Mathematical Statistics
Stockholm University

Quasi Equilibrium Approximations of the Fixation Index under Neutrality: The Finite and Infinite Island Models

Ola Hössjer
Per Erik Jorde
Nils Ryman

Research Report 2012:5

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2012:5**,
http://www.math.su.se/matstat

Quasi Equilibrium Approximations of the Fixation Index under Neutrality: The Finite and Infinite Island Models

Ola Hössjer[*]
Per Erik Jorde[†]
Nils Ryman[‡]

April 30, 2012

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 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 F_{ST} to be underestimated for large migration rates.

Key words: Effective population size, fixation index, fertilization, gamete formation, island model, migration, quasi equilibrium, reproduction.

---

[*] Department of Mathematics, Div. of Mathematical Statistics, Stockholm University.

[†] Dept. of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, P.O. Box 1066 Blindern, N-0371 Oslo, Norway. Institute of Marine Research, Flødevigen Research Station, N-4817 His, Norway.

[‡] Department of Zoology, Div. of Population Genetics, Stockholm University.

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 another 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., in preparation).

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. We thus assume that the diploid population evolves in non-overlapping generations, with the total population consisting of s islands of equal census (N) and effective (N_e) size. We let p_{it} 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

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

In this paper we work with the parameter m and refer to it as the migration rate.

The fixation index of generation t is defined as

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

where $P_t = \sum_{i=1}^{s} p_{it}/s$ is the generation t frequency of the allele in the whole population, see for instance equation (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.

An adjusted version

$$F_{ST,t}^{\mathrm{adj}} = \frac{\sum_{i=1}^{s}(p_{it} - P_t)^2}{(s-1)P_t(1-P_t)} \tag{3}$$

of $F_{ST,t}$ has $s-1$ rather than s in the denominator compared to (1). We will find below that in contrast to $F_{ST,t}$ it has an equilibrium value virtually independent of s. A closely related version of the fixation index due to Weir and Cockerham (1984) is defined as the correlation between two alleles drawn from the same island, see also Cockerham (1969, 1973). If equally large samples are taken from a subset $I \subset \{1,\ldots,s\}$ of $k < s$ islands, and the estimation error due to sampling can be ignored, the estimator of the fixation index due to Weir and Cockerham (1984) reduces to

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

where $\hat{P}_t = \sum_{i\in I} p_{it}/k$ and $S^2 = \sum_{i\in I}(p_{it} - \hat{P}_t)^2/(k-1)$ are the average and sample variance of the allele frequencies in the islands from which samples are taken. In Appendix A, we verify 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. Since the total number of subpopulations is often unknown (see for instance Wright, 1943b), it is only the adjusted fixation index that can be estimated with negligible bias. 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)

$$G_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T} \tag{4}$$

is a multiallelic 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, eqn. (12.23)). However, it lacks the intuitive excess of heterozygosity interpretation.

Therefore, the unadjusted and adjusted fixation indeces 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 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 equilibrium value

$$F_{ST}^{\text{eq}} = \frac{(1-m)^2}{2N(1-(1-m)^2) + (1-m)^2} \tag{5}$$

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 of 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_{it}, 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 (5) is modified to

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

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

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

(derived from equation (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 (5)-(7) all reduce to the well known approximation

$$F_{ST}^{\text{eq}} \approx \frac{1}{4Nm+1} \tag{8}$$

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

Finding 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 large. 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

$$F_{ST}^{\mathrm{eq}} = \frac{(1-m)^2(1-u)^2}{\frac{s}{s-1}2N(1-(1-m)^2(1-u)^2)+(1-m)^2(1-u)^2} \tag{9}$$

of the fixation index (and of G_{ST} as well) can be derived, as discussed for instance by Nei (1975), Takahata (1983), Takahata and Nei (1984) and Ryman and Leimar (2008). By taking the $u \to 0$ limit, it (9) simplifies to

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

which can be viewed as a generalization of (5) to the finite island model. Slatkin and Voelm (1991) have shown that (10) can be expressed in terms of mean coalescene times, and for small migration rates, it reduces to

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

which is an extension of (8) 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 that differs from (9)-(10) in three ways. First, we work with the neutral model $u = 0$ directly, thereby avoiding to introduce mutations and then taking the $u \to 0$ limit. Secondly, it turns out that the recursions for identical by state probabilities ignore part of the random variability of allele frequencies between islands, leading to F_{ST}^{eq} values with downward bias, as manifested by the extra $(1-m)^2$ term in the numerator of (10). 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, formula (10) suggests that all subpopulations have identical allele frequencies ($F_{ST}^{\mathrm{eq}} = 0$) when $m = 1$, which is clearly not reasonable. 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.

Thirdly, before one allele gets fixed, we demonstrate that $F_{ST,t}$ does not converge to a fixed value but exhibits random quasi equilibrium fluctuations. Indeed, the change of the fixation index from one generation to the next can be written as a sum

$$F_{ST,t+1} = E(F_{ST,t+1}) + \epsilon_{t+1}, \tag{11}$$

of a deterministic and random component, where expectation is conditional on generation t and ϵ_{t+1} is an error term satisfying $E(\epsilon_{t+1}) = 0$. The quasi equilibrium value is then obtained by putting

$$F_{ST,t} = E(F_{ST,t+1}) = F_{ST}^{\mathrm{eq}} \tag{12}$$

and solving the resulting equation with respect to F_{ST}^{eq}. In this way, F_{ST}^{eq} becomes the mean value of the (typically small) fluctuations that the fixation index exhibits under quasi equilibrium, i.e. *conditionally* on the event that no allele has been fixed in all islands.

While it is well known that $F_{ST,t}$ exhibits random variation for the finite island model, see for instance Nagylaki (1998), to the best of our knowledge, the quasistationary oscillations before fixation have not been phrased in mathematical terms before. As mentioned above, our method differs from (9)-(10) and is rather a generalization of the infinite island model approach of Sved and Latter (1977) to the finite island model.

We will utilize the approximation

$$\begin{aligned} E(F_{ST,t+1}) &\approx E\left(\sum_{i=1}^{s}(p_{i,t+1} - P_{t+1})^2\right) / \left(sE\left(P_{t+1}(1-P_{t+1})\right)\right) \\ &= E\left(\sum_{i=1}^{s}(p_{i,t+1} - P_{t+1})^2\right) / \left(sP_t(1-P_t)(1-(2N_{e,\mathrm{tot}})^{-1}\right) \\ &=: F_{ST,t+1}^{\mathrm{appr}}, \end{aligned} \tag{13}$$

of the expected fixation index of generation $t+1$, with

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

the (variance) effective size of the whole population, a quantity reveiwed by Caballero (1994), Wang and Caballero (1999) and Waples (2002). Often, the term $(2N_{e,\mathrm{tot}})^{-1}$ in (13) is so small that it can be dropped. In Appendix B, we demonstrate that $F_{ST,t+1}^{\mathrm{appr}} = AF_{ST,t} + B$ for all versions of the island model considered in this paper, with the A and B constants depending on the genetic model but not on $F_{ST,t}$. Insertion into (11) reveals that

$$F_{ST,t+1} = AF_{ST,t} + B + \epsilon_{t+1}, \tag{15}$$

becomes an autoregressive process of order 1 (Brockwell and Davis, 1987), whose mean value $F_{ST}^{\mathrm{eq}} = B/(1-A)$ is obtained from (12). We also motivate in Appendix C that (15) also holds for G_{ST}, with the same constants A and B. In particular, the quasi equilibrium value $B/(1-A)$ of G_{ST} is the same as for F_{ST}.

