Theoretical Population Biology 102 (2015) 40–59

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Metapopulation inbreeding dynamics, effective size and subpopulation differentiation—A general analytical approach for diploid organisms

Ola Hössjer [a,*], Fredrik Olsson [a], Linda Laikre [b], Nils Ryman [b]

[a] Department of Mathematics, Division of Mathematical Statistics, Stockholm University, SE 106 91 Stockholm, Sweden
[b] Department of Zoology, Division of Population Genetics, Stockholm University, SE 106 91 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 22 September 2014
Available online 12 April 2015

Keywords:
Diploidy
Dioecious population
Effective population size
Fixation index
Identity by descent recursions
Monoecious population

ABSTRACT

Motivated by problems in conservation biology we study genetic dynamics in structured populations of diploid organisms (monoecious or dioecious). Our analysis provides an analytical framework that unifies substantial parts of previous work in terms of exact identity by descent (IBD) and identity by state (IBS) recursions. We provide exact conditions under which two structured haploid and diploid populations are equivalent, and some sufficient conditions under which a dioecious diploid population can be treated as a monoecious diploid one. The IBD recursions are used for computing local and metapopulation inbreeding and coancestry effective population sizes and for predictions of several types of fixation indices over different time horizons.

1. Introduction

The effective size N_e is the most widely used parameter to quantify rate of loss of genetic variation. The concept was first introduced by Wright (1931, 1938) as the size of a homogeneous population without mutation or selection and binomial variation of offspring numbers that has the same expected change of some genetic characteristic (e.g. inbreeding) per generation as the studied one. Many versions of N_e have been developed since, as reviewed for instance by Crow and Denniston (1988), Caballero (1994), Wang and Caballero (1999), Waples (2002, 2010), and Charlesworth (2009).

One of the most important applications of N_e is conservation biology (Allendorf and Ryman, 2002; Traill et al., 2010), and the present work was initiated from practical, real life conservation and management questions. Over the last centuries the rate of extinction of species and populations has increased by three orders of magnitude as compared to "normal", background extinction rates (Pimm et al., 2014), and many natural animal and plant populations are declining in size and are becoming fragmented over space (Groom et al., 2005). Numerous empirical studies have documented loss of genetic variation in such reduced and/or fragmented populations (e.g., Larson et al., 2002, Nabata et al., 2004 and Kettle, 2014) as well as associated negative effects such as inbreeding depression (Frankham, 2005; Liberg et al., 2005; Räikkönen et al., 2006, 2009).

General conservation genetic rules of thumb for the genetically effective population sizes required to avoid excessive rates of inbreeding and drift were suggested over three decades ago (Franklin, 1980), and the so-called 50/500 rule is now widely established (Jamieson and Allendorf, 2012), suggesting an $N_e > 50$ for short term conservation and $N_e > 500$ for long term conservation. However, this rule refers to single, isolated populations. Analytical approaches for understanding and computing N_e for subdivided so-called metapopulations (Levins, 1970; Harrison and Hastings, 1996) have not been available, where separated subpopulations may vary in size and even become extinct/recolonized. We initiated work to develop such analytical means, and recently presented a general approach for modelling effective size in subdivided populations over time (Hössjer et al., 2014) under a haploid assumption typical for many population genetics models. Here, we extend this work for diploid organisms. This involves four major mathematical contributions:

First, we provide a framework for a large class of diploid (monoecious or dioecious) structured and time-varying populations in Sections 2 and 3. We use a very general definition of a structured population as one consisting of several different subpopulations,

* Corresponding author.
E-mail addresses: ola@math.su.se (O. Hössjer), fredriko@math.su.se (F. Olsson), Linda.Laikre@popgen.su.se (L. Laikre), Nils.Ryman@popgen.su.se (N. Ryman).

http://dx.doi.org/10.1016/j.tpb.2015.03.006

where individuals within each subpopulation share some features. This includes many settings that have not been dealt with before in a unified way, such as spatial subdivision into geographical subunits or demes of arbitrary size and possibly asymmetric migration, an age-structured population with sex-specific survival and birth rates for the various age classes, or a pedigree of arbitrary form. Subpopulations can either be small or large, with sizes varying from an infinitely large continent to single individuals of a pedigree (cf. Crow and Kimura, 1970, Chapter 3). It is possible in this context to consider a pedigree whose individuals are distinct subpopulations, with occasional immigration from larger "background populations". The individuals have either unknown or known sex, with male and female gametes that are distinguishable or not. Time dynamics is defined through a reproduction cycle that allows for mating, selfing or cloning. Since the number of subpopulations may vary with time, it is also possible to incorporate subpopulation extinction and recolonization into our model.

Second, we use exact matrix analytic techniques to find recursions for probabilities that two alleles share a common ancestor identity by descent (IBD) or identity by state (IBS) in Sections 2, 3 and 7. Within our studied class of subdivided diploid populations, this requires a separate treatment for pairs of genes drawn from the same individual, from different individuals of the same subpopulation, or from different subpopulations. In particular, a general way of exploiting symmetries of the model is introduced in Section 4 in order to reduce dimensionality.

Third, we compare the diploid IBD-recursions with the corresponding haploid ones in Section 5. We find exact conditions in terms of random mating, random selfing and random coalescence, under which a monoecious diploid recursion is equivalent to the corresponding haploid recursion of Hössjer et al. (2014). We also give some sufficient conditions under which a diploid dioecious population can be reduced to a diploid monoecious one, where each pair of subpopulations of males and females, that represent different geographic demes or age classes, has been replaced by one monoecious subpopulation. We believe these results are important for at least two reasons. It gives theoretical insights into how the genetic composition of a population is affected by diploidy and two sexes, and it provides guidelines when computationally more feasible models (either monoecious diploid or haploid) can be used, with little or no loss of information.

Fourth, we use the diploid IBD- and IBS-recursions to compute effective sizes in Section 6. The effective size of a subdivided population is not captured by one single number though, since migration causes inbreeding to increase at a time varying rate. We therefore define the effective size as a time varying curve, as in Hössjer et al. (2014), so that short and long term effects of genetic drift and migration are captured. In addition, the diploid framework of this paper makes it possible to treat inbreeding and coancestry effective sizes separately, which is crucial for accurate modelling of inbreeding depression. It is also possible to incorporate local and global effective sizes by varying the weights of subpopulations.

Fifth, in Section 7 we use the diploid IBD- and IBS-recursions to predict various measures of subpopulation differentiation and departures from Hardy–Weinberg proportions, over different time horizons. Wright (1943, 1951) introduced a number of fixation indices for populations that are structured in a way of being geographically subdivided. Of these F_{ST} is a measure of subpopulation differentiation that quantifies genetic variation among subpopulations (S) within the total population (T), whereas F_{IS} and F_{IT} both quantify genetic variation within individuals (I) relative to subpopulations or the total population. These fixation indices were originally defined for biallelic genes, and later generalized to multiallelic and multilocus settings by Nei (1973, 1977), Wright (1978), Chakraborty (1993) and Nagylaki (1998a). In this paper we compute predictions of F_{IS}, F_{IT} and the coefficient of gene differentiation G_{ST}, which is the multiallelic version of F_{ST}.

We conclude with a discussion in Section 8, give mathematical details and some further examples in a supplementary material SM (see Appendix A), and summarize the most important notation in Table 1.

2. Model

Consider a diploid population evolving in discrete time $t = 0, 1, 2, \ldots$. We will sometimes refer to t as a generation, although our setup is more general and incorporates overlapping generations. The population consists of s_t subpopulations at time t, which may represent geographic regions (demes), age classes or even single individuals. The model is either monoecious or dioecious, and in the latter case all individuals of a subpopulation must have the same sex.

Let N_{ti} be the local census size of subpopulation i at time t. Each individual carries two copies of a portion of DNA that is small enough to neglect recombinations. It is located at a specific point that we refer to as a gene, so that subpopulation i has $2N_{ti}$ gene copies at time point t. Backward migration is specified in terms of $B_{ti,k}$, the fraction of genes of individuals in subpopulation i at time t that originate from subpopulation k at time $t-1$. The corresponding number $2N_{ti}B_{ti,k}$ of genes from k is a non-negative integer. (Throughout the paper we use commas in order to distinguish indices of different time points.) For some models and subpopulations i, a local effective size N_{eti} at time t can be specified in order to quantify the amount of genetic drift within i if it had been isolated. For some applications the model simplifies if N_{eti} replaces N_{ti}, but for models with overlapping generations there is no natural definition of N_{eti} when i represents an adult age class. For this reason we use N_{ti} rather than N_{eti} as a generic parameter.

Three types of fertilization are possible, either the same individual passes on its two genes to the offspring, which is then a genetically identical copy of the parent (cloning), or one individual passes on two genes to the offspring, drawn randomly with replacement (selfing), or each of two distinct individuals randomly passes on one of its two genes to the offspring (mating). When $B_{ti,k} > 0$, we let $c_{ti,k}\theta_{ti,k}$, $(1-c_{ti,k})\theta_{ti,k}$ and $1-\theta_{ti,k}$ be fractions of gametes of subpopulation i and time t that originate from k, that were reproduced through cloning, selfing and mating, respectively. Notice that survival of an individual can be regarded as a special case of cloning, where the parent has no more than one offspring—itself. The overall fraction of individuals produced through selfing or cloning in subpopulation i at time t is

$$\theta_{ti} = \sum_{k=1}^{s_t} B_{ti,k}\theta_{ti,k}. \tag{1}$$

Among all $2\sum_{i=1}^{s_t} N_{ti}$ gene copies that exist at time point t, consider two distinct ones. Let f_{ti} be the probability that they are IBD when picked from the *same* individual of subpopulation i, and f_{tij} the probability that they are IBD when drawn from *different* individuals of subpopulations i and j. Several definitions of IBD are possible, but here we mean that the two genes originate from the same ancestral gene of a founder generation, whether or not any mutations have occurred since then. If follows that f_{ti} is the inbreeding coefficient of individuals of subpopulation i and time point t, whereas f_{tij} is the kinship coefficient, also referred to as the coefficient of consanguinity or coancestry of individuals from i and j, see Chapter 3 of Crow and Kimura (1970). Inbreeding and coancestry within subpopulations put a bound on the amount of coancestry between subpopulations, in that

$$0 \le f_{tij} \le \sqrt{\left[\frac{2N_{ti}-2}{2N_{ti}}f_{tii} + \frac{1}{2N_{ti}}(f_{ti}+1)\right]\left[\frac{2N_{tj}-2}{2N_{tj}}f_{tjj} + \frac{1}{2N_{tj}}(f_{tj}+1)\right]} \tag{2}$$

for all $i \neq j$, see the SM for a proof.

Table 1
Notation for parameters.

Symbol	Definition
s_t	Number of subpopulations at time t
r_t	Number of single individuals at time t
i, j, k, l	Subpopulation numbers at time t (for i, j) or $t-1$ (for k, l)
ti	Subpopulation i at time t. When $N_{ti} = 1$, it specifies uniquely one individual.
a, b	Type of gene pair denoting subpopulation origin within or between individuals at time t (for a) or $t-1$ (for b)
$\mathcal{T}_t$	Type space of all gene pairs a at time t
d_t	Number of gene pair types a at time t
N_{ti}	Local census size of subpopulation i at time t
$\mu_{(t-1)k,i}$	Forward migration rate, i.e. expected number of gametes that individuals of subpopulation k and time $t-1$ transmit to subpopulation i of time t
$\nu_{(t-1)k,i}$	Mutation probability for a gamete produced in subpopulation k at time $t-1$ and transmitted to subpopulation i of time t
$B_{ti,k}$	Backward migration probability for gametes of subpopulation i at time t to originate from k
θ_{ti}	Fraction of offspring in subpopulation i at time t produced by selfing or cloning
$\theta_{ti,k}$	Fraction of genes produced by selfing or cloning, among those of subpopulation i at time t with parents from k
$\theta^{\text{rel}}_{ti,k}$	Fraction of genes produced by selfing or cloning, among those of subpopulation i at time t with parents from k, relative to fraction of other genes in i with its parent in k
$c_{ti,k}$	Fraction of genes produced by cloning, among those of subpopulation i at time t with parents from k that were produced by selfing or cloning
$B_{ti,l\|k}$	Mating function for individuals of subpopulation i at time t when male and female gametes cannot be distinguished: Probability that one parent from k picks a spouse from l
$Q_{ti,k}$	Joint backward migration probability for two genes within an individual of subpopulation i at time t to have both parental genes within the same individual of subpopulation k
$Q_{ti,kl}$	Joint backward migration probability for two genes within an individual of subpopulation i at time t to have its two parental genes in different individuals of subpopulations k and l
$Q_{tij,kl}$	Joint backward migration probability for two genes of different individuals of subpopulations i and j at time t to have their parental genes in subpopulations k and l
$\sigma_{(t-1)k,ij}$	Roughly the coefficient of covariation of the number of gametes an individual of subpopulation k and time $t-1$ transmits to subpopulations i and j of time t
$p_{tij,k}$	Coalescence probability, for a gene pair from subpopulations i and j at time t that both have their parental genes in k, to have these parental genes within the same individual
$\mu^x_{(t-1)k,i}$	Sex-specific forward migration rate. Expected number of gametes of sex $x \in \{m, f\}$ (male or female) individuals of subpopulation k and time $t-1$ transmit to subpopulation i of time t
$B^x_{ti,k}$	Backward migration probability for gametes of sex x of subpopulation i at time t to originate from k
$\tilde{B}^x_{ti,k}$	Backward migration probability for gametes of sex x that belong to mating individuals of subpopulation i at time t to originate from k
$B^{f\|m}_{ti,l\|k}$	Mating function for individuals of subpopulation i at time t when male and female gametes can be distinguished: Probability that a male parent from k picks a female parent from l
$Q^{xy}_{tij,kl}$	Joint backward migration probability for two genes of different individuals of subpopulations i and j of time t to have sexes x and y, and their parental genes in subpopulations k and l
$\sigma^{xy}_{(t-1)k,ij}$	Roughly the coefficient of covariation of the number of gametes of sexes x and y an individual of subpopulation k and time $t-1$ transmits to subpopulations i and j of time t
$p^{xy}_{tij,k}$	Coalescence probability, for a gene pair from subpopulations i and j at time t with sexes x and y, both of whose parental genes are in k, to have the parental genes from the same individual
$\xi_{(t-1)kr,i}$	Number of gametes that individual r of subpopulation k at time $t-1$ passes on to subpopulation i the next time point
$\eta_{(t-1)kr,i}$	Number of gametes that individual r of subpopulation k at time $t-1$ produces through mating and passes on to subpopulation i the next time point
$\theta_{(t-1)kr,i}$	Number of offspring that individual r of subpopulation k at time $t-1$ produces through selfing and passes on to subpopulation i the next time point
$\xi^x_{(t-1)kr,i}$	Number of gametes of sex x that individual r of subpopulation k at time $t-1$ passes on to subpopulation i the next time point
$\eta^x_{(t-1)kr,i}$	Number of gametes of sex x that individual r of subpopulation k at time $t-1$ produces through mating and passes on to subpopulation i the next time point
f_{ti}	Inbreeding coefficient of individuals from subpopulation i at time t.
f_{tii}	Coefficient of coancestry of two different individuals from subpopulation i at time t
f_{tij}	Coefficient of coancestry of two individuals from subpopulations $i \neq j$ at time t
h_{ti}	Non-IBD (non-IBS) probability for pairs of genes in individuals of subpopulation i at time t
h_{tij}	Non-IBD (non-IBS) probability for pairs of genes from different individuals in subpopulations i, j at time t
h_{It}	Non-IBD (non-IBS) probability for pairs of genes drawn from same individual at time t
h_{St}	Non-IBD (non-IBS) probability for pairs of genes drawn from same subpopulation at time t
h_{Tt}	Non-IBD (non-IBS) probability for pairs of genes drawn from the total population at time t
$\boldsymbol{h}_t$	Column vector of non-IBD (or non-IBS) probabilities for all types of gene pairs at time t
H_{ti}	Non-IBS probability for a gene pair within an individual of subpopulation i at time t, conditionally on allele frequencies at time t
H_{tij}	Non-IBS probability for two genes from different individuals in subpopulations i and j at time t, conditionally on allele frequencies at time t
$\boldsymbol{Q}_t$	Matrix $(Q_{ta,b})$ with backward migration probabilities for pairs of genes
$\boldsymbol{D}_t$	Matrix $(D_{ta,b})$ in recursion for non-IBD probabilities
$\boldsymbol{A}_t$	Matrix $(A_{ta,b})$ in recursion for predicted non-IBS probabilities
$\boldsymbol{V}_t$	Matrix $(V_{ta,b})$ with non-mutation probabilities for pairs of genes
w_i	Weight assigned to subpopulation i
W_{ta}	Weight assigned to gene pair a at time t
$\boldsymbol{W}_t$	Row vector (W_{ta}) with weights for all gene pairs a
$N_e([t_1, t_2])$	Effective size over time interval $[t_1, t_2]$
$N_{eI}([t_1, t_2])$	Inbreeding effective size over time interval $[t_1, t_2]$
$N_{eS_d}([t_1, t_2])$	Coancestry effective size within subpopulations over time interval $[t_1, t_2]$
$N_{eD}([t_1, t_2])$	Coancestry effective size between subpopulations over time interval $[t_1, t_2]$
N_{eE}	Eigenvalue effective size

(continued on next page)

Table 1 (*continued*)

Symbol	Definition
N_{eti}	Number of breeders of subpopulation i at time t, and (approximately) local effective size of i at time t, if i is isolated
F_{STt}	Fixation index of subpopulations within total population, time t
G_{STt}	Coefficient of gene differentiation of subpopulations within total population, time t
g_{STt}	Prediction of coefficient of gene differentiation at time t
F_{ITt}	Fixation index of individuals within total population, time t
F_{ISt}	Fixation index of individuals within subpopulations, time t
f_{XYt}	Prediction of fixation index at time t ($XY \in \{IT, IS\}$)
P_{tcir}	Frequency of allele c in individual r of subpopulation i at time t
P_{tci}	Frequency of allele c in subpopulation i at time t
P_{tc}	Frequency of allele c in total population (as specified by the subpopulation weights), time t

We introduce a state space

$$\begin{aligned}\mathcal{T}_t &= \mathcal{I}_t \cup \mathcal{S}_t \cup \mathcal{D}_t \\ &= \{1, \ldots, s_t\} \cup \{ii,\ 1 \le i \le s_t, N_{ti} \ge 2\} \\ &\cup\ \{ij, 1 \le i \ne j \le s_t\}\end{aligned} \tag{3}$$

at time t that represents all possible ways to pick the two genes, either from the same individual ($\mathcal{I}_t$), from two different individuals of the same subpopulation ($\mathcal{S}_t$), or from two different subpopulations ($\mathcal{D}_t$). The state space has

$$d_t = s_t + (s_t - r_t) + s_t(s_t - 1) \tag{4}$$

elements at time t, with r_t the number of subpopulations with one single individual ($N_{ti} = 1$).

