$$

$$= (1 - m)P_{ti}^* + \frac{m}{s}\sum_{j=1}^s P_{tj}^*$$

$$\to (1 - m)P_{ti}^* + mP_t \quad \text{as } s \to \infty. \quad (B.7)$$

In the last step of (B.7) we used the Law of Large Numbers, as is easily verified by means of a variance calculation, followed by Chebyshev's Inequality.

Hence, a fraction $1 - m$ of the alleles in island i and generation $t+1$ are drawn by means of binomial sampling, and, when s is large, the remaining fraction m has exactly the same allele frequency P_t as the total population, i.e. a fixed immigrant allele frequency, as claimed.

Proof of (25). In this case (B.4) holds, with $\eta_{t+1,i} = (1 - m)(P_{ti}^* - P_{ti})$ and $\xi_{t+1} = m\sum_i (P_{ti}^* - P_t)/s$. Hence it follows from (18) and (19) that

$$d_{ii} = \frac{(1-m)^2 P_{ti}(1 - P_{ti})}{2N_e},$$

and

$$\frac{1}{s}\text{tr}(D) = \frac{(1-m)^2}{2N_e}\frac{1}{s}\sum_{i=1}^s P_{ti}(1 - P_{ti})$$

$$= \frac{(1-m)^2}{2N_e}P_t(1 - P_t)(1 - F_{ST,t}).$$

Inserting this expression into (B.5) we obtain

$$A = \frac{(1-m)^2\left(1 - \frac{s-1}{s}\frac{1}{2N_e}\right)}{1 - (2N_{e,\text{tot}})^{-1}} \quad (B.8)$$

and

$$B = \frac{s-1}{s}\frac{(1-m)^2}{2N_e}\frac{1}{1 - (2N_{e,\text{tot}})^{-1}}$$

which are then plugged into (16), so that we finally arrive at (25). □

Proof of (27). Because of (23), we have that

$$P_{t+1,i} = \sum_{j=1}^s m_{ji}P_{tij}^*$$

$$= (1 - m)P_{ti} + mP_t + \sum_{j=1}^s m_{ji}(\tilde{P}_{tj} - P_{tj})$$

$$+ \sum_{j=1}^s m_{ji}(P_{tji}^* - \tilde{P}_{tj})$$

$$= (1 - m)P_{ti} + mP_t + \frac{m}{s}\sum_{j=1}^s (\tilde{P}_{tj} - P_{tj})$$

$$+ (1 - m)(\tilde{P}_{ti} - P_{ti}) + \sum_{j=1}^s m_{ji}(P_{tji}^* - \tilde{P}_{tj}),$$

so that (B.1) holds with error term (B.4) and

$$\eta_{t+1,i} = (1-m)(\tilde{P}_{ti} - P_{ti}) + \sum_{j=1}^{s} m_{ji}(P_{tji}^{*} - \tilde{P}_{tj})$$

and $\xi_{t+1} = m \sum_{j=1}^{s}(\tilde{P}_{tj} - P_{tj})/s$. Hence

$$\frac{1}{s}\mathrm{tr}(D) \sim (1-m)^2 \left(\frac{1}{2N_e} - \frac{1}{2N}\right)\frac{1}{s}\sum_{i=1}^{s}P_{ti}(1-P_{ti})$$

$$+ \frac{1}{s}\sum_{i,j=1}^{s}m_{ji}^2 \frac{P_{tj}(1-P_{tj})}{2Nm_{ji}}$$

$$= \left((1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)+\frac{1}{2N}\right)\frac{1}{s}\sum_{i=1}^{s}P_{ti}(1-P_{ti})$$

$$= \frac{1}{2\tilde{N}}P_t(1-P_t)(1-F_{ST,t}).$$

Inserting this expression into (B.5) we find that

$$A = \frac{(1-m)^2 - \frac{s-1}{s}\frac{1}{2\tilde{N}}}{1-(2N_{e,\mathrm{tot}})^{-1}}$$