As mentioned in the introduction, the quasi equilibrium approximation of the adjusted fixation index is obtained by multiplying the quasi equilibrium approximation of the unadjusted fixation index by $(s-1)/s$.

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 $\tilde{p}_{it}$. To account for varying reproductive rate among the individuals of island i, $\tilde{p}_{it}$ may differ from p_{it} in a way quantified as

$$\frac{E\left(\tilde{p}_{it} - p_{it}\right)^2}{p_{it}(1-p_{it})} \sim \frac{1}{2N_e} - \frac{1}{2N}, \tag{16}$$

where $x \sim y$ means that x/y tends to one as the population size N gets large.

Formula (16) 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 $(2N_e)^{-1} - (2N)^{-1}$). 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 ($\tilde{p}_{it} = p_{it}$), 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.

TABLE 1 ABOUT HERE

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 (16), 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 during gamete formation (for a diploid organism this corresponds to N_e breeders). The quantity $\tilde{p}_{it}$ 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

$$\tilde{p}_{it} \sim \mathrm{Hyp}(2N, 2N_e, p_{it})/(2N_e)$$

of the gamete pool is obtained. By second moment properties of the hypergeometric distribution, it follows that (16) 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

$$\tilde{p}_{it} = \sum_{j;j \text{ has allele } 1} w_j.$$

Suppose $w \sim \mathrm{Dir}(\alpha/(2N), \ldots, \alpha/(2N))$ 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

$$\tilde{p}_{it} \sim \mathrm{Beta}\left(\alpha(1 - p_{it}), \alpha p_{it}\right).$$

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

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

as in order to satisfy (16). The degenerate case $N = N_e$ corresponds to $\alpha = \infty$, $w_j \equiv 1/(2N)$ and $\tilde{p}_{it} = p_{it}$.

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 $N = N_m + N_f$. Gamete pool i of generation t is constructed by drawing randomly *without replacement* $2N_m$ male genes out of all $2N$. 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. If $X_{it} \sim \mathrm{Hyp}(2N, 2N_m, p_{it})$ refers to the number of male genes that have the specified allele, we find that

$$\tilde{p}_{it} = \frac{1}{2}\tilde{p}_{itm} + \frac{1}{2}\tilde{p}_{itf},$$

where $\tilde{p}_{itm} = X_{it}/(2N_m)$ and $\tilde{p}_{itf} = (2Np_{it} - X_{it})/(2N_f)$ refer to the allele frequency of the male and female parts of gamete pool i and generation t. Using moment properties of the hypergeometric distribution, it can be shown that (16) holds with

$$N_e = \frac{4N_mN_f}{N},$$

in agreement with, for instance, Section 3.11 of Crow of 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 Figure 1, and described in more detail as follows:

FIGURE 1 ABOUT HERE

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^*_{it}, with

$$\frac{E\left((p^*_{it} - \tilde{p}_{it})^2 | \tilde{p}_{it}\right)}{\tilde{p}_{it}(1-\tilde{p}_{it})} \sim \frac{1}{2N}. \tag{17}$$

For the monoecious models, we achieve (17) by means of binomial sampling,

$$p^*_{it} = \frac{\mathrm{Bin}(2N, \tilde{p}_{it})}{2N}, \tag{18}$$

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

$$p^*_{it} = \frac{\mathrm{Bin}(N, \tilde{p}_{itm}) + \mathrm{Bin}(N, \tilde{p}_{itf})}{2N}. \tag{19}$$

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^*_{ijt} 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

$$p^*_{it} = \sum_{j=1}^{s} m_{ij} p^*_{ijt}.$$

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

$$p_{i,t+1} = \sum_{j=1}^{s} m_{ji} p^*_{jit}. \tag{20}$$

The mixture (20) 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

$$m_{ij} = \begin{cases} 1 - m + m/s, & i = j, \\ m/s, & i \neq j, \end{cases} \tag{21}$$

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 do *not* imply a fixed number of immigrants, showing that in general the two concepts are different.

We further assume *fixed migrant allele frequencies*

$$p_{ijt}^* = p_{it}^* \text{ for all } i, j, \tag{22}$$

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 (21)-(22) lead to

$$F_{ST}^{\text{eq}} = \frac{(1-m)^2}{\frac{s}{s-1}2N_e(1-(1-m)^2 - \frac{1}{2N_{e,\text{tot}}}) + (1-m)^2}, \tag{23}$$

a generalization of (10) to $N_e \neq N$, or, of (5) to $s < \infty$ and $N_e \neq N$.

4.2.2 Fixed migrant proportions and stochastic migrant allele frequencies.

We retain the fixed migrant proportions (21) but assume that the $2N$ genes of the pre-migration population j is 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_{jit}^*. In particular,

$$\text{Var}(p_{jit}^* | \tilde{p}_{jt}) = \frac{\tilde{p}_{jt}(1-\tilde{p}_{jt})}{2Nm_{ji}} \sim \frac{p_{jt}(1-p_{jt})}{2Nm_{ji}}, \tag{24}$$

which is achieved for the monoecious and dioecious models analogously to (18) and (19). It is shown in Appendix B that

$$F_{ST}^{\text{eq}} = \frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2 - \frac{1}{2N_{e,\text{tot}}}) + 1}, \tag{25}$$

where

$$\frac{1}{\tilde{N}} = \frac{(1-m)^2}{N_e} + \frac{2m - m^2}{N} \tag{26}$$

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 $N_e = N$, (25) reduces to none of (5)-(7), although it is much closer to (6)-(7), both of which assume stochastic immigrant allele frequencies, than to (5), 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. Figure 2 illustrates one generation cycle of this model, a detailed description of which is as follows:

FIGURE 2 ABOUT HERE

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

$$\check{p}_{it} = \sum_{j=1}^{s} m_{ji} \tilde{p}_{jt}. \tag{27}$$

In (27) we assume that the subpopulation of gamete pool j that migrates to island i consists of so many gametes that its allele frequency $\tilde{p}_{jt}$ 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 (27) by assuming equal migration proportions for the male and female alleles, so that $\check{p}_{it} = (\check{p}_{itm} + \check{p}_{itf})/2$, where $\check{p}_{itm} = \sum_{j=1}^{s} m_{ji} \tilde{p}_{jtm}$ and $\check{p}_{itf} = \sum_{j=1}^{s} m_{ji} \tilde{p}_{jtf}$.