It is mathematically more convenient to use probabilities $h_{ti} = 1 - f_{ti}$ and $h_{tij} = 1 - f_{tij}$ that two distinct genes are *not* IBD when drawn from the same individual of subpopulation i or different individuals of subpopulations i and j. Write $a \in \mathcal{T}_t$ and $b \in \mathcal{T}_{t-1}$ for an arbitrary type at times t and $t-1$, and gather all non-IBD probabilities at time t into a column vector $\boldsymbol{h}_t = (h_{ta};\ a \in \mathcal{T}_t)'$ of length d_t. Below we will derive a joint linear time recursion

$$\boldsymbol{h}_t = \boldsymbol{D}_t \boldsymbol{h}_{t-1}, \quad t = 1, 2, 3, \ldots \tag{5}$$

for all inbreeding and coancestry (within and between subpopulations) coefficients. Each element $D_{ta,b}$ of the $d_t \times d_{t-1}$ matrix $\boldsymbol{D}_t = (D_{ta,b};\ a \in \mathcal{T}_t, b \in \mathcal{T}_{t-1})$ is the probability, for a randomly drawn pair of genes of type a at time t to have different parental genes at time $t-1$ of type b. We may interpret $\boldsymbol{D}_t$ as part of a transition matrix of a structured coalescent for pairs of genes. This is a Markov chain that runs backwards in time, with absorption equivalent to coalescence on the gene level. The matrix $\boldsymbol{D}_t$ contains transition probabilities between all non-absorbing states, that is, between all pairs of genes before they coalesce.

In component form we rewrite (5) as

$$h_{ta} = \sum_{b \in \mathcal{T}_{t-1}} D_{ta,b} h_{(t-1)b} \tag{6}$$

for all $a \in \mathcal{T}_t$. It will be seen in Section 4 that the state space in (5)–(6) can be reduced, due to the symmetry requirement

$$h_{tij} = h_{tji} \tag{7}$$

on all non-IBD probabilities such that $i \ne j$. The IBD recursion in (5)–(6) is homogeneous. In Section 7 we consider an extension based on IBS sharing that is inhomogeneous due to mutations.

3. Reproduction cycle

Recursion (5) depends on the form of the reproduction cycle between time points $t-1$ and t, and in particular on how individuals are interpreted. Four major cases will be discussed:

I An essentially haploid and monoecious model, where reproduction is defined on the gene (not individual) level. Then genes are randomly paired up to form individuals at each time point, within all subpopulations, when trying to mimic a diploid model.

II A diploid model where individuals are monoecious and isogamous, and gametes cannot be morphologically distinguished as male and female.

III A diploid model where individuals are monoecious, but their male and female gametes can be morphologically distinguished.

IV A diploid model where individuals are dioecious, so that their sex is well defined and also accounted for in the model. Their male and female gametes can also be morphologically distinguished.

Since individuals of a haploid model are formed randomly from pairs of genes, when mimicking a diploid population, there is no cloning but a random amount of selfing for I, as will be specified below. Cloning may occur for models II–IV (e.g. when generations are overlapping), there is no restriction on the amount of selfing for models II and III, but no selfing occurs for model IV (although mathematically, the theory allows for this). Cases II and III are biologically equivalent, but differ in the available amount of information. This distinction is important, for instance, when migration patterns or reproductive variability between male and female gametes differ.

In the next three subsections we treat Case I, Case II and Case III + IV, respectively.

3.1. Haploid models

Haploid models for subdivided populations have been treated by many authors, starting with the seminal work of Malécot (1951). Here we include it as a point of reference for comparison with diploid models. We will assume that migration rates are fixed, so that parental subpopulations of genes from different individuals are drawn without replacement in accordance with the backward migration rates $B_{ti,k}$, see Sved and Latter (1977), Wang (1997a), Hössjer et al. (2013) and Hössjer and Ryman (2014).

When two distinct genes are drawn from the population, it suffices to compute their non-IBD probability depending on which pair of subpopulations i and j they belong to. Indeed, since genes are paired up randomly to individuals (when trying to mimic a diploid situation), the non-IBD probability for two genes within the same individual will be the same, whether they belong to the same individual or not. For this reason, we can merge states i and ii of state space (3) without loss of information, and obtain a reduced one

$$\mathcal{T}_t^{\text{hapl}} = \{ij;\ 1 \le i, j \le s_t\}$$

of size s_t^2 at time t, corresponding to all ordered pairs of subpopulations, and with hapl short for haploid. In particular, h_{tii} is the non-IBD probability of two distinct genes from i, whether they belong to the same individual or not. (In Section 4 we will consider state space reduction more generally.)

The non-IBD vector $\boldsymbol{h}_t$ has length s_t^2 and the matrix $\boldsymbol{D}_t = (D_{ta,b}) = (D_{tij,kl})$ of the non-IBD recursion is of dimension $s_t^2 \times s_{t-1}^2$, since $a = ij \in \mathcal{T}_t^{\text{hapl}}$ and $b = kl \in \mathcal{T}_{t-1}^{\text{hapl}}$. Hössjer and Ryman

(2014) provided non-IBD recursions for fixed migration rates when the number of subpopulations $s_t = s$ is constant over time. Their results can easily be extended to scenarios where s_t depends on t, with

$$D_{tij,kl} = Q_{tij,kl}(1 - p_{tij,k})^{\delta_{\{k=l\}}}, \tag{8}$$

where δ_A equals 1 if A holds, and 0 if not. The first term

$$Q_{tij,kl} = B_{ti,k}\frac{2N_{tj}B_{tj,l} - \delta_{\{i=j,k=l\}}}{2N_{tj} - \delta_{\{i=j\}}} \approx B_{ti,k}B_{tj,l} \tag{9}$$

on the right hand side of (8) is a backward migration probability for a random pair of different genes of subpopulations i and j at time t to have their parental genes from subpopulations k and l at time $t-1$. The approximation in (9) holds for large populations, and corresponds to drawing parental subpopulations multinomially with replacement. The second quantity $p_{tij,k}$ is only defined when $k = l$. It is the coalescence probability for a pair of genes from subpopulations i and j at time t, to have the same parental gene in subpopulation k, given that both parental genes originate from the same subpopulation k.

Suppose in particular that $N_{e(t-1)k} \le N_{(t-1)k}$ breeders are chosen randomly among all individuals in subpopulation k at time point $t-1$. Then different offspring choose parental genes in k randomly with replacement among the $2N_{e,t-1,k}$ genes of the breeders. This leads to

$$p_{tij,k} = \frac{1}{2N_{e(t-1)k}}, \quad \text{if } Q_{tij,kk} > 0, \tag{10}$$

so that the coalescence probability only depends on the parental subpopulation k, not the subpopulations i and j of the offspring. In particular, when subpopulation k is isolated, $N_{e(t-1)k}$ can be interpreted as a local inbreeding effective size of k between time points $t-1$ and t.

3.2. Diploid models with indistinguishable male and female gametes

Monoecious isogamous diploids correspond to Case II above and Cases 1 and 2 of the reproduction scenarios treated by Crow and Denniston (1988) for homogeneous populations, and Balloux et al. (2003) for the island model with mating, selfing and cloning. In the SM we prove that the elements of $\boldsymbol{D}_t$ are given by

$$D_{ta,b} = \begin{cases} Q_{ti,k}\dfrac{1+c_{tik}}{2}, & a = i,\ b = k, \\ Q_{ti,kl}, & a = i,\ b = kl, \\ Q_{tij,kk}(1 - p_{tii,k}), & a = ij,\ b = kk, \\ Q_{tij,kl}, & a = ij,\ b = kl,\ k \neq l, \\ Q_{tij,kk}\dfrac{p_{tij,k}}{2}, & a = ij,\ b = k. \end{cases} \tag{11}$$

The first quantity on the right hand side of (11), $Q_{ta,b}$, is a backward migration probability for a random pair of different genes of type a at time t to have their two parental genes as a pair of type b at time $t-1$, before possible coalescence events of genes to originate from the same individual, are taken into account. Coalescence events of a haploid model are always between genes, but for diploid models they occur both at the individual and gene level. The terms $(1 + c_{tik})/2$ and $1/2$ of the first and fifth equations of (11) are coalescence probabilities at the gene level, whereas $p_{ta,b}$ is a coalescence probability for individuals, defined when $a \notin \mathcal{I}_t$ and $b \in \mathcal{I}_{t-1}$. It is the probability for a pair genes of type a, from different individuals at time t, to have their two parental genes in the same individual of type $b = k$ (but not necessarily the same gene), given that both parental genes originate from the same subpopulation k. Fig. 1 illustrates backward migration and coalescence probabilities in more detail.

It is proved in the SM that fixed migration rates lead to backward migration probabilities of the form

$$\begin{aligned} Q_{ti,k} &= \theta_{ti,k}B_{ti,k}, \\ Q_{ti,kl} &= (1 - \theta_{ti,k})B_{ti,k}B_{ti,l|k}, \\ Q_{tij,kl} &= \frac{2N_{tj}B_{ti,k}B_{tj,l} - \delta_{\{i=j,k=l\}}(B_{ti,k} + Q_{ti,k}) - \delta_{\{i=j\}}Q_{ti,kl}}{2N_{tj} - 2\delta_{\{i=j\}}} \\ &\approx B_{ti,k}B_{tj,l}, \end{aligned} \tag{12}$$

for pairs of genes, where the first equation corresponds to selfing/cloning, the second to mating and the third to backward migration for pairs of genes from different individuals. The approximation in the last equation of (12) is accurate for large populations and neglects that genes are drawn without replacement.

For haploid models, backward migration probabilities for pairs of genes are uniquely determined by those for single genes, cf. (9). For diploid models we also need to know how gametes unite through mating, the so called mating rule for how the two parents' subpopulations are chosen. It can either be expressed through $Q_{ti,kl}$, the fraction of ordered pairs of gametes in i at time t, among those that united through mating, that originated from k and l respectively. Without loss of generality we may assume that $Q_{ti,kl} = Q_{ti,lk}$ whenever $k \neq l$, so that $\sum_{l=1}^{s} Q_{ti,kl} = B_{ti,k}(1 - \theta_{ti,k})$ is the total fraction of mating gametes in i whose parents lived in k. It follows from (12) that

$$B_{ti,l|k} = \frac{Q_{ti,kl}}{B_{ti,k}(1 - \theta_{ti,k})} \tag{13}$$

is the conditional probability that the homologous gene of a gamete in subpopulation i at time t is from subpopulation l, given that the parent of the first gene is from subpopulation k. Formula (13) provides an alternative (and equivalent) definition of the mating rule, which is convenient to use for

$$B_{ti,l|k} = \delta_{\{l=k\}}, \quad \text{(mating before migration)}, \tag{14}$$

i.e. when the two homologous genes of a biparental individual are required to come from the same subpopulation. Another mating rule

$$B_{ti,l|k} = \frac{\frac{2N_{ti}B_{ti,l} - \delta_{\{l=k\}}}{(2N_{ti}-1)} - \delta_{\{l=k\}}\theta_{ti,k}}{(1 - \theta_{ti,k})}, \quad \text{(random mating after migration)}, \tag{15}$$

is defined by drawing the two genes within an individual of i at time t randomly without replacement, regardless of which subpopulations k and l they were inherited from. Random mating after migration is only well defined when the numerator of (15) is non-negative for all $l = k$.

To summarize, the major difference between the two mating rules (14) and (15) is that parents from the same subpopulation mate in (14) before their offspring migrate, whereas parents first migrate in (15), then find their partner (possibly itself) independently of its subpopulation origin and reproduce. Therefore, the randomness of (15) only refers to the choice of the parents' *subpopulations*. Random mating among parents *within* subpopulations requires, for each offspring with both parents in k, that the two parents are picked randomly without replacement. This is satisfied for the model below with $N_{e(t-1)k} = N_{(t-1)k}$ breeders.

Selfing or cloning in subpopulation k is only possible if at least two genes in i at time t originate from k, i.e. $2N_{ti}B_{ti,k} > 1$. By random selfing/cloning we mean that the probability for selfing or cloning to occur is the same as for a haploid model where parental genes are drawn independently with replacement. That is, for each individual of subpopulation i at time t with both genes inherited

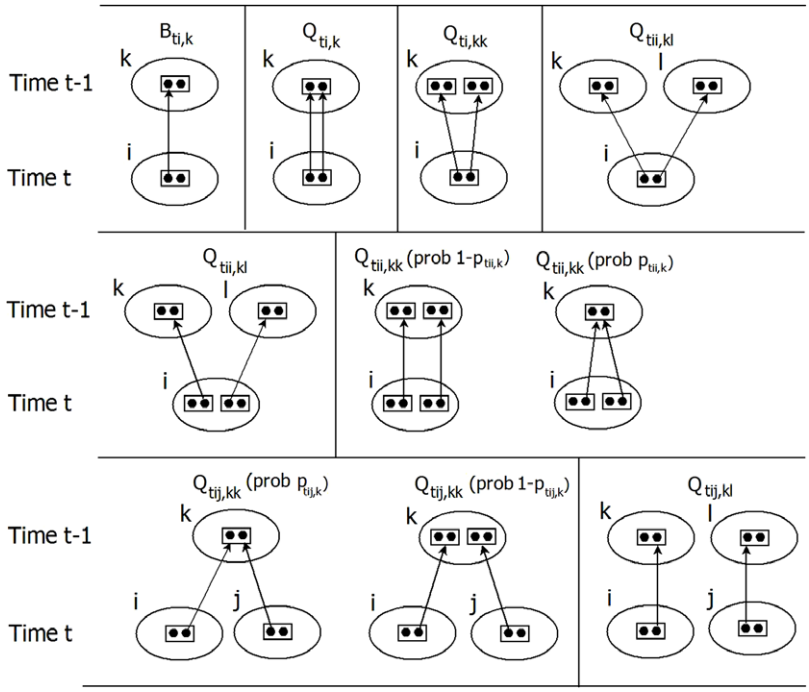

Fig. 1. Illustration of backward migration for a diploid model between time points t and $t-1$ when male and female gametes cannot be distinguished (Case II). Ellipses, rectangles and filled circles correspond to single subpopulations (i, j, k, l), individuals and genes, respectively. The upper row shows backward migration for one gene ($B_{ti,k}$) or for two genes from a single individual, the second row backward migration of two genes from two individuals within the same subpopulation, and the third row backward migration for a pair of genes from different individuals of different subpopulations. Two genes of different individuals that have been inherited from the same subpopulation k ($Q_{tii,kk}$ or $Q_{tij,kk}$) could either originate from the same individual or not, as determined by coalescence probability $p_{tii,k}$ or $p_{tij,k}$. It is assumed that $i \neq j$ and $k \neq l$.

from k, these two genes are drawn randomly with replacement from k. This can be expressed as

$$\frac{Q_{ti,k}}{Q_{ti,k}+Q_{ti,kk}} = \frac{1}{N_{(t-1)k}}, \quad \text{(random selfing/cloning)}. \tag{16}$$

Formulas (12) and (14)–(17) imply that the left hand side of (16) equals θ_{tik} when mating precedes migration, and

$$\theta^{\text{rel}}_{ti,k} = \theta_{ti,k} \Big/ \frac{2N_{ti}B_{ti,k}-1}{2N_{ti}-1} \tag{17}$$

for random mating after migration, which may be interpreted as a selfing/cloning rate from i to k, *relative* to the total fraction of other genes in i with parents from k. It follows that

$$\begin{aligned} &\text{Mating before migration, random selfing/cloning :} \\ &\theta_{ti,k} = \frac{1}{N_{(t-1)k}}, \\ &\text{Random mating after migration, random selfing/cloning :} \\ &\theta^{\text{rel}}_{ti,k} = \frac{1}{N_{(t-1)k}}. \end{aligned} \tag{18}$$

Since the fraction of individuals in i with both parental genes from k is smaller when mating is random, the overall fraction θ_{tik} of selfing/cloning among genes in i with parents from k, is lower in (18) for random mating after migration, than for mating that precedes migration.