and

$$B = \frac{s-1}{s}\frac{1}{2\tilde{N}}\frac{1}{1-(2N_{e,\mathrm{tot}})^{-1}}, \tag{B.9}$$

which in turn are inserted into (16), and this gives (27). □

Proof of (32). We have that

$$P_{t+1,i} = (1-m)\tilde{P}_{ti} + m\tilde{P}_t + (P_{t+1,i} - \check{P}_{ti})$$
$$= \bar{P}_{ti} + (1-m)(\tilde{P}_{ti} - P_{tit}) + m(\tilde{P}_t - P_t) + (P_{t+1,i} - \check{P}_{ti}),$$

where $\bar{P}_{ti} = (1-m)P_{ti} + mP_t$. Hence (B.1) holds with error term decomposable as (B.4), with

$$\eta_{t+1,i} = (1-m)(\tilde{P}_{ti} - P_{ti}) + (P_{t+1,i} - \check{P}_{ti})$$

and $\xi_{t+1} = m(\tilde{P}_t - P_t)$. It follows from (18) and (30) that

$$d_{ii} \sim (1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)P_{ti}(1-P_{ti}) + \frac{1}{2N}E\left(\check{P}_{ti}(1-\check{P}_{ti})\right)$$

$$\sim (1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)P_{ti}(1-P_{ti}) + \frac{1}{2N}\bar{P}_{ti}(1-\bar{P}_{ti}),$$

where in the last step we used that $E(\check{P}_{ti}) = \bar{P}_{ti}$ and $E((\check{P}_{ti}-\bar{P}_{ti})^2) \to 0$ as N grows. Hence

$$\frac{1}{s}\mathrm{tr}(D) \sim (1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)\frac{1}{s}\sum_{i=1}^{s}P_{ti}(1-P_{ti})$$

$$+ \frac{1}{2N}\frac{1}{s}\sum_{i=1}^{s}\bar{P}_{ti}(1-\bar{P}_{ti})$$

$$= (1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)P_t(1-P_t)(1-F_{ST,t})$$

$$+ \frac{1}{2N}P_t(1-P_t)(1-(1-m)^2 F_{ST,t})$$

$$= P_t(1-P_t)\left(\frac{1}{2\tilde{N}}-(1-m)^2\frac{1}{2N_e}F_{ST,t}\right),$$

and consequently, inserting this expression into (B.5), we find that A and B are as in (B.8) and (B.9). Plugging these formulas for A and B into (16), we arrive at (32). □

Proof of (37). It follows from (36) that

$$P_{t+1,i} = (1-m)P_{ti} + mP_t + (1-m)(\tilde{P}_{ti} - P_{ti})$$
$$+ (1-m)(P_{ti}^* - \tilde{P}_{ti}) + m(\tilde{P}_t - P_t) + m(P_{ti}' - \tilde{P}_t).$$

Hence (B.1) holds with error terms decomposable as in (B.4), with

$$\eta_{t+1,i} = (1-m)(\tilde{P}_{ti} - P_{ti}) + (1-m)(P_{ti}^* - \tilde{P}_{ti}) + m(P_{ti}' - \tilde{P}_t)$$

and $\xi_{t+1} = m(\tilde{P}_t - P_t)$. It follows from (18), (33) and (35) that

$$\frac{1}{s}\mathrm{tr}(D) \sim (1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)\cdot\frac{1}{s}\sum_{i=1}^{s}P_{ti}(1-P_{ti})$$

$$+ (1-m)^2\frac{1}{2N(1-m)}\cdot\frac{1}{s}\sum_{i=1}^{s}P_{ti}(1-P_{ti})$$

$$+ m^2\frac{1}{2N(1-m)}P_t(1-P_t)$$

$$= P_t(1-P_t)\left((1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)\right.$$

$$+ \frac{1-m}{2N}(1-F_{ST,t}) + \frac{m}{2N}\bigg)$$

$$= P_t(1-P_t)\left(\frac{1}{2\tilde{N}}-\left(\frac{1}{2\tilde{N}}-\frac{m}{2N}\right)F_{ST,t}\right).$$