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 $p_{i,t+1}$ satisfies

$$\frac{E\left((p_{i,t+1} - \check{p}_{it})^2 | \check{p}_{it}\right)}{\check{p}_{it}(1 - \check{p}_{it})} \sim \frac{1}{2N}. \tag{28}$$

For the monoecious models, (28) is achieved by means of binomial sampling, and for the dioecious model through

$$p_{i,t+1} = \frac{\text{Bin}(N, \check{p}_{itm}) + \text{Bin}(N, \check{p}_{itf})}{2N}. \tag{29}$$

4.3.1 Fixed migrant proportions.

Assume that the migrant proportions (21) 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

$$F_{ST}^{\text{eq}} = \frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2 - \frac{1}{2N_{e,\text{tot}}}) + \frac{\tilde{N}}{N_e}(1-m)^2}, \tag{30}$$

thus generalizing (7) 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 $\tilde{p}_{it}$. Denote the allele frequency of these genes by p_{it}^*, where

$$\frac{E\left((p_{it}^* - \tilde{p}_{it})^2|\tilde{p}_{it}\right)}{\tilde{p}_{it}(1-\tilde{p}_{it})} \sim \frac{1}{2N(1-m)}. \tag{31}$$

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

$$\tilde{p}_t = \sum_{i=1}^{s} w_i \tilde{p}_{it}. \tag{32}$$

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 h_{it}', where

$$\frac{E\left((h_{it}' - \tilde{p}_t)^2|\tilde{p}_t\right)}{\tilde{p}_t(1-\tilde{p}_t)} \sim \frac{1}{2Nm}. \tag{33}$$

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

$$p_{i,t+1} = (1-m)p_{it}^* + mh_{it}' \tag{34}$$

of island i and generation $t+1$.

Notice that (31) and (33) 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 (19) and (29).

It is shown in Appendix B that when all islands contribute equally to the combined gamete pool (so called *fixed migrant proportions*, $w_i = 1/s$), (6) generalizes to

$$F_{ST}^{\mathrm{eq}} = \frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2 - \frac{1}{2N_{e,\mathrm{tot}}}) + 1 - \frac{\tilde{N}}{N}m}. \tag{35}$$

This is the natural counterpart of a fixed number of immigrants with stochastic immigrant allele frequency (cf. (6)) 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

TABLE 2 ABOUT HERE.

Expressions for F_{ST}^{eq}; 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 notation 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 we included 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_{ST}^{eq} numerically as function of m (Table 3), N (Table 4) or s (Table 5). All F_{ST}^{eq} formulas for the finite island model need $N_{e,\mathrm{tot}}$ in (14) to be specified. Often this term can be dropped with good accuracy, but more refined choices of $N_{e,\mathrm{tot}}$ are described in the appendix, with (D.1) as the default choice in all tables.

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 $p_{i0} = 0.5$, $i = 1, \ldots, s$, 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, the remaining, non-migrants, were similarly drawn by 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 (32) 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.

TABLES 3-5 ABOUT HERE.

The three stochastic immigrant allele frequency models FM/FS, MF/FF, and I/F have almost identical values of F_{ST}^{eq}, 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 part of Table 4. When the immigrant allele frequency is stochastic and the migration rate high, it turns out that F_{ST}^{eq} 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}^{eq}, 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.

Figure 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 $F_{ST,t}^{\mathrm{adj}}$ is plotted (solid lines) as function of t, together with the average $F_{ST,t}$ (dashed line), estimated from 10 000 simulations, and the quasi equilibrium value $F_{ST}^{\mathrm{eq,adj}}$ (horisontal dotted line). It is seen that $F_{ST,t}^{\mathrm{adj}}$ 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.

FIGURE 3 ABOUT HERE.

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.

TABLE 6 ABOUT HERE.

6 Discussion

6.1 Summary and extensions

In this paper, we have introduced a novel quasi equilibrium approach for computing the fixation index F_{ST}^{eq} of spatial mutation free models that exhibit migration between subpopulations. This approach is suitable for the relatively short time spans encountered e.g. in conservation genetics and requires that the mutation rate is of smaller order than the migration 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}^{eq} 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}^{eq} 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}^{eq}. This is caused by the rather unrealistic fixed allele frequency of immigrants, as manifested by an extra term $(1-m)^2$ in the numerator of the defining

equation of F_{ST}^{eq}. Similar values of F_{ST}^{eq} as for FM/FF are obtained with the traditional identical-by-state probabilities approach. The difference in F_{ST}^{eq} 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}^{eq} depends not only on N_e, but also on N, for all parameter combinations. However, the dependence of F_{ST}^{eq} on N is only substantial 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}^{eq} is concerned. The same can be said for the total variance effective population size $N_{e,\text{tot}}$.

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 $N_{e,\text{tot}}$ by the temporal method when subpopulation structure is ignored (Ryman et al., in preparation). Second, one may introduce an extra source of variability by allowing for randomly varying migrant proportions, thereby increasing F_{ST}^{eq}. Third, 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. We plan to do the same in a forthcoming paper, using instead the quasi equilibrium approach.

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, but also on N, whenever $m > 0$, for all reproduction models except FM/FF, i.e. the original island model of Wright. Intuitively, we may explain this as follows: Recall the discussion below (16) that the genetic drift of an isolated island ($m = 0$) can be decomposed into two variance components due to gamete formation (size proportional to $(2N_e)^{-1} - (2N)^{-1}$) and fertilization (size proportional to $(2N)^{-1}$), although the total size of the genetic drift (proportional to $1/(2N_e)$) is unaffected by this division. On the other hand, when migration is included ($m > 0$), this division into gamete formation and fertilization affects the total genetic drift and hence also F_{ST}^{eq}.

For instance, when migration precedes reproduction it is the gametes that migrate. It is then intuitively reasonable that the part of the genetic drift due to gamete formation will affect the equilibrium F_{ST}. When fertilization precedes migration, this is intuitively less clear though, but nevertheless true if migrant allele frequencies are stochastic, so that the genes of the migrating individuals are drawn independently from the gamete pool of their respective source islands. On the other hand, if the migrating individuals have the same allele frequencies, division of the total genetic drift into gamete formation and fertilization becomes unimportant, so that the equilibrium F_{ST} only depends on N_e.

On a more mathematical level, we may explain why $\tilde{N}$ rather than N_e appears in the formulas for F_{ST}^{eq} for all reproduction scenarios except FM/F. It turns out that the deterministic part of the recursion (15) has the form

$$F_{ST,t+1}^{\mathrm{appr}} \approx (1-m)^2 F_{ST,t} + B \tag{36}$$

when $N_{e,\mathrm{tot}}^{-1}$ is negligible, as deduced from (B.5) in Appendix B. The right hand side of (36) 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.

In general, the genetic drift term B will to some extent depend on $F_{ST,t}$. However, it follows from the calculations in Appendix B that

$$B \approx \frac{s-1}{s}\left\{(1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)+1\cdot\frac{1}{2N}\right\} = \frac{s-1}{s}\frac{1}{2\tilde{N}} \tag{37}$$

is sufficiently accurate, for reproduction scenarios FM/FS, MF/FF and I/F, in order to explain the main features of the recursion (36). 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,

$$\begin{array}{rcl} B & \approx & \frac{s-1}{s}\left\{(1-m)^2\left(\frac{1}{2N_e}-\frac{1}{2N}\right)+(1-m)^2\cdot\frac{1}{2N}\right\} \\ & = & \frac{s-1}{s}\frac{(1-m)^2}{2N_e}. \end{array} \tag{38}$$

The only difference compared to (37) 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 $F_{ST,t}$.