In order to find general expressions for the coalescence probabilities $p_{tij,k}$, we need to look forward in time and model reproduction between time points $t-1$ and t explicitly. We will generalize formulas for the coalescence probability in Wang (1996a, 1997a), and to this end we introduce the total number of gametes

$$\xi_{(t-1)kr,i} = 2\theta_{(t-1)kr,i} + \eta_{(t-1)kr,i} \tag{19}$$

that individual $r = 1, \ldots, N_{(t-1)k}$ of subpopulation k and time $t-1$ passes on to i the next time point t. Of these gametes $2\theta_{(t-1)kr,i}$ are obtained through cloning or selfing, and $\eta_{(t-1)kr,i}$ through mating. It follows that

$$\begin{aligned} 2N_{ti}B_{ti,k} &= \sum_{r=1}^{N_{t-1,k}} \xi_{(t-1)kr,i}, \\ N_{ti}B_{ti,k}\theta_{ti,k} &= \sum_{r=1}^{N_{(t-1)k}} \theta_{(t-1)kr,i}, \\ 2N_{ti}B_{ti,k}(1-\theta_{ti,k}) &= \sum_{r=1}^{N_{(t-1)k}} \eta_{(t-1)kr,i}. \end{aligned} \tag{20}$$

We assume that $\{\theta_{(t-1)kr,i}, \eta_{(t-1)kr,i}\}_{i=1}^{s}$ are exchangeable random vectors for $r = 1, \ldots, N_{(t-1)k}$. It will simplify formulas to introduce

$$\begin{aligned} \mu_{(t-1)k,i} &= E(\xi_{(t-1)k1,i}) = 2N_{ti}B_{ti,k}/N_{(t-1)k}, \\ \sigma_{(t-1)k,ij} &= E\left[\xi_{(t-1)k1,i}(\xi_{(t-1)k1,j} - \delta_{\{i=j\}})\right]/(\mu_{(t-1)k,i}\mu_{(t-1)k,j}), \end{aligned} \tag{21}$$

where $\mu_{(t-1)k,i}$ is the expected number of gametes that each individual in k at time $t-1$ passes on to i the next time point, and $\sigma_{(t-1)k,ij}$ is related but not equivalent to a squared coefficient of covariation of the number of gametes that individuals in k pass on to subpopulations i and j. The coalescence probability

$$\begin{aligned} p_{tij,k} &= \frac{\sigma_{(t-1)k,ij} - \delta_{\{i=j\}}\frac{\theta_{ti,k}}{\mu_{(t-1)k,i}}}{N_{(t-1)k} - \frac{2-(1-\theta_{ti,k})(1-B_{ti,k|k})}{\mu_{(t-1)k,i}}} \\ &\approx \frac{\sigma_{(t-1)k,ij} - \delta_{\{i=j\}}\frac{\theta_{ti,k}}{\mu_{(t-1)k,i}}}{N_{(t-1)k}}, \end{aligned} \tag{22}$$

see the SM for a proof. It is essentially inversely proportional to the local census size $N_{(t-1)k}$ of the subpopulation k in which the individual lineages merge, with a constant of proportionality $\sigma_{(t-1)k,ij} - \delta_{\{i=j\}}\theta_{ti,k}/\mu_{(t-1)k,i}$ that can be interpreted as a coalescence rate when time is measured in units of $N_{(t-1)k}$.

For models where the coalescence probabilities in (22) only depend on the subpopulation k of the parent, but not on the subpopulations i and j of the two children, we may define a local effective size $N_{e(t-1)k}$ of k at time $t-1$ through

$$p_{tij,k} = \frac{1}{N_{e(t-1)k}} \quad \text{if } Q_{tij,kk} > 0. \tag{23}$$

A sufficient condition for (23) to hold is that $N_{e(t-1)k} \leq N_{(t-1)k}$ individuals are chosen randomly as breeders among all individuals in subpopulation k at time point $t-1$. Then single parents in k that reproduced through selfing, or mating parents in k that have their spouse in another subpopulation are chosen randomly with replacement among these breeders in k, for different progeny of the next time point. If, on the other hand, pairs of parents within k mate, they are chosen randomly without replacement among breeders in k for each offspring. Then different offspring pick their parents independently. It is shown in the SM that (23) follows from (22) under these assumptions. On the other hand, (23) does not hold for models with overlapping generations.

3.3. Diploid models with distinguishable male and female gametes

In this subsection we treat Cases III and IV jointly. Case III is relevant for several plant and some animal species, with male or female gametes distinguished and allowed to migrate independently between subpopulations before fertilization, possibly with different migration rates, see for instance Wang (1997a). Case IV, on the other hand, is relevant for mammals and many other species, where individuals have a specific sex. This can be modelled so that each subpopulation has all its individuals of the same known sex, see for instance Caballero and Hill (1992b) and Nagylaki (1995) for homogeneous populations, and Wang (1997b) for island models.

It is shown in the SM that the non-inbreeding recursion (5) for a diploid population in which male and female gametes are distinguished is

$$D_{ta,b} = \begin{cases} Q_{ti,k}\dfrac{1+c_{ti,k}}{2}, & a=i,\ b=k, \\ Q_{ti,kl}, & a=i,\ b=kl, \\ Q^{mm}_{tij,kl}(1-p^{mm}_{tij,k}) + Q^{mf}_{tij,kl}(1-p^{mf}_{tij,k}) & \\ \quad + Q^{fm}_{tij,kl}(1-p^{fm}_{tij,k}) + Q^{ff}_{tij,kl}(1-p^{ff}_{tij,k}), & a=ij,\ b=kk, \\ Q^{mm}_{tij,kl} + Q^{mf}_{tij,kl} + Q^{fm}_{tij,kl} + Q^{ff}_{tij,kl}, & a=ij,\ b=kl,\ k\neq l, \\ Q^{mm}_{tij,kl}\dfrac{p^{mm}_{tij,k}}{2} + Q^{mf}_{tij,kl}\dfrac{p^{mf}_{tij,k}}{2} & \\ \quad + Q^{fm}_{tij,kl}\dfrac{p^{fm}_{tij,k}}{2} + Q^{ff}_{tij,kl}\dfrac{p^{ff}_{tij,k}}{2}, & a=ij,\ b=k, \end{cases} \tag{24}$$

i.e. a bit more complicated than the corresponding formula (11) when gametes are indistinguishable. The quantity $Q^{xy}_{tij,kl}$ is the probability that a pair of genes picked from different individuals of subpopulations i and j at time t have sexes x and y and migrated from k and l respectively, and $p^{xy}_{tij,k}$ is the probability that a randomly chosen pair of x and y gametes from different individuals of subpopulations i and j at time t have the same parent, but not necessarily the same parental gene, given that both parental genes are from k.

In order to define (24) more explicitly, we write the terms of (19) as a sum

$$\begin{aligned} \xi_{(t-1)kr,i} &= \xi^m_{(t-1)kr,i} + \xi^f_{(t-1)kr,i}, \\ \eta_{(t-1)kr,i} &= \eta^m_{(t-1)kr,i} + \eta^f_{(t-1)kr,i}, \end{aligned} \tag{25}$$

of the total or mating number of male and female gametes passed on by individual r of subpopulation k at time $t-1$ to subpopulation i of the next time point. When the sex of r is known, or when r has a well defined sex that is unknown, at most one of $\eta^m_{(t-1)kr,i}$ and $\eta^f_{(t-1)kr,i}$ is nonzero.

Introduce the sex specific backward migration rates

$$\begin{aligned} B^x_{ti,k} &= \sum_{r=1}^{N_{(t-1)k}} \xi^x_{(t-1)kr,i}/N_{ti}, \\ \tilde{B}^x_{ti,k} &= \sum_{r=1}^{N_{(t-1)k}} \eta^x_{(t-1)kr,i}/\left(N_{ti}(1-\theta_{ti})\right), \end{aligned}$$

i.e. the fraction of x gametes of subpopulation i at time t whose parental genes are from k, either among all gametes or among those that mate. It follows from (20) that the corresponding non-sex specific backward migration rate is

$$B_{ti,k} = \frac{1}{2}(B^m_{ti,k} + B^f_{ti,k}). \tag{26}$$

As in the previous subsection, pairs of gametes produced by selfing or cloning migrate together, but gametes produced by mating migrate according to a mating function $Q_{ti,kl} = Q^{mf}_{ti,kl}$ that equals the fraction of individuals in subpopulation i at time t, whose homologous male and female gametes originate from subpopulations k and l respectively at time point $t-1$. It follows that

$$\begin{aligned} \sum_{l=1}^{s_t} Q_{ti,kl} &= \sum_{r=1}^{N_{(t-1)k}} \eta^m_{(t-1)kr,i}/N_{ti} = (1-\theta_{ti})\tilde{B}^m_{ti,k}, \\ \sum_{l=1}^{s_t} Q_{ti,lk} &= \sum_{r=1}^{N_{(t-1)k}} \eta^f_{(t-1)kr,i}/N_{ti} = (1-\theta_{ti})\tilde{B}^f_{ti,k}, \end{aligned}$$

and we let

$$B^{f|m}_{ti,l|k} = \frac{Q_{ti,kl}}{(1-\theta_{ti})\tilde{B}^m_{ti,k}} \tag{27}$$

be the conditional probability that the gene of an individual in subpopulation i and time t with a female parent originates from subpopulation l, given that the other homologous gene with a male parent, originates from subpopulation k. In the same way, we define the conditional probability $B^{m|f}_{ti,l|k}$ of a female parent from k to choose its male spouse from subpopulation l. The mating schemes

$$B^{y|x}_{ti,l|k} = \begin{cases} \tilde{B}^y_{ti,l}, & \text{random mating after migration,} \\ \delta_{\{l=k\}}, & \text{mating before migration,} \end{cases} \tag{28}$$

correspond to independence and identity between the parents' subpopulations. Other mating functions for two-sex models (Case IV) will be defined in Section 5 for age-structured populations, and for pedigrees. Gasbarra et al. (2005) consider a two-sex model with a mating rule defined in terms of a Polya urn, both for homogeneous and structured populations. It can be used to control the degree of monogamy, either of males or females, but requires joint analysis of more than two genes.

The backward migration probabilities for pairs of genes can now be written as Eq. (29) given in Box I where the first equation follows by the definition of $Q_{ti,k}$, $\theta_{ti,k}$ and $B_{ti,k}$, the second equation follows from (27) and the last one is proved in the SM. As in the previous subsection, the approximation of the last line is accurate for large population sizes.

In order to define coalescence probabilities, it is convenient to introduce

$$\begin{aligned} \mu^x_{(t-1)k,i} &= E(\xi^x_{(t-1)k1,i}) = N_{ti}B^x_{ti,k}/N_{(t-1)k}, \\ \sigma^{xy}_{(t-1)k,ij} &= E\left[\xi^x_{(t-1)k1,i}(\xi^y_{(t-1)k1,j} - \delta_{\{i=j,x=y\}})\right]/(\mu^x_{(t-1)k,i}\mu^y_{(t-1)k,j}), \end{aligned} \tag{30}$$

$$
\begin{aligned}
Q_{ti,k} &= \theta_{ti,k} B_{ti,k},\\
Q_{ti,kl} &= (1-\theta_{ti})\tilde{B}^{m}_{ti,k} B^{f|m}_{ti,l|k},\\
Q^{xy}_{tij,kl} &= \frac{1}{4}\frac{N_{ti}B^{x}_{ti,k}B^{y}_{tj,l} - \delta_{\{i=j,k=l,x=y\}}B^{x}_{ti,k} - \delta_{\{i=j,k=l,x\neq y\}}B_{ti,k}\theta_{ti,k} - \delta_{\{i=j,x\neq y\}}(1-\theta_{ti})\tilde{B}^{x}_{ti,k}}{N_{ti}-\delta_{\{i=j\}}}\\
&\approx \frac{1}{4}B^{x}_{ti,k}B^{y}_{tj,l},
\end{aligned}
\tag{29}
$$

Box I.

with similar interpretations as in (21). It is shown in the SM that

$$
\begin{aligned}
p^{xx}_{tii,k} &= \frac{\sigma^{xx}_{(t-1)k,ii}}{N_{(t-1)k} - \frac{1}{\mu^{xx}_{(t-1)k,i}}} \approx \frac{\sigma^{xx}_{(t-1)k,ii}}{N_{(t-1)k}},\\
p^{xy}_{tii,k} &\overset{x\neq y}{=} \frac{\sigma^{xy}_{(t-1)k,ii} - \frac{\mu_{(t-1)k,i}\theta_{ti,k}}{2\mu^{m}_{(t-1)k,i}\mu^{f}_{(t-1)k,i}}}{N_{(t-1)k} - \frac{\mu_{(t-1)k,i}(1-(1-\theta_{ti,k})(1-B_{ti,k|k}))}{2\mu^{m}_{(t-1)k,i}\mu^{f}_{(t-1)k,i}}}\\
&\approx \frac{\sigma^{xy}_{(t-1)k,ii} - \frac{\mu_{(t-1)k,i}\theta_{ti,k}}{2\mu^{m}_{(t-1)k,i}\mu^{f}_{(t-1)k,i}}}{N_{(t-1)k}},\\
p^{xy}_{tij,k} &\overset{i\neq j}{=} \frac{\sigma^{xy}_{(t-1)k,ij}}{N_{(t-1)k}},
\end{aligned}
\tag{31}
$$

where $B_{ti,k|k}$ is the fraction of all gametes in subpopulation i at time t with parental genes in subpopulation k that have their homologous gene originating from k as well, given that they reproduce through mating.

We mentioned in Section 2 that subpopulations k are sex-specific for dioecious models (Case IV). Then it is often not needed to distinguish coalescence probabilities $p^{xy}_{tij,k}$ for those gamete pairs xy that could have originated from k. If additionally $p^{xy}_{tij,k}$ only depends on the parental subpopulation k, not the subpopulations i and j of the two offspring gametes, we write

$$
p^{xy}_{tij,k} = \frac{1}{N_{e(t-1)k}} \quad \text{if } Q^{xy}_{tij,kk} > 0, \tag{32}
$$

where $N_{e(t-1)k}$ is a local effective size of k at time $t-1$. A sufficient condition for (32) to hold is that parents are chosen randomly among $N_{e(t-1)k} \leq N_{(t-1)k}$ breeders in subpopulation k at time $t-1$, whether they reproduce through mating, selfing or cloning, see the SM for a proof. On the other hand, we will find in Example 5 of Section 5 that (32) does not hold for a dioecious age-structured model.

It follows after some computations from (24), (32) and (26), that

$$
\begin{aligned}
D_{ti,k} &= \theta_{ti}B_{ti,k}\frac{1+c_{ti,k}}{2},\\
D_{ti,kl} &= (1-\theta_{ti})\tilde{B}^{m}_{ti,k}B^{f|m}_{ti,l|k},\\
D_{tij,kl} &= \frac{N_{ti}B_{ti,k}B_{tj,l} - \delta_{\{i=j,k=l\}}\frac{1}{2}B_{ti,k}(1+\theta_{ti,k}) - \delta_{\{i=j\}}(1-\theta_{ti})\frac{1}{2}\tilde{B}^{m}_{ti,k}B^{f|m}_{ti,l|k}}{N_{ti}-\delta_{\{i=j\}}}\\
&\quad\times\left[1-\frac{1}{N_{e(t-1)k}}\right]^{\delta_{\{k=l\}}},\\
D_{tij,k} &= \frac{N_{ti}B_{ti,k}B_{tj,k} - \delta_{\{i=j\}}\frac{1}{2}B_{ti,k}(1+\theta_{ti,k}) - \delta_{\{i=j\}}(1-\theta_{ti})\frac{1}{2}\tilde{B}^{m}_{ti,k}B^{f|m}_{ti,k|k}}{N_{ti}-\delta_{\{i=j\}}}\\
&\quad\cdot\frac{1}{2N_{e(t-1)k}}.
\end{aligned}
\tag{33}
$$

4. State space reduction

Hössjer et al. (2014) gave general conditions for reducing the state space of haploid models. It is possible to apply the same technique to diploid populations, and divide the d_t states of (3)–(4) into $\bar{d}_t$ groups or equivalence classes

$$
\mathcal{T}_t = \bigcup_{\alpha=1}^{\bar{d}_t}\mathcal{T}_{t\alpha},
$$

where $\mathcal{T}_{t\alpha}$ contains the states of class α of $\mathcal{T}_t$. Under certain conditions

$$
h_{ta} = \bar{h}_{t\alpha} \quad \text{for all } a \in \mathcal{T}_{t\alpha} \tag{34}
$$

holds for $\alpha = 1,\ldots,\bar{d}_t$ and $t = 0,1,2,\ldots$, with a recursion (5) that can be written in terms of the reduced column vectors $\bar{\boldsymbol{h}}_t = (\bar{h}_{t\alpha};\ \alpha = 1,\ldots,\bar{d}_t)'$ as

$$
\bar{\boldsymbol{h}}_t = \bar{\boldsymbol{D}}_t\bar{\boldsymbol{h}}_{t-1}, \quad t = 1,2,3,\ldots, \tag{35}
$$

for a matrix $\bar{\boldsymbol{D}}_t = (\bar{D}_{t\alpha,\beta})$ of dimension $\bar{d}_t \times \bar{d}_{t-1}$.