Inserting this expression into (B.5), we find that

$$A = \frac{(1-m)^2 - \frac{s-1}{s}\left(\frac{1}{2\tilde{N}}-\frac{m}{2N}\right)}{1-(2N_{e,\mathrm{tot}})^{-1}}$$

and B is given by (B.9). Finally, (32) is obtained by inserting these expressions for A and B into (16). □

Appendix C. Quasi equilibrium, multiple loci and alleles

First, we motivate that the recursion formula (13) is most relevant also for studying time dynamics and quasi equilibrium properties of its multiallelic extension G_{ST} defined in (5).

Assume there are genetic markers $x = 1, \ldots, n$, with the x:th marker having alleles $a = 1, \ldots, n(x)$. Let $P_{ti}(x, a)$ and $P_t(x, a) = \sum_{i=1}^{s}P_{ti}(x, a)/s$ refer to the frequency of allele a at marker x in island i and the whole population respectively. A locus and allele specific version of the exact fixation index (1) in generation t is

$$F_{ST,t}(x, a) = \frac{\sum_{i=1}^{s}(P_{ti}(x,a) - P_t(x,a))^2}{sP_t(x,a)(1-P_t(x,a))} = \frac{Q_t(x,a)}{sP_t(x,a)(1-P_t(x,a))}.$$

It follows from Wright (1978) that the value of G_{ST} at generation t, can be written as

$$G_{ST,t} = \frac{\sum_{x,a}Q_t(x,a)}{s\sum_{x,a}P_t(x,a)(1-P_t(x,a))} = \sum_{x,a}\omega_t(x,a)F_{ST,t}(x,a), \tag{C.1}$$

with weights $\omega_t(x, a) \propto P_t(x, a)(1 - P_t(x, a))$ normalized to sum to one. See also Nei (1977) and Chakraborty and Leimar (1987) for discussions on the relation between F_{ST} and G_{ST}.

In order to study the time dynamics of $G_{ST,t+1}$ conditionally on generation t, we write

$$G_{ST,t+1} = E(G_{ST,t+1}) + \epsilon_{t+1}, \tag{C.2}$$

where the expected value is conditional on generation t, although this is not written out. Analogously to (12), ϵ_{t+1} is an error term

satisfying $E(\epsilon_{t+1}) = 0$. When the number of loci n is large, we have, to a good approximation, that

$$
\begin{aligned}
&E(G_{ST,t+1}) \\
&\sim \sum_{x,a} E(Q_{t+1}(x,a)) \Big/ \left(s \sum_{x,a} E\left(P_{t+1}(x,a)(1 - P_{t+1}(x,a))\right) \right) \\
&= \sum_{x,a} E(Q_{t+1}(x,a)) \Big/ \left(s \sum_{x,a} P_t(x,a) \right. \\
&\quad \times \left. (1 - P_t(x,a))(1 - 1/(2N_{e,\text{tot}})) \right) \\
&= \sum_{x,a} \omega_t(x,a) E(Q_{t+1}(x,a)) / \left(s P_t(x,a)\,(1 - P_t(x,a)) \right. \\
&\quad \times \left. (1 - 1/(2N_{e,\text{tot}})) \right) \\
&\approx \sum_{x,a} \omega_t(x,a)(A F_{ST,t}(x,a) + B) \\
&= A \sum_{x,a} \omega_t(x,a) F_{ST,t}(x,a) + B \sum_{x,a} \omega_t(x,a) \\
&= A G_{ST,t} + B. \tag{C.3}
\end{aligned}
$$