By comparing the right hand sides of (37) and (38) and solving (36) and (12), we find that the overall effect on F_{ST}^{eq} of having a fixed immigrant allele frequency is that N_e replaces $\tilde{N}$ in the denominator of F_{ST}^{eq} and a multiplicative term $(1-m)^2$ appears in the numerator of F_{ST}^{eq}. Hence, the appearance of N_e in the well-known expression (8) 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 (1989), where similar issues are discussed when temporal rather than spatial variation of allele frequencies are tested.

First, formula (10) implies $F_{ST}^{\text{eq}}=0$ 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 reveal that

$$F_{ST}^{\text{eq}} \overset{m=1}{\approx} \frac{s-1}{s}\frac{1}{2N} \Longrightarrow F_{ST}^{\text{eq,adj}} \overset{m=1}{\approx} \frac{1}{2N} \tag{39}$$

regardless of the value of N_e. It may be the case, for a well powered test based on a large sample sizes, that the overly optimistic assumption $F_{ST}^{\text{eq}}=0$ 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. On the other hand, Knudsen et al. (2011) report $\hat{F}_{ST} = 0.0037$, with a 95% CI of $(0.0017, 0.0060)$, between two Atlantic cod populations. In this case the estimated census size of one of the populations is $\hat{N} = 1391$ using capture-recapture methods, and $1/(2\hat{N}) = 0.00036$ is too small to affect the outcome of the test.

Second, formulas (6)-(7), which can be inferred from Sved and Latter (1977), imply $F_{ST}^{\text{eq}} = 1/(2N)$ under panmixia when $s = \infty$ and $N_e = N$. From this it would tempting to assume $F_{ST}^{\text{eq,adj}} = 1/(2N_e)$. However, (39) reveals that this assumption would lead to *too few* rejections of panmixia, and hence too few findings of population substructure. For instance, Palm et al. (2003) performed an empirical genetic monitoring study on brown trout (Salmo trutta) in central Sweden along a 3 km long stream. Samples of approximately 100 fish were collected annually for 20 years from a well defined part of the stream, giving estimates of quite a large genetic drift for the whole population and hence a small N_e. Although the statistical power for detecting population subdivision may be poor (Ryman et al., 2006, Waples and Gaggiotti, 2006), they obtained strikingly uniform allele frequencies, indicating a very small fixation index, when sampling over the entire 3 km section of the stream (unpublished). This seems to indicate that panmixia cannot be rejected. However, in view of (39), we cannot exclude panmixia before N has been estimated, since the census size may very well be much larger than N_e. For marine species it is indeed believed that the total census size is often several orders of magnitude larger than the total effectie size $N_{e,\text{tot}}$ (Hauser and Carvalho, 2008). However, it is not yet settled how much of this discrepancy is due to artifacts of the estimation procedure (Ryman et al., 2012). Therefore, more research is needed.

Acknowledgements

Ola Hössjer's research was supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine, Per Erik Jorde's research by the Research Council of Norway, and Nils Ryman's research by support from the Swedish Research Council, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), and from the BONUS Baltic Organisations' Network for Funding Science EEIG (the BaltGene research project).

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

We let $E(\cdot)$ denote expectation when I is a randomly selected subset of $\{1,\ldots,s\}$ and use the approximation

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

Writing $\delta_{it} = p_{it} - P_t$, we find, after some computations, that

$$\begin{array}{rcl} S^2 & = & \frac{1}{k-1}\sum_{i\in I}\left(\delta_{it} - \frac{1}{k}\sum_{j\in I}\delta_{jt}\right)^2 \\ & = & \frac{1}{k}\sum_{i\in I}\delta_{it}^2 - \frac{2}{k(k-1)}\sum_{(i,j)\in I; i<j}\delta_{it}\delta_{jt}. \end{array}$$

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

$$\begin{array}{rcl} E(S^2) & = & E(\delta_{I_1t}^2) - E(\delta_{I_1t}\delta_{I_2t}) \\ & = & \frac{s}{s-1}E(\delta_{I_1t}^2) \\ & = & \frac{s}{s-1}P_t(1-P_t)F_{ST,t} \\ & = & P_t(1-P_t)F_{ST,t}^{\mathrm{adj}} \end{array} \tag{A.2}$$

and

$$\begin{array}{rcl} E(\hat{P}_t(1-\hat{P}_t)) & = & P_t(1-P_t) - E\left(\frac{1}{k}\sum_{i\in I}\delta_{it}\right)^2 \\ & = & P_t(1-P_t) - \frac{1}{k}E(\delta_{I_1t}^2) - \frac{k-1}{k}E(\delta_{I_1t}\delta_{I_2t}) \\ & = & P_t(1-P_t) + \frac{(k-1)/(s-1)-1}{k}E(\delta_{I_1t}^2) \\ & = & P_t(1-P_t)\left(1 - \frac{s-k}{k(s-1)}F_{ST,t}\right). \end{array} \tag{A.3}$$

Inserting (A.2) and (A.3) into (A.1), we arrive at

$$E(\hat{F}_{ST,t}) \approx \frac{F_{ST,t}^{\mathrm{adj}}}{1 + \frac{F_{ST,t}}{s-1}}. \tag{A.4}$$

The right hand side of (A.4) essentially equals $F_{ST,t}^{\mathrm{adj}}$ when either $F_{ST,t}$ is small or s is large. ☐

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

In order to compute F_{ST}^{eq}, we need a general formula for $F_{ST,t+1}^{\mathrm{appr}}$ to insert into (12). To this end, we consider models for which the allele frequency of island i and generation $t+1$ can be expressed recursively as

$$p_{i,t+1} = (1-m)p_{it} + mP_t + \varepsilon_{it}, \tag{B.1}$$

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 ε_{it}. Assuming that $E(\varepsilon_{it})=0$ and putting $\mathrm{Cov}(\varepsilon_{it},\varepsilon_{jt})=\sigma_{ij}$, we can rewrite the recursion (13) as

$$\left(1-(2N_{e,\mathrm{tot}})^{-1}\right)F_{ST,t+1}^{\mathrm{appr}}=(1-m)^2F_{ST,t}+\frac{\frac{s-1}{s^2}\mathrm{tr}(\Sigma)-\frac{1}{s^2}\sum_{i\neq j}\sigma_{ij}}{P_t(1-P_t)}, \tag{B.2}$$

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

$$p_{i,t+1}-P_{t+1}=(1-m)(p_{it}-P_t)+\varepsilon_{it}-\varepsilon_t, \tag{B.3}$$