Suppose (34) is satisfied when $t = 0$. Then the crucial condition for (34)–(35) to hold for $t = 1,2,\ldots$ is that

$$
\bar{D}_{t\alpha,\beta} := \sum_{b\in\mathcal{T}_{(t-1)\beta}} D_{ta,b}, \quad \text{for all } a \in \mathcal{T}_{t\alpha}, \tag{36}
$$

and $1 \leq \alpha \leq \bar{d}_t,\ 1 \leq \beta \leq \bar{d}_{t-1}$.

In view of (7), it is possible to merge pairs

$$
\mathcal{T}_{t\alpha} = \{ij, ji\} \tag{37}
$$

of distinct subpopulations of any model, giving a reduced state space of size $\bar{d}_t = 2s_t - r_t + s_t(s_t-1)/2$. We have preferred to use the larger unreduced state space in Sections 2–3, since it makes some formulas simpler. For practical implementation, it is recommended though to apply any possible dimension reduction, at least when d_t is large. We will see in the next section that $\bar{d}_t$ can be chosen very small for models with inbuilt symmetries.

Laporte and Charlesworth (2003) use a strategy for dimensionality reduction where averages $\bar{h}_{t\alpha}$ of h_{ta} are computed over disjoint sets $\mathcal{T}_{t\alpha}$ of types. Such averaging involves some loss of information in general, except when (34) holds.

5. Examples of non-IBD recursions

Non-IBD recursions for pairs of genes are of central importance for computing the effective size of a population over different time horizons. In order to illustrate the diversity of our framework, we derive non-IBD recursions in this section for a homogeneous population, an island–continent model, island models, age-structured populations and pedigrees with background populations. We also give conditions under which non-IBD recursions are equivalent for certain haploid, diploid monoecious and dioecious populations.

Example 1 (*Homogeneous Population*)**.** In this first example we show that the IBD recursion of Wang (1996a) is a special case of Case II, for an isolated, monoecious and diploid population with an arbitrary amount of selfing, no cloning, and indistinguishable gametes. Since $s_t = 1$, we may drop any index referring to subpopulations. The population size varies over time, and the state space

$\mathcal{T}_t = \{1, 11\}$ at all time points if $N_t \geq 2$. The two states correspond to pairs of genes within (1) or between (11) individuals. Under these conditions, (6)–(12) reduce to the recursion

$$h_{t1} = \frac{\theta_t}{2} h_{(t-1)1} + (1-\theta_t) h_{(t-1)11},$$
$$h_{t11} = \frac{1}{2N_{e(t-1)}} h_{(t-1)1} + \left(1 - \frac{1}{N_{e(t-1)}}\right) h_{(t-1)11} \tag{38}$$

of Wang (1996a), with

$$p_t = \frac{1}{N_{e(t-1)}} = \frac{\frac{\mathrm{Var}(\xi_{(t-1)1})}{\mu_{t-1}} + \mu_{t-1} - 1 - \theta_t}{2(N_t - 1)}$$

as in (22), since $2N_t = \mu_{t-1} N_{t-1}$. In particular, the two recursions in (38) are identical when $\theta_t = 1/N_{e(t-1)}$, so that inbreeding within individuals is the same as coancestry between them ($f_{t1} = f_{t11}$). If $N_{e(t-1)} = N_{t-1}$ as well, we have random selfing (16), and the diploid model cannot be distinguished from a haploid one (Case I) of the same size. Mating between individuals need not be random, but for the breeders model above (23) with $N_{e(t-1)} = N_{t-1}$, it is. □

Example 2 (*Diploid and Haploid Models*)**.** We will generalize the findings of the previous example to subdivided populations. That is, we give conditions under which the non-IBD recursion of a haploid model (Case I) is equivalent to the corresponding non-IBD recursion for a diploid, monoecious and isogamous model (Case II) when cloning is absent, whereas selfing, mating and coalescence are all random. In order to distinguish the two recursions, we write h^{hapl}_{tij} for the non-IBD probability of a pair of distinct genes of the haploid model, picked from subpopulations i and j at time t, and h^{mono}_{ta} for the non-IBD probability of a gene pair $a \in \mathcal{T}_t$ of the diploid model.

It follows from (8)–(10) that

$$h^{\mathrm{hapl}}_{tij} = \sum_{k,l=1}^{s_t} B_{ti,k} \frac{2N_{tj} B_{tj,l} - \delta_{\{i=j,k=l\}}}{2N_{tj} - \delta_{\{i=j\}}} \times \left(1 - \frac{1}{2N_{e(t-1)k}}\right)^{\delta_{\{k=l\}}} h^{\mathrm{hapl}}_{(t-1)kl}. \tag{39}$$

Since the haploid model has fewer states than the diploid one, we will compare (39) with a certain state space reduced version of the diploid recursion. We use that the probability is $1/(2N_{ti} - 1)$ and 0 that a randomly chosen gene pair from i and j is from the same individual, when either $i = j$ or not. Therefore, the non-IBD probability of a randomly chosen gene pair of the diploid model is

$$\bar{h}_{tij} = \begin{cases} \frac{1}{2N_{ti}-1} h^{\mathrm{mono}}_{ti} + \frac{2N_{ti}-2}{2N_{ti}-1} h^{\mathrm{mono}}_{tii}, & i = j, \\ h^{\mathrm{mono}}_{tij}, & i \neq j. \end{cases} \tag{40}$$

In order to compare the haploid and diploid recursions, we will consider the following four conditions, as indicated above:

1. There is no cloning, i.e. $c_{tik} \equiv 0$,
2. Selfing is random (16),
3. Coalescence is random, so that (23) holds with $N_{e(t-1)k} = N_{(t-1)k}$.
4. Mating is random and after migration (15).

It is shown in the SM that (40) obeys the haploid recursion (39) if conditions 1–3 hold. Consequently, if $\bar{h}_{tij} = h^{\mathrm{hapl}}_{tij}$ when $t = 0$ and 1–3 are satisfied, then this identity will hold for $t = 1, 2, \ldots$ as well, regardless of whether mating takes place before migration or not.

If random mating after migration (condition 4) is added, it follows that i and ii can be merged into one single state ii in accordance with (34), with

$$\bar{h}_{tii} = h^{\mathrm{mono}}_{ti} = h^{\mathrm{mono}}_{tii}, \tag{41}$$

for all $t \geq 1$, regardless of the initial conditions on h_{0a}, see the SM for details. The haploid and diploid models of the present type are therefore equivalent under conditions 1–4. In Example 8 we motivate that (41) still holds if condition 2 is modified so that selfing is random among the breeders ($1/N_{e(t-1)k}$ on the right hand side of (16)), and condition 3 so that $N_{e(t-1)k} = N_{(t-1)k}$ is not required. □

Example 3 (*One and Two-Sex Island–Continent Model*)**.** In order to exemplify a dioecious model (Case IV) we consider a population with non-overlapping generations and 2 demes, of which the first is a small island, which is mostly isolated but occasionally receives one or more immigrants from the second deme, an infinitely sized continent without inbreeding, see Fig. 2 for an illustration. In the SM we compare non-IBD recursions for such a dioecious model without cloning and selfing, where mating within the island is random and occurs before migration, with a monoecious diploid model (Case II) that has no cloning but possibly selfing, and an effective number of individuals at the island that is the harmonic mean of the effective number of males and females. Finally, we consider a haploid model (Case I), where coancestry and inbreeding is not distinguished.

In the SM we show that 5 subpopulations are needed for the dioecious population in order to capture inbreeding and coancestry within/between males and females of the island. When sexes are not distinguished, it suffices with 2 subpopulations in order to describe inbreeding and coancestry of individuals from the island. We prove that these dioecious and diploid monoecious models are equivalent when the island is isolated. On the other hand, the two models may differ substantially when the island receives immigrants from the continent, and either the island population is very small, or the sex ratio differs between immigrants and the island.

The haploid model needs only 1 subpopulation in order to capture coancestry/inbreeding within the island. It follows from Example 2 that it is equivalent to the diploid monoecious model when coalescence and selfing is random. The two models may differ quite substantially, though, when the selfing rate is changed, see the SM for details. □

Example 4 (*Island Model with Sex of Gametes Distinguished or Not*)**.** In order to show that Cases II and III may differ, we consider the island model (Wright, 1943; Maruyama, 1970), a population for which all demes are treated symmetrically. The census sizes $N_{ti} = N \geq 2$ are the same for all $s_t = s \geq 2$ demes at all generations, and backward migration

$$B_{ti,k} = B_{i,k} = \begin{cases} 1 - M, & k = i, \\ M/(s-1), & k \neq i, \end{cases} \tag{42}$$

is also time invariant and the same between all pairs of different demes. The migration rate M quantifies the proportion of individuals whose parents originate from another deme. We assume that (23) holds with $N_{eti} = N_e$ for all demes and generations. This number N_e is the constant local effective size that each deme would have if it was isolated. We further assume a non-mating probability $\theta_{ti} = \theta$ in (1) that is independent of generation t and deme i, with a cloning rate $c_{tik} \equiv c$ between all pairs of demes i and k, so that time index can be dropped.

It is possible to reduce the state space from $d = 2s + s(s-1)$ to $\bar{d} = 3$ components

$$\mathcal{T}_1 = \mathcal{I},$$
$$\mathcal{T}_2 = \mathcal{S}, \tag{43}$$
$$\mathcal{T}_3 = \mathcal{D},$$

corresponding to inbreeding, coancestry of individuals within the same deme, and coancestry of individuals of different demes.

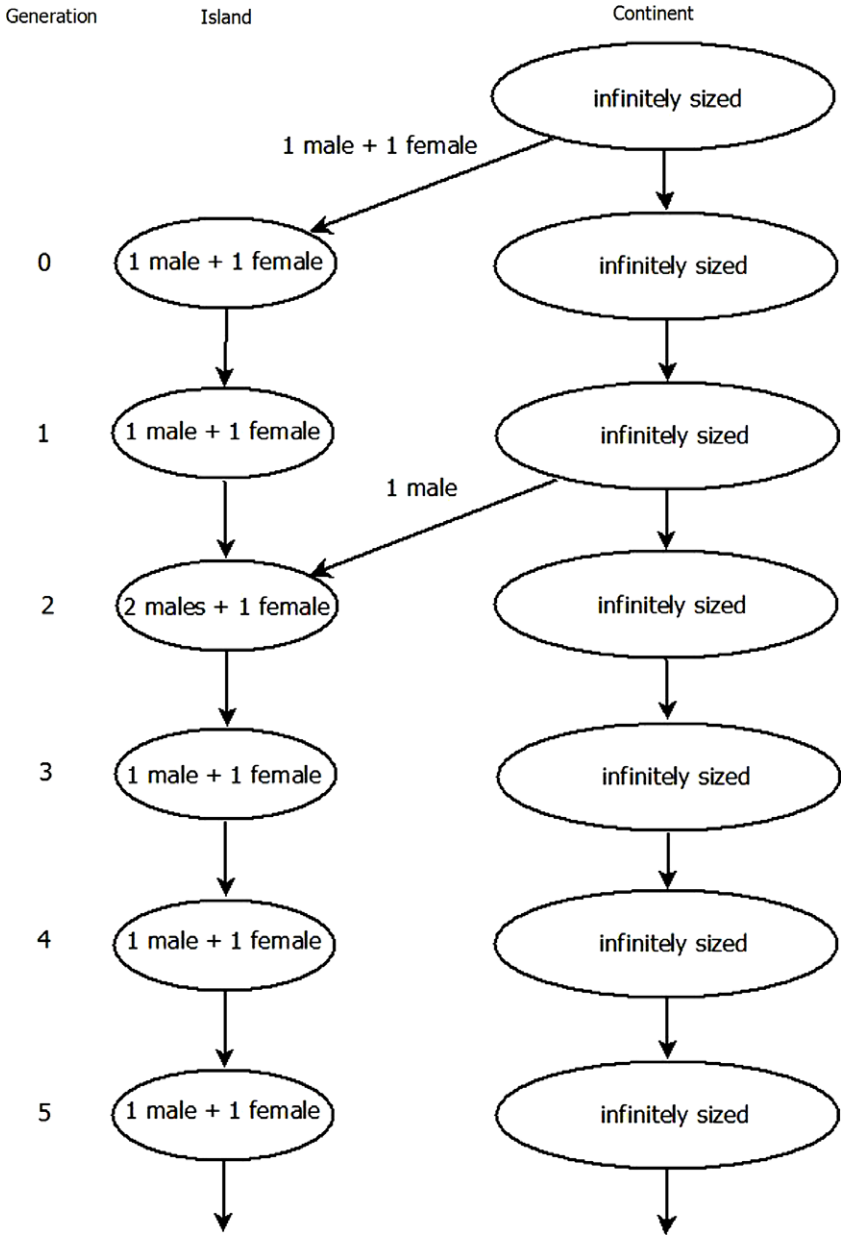

Fig. 2. Illustration of Scenario I (of Tables S1 and S2 in the SM) for an island–continent model with non-overlapping generations (Case IV) and one couple immigrating to the island in generation 0. The island remains with one male and one female in the following generations, except for an additional male that immigrates at $t = 2$. Scenario II differs in that nine males and one female immigrates from the continent at $t = 0$. Then the island remains with nine males and one female, except for an additional couple of one male and one female that immigrates at $t = 2$. See text and Tables S1 and S2 of the SM.

In the SM we compare the time recursion (35) of $\bar{\boldsymbol{h}}_t = (\bar{h}_{t1}, \bar{h}_{t2}, \bar{h}_{t3})'$ for two models. The first of these has isogamous individuals (Case II) that reproduce before they migrate. The second model is for plants with unknown sex, but with females gametes distinguishable from pollen (Case III). Only pollen migrate, and reproduction is different between male and female gametes as well. It is shown how these different migration and reproduction features affect the two recursions. □

Example 5 (*Separate Sexes with Overlapping Generations*)**.** We will show how a two-sex population with overlapping generations can be incorporated into Case IV. We drop time index, since all parameters are assumed to be time invariant. The dioecious population has L age classes and $s = 2L$ subpopulations, with subpopulation $i = 1, \ldots, L$ corresponding to age class i among males, and subpopulation $i = L+1, \ldots, s$ to age class $i-L$ among females. The newborns in subpopulations $i = 1$ and $L+1$ are produced through mating, so that $\theta_1 = \theta_{L+1} = 0$, and any adult age class i through survival from subpopulation $i-1$, so that $\theta_i = \theta_{i,i-1} = c_{i,i-1} = 1$. Let $\mu_{k,i}$ be the expected number of gametes transmitted by individuals of subpopulation k to subpopulation i of the next time point. Then $\mu_{k,1}$ ($\mu_{k,L+1}$) is the expected number of matings by an individual of subpopulation k that result in a newborn male (female), $\mu_{k,k+1}$ is twice the survival probability of individuals in k and all other $\mu_{k,i}$ equal to 0. Constant subpopulation sizes over time requires $2N_i = \sum_{k=1}^{s} N_k \mu_{k,i}$ for $i = 1, \ldots, s$. The backward migration rates (26) are

$$B_{i,k} = \begin{cases} \dfrac{N_k \mu_{k,i}}{2N_i}, & i \in \{1, L+1\}, \\ \delta_{\{k=i-1\}}, & i \notin \{1, L+1\}. \end{cases} \tag{44}$$

By keeping track of which subpopulations that are male and female, we obtain sex specific forward and backward migration rates

$$\mu^m_{k,i} = \begin{cases} \mu_{k,i}, & k \in \{1,\dots,L\},\ i \in \{1, L+1\}, \\ \frac{1}{2}\mu_{k,i}, & k \notin \{L, s\},\ i = k+1, \\ 0, & \text{otherwise}, \end{cases} \tag{45}$$

$$B^m_{i,k} = \begin{cases} \delta_{\{k=i-1\}}, & i \notin \{1, L+1\}, \\ \delta_{\{k\le L\}} \dfrac{N_k \mu_{k,i}}{N_i}, & i \in \{1, L+1\}, \end{cases}$$

for males, and

$$\mu^f_{k,i} = \begin{cases} \mu_{k,i}, & k \in \{L+1,\dots,s\},\ i \in \{1, L+1\}, \\ \frac{1}{2}\mu_{k,i}, & k \notin \{L, s\},\ i = k+1, \\ 0, & \text{otherwise}. \end{cases} \tag{46}$$

$$B^f_{i,k} = \begin{cases} \delta_{\{k=i-1\}}, & i \notin \{1, L+1\}, \\ \delta_{\{k\ge L+1\}} \dfrac{N_k \mu_{k,i}}{N_i}, & i \in \{1, L+1\} \end{cases}$$

for females.

For simplicity we assume that $N_{ek} \le N_k$ individuals act as breeders in each subpopulation, independently of whether they survive or not. The coalescence probabilities $p^{xy}_{ij,k}$ are only well defined when it is possible for an x gamete from i and a y gamete from j to have been transmitted from the same individual of k the time point before. It is shown in the SM that the well defined coalescence probabilities equal $1/N_{ek}$ for two offspring in $i, j \in \{1, L+1\}$ to have the same parent from k, $1/N_k$ for a progeny in $i \in \{1, L+1\}$ to have a certain adult in age class $j = k+1$ as parent in age class k the previous time point, and 0 for two different adults of the same age group $i = j = k+1$ to have been the same individual in age class k the time point before. This implies that N_{ek} is *not* a local effective size of k, since (32) is violated.