In the first step of (C.3) we used the Law of Large Numbers to deduce that the coefficient of variation of the numerator *and* denominator of $G_{ST,t+1}$ tend to zero as $n \to \infty$, and $\sim$ here means that the ratio of the two expressions to the right and left tend to 1 as n grows. In the second step we used that

$$
E((P_{t+1}(x,a) - P_t(x,a))^2) = \frac{P_t(x,a)(1 - P_t(x,a))}{2N_{e,\text{tot}}}, \tag{C.4}
$$

holds for all alleles a at all loci x, in analogy with (14). In the third step we used the definition of the weight $\omega_t(x,a)$, and in the fourth step we utilized that the approximation of the last line in (13) is valid for all x, a, with the same constants A and B. The reason is that these constants only depend on parameters of the reproduction cycle, not on the allele frequencies. Finally, in the last steps, we used (C.1).

Inserting (C.3) as an approximation of $E(G_{ST,t+1})$ into (C.2), we find that

$$
G_{ST,t+1} \approx A G_{ST,t} + B + \epsilon_{t+1} \tag{C.5}
$$

is approximately an autoregressive process of order 1 with identical parameters A and B as in (15). A consequence of (C.5) is that the quasi equilibrium approximation (17) is the same as in the one locus case (16). However, (17) is a *better* approximation, the larger the number of loci n is, because of the first step in (C.3).

A Taylor expansion of (C.1) yields an error term

$$
\begin{aligned}
\epsilon_{t+1} = &\frac{1}{sC_2} \sum_{x,a} (Q_{t+1}(x,a) - E(Q_{t+1}(x,a))) \\
&- \frac{C_1}{sC_2^2} \sum_{x,a} (P_{t+1}(x,a)(1 - P_{t+1}(x,a)) \\
&- E\left(P_{t+1}(x,a)(1 - P_{t+1}(x,a))\right)),
\end{aligned}
$$

with $C_1 = \sum_{x,a} E(Q_{t+1}(x,a))$ and $C_2 = \sum_{x,a} E(P_{t+1}(x,a)(1 - P_{t+1}(x,a)))$. From covariance expressions of AR(1) processes (see Brockwell and Davis (1987)), it follows that approximately

$$
\text{Cov}(G_{ST,t}, G_{ST,t+\tau}) = \frac{\text{Var}(\epsilon_t)}{1 - A^2} A^{|\tau|},
$$

provided $\text{Var}(\epsilon_t)$ varies slowly with t. As a rule of thumb, $\text{Var}(G_{ST,t})$ will be a decreasing function of s, m and n_{eff}, where n_{eff} is the effective number of loci in linkage equilibrium. $\quad\square$

Appendix D. Choice of $N_{e,\text{tot}}$

The simplest approximation $N_{e,\text{tot}} = \infty$ typically works well, but it underestimates F_{ST}^{eq} when s and N_e are both small. Often $N_{e,\text{tot}} = sN_e$ works better, but it gives too large values of F_{ST}^{eq} when close to 1. Instead, a more accurate approximation is

$$
N_{e,\text{tot}} = \begin{cases} sN_e/(1 - F_{ST}^{\text{appr}}), & \text{for FM/F and FM/FS,} \\ sN_e \Big/ \left(1 - F_{ST}^{\text{appr}} + (N_e/N)(1 - (1 - m)^2)F_{ST}^{\text{appr}} \right), \\ \quad \text{for MF/FF,} \\ sN_e/(1 - (1 - N_e m/N)F_{ST}^{\text{appr}}), & \text{for I/F.} \end{cases} \tag{D.1}
$$

The upper part of (D.1) was derived by Wright (1943a) for the island model and by Wang and Caballero (1999, Eq. (15)) for more general subdivided populations with equally large and productive subpopulations and where fertilization precedes migration. On the other hand, $N_{e,\text{tot}}$ gets lower than predicted by the upper part of (D.1) if the productivity of the subpopulations varies (Whitlock and Barton, 1997). The two lower equations in (D.1) seem new and are derived in Hössjer and Ryman (submitted for publication).