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

$$\begin{array}{rcl}\left(1-(2N_{e,\mathrm{tot}})^{-1}\right)F_{ST,t+1}^{\mathrm{appr}} & = & \sum_{i=1}^s E\left((1-m)(p_{it}-P_t)+\varepsilon_{it}-\varepsilon_t\right)^2/(sP_t(1-P_t))\\ & = & (1-m)^2F_{ST,t}+\sum_{i=1}^s E\left(\varepsilon_{it}-\varepsilon_t\right)^2/(sP_t(1-P_t)),\end{array}$$

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

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

$$\varepsilon_{it}=\eta_{it}+\xi_t, \tag{B.4}$$

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

$$\left(1-(2N_{e,\mathrm{tot}})^{-1}\right)F_{ST,t+1}^{\mathrm{appr}}=(1-m)^2F_{ST,t}+\frac{s-1}{s}\cdot\frac{\frac{1}{s}\mathrm{tr}(D)}{P_t(1-P_t)}. \tag{B.5}$$

Indeed, (B.4) implies that

$$p_{i,t+1}-P_{t+1}=(1-m)(p_{it}-P_t)+\eta_{it}-\eta_t, \tag{B.6}$$

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

It turns out that (B.4)-(B.5) is applicable for all breeding, fertilization and migration scenarios of the island model considered in this paper. □

Fixed migrant proportions and allele frequencies imply fixed immigrant allele frequencies. From (20), (21), and (22) it follows that the allele frequency of island i after a completed reproduction cycle $t\to t+1$ is

$$\begin{array}{rcl}p_{i,t+1} & = & (1-m)p_{iit}^*+\frac{m}{s}\sum_{j=1}^s p_{jit}^*\\ & = & (1-m)p_{it}^*+\frac{m}{s}\sum_{j=1}^s p_{jt}^*\\ & \to & (1-m)p_{it}^*+mP_t \text{ as } s\to\infty.\end{array} \tag{B.7}$$

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 (23). In this case (B.4) holds, with $\eta_{it} = (1-m)(p_{it}^* - p_{it})$ and $\xi_t = m\sum_i (p_{it}^* - P_t)/s$. Hence it follows from (16) and (17) that

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

and

$$\begin{aligned} \tfrac{1}{s}\mathrm{tr}(D) &= \tfrac{(1-m)^2}{2N_e}\tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) \\ &= \tfrac{(1-m)^2}{2N_e}P_t(1-P_t)(1-F_{ST,t}). \end{aligned}$$

Inserting this expression into (B.5), putting $F_{ST,t+1} = F_{ST,t} = F_{ST}^{\mathrm{eq}}$ and solving for F_{ST}^{eq} we arrive at (23). $\square$

Proof of (25). Because of (21), we have that

$$\begin{aligned} p_{i,t+1} &= \textstyle\sum_{j=1}^s m_{ji}p_{ijt}^* \\ &= (1-m)p_{it} + mP_t + \textstyle\sum_{j=1}^s m_{ji}(\tilde{p}_{jt} - p_{jt}) + \sum_{j=1}^s m_{ji}(p_{jit}^* - \tilde{p}_{jt}) \\ &= (1-m)p_{it} + mP_t + \tfrac{m}{s}\textstyle\sum_{j=1}^s(\tilde{p}_{jt} - p_{jt}) + (1-m)(\tilde{p}_{it} - p_{it}) \\ &+ \textstyle\sum_{j=1}^s m_{ji}(p_{jit}^* - \tilde{p}_{jt}), \end{aligned}$$

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

$$\eta_{it} = (1-m)(\tilde{p}_{it} - p_{it}) + \sum_{j=1}^s m_{ji}(p_{jit}^* - \tilde{p}_{jt})$$

and $\xi_t = m\sum_{j=1}^s(\tilde{p}_{jt} - p_{jt})/s$. Hence

$$\begin{aligned} \tfrac{1}{s}\mathrm{tr}(D) &\sim (1-m)^2\left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right)\tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) + \tfrac{1}{s}\sum_{i,j=1}^s m_{ji}^2\tfrac{p_{jt}(1-p_{jt})}{2Nm_{ji}} \\ &= \left((1-m)^2\left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) + \tfrac{1}{2N}\right)\tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) \\ &= \tfrac{1}{2\tilde{N}}P_t(1-P_t)(1-F_{ST,t}). \end{aligned}$$

Inserting this expression into (B.5) and solving for F_{ST}^{eq} we arrive at (25). $\square$

Proof of (30). We have that

$$\begin{aligned} p_{i,t+1} &= (1-m)\tilde{p}_{it} + m\tilde{p}_t + (p_{i,t+1} - \check{p}_{it}) \\ &= \bar{p}_{it} + (1-m)(\tilde{p}_{it} - p_{it}) + m(\tilde{p}_t - p_t) + (p_{i,t+1} - \check{p}_{it}), \end{aligned}$$

where $\bar{p}_{it} = (1-m)p_{it}+mp_t$. Hence (B.1) holds with error term decomposable as (B.4), with

$$\eta_{it} = (1-m)(\tilde{p}_{it} - p_{it}) + (p_{i,t+1} - \check{p}_{it})$$

and $\xi_t = m(\tilde{p}_t - p_t)$. It follows from (16) and (28) that

$$\begin{aligned} d_{ii} &\sim (1-m)^2 \left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) p_{it}(1-p_{it}) + \tfrac{1}{2N} E\left(\check{p}_{it}(1-\check{p}_{it})\right) \\ &\sim (1-m)^2 \left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) p_{it}(1-p_{it}) + \tfrac{1}{2N}\bar{p}_{it}(1-\bar{p}_{it}), \end{aligned}$$

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

$$\begin{aligned} \tfrac{1}{s}\mathrm{tr}(D) &\sim (1-m)^2 \left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) \tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) + \tfrac{1}{2N}\tfrac{1}{s}\textstyle\sum_{i=1}^s \bar{p}_{it}(1-\bar{p}_{it}) \\ &= (1-m)^2 \left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) P_t(1-P_t)(1-F_{ST,t}) \\ &\quad + \tfrac{1}{2N} P_t(1-P_t)(1-(1-m)^2 F_{ST,t}) \\ &= P_t(1-P_t)\left(\tfrac{1}{2\tilde{N}} - (1-m)^2 \tfrac{1}{2N_e} F_{ST,t}\right). \end{aligned}$$

Inserting this expression into (B.5) and solving for F_{ST}^{eq} we arrive at (30). □

Proof of (35). It follows from (34) that

$$p_{i,t+1} = (1-m)p_{it}+mP_t+(1-m)(\tilde{p}_{it}-p_{it})+(1-m)(p^*_{it}-\tilde{p}_{it})+m(\tilde{p}_t-p_t)+m(h'_{it}-\tilde{p}_t).$$

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

$$\eta_{it} = (1-m)(\tilde{p}_{it} - p_{it}) + (1-m)(p^*_{it} - \tilde{p}_{it}) + m(h'_{it} - \tilde{p}_t)$$

and $\xi_t = m(\tilde{p}_t - p_t)$. It follows from (16), (31) and (33) that

$$\begin{aligned} \tfrac{1}{s}\mathrm{tr}(D) &\sim (1-m)^2 \left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) \cdot \tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) \\ &+ (1-m)^2 \tfrac{1}{2N(1-m)} \cdot \tfrac{1}{s}\textstyle\sum_{i=1}^s p_{it}(1-p_{it}) \\ &+ m^2 \tfrac{1}{2N(1-m)} P_t(1-P_t) \\ &= P_t(1-P_t)\left((1-m)^2\left(\tfrac{1}{2N_e} - \tfrac{1}{2N}\right) + \tfrac{1-m}{2N}(1-F_{ST,t}) + \tfrac{m}{2N}\right) \\ &= P_t(1-P_t)\left(\tfrac{1}{2\tilde{N}} - (\tfrac{1}{2\tilde{N}} - \tfrac{m}{2N})F_{ST,t}\right). \end{aligned}$$

Inserting this expression into (B.5) and solving for F_{ST}^{eq} we arrive at (30). □

Appendix C Quasi equilibrium, multiple loci/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 (4).