The mating function $B^{f|m}_{i,l|k}$ for newborns $i \in \{1, L+1\}$ is the probability for males of age $k \in \{1,\dots,L\}$ to mate a female of age $l - L \in \{1,\dots,L\}$. Then

$$B^{f|m}_{i,l|k} = \begin{cases} B^f_{i,l}, & \text{mating independently of age}, \\ \delta_{\{l=k+L\}}, & \text{same age mating}, \end{cases}$$

are two opposite scenarios where males choose a female either regardless of age or only of their own age. The latter is a kind of assortative mating which requires $B^m_{i,k} = B^f_{i,k+L}$ for $k = 1,\dots,L$. By symmetry, we may also write mating as probabilities $B^{m|f}_{i,l|k}$ of females to choose males of different age. It is assumed in any case the spouse is picked randomly and independently for each mating among the breeders within the chosen age group. A more sophisticated model where each individual has a maximum number of spouses (Balloux and Lehmann, 2003) requires subpopulations in terms of single individuals.

After this model specification, we obtain the non-inbreeding recursion (6) by inserting (44)–(46) and the coalescence probabilities into (24) and (29). □

Example 6 (*Pedigree with Background Populations*)**.** We will show how to make inbreeding calculations of a pedigree that is influenced by a background population. The whole metapopulation belongs to Case IV and has T generations $t = 0, 1, \dots, T-1$. In each generation t there are r_t individuals of known sex that belong to the pedigree, each of which represents a subpopulation i with $N_{ti} = 1$. If at least one individual of a non-founder generation ($t > 0$) is an immigrant with unknown parents, we add one male and one female background population for generations $t = 0, 1, \dots, \tau-1$, where $0 \le \tau \le T-1$ is the generation of the youngest immigrant. The number of subpopulations in each generation is then

$$s_t = \begin{cases} r_t + 2, & t = 0,\dots,\tau-1, \\ r_t, & t = \tau,\dots,T-1. \end{cases}$$

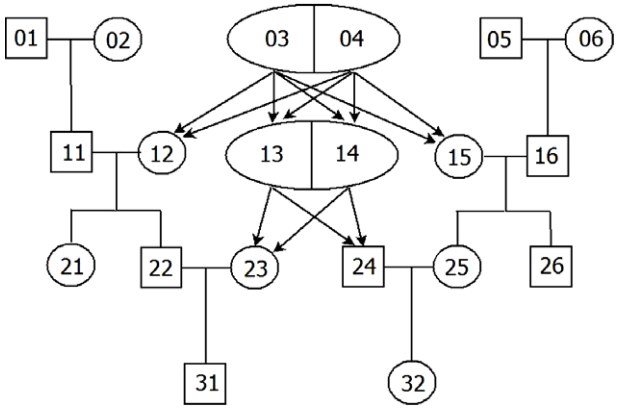

Fig. 3. Four generation pedigree with immigration from a non-pedigreed, background population displayed over two generations (ellipses) with a male (to the left; 03 and 13) and female (to the right; 04, 14) segment, respectively. The pedigree has sixteen individuals, with males and females depicted as squares and circles, respectively. Three of the females (12, 15, 23) and one of the males (24) are immigrants from the background population. Notations are as in the text, with $T = 4$, $\tau = 2$, $r_0 = r_1 = 4$, $r_2 = 6$, $r_3 = 2$, $s_0 = s_1 = s_2 = 6$ and $s_3 = 2$.

Write ti to denote subpopulation number i within generation t, and let bp refer to all subpopulations ti that represent a (male or female) background population. It is assumed that each $ti \in$ bp has $2 \le N_{ti} \le \infty$ individuals of which $2 \le N_{eti} \le N_{ti}$ are breeders, so that (32)–(33) hold. All remaining subpopulations $ti \notin$ bp are single individuals with coalescence probabilities 1, so that $N_{eti} = 1$.

We assume there is no cloning/selfing ($\theta_{ti} \equiv 0$), so that all individuals in the pedigree have two parents. For each subpopulation ti of a non-founder generation $t > 0$, we let $m_t(i)$ and $f_t(i)$ refer to the male and female subpopulations of generation $t-1$ to which the two parents belong. Depending on whether a single individual ti is an immigrant or not, either both or none of $(t-1)m_t(i)$ and $(t-1)f_t(i)$ are background populations. Fig. 3 shows a hypothetical four generation pedigree with sixteen individuals, of which four are immigrants. The remaining four subpopulations represent the continent of the first two generations, with one male and one female background population per generation. It is inspired by the Swedish wild wolf population, which has been isolated for a long time, with a known pedigree. It occasionally received immigrants from the continent—the Finnish wolf population and populations further east (Laikre et al., 2013).

In the SM we use two different methods to compute the non-IBD probabilities $\bar{h}_{t\alpha} = 1 - \bar{f}_{t\alpha}$ for all types α of gene pairs of a reduced state space (37) that does not keep track of the order of different subpopulations. The first method uses recursion (35) for the whole vector $\bar{\boldsymbol{h}}_t$ of non-IBD probabilities. The second method extends path analysis of Wright (1951, 1965) to situations where founders are possibly related and the impact of background populations is accounted for. We illustrate the latter approach by computing the inbreeding coefficient $\bar{f}_{31}$ of individual $ti = 31$ in Fig. 3. Its two parents have no known common ancestors, but the mother and the grandmother on the father's side are both immigrants from the continent, so there are two paths having the male background population 03 of the founder generation as "common ancestor", and similarly for the other female background population 04. Going through all possible lineages from the mother and father of 31

to the founder generation, one obtains

$$\bar f_{31} = \frac{1}{16}\cdot\frac{1+\bar f_{03}}{N_{e03}} + \frac{1}{16}\cdot\frac{1+\bar f_{04}}{N_{e04}} + \frac{1}{8}\left(1-\frac{1}{N_{e03}}\right)\bar f_{033} + \frac{1}{8}\left(1-\frac{1}{N_{e04}}\right)\bar f_{044} + \frac{1}{8}\bar f_{013} + \frac{1}{8}\bar f_{014} + \frac{1}{8}\bar f_{023} + \frac{1}{8}\bar f_{024},$$

where, for instance, $\bar f_{03}$ and N_{e03} are the inbreeding coefficient and local effective size of background population 03, and $\bar f_{013}$ is the coancestry coefficient between individual 01 and background population 03. Suppose for instance that all individuals of the founder generation are unrelated ($\bar f_{0ij} = 0$ for all $i \neq j$, $\bar f_{033} = \bar f_{044} = 0$) and that those from a background population are not inbred ($\bar f_{03} = \bar f_{04} = 0$). Then the inbreeding coefficient of 31 ranges between $\bar f_{31} = 0$ and $\bar f_{31} = 1/16 + 1/16 = 1/8$, depending on whether the background population has no genetic drift ($N_{e03} = N_{e04} = \infty$) or maximal genetic drift caused by one male and one female breeder ($N_{e03} = N_{e04} = 1$).

Since the number of terms in path analysis grows rapidly with number of generations, it is often computationally more efficient to find inbreeding coefficients by means of recursion (35). A similar approach has been used by Wakeley et al. (2012) to compute coalescence probabilities on pedigrees. Our model is an extension, with background populations included. See also Cannings et al. (1978) and Koski and Noble (2011) for more general probability recursions on pedigrees and networks. □

6. Effective population size

In the following two sections we assume a diploid population with a fixed number $s = s_t$ of subpopulations of size $N_{ti} \geq 2$, so that the spate space $\mathcal{T}_t = \mathcal{T}$ in (3) is time invariant. In order to define effective size, we will assign weights W_{ta} to all $a \in \mathcal{T}$. We may think of sampling two genes at random from the population at time t, with W_{ta} the probability that the gene pair has type a, so that

$$\sum_{a\in\mathcal{T}} W_{ta} = 1.$$

In this section we assume that the two genes are drawn without replacement, since this is the most common approach for the inbreeding and other types of effective sizes we consider, but in the next section we assume they are drawn with replacement, since this is needed for the measures of subpopulation differentiation that we use.

Let $\boldsymbol{W}_t = (W_{ta};\ a \in \mathcal{T})$ be the row vector of sampling probabilities, and

$$h_t = \sum_{a\in\mathcal{T}} W_{ta}h_{ta} = \boldsymbol{W}_t\boldsymbol{h}_t = \boldsymbol{W}_t\boldsymbol{D}_t\cdot\cdots\cdot\boldsymbol{D}_1\boldsymbol{h}_0 = \bar{\boldsymbol{W}}_t\bar{\boldsymbol{h}}_t \tag{47}$$

the probability, for sampling scheme $\boldsymbol{W}_t$, that a randomly chosen pair of genes from time t is not IBD, given information available at time 0. The right hand side of (47) contains the state space reduced vector (cf. Section 4) of non-IBD probabilities, and $\bar{\boldsymbol{W}}_t = (\bar W_{t\alpha};\ \alpha = 1,\dots,\bar d)$ is the corresponding weight vector, with elements $\bar W_{t\alpha} = \sum_{a\in\mathcal{T}_\alpha} W_{ta}$.

As in Hössjer et al. (2014), define the effective size over time interval $[t_1, t_2]$ as

$$N_e([t_1,t_2]) = \begin{cases} \dfrac{1}{2\left(1-(h_{t_2}/h_{t_1})^{1/(t_2-t_1)}\right)}, & \text{if } h_{t_2} < h_{t_1},\\ \text{NaN}, & \text{if } h_{t_2} \geq h_{t_1}.\end{cases} \tag{48}$$

Notice that N_e is undefined (written as NaN, Not a Number) when the non-IBD probability h_t does not decrease over the time interval $[t_1, t_2]$. We will see below that (48) depends on the subpopulation weighting scheme, the migration pattern between subpopulations and how much reproduction varies within each subpopulation that has a positive weight. Depending on how subpopulation weights are chosen, (48) incorporates both local and global effective sizes. It was argued in Hössjer et al. (2014), that for subdivided populations (particularly systems with low levels of migration) one should not report one single effective size, but rather compute N_e for several time intervals of varying length. At one extreme, $N_e[t_1, t_1+1]$ is the instantaneous effective size at time t_1. At the other extreme, if migration is possible back and forth between all pairs of subpopulations in some number of time steps, and if population characteristics vary cyclically with period τ, the long term limit

$$\lim_{t_2\to\infty} N_e([t_1,t_2]) = N_{eE} = \frac{1}{2(1-\lambda_{\max}(\boldsymbol{D}_\tau\cdot\cdots\cdot\boldsymbol{D}_1)^{1/\tau})} \tag{49}$$

of the effective size exists, with $\lambda_{\max}(\cdot)$ the largest positive eigenvalue of a non-negative and irreducible matrix. The limit in (49) coincides with the eigenvalue effective size N_{eE} of Crow (1954) and Ewens (1982), see Whitlock and Barton (1997), Pollak (2002), Hössjer et al. (2014) and Hössjer (in press) for details.

The sampling scheme $\boldsymbol{W}_t$ gives great flexibility in defining various types of effective size. We will define a large class of such schemes (see Table 2), which all require that non-negative weights $w_1,\dots,w_s$ with $\sum_{i=1}^s w_i = 1$ are assigned to all subpopulations. These weights may be equal, proportional to size or to the long term reproductivity of subpopulations. They may also be local, so that a single subpopulation or a group of subpopulations has a total weight of one.

The first sampling scheme T chooses randomly a pair of genes from the total set of subpopulations with positive weights w_i. This can be conceptualized as an urn, where the genes of subpopulation i have sampling probabilities proportional to $w_i/(2N_{ti})$. When the first gene is not put back into the urn, this gives a weight vector $\boldsymbol{W}_{Tt} = (W_{Tta};\ a \in \mathcal{T})$ of dimension $1\times(s+s^2)$ whose elements

$$W_{Tta} = \begin{cases} w_i^2/\{2N_{ti}[1-w_i/(2N_{ti})]\}, & a = i,\\ w_i^2(1-1/N_{ti})/[1-w_i/(2N_{ti})], & a = ii,\\ w_iw_j/[1-w_i/(2N_{ti})], & a = ij,\ i\neq j,\end{cases} \tag{50}$$

are probabilities of choosing all possible types a of gene pairs. The probability of choosing $a = ij$ when $i \neq j$ is a product of w_i, the probability that the first gene is chosen from i, and $w_j/[1-w_i/(2N_{ti})]$, the probability that the second gene is chosen from j when the first one from i has been removed. The other two equations of (50) are derived similarly. A slight drawback of this scheme is asymmetry for $a = ij$ and $a = ji$. That is, if w_i and w_j are both positive and $N_{ti} > N_{tj}$ we have $W_{Ttij} < W_{Ttji}$, although the difference is minor for large populations. A symmetrized version of scheme T appears in the next section.

In order to illustrate state space reduction from (50), consider the symmetric island model of Example 4. Recall that $N_{ti} = N$ and only the $\bar d = 3$ components of (43) are needed. The sampling probabilities in (50) correspond to drawing two genes from the entire population without replacement, when uniform weights $w_i = 1/s$ are used. Since they do not depend on time we remove time index t, use the definition of $\bar W_{Tt\alpha} = \bar W_{T\alpha}$ above (48) and write

$$\begin{aligned}\bar W_{T1} &= 1/\{2sN[1-1/(2sN)]\},\\ \bar W_{T2} &= (1-1/N)/\{s[1-1/(2sN)]\},\\ \bar W_{T3} &= (1-1/s)/[1-1/(2sN)]\end{aligned} \tag{51}$$

for the probabilities of sampling two genes from the same individual, from different individuals within the same island, and from different islands respectively.

If only coancestry of individuals from different subpopulations of a general population is of concern, and at least two w_i are strictly

Table 2
Different schemes for sampling a pair of genes from the population.

Scheme	Pair of genes drawn	Application	Effect of replacement?
T	From the *T*otal set of subpopulations with positive weights w_i.	N_{eV}, haploid N_{eI}, g_{ST}, f_{IT}	Yes/Yes
D	From *D*ifferent subpopulations.	N_{eD}	No/No
S	From the *S*ame subpopulation.	g_{ST}, f_{IS}	Yes/Yes
S_d	From the *S*ame subpopulation, but *d*ifferent individuals.	N_{eS_d}	No/No
I	From the same *I*ndividual.	Diploid N_{eI}, f_{IT}, f_{IS}	Yes/No

For any scheme, the two genes can be sampled with or without replacement. The global or local population sizes refer to variance (N_{eV}), inbreeding (N_{eI}) and coancestry (N_{eS_d}, N_{eD}) effective sizes. The last column refers to whether drawing the two genes with our without replacement has any effect on the scheme itself and its weight vector $\boldsymbol{W}_t$.

positive, we get a second sampling scheme D with a weight vector $\boldsymbol{W}_D = (W_{Da};\ a \in \mathcal{T})$, where

$$W_{Da} = \begin{cases} 0, & a = i, \\ 0, & a = ii, \\ w_i w_j/(1 - w_1^2 - \cdots - w_s^2), & a = ij,\ i \neq j. \end{cases}$$

The third sampling scheme S_d is for coancestry of different individuals from the same subpopulation. Its weight vector $\boldsymbol{W}_{S_d} = (W_{S_d a};\ a \in \mathcal{T})$ has elements

$$W_{S_d a} = \begin{cases} 0, & a = i, \\ w_i, & a = ii, \\ 0, & a = ij,\ i \neq j. \end{cases} \tag{52}$$

When inbreeding within individuals is of interest, we consider a fourth scheme I, where first a subpopulation i is drawn with probability w_i, then two different genes are picked from the same, randomly chosen individual within this subpopulation. This gives a weight vector $\boldsymbol{W}_I = (W_{Ia};\ a \in \mathcal{T})$ with components

$$W_{Ia} = \begin{cases} w_i, & a = i, \\ 0, & a = ii, \\ 0, & a = ij,\ i \neq j. \end{cases} \tag{53}$$

Writing $\boldsymbol{W}_t$ for the weight vector associated with any of the four schemes $\boldsymbol{W}_{Tt}$, $\boldsymbol{W}_D$, $\boldsymbol{W}_{S_d}$ or $\boldsymbol{W}_I$, we get different types of effective sizes (48). Similarly as in Wang (1997a,b) we refer to $N_e = N_{eI}$ as an inbreeding effective size for weighting scheme I, a coancestry effective size $N_e = N_{eS_d}$ within subpopulations for sampling scheme S_d, and a coancestry effective size $N_e = N_{eD}$ between subpopulations for sampling scheme D. With this definition, $N_{eI}([t_1, t_2])$ quantifies the rate of increased inbreeding within *individuals* over $[t_1, t_2]$. It is important to emphasize that this is a diploid quantity, and not the same as $N_{eI}([t_1, t_2])$ for a haploid model (Case I). Although the haploid $N_{eI}([t_1, t_2])$ is defined as in (47)–(48), the non-IBD probabilities h_{ta} for a haploid model cannot distinguish between pairs of genes drawn within or between individuals, as discussed in Section 3.1. The haploid N_{eI} for a structured population is usually defined by drawing two genes randomly from the whole population, so that $w_i \equiv 1/s$.

