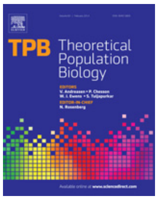
Theoretical Population Biology 102 (2015) 40–59

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