According to (D.1), F_{ST}^{appr} and $N_{e,\text{tot}}$ have to be computed jointly as functions s, m, N_e and N. This requires an iterative procedure. For instance, one may start inserting $F_{ST}^{\text{appr}} = 0$ into (D.1) in order to compute $N_{e,\text{tot}}$, which is plugged into the appropriate entry of Table 2 in order to compute F_{ST}^{appr}, which is inserted again into (D.1) etc., iterating this procedure until convergence. $\quad\square$

References

Allendorf, F.W., Phelps, S.R., 1981. Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences 38, 1507–1514.

Balloux, F., 2001. EASYPOP (version 1.7): a computer program for population genetics simulations. Journal of Heredity 92, 301–302.

Beerli, P., 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Molecular Ecology 13, 827–836.

Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer-Verlag, New York.

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

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, NJ, p. 420.

Chakraborty, R., Nei, M., 1982. Genetic differentiation of quantitative characters between populations of species, I. Mutation and random genetic drift. Genetical Research Cambridge 39, 303–314.

Charlesworth, B., 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15 (5), 538–543.

Chesser, R.K., Rhodes, O.E., Sugg, D.W., Schabel, A., 1993. Effective sizes for subdivided populations. Genetics 135, 1221–1232.

Cockerham, C.C., 1969. Variance of gene frequencies. Evolution 23, 72–84.

Cockerham, C.C., 1973. Analyses of gene frequencies. Genetics 74, 679–700.

Cockerham, C.C., Weir, B.S., 1987. Correlations, descent measures: drift with migration and mutation. Proceedings of the National Academy of Sciences of the United States of America 84, 8512–8514.

Crow, J.F., Aoki, K., 1984. Group selection for polygenic behavioral trait: estimating the degree of population subdivision. Proceedings of the National Academy of Sciences of the United States of America 81, 6073–6077.

Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. The Blackburn Press, Caldwell, New Jersey, USA.

Hauser, L., Carvalho, G.R., 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9, 333–362.

Hössjer, O., 2012. Spatial autocorrelation for subdivided populations with invariant migration schemes, Research report 2012:11, Department of Mathematics, Stockholm University (submitted for publication).

Hössjer, O., Ryman, N., 2012. Quasi equilibrium, variance effective population size and fixation index for models with spatial structure (submitted for publication).

Latter, B.D.H., 1973. The island model of population differentiation: a general solution. Genetics 73, 147–157.

Latter, B.D.H., Sved, J.A., 1981. Migration and mutation in stochastic models of gene frequency change, II. Stochastic migration with a finite number of islands. Journal of Mathematical Biology 13, 95–104.

Li, W.-H., 1976. Effect of migration on genetic distance. The American Naturalist 110, 841–847.

Malécot, G., 1951. Un traitement stochastique des problémes linéaires (mutation, linkage, migration) en génétique de population. Ann. Univ. Lyon., Sci A 14, 79–117.

Maruyama, T., 1970. Effective number of alleles in subdivided populations. Theoretical Population Biology 1, 273–306.

Nagylaki, T., 1980. The strong migration limit in geographically structured populations. Journal of Mathematical Biology 9, 101–114.

Nagylaki, T., 1983. The robustness of neutral models of geographical variation. Theoretical Population Biology 23, 268–294.

Nagylaki, T., 1995. The inbreeding effective population number in dioecious populations. Genetics 139, 473–485.

Nagylaki, T., 1998. Fixation indices in subdivided populations. Genetics 148, 1325–1332.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70, 3321–3323.

Nei, M., 1975. Molecular Population Genetics and Evolution. North-Holland, New York.

Nei, M., 1977. *F*-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics 41, 225–231.

Nei, M., Chakravarti, A., 1977. Drift variances of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. Theoretical Population Biology 11, 307–325.