For the haploid symmetric island model (see Section 4.6.3 of Durrett, 2008, and references therein) we can merge the first two states of (43) and only distinguish whether two genes are drawn from the same ($\mathcal{T}_1^{\text{hapl}} = \mathcal{A} \cup \mathcal{S}$) or different ($\mathcal{T}_2^{\text{hapl}} = \mathcal{D}$) islands. Uniform sampling without replacement from the entire population is then equivalent to sampling scheme in (51), provided that its first two equations are merged. The resulting sampling probabilities are

$$\bar{W}_{T1} = (1/s)[1 - 1/(2N)]/[1 - 1/(2sN)],$$
$$\bar{W}_{T2} = (1 - 1/s)/[1 - 1/(2sN)]$$

for two genes drawn from the same and different islands. See also Hössjer et al. (2014) for weighting schemes of more general haploid models.

The original definition of N_e in Wright (1931) is an inbreeding effective size of a diploid, monoecious, isogamous and homogeneous ($s = 1$) population over one generation $[t, t+1]$, which does not relate to changed inbreeding within individuals. It is more related to a haploid than to a diploid $N_{eI}([t, t+1])$, see Hössjer et al. (2014). It was also shown in that paper that the variance effective size N_{eV} corresponds to (48) with weights W_{ta} obtained from a version (56) of sampling scheme T where genes are drawn with replacement. It follows that N_{eV} is more closely related to a haploid than to a diploid N_{eI}.

Example 7 (*Homogeneous Populations*)**.** We will compute the eigenvalue effective size of the homogeneous population of Example 1 (Case II), when the coalescence probability $p = 1/N_e$, selfing rate θ and cloning fraction c are time invariant. The non-IBD probabilities satisfy a recursion (5), with

$$\boldsymbol{D}_t = \boldsymbol{D} = \begin{pmatrix} \frac{1+c}{2} \cdot \theta & 1 - \theta \\ \frac{1}{2N_e} & 1 - \frac{1}{N_e} \end{pmatrix}.$$

The largest eigenvalue of $\boldsymbol{D}$ can be found by solving a quadratic equation, and it follows from (49) (with period $\tau = 1$) that

$$N_{eE} = \frac{1}{1 + \frac{1}{N_e} - \frac{(1+c)\theta}{2} - \sqrt{\left(1 - \frac{1}{N_e} - \frac{(1+c)\theta}{2}\right)^2 + \frac{2(1-\theta)}{N_e}}}$$
$$= \begin{cases} N_e \dfrac{1 - \frac{1+c}{2}\theta}{1 - c\theta} + o(N_e), & \theta \gg N_e^{-1}, \\ N_e, & \theta = N_e^{-1},\ c = 0, \\ N_e + \dfrac{1}{2} + o(1), & \theta = 0, \end{cases} \tag{54}$$

where the first and third equations on the right hand side of (54) are asymptotic as $N_e \to \infty$. It is well known that the coalescence effective size is $N_e(1 - \theta/2)$ in the absence of cloning (Pollak, 1987; Nordborg and Donnelly, 1997), and the first equation of (54) generalizes this to arbitrary cloning rates, since the eigenvalue and coalescence effective sizes are asymptotically equivalent for a large homogeneous population Hössjer (in press). The second equation assumes random selfing. The population then behaves as a haploid Wright Fisher model with effective size N_e, in accordance with Example 2. Finally, it is well known that the absence of selfing adds 1/2 to the effective size, see for instance Caballero (1994), Balloux (2004) and Waples (2010). □

Example 8 (*Local Effective Size*)**.** We will give sufficient conditions under which the effective size $N_e([t-1, t])$ in (48) agrees with the local effective size $N_{e(t-1)i}$ of a certain subpopulation i at time $t - 1$, derived from a coalescence probability that satisfies (23). Our framework is a diploid and isogamous population with s subpopulations (Case II). In the SM we prove that

$$N_{eX}([t-1, t]) = N_{e(t-1)i}, \tag{55}$$

for any of the three sampling schemes $X \in \{T, S_d, I\}$ in (50), (52), (53), under the following conditions: (i) Subpopulation i is isolated over interval $[t-1, t]$, so that $B_{ti,i} = 1$, and (ii) local weights

($w_i = 1$) for i are used. The first thing to notice is that (55) does not hold if (i) fails and i receives immigrants from at least one other subpopulation. Inbreeding/coancestry within i at time t is then not only affected by genetic drift within i, but also by migration from other subpopulations, whose levels of inbreeding/coancestry may differ from that of i. But in order for (55) to hold, we must also require (iii) no cloning, random selfing among breeders ((16) holds with $1/N_{e(t-1)k}$ on the right hand side) and random mating after migration (15). This condition (iii) implies that the population can be reduced to a haploid one, see the SM for details. □

Example 9 (*Two Subpopulations with Asymmetric Migration*)**.** We will plot $t \to N_{el}([0,t])$ for the monoecious, isogamous and diploid population of Fig. 4 with non-overlapping generations (Case II), and four migration scenarios ((a)–(d)). The system has time invariant parameters, so that index t is dropped, and two subpopulations of sizes $N_1 = 200$ and $N_2 = 20$. The reproduction cycle has no cloning/selfing, and the local effective size of each subpopulation is $N_{ei} = N_i$ in (23). Individuals mate before they migrate, with n_1 of the offspring migrating from the larger and n_2 from the smaller subpopulation per generation.

Fig. 5 shows time profiles $t \to N_{el}([0,t])$ of the inbreeding effective size, and the constant eigenvalue effective size N_{eE}, for all four migration scenarios n_1, n_2 when there is no inbreeding or coancestry in generation 0 ($f_{0a} = 0$ for all a). In the upper left subplot (a) the two local subpopulations are isolated, and the two solid curves depict $N_{el}([0,t])$ for each subpopulation i (using weight $w_i = 1$). Since there is no coancestry at $t = 0$ and no selfing, there will be no inbreeding within individuals at time $t = 1$, and therefore both local curves start with $N_{el}([0,1]) = \infty$. Then they converge towards their asymptotic limits $N_i + 0.5$ for subpopulations $i = 1, 2$, in agreement with (54). The corresponding local curves $N_{el}([1,t])$ that start at $t = 1$ have no transient effects. They are essentially constant for each subpopulation, since there is some coancestry between individuals at $t = 1$. For the global population, the dashed $N_{el}([0,t])$ curve in subplot (a) has both subpopulations weighted equally ($w_1 = w_2 = 0.5$). This global curve converges extremely slowly towards a value close to $N_1 + 0.5 = 200.5$, with $N_{el}([0,500]) = 129.2$ and $N_{el}([0,1000]) = 157.1$. This limit (200.5), which may be of more mathematical than biological interest, is determined by the larger subpopulation 1 since its rate of increased inbreeding is much slower than for the smaller subpopulation 2, and therefore 1 determines the long term rate at which the non-IBD probability h_t tends to zero. By a similar argument, it can be shown that the global coancestry effective size curve $N_{eS_d}([0,t])$ has the same limit close to 200.5 when both subpopulations are weighted equally. On the other hand, the coancestry effective size curve $N_{eD}([0,t])$ between the two subpopulations will be infinite. The reason is that different alleles will become fixed in the two subpopulations, causing the non-IBD probabilities $h_{t12} = 1$ between the two subpopulations remain at their initial $t = 0$ value. Since the population is also demographically constant with period $\tau = 1$, the largest eigenvalue $\lambda_{\max}(\boldsymbol{D}_1)$ in (49) is 1, so that the eigenvalue effective size of the whole population is $N_{eE} = \infty$. Since the two subpopulations are isolated, the long term ($t \to \infty$) limit of the effective size $N_e([0,t])$ does not necessarily equal N_{eE}, as the first part of (49) stipulates. In this example (49) only holds for $N_e = N_{eD}$, but not for $N_e = N_{el}$ or $N_e = N_{eS_d}$.

In the upper right plot (b), the smaller subpopulation (N_2) is a source, with one emigrant per generation. In this case the limit of the local $N_{el}([0,t])$ curve of the source population is close to 20.5. This agrees with (54), since inbreeding within 2 will vary over time as for an isolated population of size 20. The limit of the local $N_{el}([0,t])$ curve of the larger sink population and the limiting global $N_{el}([0,t])$ for the two populations combined are both $N_{eE} = 100$. This value of N_{eE} agrees with a more general formula $N_{eE} = \max(N_1/(2n_2), N_2)$ derived in Hössjer (in press) for a haploid model with two subpopulations of which the second is a source ($n_1 = 0$) and the number of migrants from 2 to 1 varies binomially around n_2. In the lower left subplot (c), the larger population is the source, with one migrant per generation, and all three $N_{el}([0,t])$ curves converge towards $N_{eE} = 200.5$, which is close to the analogous haploid formula $N_{eE} = \max(N_1, N_2/(2n_1))$. Finally, in the lower right subplot (d) both populations exchange one migrant per generation. Similarly to the lower left subplot, all three $N_{el}([0,t])$ curves converge to $N_{eE} = 225.0$, in agreement with (49). □

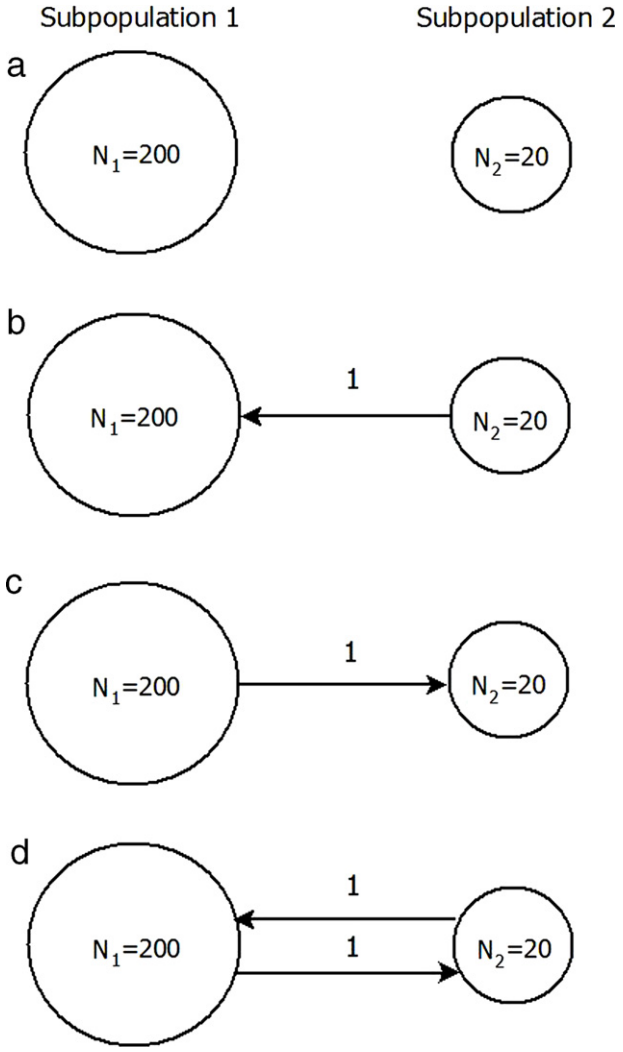

Fig. 4. Four migration scenarios (a)–(d) for a diploid and monoecious population with two subpopulations (1 and 2) of sizes $N_1 = 200$ and $N_2 = 20$, and non-overlapping generations (Case II with no selfing). The arrows illustrate the number of migrants n_i per generation from each subpopulation i, and the four scenarios correspond to (a) complete isolation ($n_1 = n_2 = 0$), (b) the small population being a source population ($n_1 = 0, n_2 = 1$), (c) the large population being a source population ($n_1 = 1, n_2 = 0$), and (d) both populations exchanging migrants ($n_1 = n_2 = 1$).

7. Subpopulation differentiation

Multiallelic versions of fixation indices are defined in terms of allele and genotype frequencies (Nei, 1975), which facilitates their estimation from data. We will focus in this section on IBS-sharing, since it is closely connected to allele frequencies. As in the previous section we let W_{ta} be the probability that a sampled gene pair is of type a, but in order to express non-IBS probabilities more easily in terms of allele frequencies, it is assumed that genes are drawn *with*

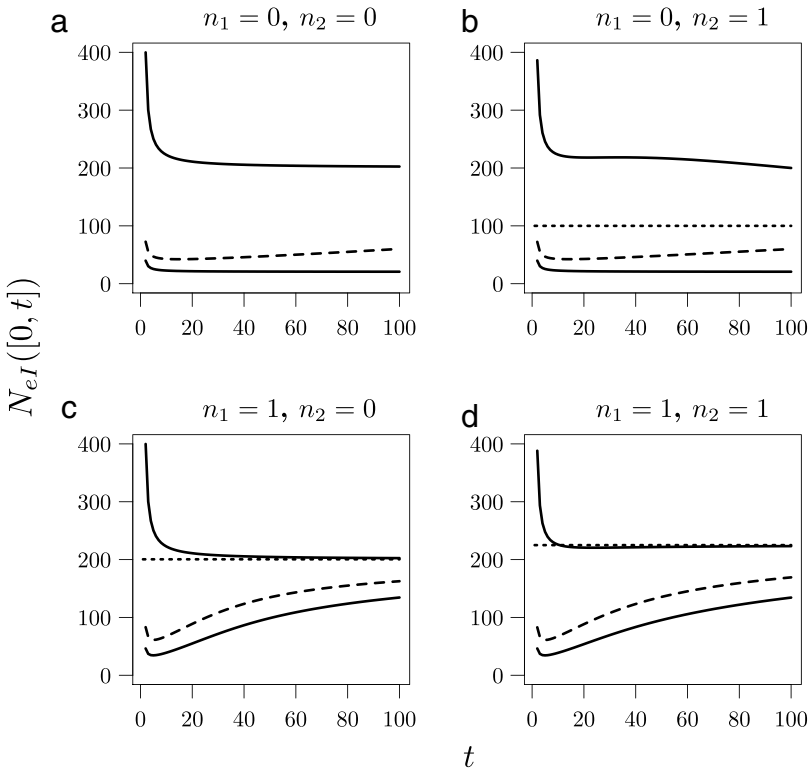

Fig. 5. Plots of inbreeding effective size $N_{eI}([0,t])$ versus time t for the population of Fig. 4, i.e. a diploid and monoecious system with non-overlapping generations and two subpopulations of sizes $N_1 = 200$ and $N_2 = 20$, with n_i individuals migrating from i in each generation (Case II without selfing). The two solid curves are N_{eI} curves for each subpopulation i (using weights $w_i = 1$), and the dashed curves depict N_{eI} for both subpopulations combined (based on weights $w_1 = w_2 = 1/2$). There is no inbreeding or cloning, mating (14) is before migration, the local effective sizes in (23) are $N_{ei} = N_i$, and there is no inbreeding at time $t = 0$ ($f_{0a} = 0$ for all $a \in \mathcal{T}$). The dotted horizontal line shows N_{eE}, except in the upper left subplot, where $N_{eE} = \infty$.

replacement. We consider three sampling schemes. The sampling probabilities

$$W_{Tta} = \begin{cases} w_i^2/N_{ti}, & a = i, \\ w_i^2(1 - 1/N_{ti}), & a = ii, \\ w_i w_j, & a = ij,\ i \neq j \end{cases} \tag{56}$$

of the first scheme T are modified slightly compared to (50), since genes are drawn with replacement. The second subpopulation based sampling scheme S has the subpopulation i picked at first with probability w_i, then two genes are drawn randomly with replacement from this subpopulation. Its weight vector $\boldsymbol{W}_{St} = (W_{Sta};\ a \in \mathcal{T})$ has components

$$W_{Sta} = \begin{cases} w_i/N_{ti}, & a = i, \\ w_i(1 - 1/N_{ti}), & a = ii, \\ 0, & a = ij,\ i \neq j. \end{cases} \tag{57}$$

The two schemes in (56)–(57) are the same as for a haploid population (Hössjer et al., 2014), since genes are drawn randomly regardless of which individuals they belong to. Only the third scheme I accounts diploidy. It has weights W_{Ia} as in (53), with subpopulations chosen at random with probabilities w_i, then an individual is picked at random within the chosen subpopulation, and finally two genes are drawn with replacement from this individual.