Assume there are genetic markers $l = 1, \dots, L$, with the l:th marker having alleles $q = 1, \dots, n_l$. Let p_{it}^{lq} and $P_t^{lq} = \sum_{i=1}^{s} p_{it}^{lq}/s$ refer to the frequency of allele q at marker l in island i and the whole population respectively. Locus and allele specific versions of the exact fixation index (1) in generation t and the approximate fixation index of generation $t+1$ are

$$F_{ST,t}^{lq} = \frac{\sum_{i=1}^{s}(p_{it}^{lq} - P_t^{lq})^2}{sP_t^{lq}(1-P_t^{lq})} =: \frac{Q_t^{lq}}{sP_t^{lq}(1-P_t^{lq})}$$

and

$$F_{ST,t+1}^{lq,\mathrm{appr}} = \frac{E(Q_t^{lq})}{sP_t^{lq}(1-P_t^{lq})\,(1-(2N_{e,\mathrm{tot}})^{-1})} \tag{C.1}$$

respectively, using

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

in analogy with (14). It has been shown by Wright (1978) that

$$G_{ST,t} = \frac{\sum_{lq} Q_t^{lq}}{s\sum_{lq} P_t^{lq}(1-P_t^{lq})} = \sum_{t}^{lq} \omega_t^{lq} F_{ST,t}^{lq}, \tag{C.3}$$

with weights $\omega_t^{lq} \propto P_t^{lq}(1-P_t^{lq})$ 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

$$G_{ST,t+1} = E(G_{ST,t+1}) + \epsilon_{t+1}, \tag{C.4}$$

where, analogously to (11), ϵ_{t+1} is an error term satisfying $E(\epsilon_{t+1}) = 0$. Then, when the number of loci L is large, we have, to a good approximation, that

$$\begin{array}{rcl} E(G_{ST,t+1}) & \sim & \sum_{lq} E(Q_{t+1}^{lq}) / \left(s\sum_{lq} E\left(P_{t+1}^{lq}(1-P_{t+1}^{lq})\right)\right) \\ & \sim & \sum_{lq} E(Q_{t+1}^{lq}) / \left(s\sum_{lq} P_t^{lq}(1-P_t^{lq})(1-1/(2N_{e,\mathrm{tot}}))\right) \\ & = & \sum_{lq} \omega_t^{lq} F_{ST,t+1}^{lq,\mathrm{appr}} \\ & =: & G_{ST,t+1}^{\mathrm{appr}}. \end{array} \tag{C.5}$$

In the first step of (C.5) we used the Law of Large Numbers, in the second step (C.2), and in the third step we used the definitions of $F_{ST,t+1}^{lq,\mathrm{appr}}$ and ω_t^{lq} in (C.1) and (C.3) respectively.

It follows from one locus calculations that

$$F_{ST,t+1}^{lq,\mathrm{appr}} = AF_{ST,t}^{lq} + B \tag{C.6}$$

will satisfy the same recursions, independently of l and q, where A and B are constants depending only on the reproduction cycle model, and not on $F_{ST,t}^{lq}$. Hence, it follows from (C.3) and (C.5) that the mean approximation of the coefficient of gene differentiation satisfies the same recursion

$$G_{ST,t+1}^{lq,\mathrm{appr}} = AG_{ST,t} + B \tag{C.7}$$

as (C.6). Inserting (C.7) as an approximation of $E(G_{ST,t+1})$ into (C.4), if follows that

$$G_{ST,t+1} = AG_{ST,t} + B + \epsilon_{t+1} \tag{C.8}$$

becomes an autoregressive process of order 1 with identical parameters A and B as in (15). In particular, the quasi equilibrium value $B/(1-A)$ is the same as in the one locus case. A Taylor expansion of (C.3) yields an error term

$$\epsilon_{t+1} = \frac{1}{sC_2}\sum_{lq}\left(Q_{t+1}^{lq} - E(Q_{t+1}^{lq})\right) - \frac{C_1}{sC_2^2}\sum_{lq}\left(P_{t+1}^{lq}(1-P_{t+1}^{lq}) - E\left(P_{t+1}^{lq}(1-P_{t+1}^{lq})\right)\right),$$

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

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

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

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

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

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

The upper part of (D.1) was derived by Wright (1943a) for the island model and by Wang and Caballero (1999, eqn. (15)) for more general subdivided populations with equally large and productive subpopulations and where fertilization precedes migration. On the other hand, $N_{e,\mathrm{tot}}$ 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 a working paper.

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

References

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

Brockwell, P.J. and 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. and Nei, M. 1982. Genetic differentiation of quantitative characters between populations of species. I. Mutation and random genetic drift. *Genetical Research Cambridge* **39**, 303-314.

Chakraborty, R. and Leimar, O. 1987. Genetic variation within a subdivided population. In *Population genetics and fishery management*, Ryman, N. and Utter, R. eds. Washington Sea Grant Program, Seattle, WA. Reprinted 2009 by The Blackburn Press, Caldwell, NJ, 420 pp.

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

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

Cockerhem, C.C. 1973. Analyses of gene frequencies. *Genetics* **74**, 679-700.

Crow, J.F. and Kimura, M. 1970. *An introduction to population genetics theory.* The Blackburn Press, Caldwell, New Jersey, USA.

Cockerham, C.C. and Weir, B.S. (1987). Correlations, descent measures: Drift with migration and mutation. *Proc. Natl. Acad. Sci. USA* **84**, 8512-8514.

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

Knutsen, H., Olsen, E.M., Jorde, P.E., Espeland, S.H. and André, C. 2011. Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A coast study of coastal Atlantic cod. *Molecular Ecology* **20**(4), 768-783.

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

Li, W.-H. 1976. Effect of migration on genetic distance. *Am. Nat.* **110**, 841-847.

Malécot, G. 1951. Un traitement stochastique des problemes lineaires (mutation, linkage, migration) en Genetique de Population. Ann. Univ. Lyon, Sci., Sect. A **14**, 79-117.

Maruyama, T. 1970. Effective number of alleles in subdivided populations. *Theor. Pop. Biol.* **1**, 273-306.