Nei, M., Chakravarti, A., Tateng, Y., 1977. Mean and variance of F_{ST} in a finite number of incompletely isolated populations. Theoretical Population Biology 11, 291–306.

Nei, M., Chesser, R.K., 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47, 253–259.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford.

Nei, M., Tajima, F., 1983. Genetic drift and estimation of effective population size. Genetics 98, 625–640.

Olsson, F., Hössjer, O., Laikre, L., Ryman, N., 2012. Variance effective size of populations in which size and age compositon fluctuate (submitted for publication).

Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142, 1357–1362.

Ryman, N., Allendorf, F.W., Jorde, P.E., Laikre, L., Hössjer, O., 2012. Samples from subdivided populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation (submitted for publication).

Ryman, N., Leimar, O., 2008. Effect of mutation on genetic differentiation among nonequilibrium populations. Evolution 62 (9), 2250–2259.

Sampson, K., 2006. Structured coalescent with nonconservative migration. Journal of Applied Probability 43, 351–362.

Sawyer, S., 1976. Results for the stepping-stone model for migration in population genetics. Annals of Probability 4, 699–728.

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462.

Slatkin, M., 2005. Seeing ghosts: the effect of unsampled populations on migration rate estimated for sampled populations. Molecular Ecology 14, 67–73.

Slatkin, M., Voelm, L., 1991. F_{ST} in a hierarchical model. Genetics 127, 627–629.

Sved, J.A., Latter, B.D.H., 1977. Migration and mutation in stochastic models of gene frequency change. Journal of Mathematical Biology 5, 61–73.

Takahata, N., 1983. Gene identity and genetic differentiation of populations in the finite island model. Genetics 104, 497–512.

Takahata, N., Nei, M., 1984. F_{ST} and G_{ST} statistics in the finite island model. Genetics 107, 501–504.

Wang, J., 1997a. Effective size and *F*-statistics of subdivided populations, I. Monoecious species with partial selfing. Genetics 146, 1453–1463.

Wang, J., 1997b. Effective size and *F*-statistics of subdivided populations, II. Dioecious species. Genetics 146, 1465–1474.

Wang, J., Caballero, A., 1999. Developments in predicting the effective size of subdivided populations. Heredity 82, 212–226.

Waples, R.S., 1989a. A generalized approach for estimating effective population size from temporal changes of allele frequency. Genetics 121, 379–391.

Waples, R.S., 1989b. Temporal variation in allele frequencies: testing the right hypothesis. Evolution 43 (6), 1236–1251.

Waples, R.S., 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Heredity 89 (5), 438–450.

Waples, R.S., 2002. Definition and estimation of effective population size in the conservation of endangered species. In: Beissinger, S.R., McCullogh, D.R. (Eds.), Population Viability Analysis. pp. 147–168.

Waples, R.S., Do, C., 1994. Genetic risk associated with supplementation of pacific salmonids: captive broodstock programs. Canadian Journal of Fisheries and Aquatic Sciences 51 (Suppl. 1), 310–329.

Waples, R.S., Faulkner, R., 2009. Modelling evolutionary processes in small populatins: not as ideal as you think. Molecular Ecology 18, 1834–1847.

Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38 (6), 1358.

Whitlock, M.C., Barton, N.H., 1997. The effective size of a subdivided population. Genetics 146, 427–441.

Wright, S., 1921. Systems of mating, I–V. Genetics 6, 111–178.

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

Wright, S., 1943a. Isolation by distance. Genetics 28, 114–138.

Wright, S., 1943b. An analysis of local variability of flower color in Linanthus Parraye. Genetics 28, 139–156.

Wright, S., 1951. The genetical structure of populations. Annals of Eugenics 15, 323–354.

Wright, S., 1978. Variability Within and Among Genetic Populations. In: Evolution and the Genetics of Populations, vol. 4. University of Chicago Press, Chicago, London.