Let H_{ta} be the probability that a gene pair of type a at time t, chosen randomly *with* replacement, is not IBS. For all three schemes, define gene diversities

$$H_{It} = 2\sum_a W_{Ia}H_{ta} = 2\sum_i w_i H_{ti},$$

$$H_{St} = \sum_a W_{Sta}H_{ta} = \sum_i w_i \left[\frac{1}{N_{ti}}H_{ti} + \left(1 - \frac{1}{N_{ti}}\right)H_{tii}\right],$$

$$H_{Tt} = \sum_a W_{Tta}H_{ta} = \sum_i w_i^2 \left[\frac{1}{N_{ti}}H_{ti} + \left(1 - \frac{1}{N_{ti}}\right)H_{tii}\right] + \sum_{i \neq j} w_i w_j H_{tij},$$

so that H_{St} (H_{Tt}) is the probability that two randomly chosen genes, drawn with replacement according to scheme S (T), are not IBS, and H_{It} is twice this probability for scheme I. The extra term 2 of H_{It} compensates for that genes are drawn with replacement within individuals. In order to express the three gene diversities in terms of allele frequencies, assume there are n_t different alleles at time t, and let $P_{tcir} \in \{0, 0.5, 1\}$ be the frequency of allele $c \in \{1, \ldots, n_t\}$ in individual $r \in \{1, \ldots, N_{ti}\}$ of subpopulation i at time t. Averaging these frequencies over all individuals in i, $P_{tci} = \sum_r P_{tcir}/N_{ti}$ is the frequency of allele c in subpopulation i at time t. Another average, taken over subpopulations, gives the frequency $P_{tc} = \sum_i w_i P_{tci}$ of allele c in the entire population at time t, when subpopulations are weighted as w_i. It follows after some calculations that

$$\begin{aligned} H_{It} &= 2\sum_{i=1}^{s} w_i \sum_{1 \leq c < d \leq n_t} \sum_{r=1}^{N_{ti}} 2P_{tcir}P_{tdir}/N_{ti}, \\ H_{St} &= \sum_{i=1}^{s} w_i \sum_{1 \leq c < d \leq n_t} 2P_{tci}P_{tdi}, \\ H_{Tt} &= \sum_{1 \leq c < d \leq n_t} 2P_{tc}P_{td}. \end{aligned} \tag{58}$$

The fixation indices

$$\begin{aligned} G_{STt} &= (H_{Tt} - H_{St})/H_{Tt}, \\ F_{ISt} &= (H_{St} - H_{It})/H_{St}, \\ F_{ITt} &= (H_{Tt} - H_{It})/H_{Tt}, \end{aligned} \tag{59}$$

are derived from the gene diversities in (58). The definitions in (59) also agree with those in Nei (1977), Chakraborty (1993) and in particular Eqs. (28), (31) and (33) of Nagylaki (1998a).

Using the characterization in (58) of non-IBS probabilities in terms of allele frequencies, we prove in the SM that

$$\begin{aligned} &0 \le G_{STt} \le 1,\\ &-1 \le F_{ISt} \le 1,\\ &-1 \le F_{ITt} \le 1. \end{aligned} \tag{60}$$

These inequalities hold since genes are drawn with replacement, and since an extra factor 2 was added in the definition of H_{It}. We also prove in the SM that

$$\begin{aligned} F_{ISt} &= 0,\\ G_{STt} &= F_{ITt}, \end{aligned} \tag{61}$$

when genotypes of all subpopulations i with $w_i > 0$ conform with Hardy–Weinberg proportions. Negative values of F_{ISt} signify an excess of heterozygous individuals compared to HW proportions. See also Cockerham (1969, 1973), Wang (1997a) and Balloux (2004) for a slightly different definition of fixation indices, where schemes S_d and D are used instead of S and T in (59). This implies for instance that $F_{IS_{dt}}$ is slightly negative under HW proportions, although very marginally so for large populations.

Assume that allele frequencies are known at time $t = 0$, but not for $t > 0$. With some abuse of notation we let $h_{ta} = E_0(H_{ta})$ for $t > 0$ be a prediction of H_{ta}, i.e. the probability that a randomly chosen pair of genes of type a at time t are not IBS given that they are drawn with replacement and we only have information about allele frequencies at time 0. (Here E_0 denotes expectation given allele frequencies at time 0.) Analogous predictions of the three quantities in (58) are

$$\begin{aligned} h_{It} &= E_0(H_{It}) = 2\boldsymbol{W}_I\boldsymbol{h}_t,\\ h_{St} &= E_0(H_{St}) = \boldsymbol{W}_{St}\boldsymbol{h}_t,\\ h_{Tt} &= E_0(H_{Tt}) = \boldsymbol{W}_{Tt}\boldsymbol{h}_t, \end{aligned} \tag{62}$$

with $\boldsymbol{h}_t = (h_{ta};\ a \in \mathcal{T})'$ a column vector of non-IBS probabilities, and $\boldsymbol{W}_{Tt}$, $\boldsymbol{W}_{St}$ and $\boldsymbol{W}_{It}$ row vectors of sampling probabilities for all three schemes. The corresponding predicted fixation indices are obtained by replacing the unknown gene diversities in (59) by the predicted ones in (62), so that

$$\begin{aligned} g_{STt} &= (h_{Tt} - h_{St})/h_{Tt},\\ f_{ISt} &= (h_{St} - h_{It})/h_{St},\\ f_{ITt} &= (h_{Tt} - h_{It})/h_{Tt}. \end{aligned} \tag{63}$$

The first of these quantities is defined in Nei (1975), Nei et al. (1977), Hössjer and Ryman (2014) and Hössjer et al. (2014). The inequalities

$$\begin{aligned} &0 \le g_{STt} \le 1,\\ &-1 \le f_{ISt} \le 1,\\ &-1 \le f_{ITt} \le 1 \end{aligned}$$

follow easily from the corresponding inequalities in (60). The analogue of (61) does *not* hold for f_{ISt} and g_{STt} when the diploid model is equivalent to a haploid one, i.e. when selfing, coalescence and mating are random, as in Example 2. This can be viewed as a Levene correction (Crow and Kimura, 1970) for a structured haploid population. In the SM we use two different methods to prove that the predicted fraction of heterozygots exceeds HW proportions, so that

$$\begin{aligned} h_{St} &< h_{It},\\ f_{ISt} &< 0,\\ f_{ITt} &< g_{STt}. \end{aligned} \tag{64}$$

To compute the quantities in (63), we need expressions for $\boldsymbol{h}_t$. This was done recursively in Hössjer and Ryman (2014) and Hössjer et al. (2014) for haploid populations and predictions of G_{STt}. Here we extend this approach to diploid populations in order to find predictions for all quantities of (59). Since IBS sharing is affected by mutations, we must take the mutation probability of each copying event into account. An infinite alleles model (Kimura, 1971) is assumed, so that each mutation creates a new allele, never seen before, and that different mutation events are independent. Let $\nu_{(t-1)k,i}$ be the mutation probability when a gene of subpopulation k at time $t-1$ is transferred or copied to subpopulation i of the next time point t. For instance, we may have $\nu_{(t-1)k,i} = \nu > 0$ if $k \to i$ corresponds to genes copied during fertilization, and $\nu_{(t-1)k,i} = 0$ for genes of surviving adults of a model with overlapping generations. Introduce $\boldsymbol{V}_t = (V_{ta,b};\ a, b \in \mathcal{T})$, a square matrix of order $s^2 + s$, with elements

$$V_{ta,b} = \begin{cases} 1-(1-\nu_{(t-1)k,i})^2, & a = i, b = k,\\ 1-(1-\nu_{(t-1)k,i})(1-\nu_{(t-1)l,i}), & a = i, b = kl,\\ 1-(1-\nu_{(t-1)k,i})(1-\nu_{(t-1)k,j}), & a = ij, b = k,\\ 1-(1-\nu_{(t-1)k,i})(1-\nu_{(t-1)l,j}), & a = ij, b = kl \end{cases}$$

that equal the probability that none of the two genes involved in a transfer from type b to type a mutates between time $t-1$ and t. It is shown in the SM that

$$\begin{aligned} \boldsymbol{h}_t &= (\boldsymbol{V}_t \odot \boldsymbol{A}_t)\boldsymbol{h}_{t-1} + \{[\boldsymbol{I} - (\boldsymbol{V}_t \odot \boldsymbol{Q}_t)]\,\boldsymbol{1}\} \odot \boldsymbol{e}\\ &\overset{\nu_{t-1,ki}\equiv\nu}{=} (1-\nu)^2\boldsymbol{A}_t\boldsymbol{h}_{t-1} + \left[1-(1-\nu)^2\right]\boldsymbol{e}, \end{aligned} \tag{65}$$

for $t > 0$. Here $\boldsymbol{I}$, $\boldsymbol{Q}_t = (Q_{ta,b};\ a, b \in \mathcal{T})$, $\boldsymbol{A}_t = (A_{ta,b};\ a, b \in \mathcal{T})$ are square matrices of order $s^2 + s$, with $\boldsymbol{I}$ an identity matrix, $Q_{ta,b}$ defined in Section 3,

$$A_{ta,b} = 0.5^{\delta_{\{a\in\mathcal{I}\}}} D_{ta,b} 2^{\delta_{\{b\in\mathcal{I}\}}} \tag{66}$$

a modified version of $D_{ta,b}$ that accounts for genes being drawn with replacement, and $\odot$ denotes an elementwise product of two matrices of equal dimensionality. The column vector $\boldsymbol{1}$ in (65) consists of s^2+s ones, and $\boldsymbol{e} = (e_a;\ a \in \mathcal{T})$ is another column vector of the same length with elements $e_a = 0.5^{\delta_{\{a\in\mathcal{I}\}}}$, the probability that two genes of type a, drawn with replacement, are distinct. State space reduction can be achieved for (65) in a similar way as in Section 4.

To summarize, we first iterate (65) $t-1$ times in order to compute $\boldsymbol{h}_t$, then use weight vectors $\boldsymbol{W}_{Tt}$, $\boldsymbol{W}_{St}$ or $\boldsymbol{W}_I$ to find the predicted fixation indices (62) for all three schemes T, S and I, and finally we insert these quantities into (63).

Example 10 (*Two Subpopulations with Asymmetric Migration*). Fig. 6 shows predicted fixation indices for the system of Example 9 with two subpopulations. The four subplots represent the same migration scenarios as in Figs. 4 and 5. Both subpopulations are genetically identical at time $t = 0$, but then they gradually drift apart. When both subpopulations are isolated (upper left, Fig. 6(a)), it can be shown outside of the figure that g_{STt} converges slowly towards 1, with $g_{ST1000} = 0.921$; when the smaller subpopulation is a source (upper right, Fig. 6(b)), it can be seen outside of the figure again, that g_{STt} first increases to a maximum of 0.266, attained around $t = 121$, and then it gradually decreases towards a limit of 0.200. When the large subpopulation is a source (lower left, Fig. 6(c)) or when both subpopulations exchange one migrant per generation (lower right, Fig. 6(d)), the corresponding g_{STt} curves converge more quickly towards limits 0.063 and 0.057 respectively.

The f_{ISt} curves are depicted for both subpopulations, in all subplots. Since there is no selfing/cloning, all these curves converge towards negative values, indicating an excess of heterozygots compared to HW proportions. This is more evident in the smaller subpopulation, with limiting values $f_{IS\infty}$ equal to -0.051 (upper left,

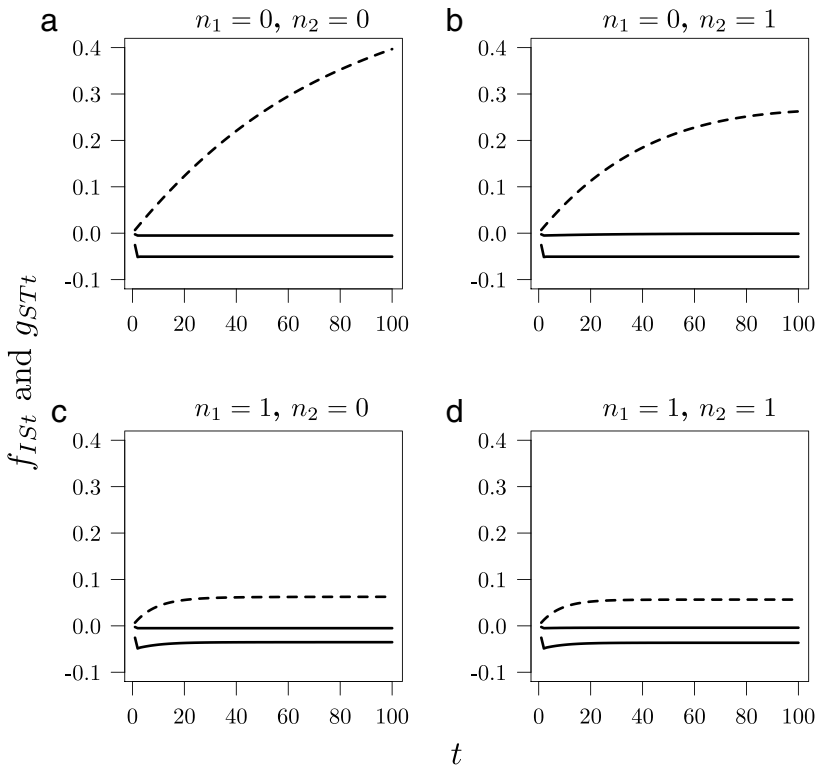

Fig. 6. Plots of predicted fixation indices f_{ISt} and g_{STt} versus time t for the population of Fig. 4 with two subpopulations of sizes $N_1 = 200$ and $N_2 = 20$, non-overlapping generations, n_i individuals migrating from i in each generation (Case II without selfing). There are no mutations ($\nu = 0$), inbreeding or cloning, mating (14) before migration, local effective size $N_{ei} = N_i$ in (23) and no inbreeding at time $t = 0$ ($f_{0a} = 0$ for all $a \in \mathcal{T}$). The two solid curves show f_{ISt} for each subpopulation i (with weights $w_i = 1$), and the smaller $i = 2$ subpopulation has consistently more negative values of f_{ISt}. The dashed curves depict g_{STt} (using weights $w_1 = w_2 = 1/2$).

Fig. 6(a)), −0.051 (upper right, 6(b)), −0.035 (lower left, 6(c)) and −0.036 (lower right, 6(d)). The corresponding limits for the larger subpopulation are −0.005, −0.0025, −0.005 and −0.004, indicating that migrants from the smaller subpopulation tend to reduce the heterozygous excess slightly. □

8. Discussion

This paper provides a general framework for identity by descent and state recursions of a monoecious or dioecious diploid and subdivided population, with applications to long and short term genetic changes (effective size), genetic differentiation between subpopulations and departures from Hardy–Weinberg proportions within each subpopulation. Our work extends that of Hössjer et al. (2014), where haploid populations are treated. The diploid model incorporates various types of reproduction, including mating, selfing and cloning, and substructures such as age classes, geographic demes and pedigrees, with immigration into a pedigreed population from population(s) more or less genetically related to the recipient, pedigreed population. The subpopulations may be of different size and their number may also vary over time, including subpopulation extinction and recolonization.

Many of our results can be viewed as generalizations of IBD- and IBS-recursions for various types of models, such as haploid and monoecious populations with geographic structure or age classes (Malécot, 1951; Felsenstein, 1971; Nagylaki, 1980, 2000; Slatkin, 1991; Whitlock and Barton, 1997; Hössjer et al., 2014), diploid and unstructured populations (Hill, 1979; Crow and Denniston, 1988; Caballero and Hill, 1992a,b; Caballero, 1995; Nagylaki, 1995; Wang, 1995), geographically subdivided diploid populations with specific migration patterns, such as stepping stone models (Sawyer, 1976), monoecious island models (Wang, 1997a; Balloux and Lehmann, 2003), more general monoecious models with binomially varying backward migration (Nagylaki, 1983, 2000), dioecious island models (Chesser, 1991; Chesser et al., 1993; Wang, 1997b; Hössjer et al., 2013), and dioecious age-structured isolated populations (Johnson, 1977; Emigh and Pollak, 1979; Engen et al., 2005).

Our work reveals that the difference between haploidy and diploidy is more pronounced for population systems with one or more small subpopulations than for large ones. This is most evident for pedigree analysis, where each individual can be thought of as a subpopulation. Matrix analytic methods are well suited to quantify this difference, since they are exact and computationally feasible for small populations. This includes not only IBD- and IBS-recursions, but also, for instance, recursions for the number of lineages of an ancestral tree, as studied by Wooding and Rogers (2002) within a haploid framework.

A special feature of our diploid framework is the mating function. It is the distribution of the two parental subpopulations of biparental individuals, and an important tool for finding IBD-probability recursions of diploid populations. This concept, which has no analogue for haploid populations, gives great flexibility in modelling various types of mating scenarios. For populations with geographic substructure and non-overlapping generations, the mating function distinguishes reproduction cycles that differ in their order of fertilization and migration. For age-structured dioecious models it determines the age-preference of mating couples, and when subpopulations are single individuals, organized into a known pedigree, the mating function defines the parents of all non-founders.

The distinction between haploidy and diploidy is crucial for computing the inbreeding effective size, since this notion has a very different meaning when individuals are distinguished in the model or not. Only with a diploid framework is it possible to compute an inbreeding effective size that quantifies the rate at which

inbreeding increases within individuals. This is important in conservation biology because negative effects of inbreeding, so-called inbreeding depression, is expressed and empirically quantified at individual and/or population level and is associated with particular rates of inbreeding accumulation quantified by the inbreeding effective size. Our diploid framework also gives criteria to assess expected departures from Hardy–Weinberg proportions for general structured populations.

From the perspective of conservation biology and conservation genetics the major contributions from this present study in combination with that of Hössjer et al. (2014) include providing analytical means for (i) predicting inbreeding effective size over short and long term in separate subpopulations of a population system as well as for the total system (metapopulation), (ii) computing population differentiation, measured as G_{ST} of such systems, and (iii) performing such computations in situations where the subpopulations vary in size both among populations and within populations (including extinction and recolonization of separate subpopulations); the subpopulations may (iv) be connected through various rates of gene flow and, (v) coancestry among and within them may vary, as well. These important parts of our framework have been implemented in an R-based software, Genetic Exploration of Structured Populations (GESP), which we are in the process of presenting as a separate publication directed towards conservation biologists and conservation geneticists (Olsson et al., 2015). Forthcoming applications of the framework presented here also include using it as applied in the GESP software for exploring practical conservation genetic management situations including that of the highly inbred Swedish wolf population (Laikre et al., 2013) and of metapopulations of Baltic Sea species (Laikre et al. in prep.).