Nagylaki, T. 1980. The strong migration limit in geographically structured populations. *J. Math. Biol.* **9**, 101-114.

Nagylaki, T. 1983. The robustness of neutral models of geographical variation. *Theor. Pop. Biol.* **23**, 268-294.

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

Nagylaki, T. 1998. Fixation indeces in subdivided populations. *Genetics* **148**, 1325-1332.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **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. and Kumar, S. 2000. *Molecular evolution and phylogenetics.* Oxford University Press, Oxford.

Palm, S., Laikre, L., Jorde, P.E., Ryman, N. 2003. Effective population size and temporal genetic change in stream resident brown trout (Salmo trutta L.). *Conservation Genetics* **4**, 249-264.

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

Ryman, N., Hössjer, O., Allendorf, F.W., Jorde, P.E. and Laikre, L. 2012. Samples from structured populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. In preparation.

Ryman, N., Palm, S., André, C., Carvalho, G.R., Dahlgren, T.G., Jorde, P.E., Laikre, L., Larsson, L.C., Palmé, A., Ruzzante, D.E. 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology* **15**, 231-245.

Ryman, N. and 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. *J. Appl. Prob.* **43**, 351-362.

Sawyer, S. 1976. Results for the stepping-stone model for migration in population genetics. *Ann. Prob.* **4**, 699-728.

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

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

Sved, J.A. and Latter, B.D.H. 1977. Migration and mutation in stochastic models of gene frequency change. *J. Math. 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. and 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. Dioescious species. *Genetics* **146**, 1465-1474.

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

Waples, R.S. 1989. Temporal variation in allele frequencies: Testing the right hypothesis. *Evolution* **43**(6), 1236-151.

Waples, R.S. 2002. Definition and estimation of effective population size in the conservation of endangered species. In *Population Viability Analysis*, eds. Beissinger, S.R. and McCullogh, D.R, pp. 147-168.

Waples R.S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

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

Whitlock, M.C. and 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.* Evolution and the genetics of populations, vol. 4. University of Chicago Press, Chicago-London.

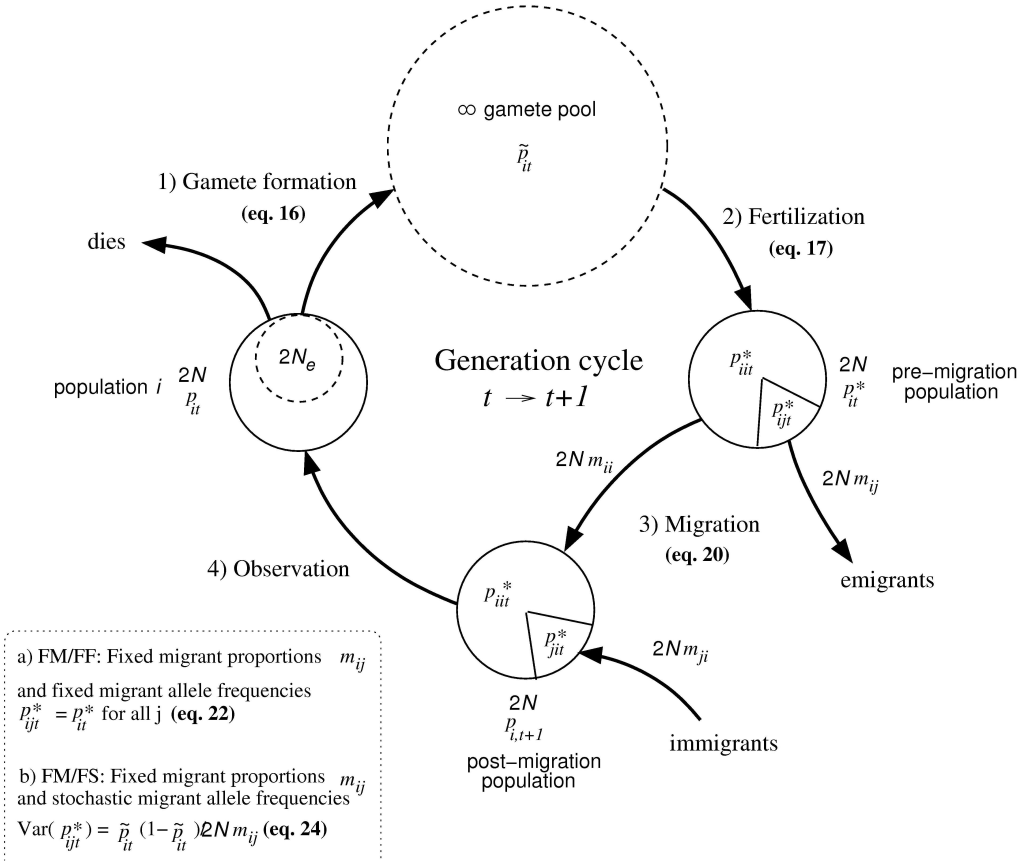

Figure 1: One generation cycle when fertilization precedes migration, with diploid migrants.

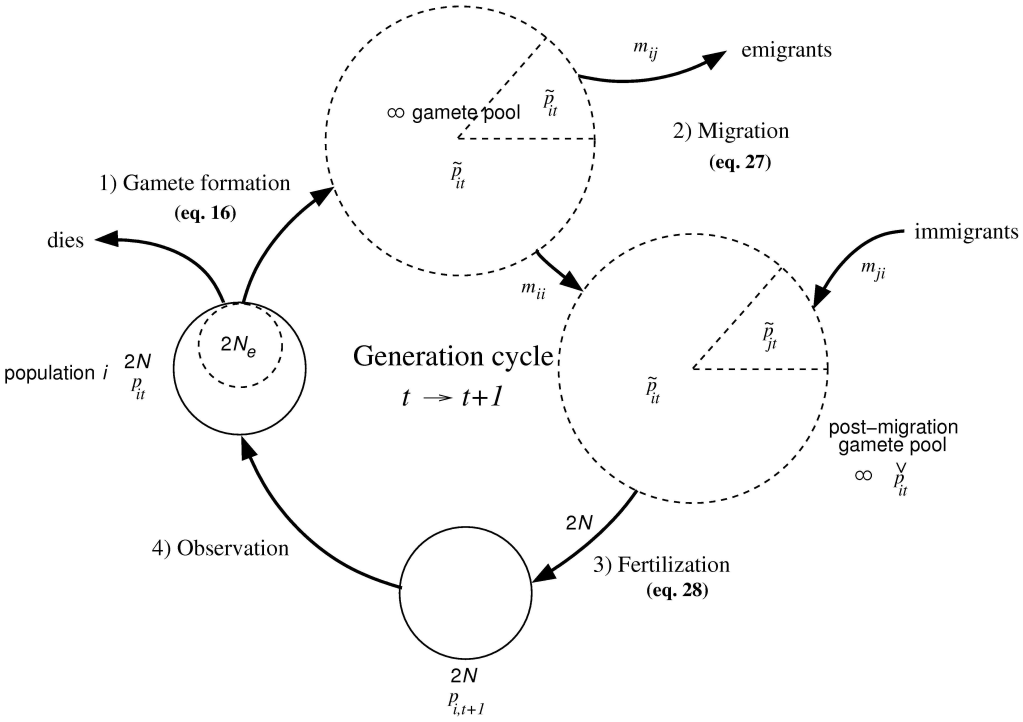

Figure 2: One generation cycle when migration precedes fertilization, with haploid migrants. Fixed migrant proportions m_{ij} and fixed migrant allele frequencies $\tilde{p}_{it}$ (MF/FF).

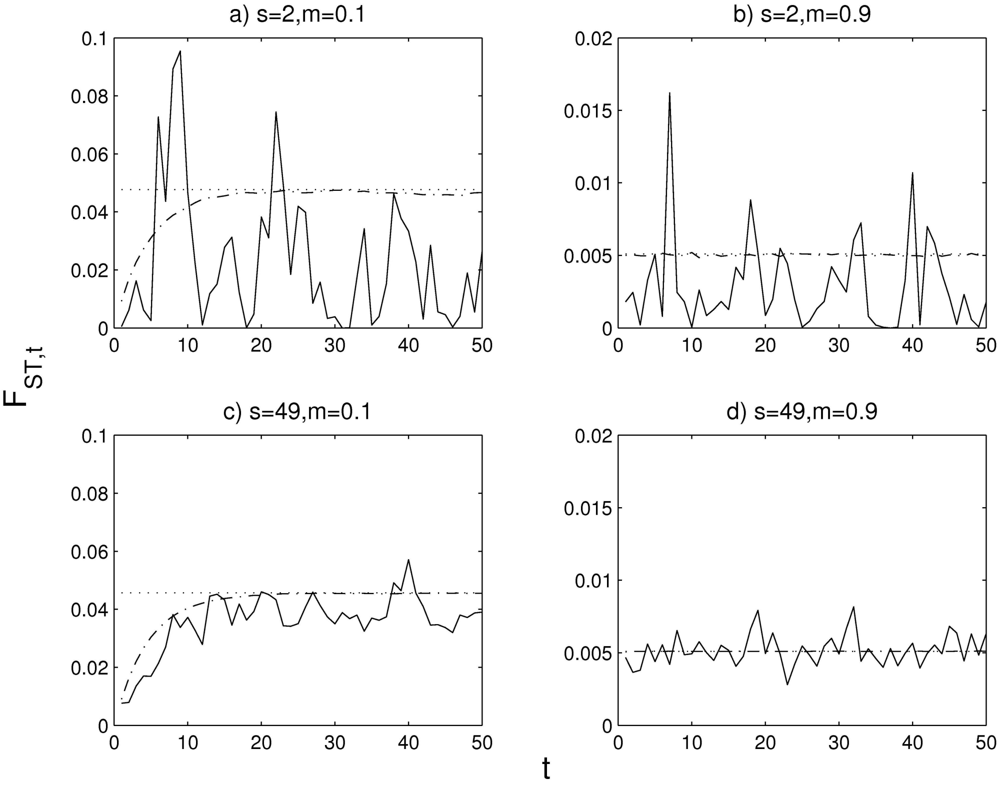

Figure 3: Plots of $F^{\mathrm{adj}}_{ST,t}$ as function of t when $N_e = 50$, $N = 100$ and a) $s = 2$, $m = 0.1$,, b) $s = 2$, $m = 0.9$, c) $s = 49$, $m = 0.1$ and d) $s = 49$, $m = 0.9$. The solid lines show $F^{\mathrm{adj}}_{ST,t}$ from one simulation, the dash-dotted lines are averages of $F^{\mathrm{adj}}_{ST,t}$ from 10 000 simulations and the dotted horisontal line corresponds to the limit as t increases, that is, the adjusted F^{eq}_{ST}. 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 $F^{\mathrm{adj}}_{ST,t}$ 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 1: **Notation for allele frequencies of various groups for alleles within a reproduction cycle.**

Symbol	Group	Scenario
p_{it}	Individuals of island i before reproduction cycle.	All
$\tilde{p}_{it}$	Gamete pool of island i, before migration.	All
p^*_{it}	Individuals after fertilization from gamete pool of island i before migration.	FM,I
$\check{p}_{it}$	Gamete pool of island i, after migration.	MF
$\tilde{p}_t$	Combined merged gamete pool	I
$\hat{p}_{it}$	Individuals of island i after fertilization from combined gamete pool.	I

Right: 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).

Table 2: **Quasi equilibrium values of F_{ST} for the neutral island model.**

s	Reproduction	$N_e = N$	$N_e \neq N$
∞	FM/FF	$\frac{(1-m)^2}{2N(1-(1-m)^2)+(1-m)^2}$ (5)	$\frac{(1-m)^2}{2N_e(1-(1-m)^2)+(1-m)^2}$
∞	FM/FS	$\frac{1}{2N(1-(1-m)^2)+1}$	$\frac{1}{2\tilde{N}(1-(1-m)^2)+1}$
∞	MF/FF	$\frac{1}{2N(1-(1-m)^2)+(1-m)^2}$ (7)	$\frac{1}{2\tilde{N}(1-(1-m)^2)+\frac{\tilde{N}}{N_e}(1-m)^2}$
∞	I/F	$\frac{1}{2N(1-(1-m)^2)+1-m}$ (6)	$\frac{1}{2\tilde{N}(1-(1-m)^2)+1-\frac{\tilde{N}}{N}m}$
$< \infty$	FM/FF	$\frac{(1-m)^2}{\frac{s}{s-1}2N(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+(1-m)^2}$	$\frac{(1-m)^2}{\frac{s}{s-1}2N_e(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+(1-m)^2}$ (23)
$< \infty$	FM/FS	$\frac{1}{\frac{s}{s-1}2N(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+1}$	$\frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+1}$ (25)
$< \infty$	MF/FF	$\frac{1}{\frac{s}{s-1}2N(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+(1-m)^2}$	$\frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+\frac{\tilde{N}}{N_e}(1-m)^2}$ (30)
$< \infty$	I/F	$\frac{1}{\frac{s}{s-1}2N(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+1-m}$	$\frac{1}{\frac{s}{s-1}2\tilde{N}(1-(1-m)^2-\frac{1}{2N_{e,\text{tot}}})+1-\frac{\tilde{N}}{N}m}$ (35)

Results 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. $N_{e,\text{tot}}$ 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. (26), 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}^{eq} is most pronounced for large migration rates, whereas F_{ST}^{eq} only depends on N_e for FM/FF. Model FM/FF has systematically lower values of F_{ST}^{eq} due to the $(1-m)^2$ term of the numerator.

Table 3: **Values of $F_{ST}^{\mathbf{eq}}$ 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 $F_{ST}^{\mathbf{eq}}$ 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 $F_{ST}^{\mathbf{eq}}$ 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 $G_{ST}^{\mathbf{eq}}$ and $G_{ST,t}$ for varying m.**

m	FM/FF	FM/FS	MF/FF	I/F	Sim1	Sim2	Sim 3	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 $\hat{p}_{it}$, 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_{it} we have removed this finite sample correction.