Several extensions are possible. First, other forces of genetic change, such as recombination and selection, could be included. Another possibility is to study the effect of mutations not only for fixation indices, but also for effective sizes (Wakeley and Sargsyan, 2009).

Second, we assumed in (12) and (29) that backward migration rates are fixed, so that the parental subpopulations of distinct genes are chosen without replacement. Other possibilities include multinomial backward migration, where genes of different individuals choose their parental subpopulations independently with replacement, see for instance Sawyer (1976), Nagylaki (1983), Balloux and Lehmann (2003) and Hössjer et al. (2014). Its actual migration rates vary multinomially around their expected values, so that the large population approximations in (12) and (29) are exact. Other possibilities with more random variation of migration rates include models of Whitlock and Barton (1997) and Dirichlet multinomial backward migration in Hössjer (in press). Wang (1996b) obtained exact IBD recursions for a dioecious model with partial sib mating. It can be interpreted as a dioecious island model, where males and females of each sibship are two separate islands, whose random sizes correspond to stochastic backward migration rates. It would be of interest to include all these models into a more general diploid framework of stochastic migration.

Third, we have considered genetic variation at one single locus. Extended models for multiple loci have the potential to generate general expressions for linkage disequilibrium measures of effective size (Hill, 1981; Waples and England, 2011; Waples et al., 2014) as well as the mean and variance of shared IBD segments (Carmi et al., 2013) for populations with general substructure.

Fourth, IBD dynamics for more than two alleles has mostly been analysed in the limit of large populations. This is closely related to finding the marginal distributions of their ancestral tree, which in the haploid case is a Kingman coalescent (Kingman, 1982) under strong migration, where migration rates between subpopulations is faster than the coalescence rates within them (Notohara, 1993, Nordborg and Krone, 2002, Sagitov and Jagers, 2005 and Hössjer, 2011). A more complicated structured coalescent limit is obtained when the limiting coalescence rates within and migration rates between subpopulations are of the same order (Notohara, 1990; Herbots, 1997) or when the number of subpopulations is large (Wakeley, 1998). The limiting ancestry of a large diploid population is also a Kingman coalescent whether it is homogeneous (Möhle, 1998) or age-structured with a strong migration limit (Pollak, 2011). The main impact of diploidy is then the coalescence effective size, a single number that provides the correct rescaling of time (Sjödin et al., 2005; Wakeley and Sargsyan, 2009). Nagylaki (1998b) showed that the ancestral trees of diploid and haploid populations often behave similarly, also when migration and coalescence rates are of the same order. Whereas many of these results are asymptotic, it would be of interest to develop a general exact coalescence framework for diploid populations.

Fifth, our matrix analytic framework makes it possible to compute inbreeding coefficients and coalescence probabilities for pairs of genes in populations that are mixtures of known pedigrees and "background populations", i.e. populations from which individuals may be drawn to create "immigrants" into the pedigree. We believe this could be useful for computing reproductive values of ancestors and coalescence theory for pedigrees (Chang, 1999; Derrida et al., 2000; Wakeley et al., 2012), for approximating of the length distribution of shared IBD-segments (Carmi et al., 2013) and for increasing power to detect the degrees of relationship (Huff et al., 2011; Li et al., 2014).

Acknowledgements

Ola Hössjer was financially supported by the Swedish Research Council, contract nr. 621-2013-4633. Linda Laikre and Nils Ryman were funded by grants from the Swedish Environmental Protection Agency, contract nr. 11/116, the Swedish Research Council, contract nr. 621-2011-3715, the Swedish Research Council Formas, 215-2012-1550, and the BONUS project BAMBI, the joint Baltic Sea research and development programme (Art 185), funded jointly from the European Union's Seventh Programme for research, technological development and demonstration and from the Swedish Research Council Formas. The authors also wish to thank four anonymous reviewers for valuable comments that considerably improved the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.tpb.2015.03.006.

References

Allendorf, F.W., Ryman, N., 2002. The role of genetics in population viability analysis. In: Bessinger, S.R., McCullogh, D.R. (Eds.), Population Viability Analysis. The University of Chicago Press, Chicago.
Balloux, F., 2004. Heterozygote excess in small populations and the heterozygote excess effective size. Evolution 58 (9), 1891–1900.
Balloux, F., Lehmann, L., 2003. Random mating with a finite number of matings. Genetics 165, 2313–2315.
Balloux, F., Lehmann, L., de Meeûs, T., 2003. The population genetics of clonal and partially clonal diploids. Genetics 164, 1635–1644.
Caballero, A., 1994. Developments in the prediction of effective population size. Heredity 73, 657–679.
Caballero, A., 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139, 1007–1011.
Caballero, A., Hill, W.G., 1992a. Effective size of nonrandom mating populations. Genetics 130, 909–916.
Caballero, A., Hill, W.G., 1992b. A note on the inbreeding effective population size. Evolution 46 (6), 1969–1972.
Cannings, C., Thompson, E.A., Skolnick, M.H., 1978. Probability functions on complex pedigrees. Adv. Appl. Probab. 10, 26–61.

Carmi, S., Palamara, P.F., Vacic, V., Lencz, T., Darvasi, A., Pe'er, I., 2013. The variance of identity-by descent sharing in the Wright–Fisher model. Genetics 193, 911–928.
Chakraborty, R., 1993. Analysis of genetic structure of populations: meaning, methods and implications. In: Majuner, P.P. (Ed.), Human Population Genetics. Plenum Press, New York, pp. 189–205.
Chang, J.T., 1999. Recent common ancestors of all present-day individuals. Adv. Appl. Probab. 31, 1002–1026.
Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. Nature Rev. Genet. 10, 195–205.
Chesser, R.K., 1991. Gene diversity and female philopatry. Genetics 129, 573–583.
Chesser, R.K., Rhodes, O.E., Sugg, D.W., Schnabel, 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. Analysis of gene frequencies. Genetics 74, 679–700.
Crow, J.F., 1954. Breeding structure of populations. II. Effective population number. In: Statistics and Mathematics in Biology. Iowas State Coll. Press, Ames, Iowa, pp. 543–556.
Crow, J.F., Denniston, C., 1988. Inbreeding and variance effective population sizes. Evolution 42 (3), 482–495.
Crow, J.F., Kimura, M., 1970. An Introduction to Population Genetics Theory. The Blackburn Press, New Jersey, USA.
Derrida, B., Manrubia, S.C., Zannette, D.H., 2000. On the genealogy of a population of biparental individuals. J. Theoret. Biol. 203, 303–315.
Durrett, R., 2008. Probability Models for DNA Sequence Evolution, 2 ed. Springer, New York.
Emigh, T.H., Pollak, E., 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theor. Popul. Biol. 15, 86–107.
Engen, S., Lande, R., Saether, B.-E., 2005. Effective size of a fluctuating age-structured population. Genetics 170, 941–954.
Ewens, W.J., 1982. On the concept of effective population size. Theor. Popul. Biol. 21, 373–378.
Felsenstein, J., 1971. Inbreeding and variance effective numbers in populations with overlapping generations. Genetics 68, 581–597.
Frankham, R., 2005. Genetics and extinction. Biol. Cons. 126, 131–140.
Franklin, I.R., 1980. Evolutionary change in small populations. In: Soulé, M.E., Wilcox, B.A. (Eds.), Conservation Biology: an Evolutionary-Ecological Perspective. Sinauer Associates, pp. 135–150.
Gasbarra, D., Sillanpää, M.J., Arjas, E., 2005. Backward simulation of ancestors of sampled individuals. Theor. Popul. Biol. 67, 75–83.
Groom, M.J., Meffe, G.K., Carroll, C.R., 2005. Principles of Conservation Biology, third ed. Sinauer Associated, Inc., Sunderland, Massachusetts, USA.
Harrison, S., Hastings, A., 1996. Genetic and evolutionary consequences of metapopulation structure. Trends Ecol. Evol. 11, 180–183.
Herbots, H.M., 1997. The structured coalescent. In: Donnelly, P., Tavaré, S. (Eds.), Progress in Population Genetics and Human Evolution. Springer-Verlag, New York, pp. 231–255.
Hill, W.G., 1979. A note of effective population size with overlapping generations. Genetics 92, 317–322.
Hill, W.G., 1981. Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38, 209–216.
Hössjer, O., 2011. Coalescence theory for a general class of structured populations with fast migration. Adv. Appl. Probab. 43 (4), 1027–1047.
Hössjer, O., 2014. On the eigenvalue effective size of structured populations. J. Math. Biol. http://dx.doi.org/10.1007/s00285-014-0832-5. (in press) Available online.
Hössjer, O., Jorde, P.E., Ryman, N., 2013. Quasi equilibrium approximations of the fixation index of the island model under neutrality. Theor. Popul. Biol. 84, 9–24.
Hössjer, O., Olsson, F., Laikre, L., Ryman, N., 2014. A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. Math. Biosci. 258, 113–133. http://dx.doi.org/10.1016/j.mbs.2014.10.001.
Hössjer, O., Ryman, N., 2014. Quasi equilibrium, variance effective population size and fixation index for models with spatial structure. J. Math. Biol. 69 (5), 1057–1128. http://dx.doi.org/10.1007/s00285-013-0728-9.
Huff, C.D., Witherspoon, D.J., Simonson, T.S., Zing, J., Waktins, W.S., et al., 2011. Maximum-likelihood estimation of recent shared ancestry (ESRA). Genome Res. 21, 768–774.
Jamieson, I.G., Allendorf, F.W., 2012. How does the 50/500 rule apply to MVPs? Trends Ecol. Evol. 27, 578–584.
Johnson, D.L., 1977. Inbreeding in populations with overlapping generations. Genetics 87, 581–591.
Kettle, C.J., 2014. Fragmentation genetics in tropical ecosystems: from fragmentation genetics to fragmentation genomic. Conserv. Genet. 15, 1265–1268.
Kimura, M., 1971. Theoretical foundations of population genetics at the molecular level. Theor. Popul. Biol. 2, 174–208.
Kingman, J.F.C., 1982. The coalescent. Stochastic Process. Appl. 13, 235–248.
Koski, T., Noble, J., 2011. Bayesian Networks: An Introduction. John Wiley and Sons.
Laikre, L., Jansson, M., Allendorf, F.W., Jakobsson, S., Ryman, N., 2013. Hunting effects on favourable conservation status of highly inbred Swedish wolves. Conserv. Biol. 27, 248–253.
Laporte, V., Charlesworth, B., 2003. Effective population size and population subdivision in demographically structured populations. Genetics 162, 501–519.
Larson, S., Jameson, R., Etnier, M., Bentzen, P., 2002. Loss of genetic diversity in sea otters (Enhydra lutris) associated with the fur trade of the 18th and 19th centuries. Mol. Ecol. 11, 1899–1903.
Levins, R., 1970. Extinction. In: Gesternhaber, M. (Ed.), Some Mathematical Problems in Biology. American Mathematical Society, Providence, Rhode Island, pp. 77–107.
Li, H., Glusman, G., Hu, H., Shankaracharya, , Caballero, J., Hubley, R., Witherspoon, D., et al., 2014. Relationship estimation from whole genome data. PLoS Genet. 10 (1), e1004144.
Liberg, O., Andrén, H., Pedersen, C.-H., Sand, H., Sejberg, D., Wabakken, P., Åkesson, M., Bensch, S., 2005. Severe inbreeding depression in a wild wolf Canis lupus population. Biol. Lett. 1, 17–20.
Malécot, G., 1951. Un treatment stochastique des problemès linairés (mutation, linkage, migration) en géneétique de populations. Ann. Univ. Lyon A 14, 79–117.
Maruyama, T., 1970. Effective number of alleles in subdivided populations. Theor. Popul. Biol. 1, 273–306.
Möhle, M., 1998. Coalescent results for two-sex population models. Adv. Appl. Probab. 30, 513–520.
Nabata, D., Masuda, R., Takahashi, O., 2004. Bottleneck effects on the sika deer Cervus Nippon population in Hokkaido, revealed by ancient DNA analysis. Zool. Sci. 21, 473–481.
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 geographic variation. Theor. Popul. Biol. 24, 268–294.
Nagylaki, T., 1995. The inbreeding effective population number in dioecious populations. Genetics 139, 473–485.
Nagylaki, T., 1998a. Fixation indeces in subdivided populations. Genetics 148, 1325–1332.
Nagylaki, T., 1998b. The expected number of heterozygous sites in a subdivided population. Genetics 149, 1599–1604.
Nagylaki, T., 2000. Geographical invariance and the strong-migration limit in subdivided populations. J. Math. Biol. 41, 123–142.
Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. US 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. Ann. Hum. Genet. 41, 225–233.
Nei, M., Chakravarti, A., Tateno, Y., 1977. Mean and variance of F_{ST} in a finite number of incompletely isolated populations. Theor. Popul. Biol. 11, 291–306.
Nordborg, M., Donnelly, P., 1997. The coalescent process with selfing. Genetics 146, 1185–1195.
Nordborg, M., Krone, S., 2002. Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin, M., Veuille, M. (Eds.), Modern Development in Theoretical Population Genetics. Oxford Univ. Press, pp. 194–232.
Notohara, M., 1990. The coalescent and the genealogical process in geographically structured populations. J. Math. Biol. 29, 59–75.
Notohara, M., 1993. The strong-migration limit for the genealogical process of geographically structured populations. J. Math. Biol. 31, 115–122.
Olsson, F., Laikre, L., Hössjer, O., Ryman, N., 2015. GESP—A program for genetic exploration of structured populations. Manuscript.
Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344 (6187), 987–997. http://dx.doi.org/10.1126/science.1246752.
Pollak, E., 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics 117, 353–360.
Pollak, E., 2002. Eigenvalue effective population numbers for populations that vary cyclically in size. Math. Biosci. 177–178, 11–24.
Pollak, E., 2011. Coalescent theory for age-structured random mating populations with two sexes. Math. Biosci. 233, 126–134.
Räikkönen, J., Bignert, A., Mortensen, P., Fernholm, B., 2006. Congenital defects in a highly inbred wild wolf population (Canis lupus). Mamm. Biol. 71, 65–73.
Räikkönen, J., Vucetich, J.A., Peterson, R.O., Nelson, M.P., 2009. Congenital bone deformities and the inbred wolves (Canis lupus) of Isle Royale. Biol. Cons. 142, 1025–1031.
Sagitov, S., Jagers, P., 2005. The coalescent effective size of age-structured populations. Ann. Appl. Probab. 15 (3), 1778–1797.
Sawyer, S., 1976. Results for the stepping stone model for migration in population genetics. Ann. Probab. 4, 699–728.
Sjödin, P., Kaj, I., Krone, S., Lascoux, M., Nordborg, M., 2005. On the meaning and existence of an effective population size. Genetics 169, 1061–1070.
Slatkin, M., 1991. Inbreeding coefficients and coalescent times. Genet. Res. Camb. 58, 167–175.
Sved, J.A., Latter, B.D.H., 1977. Migration and mutation in stochastic models of gene frequency change. J. Math. Biol. 5, 61–73.
Traill, L.W., Brook, B.W., Frankham, R.R., Bradshaw, C.J.A., 2010. Pragmatic population viability targets in a rapidly changing world. Biol. Conserv. 143, 28–34.
Wakeley, J., 1998. Segregating sites in Wright's island model. Theor. Popul. Biol. 53, 166–174.
Wakeley, J., King, L., Low, B.S., Ramachandran, S., 2012. Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. Genetics 190, 1433–1445.
Wakeley, J., Sargsyan, O., 2009. Extensions of the coalescent effective population size. Genetics 181, 341–345.
Wang, J., 1995. Exact inbreeding coefficient and effective size of finite populations under partial sib mating. Genetics 140, 357–363.

Wang, J., 1996a. Inbreeding and variance effective sizes for nonrandom mating populations. Evolution 50 (5), 1786–1794.
Wang, J., 1996b. Exact inbreeding coefficient and effective size of finite populations under partial sib mating. Genetics 140, 357–363.
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., 2002. Definition and estimation of effective population size in the conservation of endangered species. In: Beissinger, S.R., McCullogh, D.R. (Eds.), Populations Viability Analysis. The University of Chicago Press, Chicago.
Waples, R.S., 2010. Spatial–temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Mol. Ecol. Resour. 10, 785–796.
Waples, R.S., Antao, T., Luikart, G., 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics http://dx.doi.org/10.1534/genetics.114.164822.
Waples, R.S., England, P.R., 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. Genetics 189, 633–644.
Whitlock, M.C., Barton, N.H., 1997. The effective size of a subdivided population. Genetics 146, 427–441.
Wooding, S., Rogers, A., 2002. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics 161, 1641–1650.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.
Wright, S., 1938. Size of population and breeding structure in relation to evolution. Science 87, 430–431.
Wright, S., 1943. Isolation by distance. Genetics 28, 114–138.
Wright, S., 1951. The general structure of populations. Ann. Eugenics 15, 323–354.
Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19, 395–420.
Wright, S., 1978. Evolution and Genetics of Populations. Vol IV. Variability within and Among Natural Populations. University of Chicago Press, Chicago.