



A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time



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 23 November 2013

Revised 9 September 2014

Accepted 7 October 2014

Available online 20 October 2014

2000 MSC:

92D25

60F99

Keywords:

Effective population size

Coefficient of gene differentiation

Matrix analytic methods

Subpopulation weights

State space reduction

ABSTRACT

The main purpose of this paper is to develop a theoretical framework for assessing effective population size and genetic divergence in situations with structured populations that consist of various numbers of more or less interconnected subpopulations. We introduce a general infinite allele model for a diploid, monoecious and subdivided population, with subpopulation sizes varying over time, including local subpopulation extinction and recolonization, bottlenecks, cyclic census size changes or exponential growth. Exact matrix analytic formulas are derived for recursions of predicted (expected) gene identities and gene diversities, identity by descent and coalescence probabilities, and standardized variances of allele frequency change. This enables us to compute and put into a general framework a number of different types of genetically effective population sizes (N_e) including variance, inbreeding, nucleotide diversity, and eigenvalue effective size. General expressions for predictions (g_{ST}) of the coefficient of gene differentiation G_{ST} are also derived. We suggest that in order to adequately describe important properties of a subdivided population with respect to allele frequency change and maintenance of genetic variation over time, single values of g_{ST} and N_e are not enough. Rather, the temporal dynamic patterns of these properties are important to consider. We introduce several schemes for weighting subpopulations that enable effective size and expected genetic divergence to be calculated and described as functions of time, globally for the whole population and locally for any group of subpopulations. The traditional concept of effective size is generalized to situations where genetic drift is confounded by external sources, such as immigration and mutation. Finally, we introduce a general methodology for state space reduction, which greatly decreases the computational complexity of the matrix analytic formulas.

© 2014 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

Determining the amount of genetic variation within and between populations and the rate of loss of genetic variation is of fundamental importance in evolutionary and conservation genetics, and crucial parameters in this respect include the genetically effective population size (N_e) and the coefficient of gene differentiation (G_{ST}). Nei [64] introduced G_{ST} as a multiallelic and multilocus extension of the fixation index F_{ST} of Wright [115,116] and it quantifies the proportion of genetic variation that is explained by genetic differences between populations. The effective size is the size of an ideal homogeneous population without mutations or selection, 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

the concept was first introduced by Wright [113,114], as reviewed e.g. by Ewens [21], Crow and Denniston [17], Orrive [76], Caballero [7], Wang and Caballero [103], Waples [107], and Charlesworth [11]. Over the years, N_e has become an indispensable tool in conservation biology for identifying population sizes necessary for short and long term conservation of e.g. endangered species and populations [2,27,93,96].

Most models for N_e refer to a single population of constant size, and rules of thumb in conservation genetics are often based on such models assuming single, isolated populations [1]. In real life, however, populations are rarely isolated but are subject to gene flow among more or less isolated subpopulations of varying size that are dispersed over a particular geographic area.

The main purpose of the present paper is to develop theoretical means for assessing effective population size and genetic divergence in a situation with substructured populations that consist of various numbers of more or less interconnected subpopulations whose size can vary over space and time. Specifically, this analytical work was prompted by a practical, real life case – the conservation genetic situation of the Swedish wild wolf population. The Swedish wolf

* Corresponding author. Tel.: +46 70 672 12 18.

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.mbs.2014.10.001>

0025-5564/© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

population is highly inbred and has been almost completely isolated for several decades. Conservation genetics research has addressed the severe genetic situation (e.g. [46,52,80]), specifically stressing the need for breaking isolation and creating an interconnected population system where the Swedish wolf population can exchange genes with the Finnish population and populations further east [29,47]. Politicians and managers have now also realized this need and various ways of creating gene flow are discussed [53]. However, the necessary properties of such a substructured population in order to meet conservation genetic goals with respect to rate of inbreeding and genetic drift are unclear, because theory for metapopulation effective population size relevant for practical situations such as the current one has been missing. In this paper, we develop the mathematical framework for addressing such issues and several steps are involved in this process, including 1) generalized recursion formulas for a number of genetic quantities, 2) new analytical tools for reporting the time dynamics of N_e and forecasts of G_{ST} , 3) a novel class of subpopulation weights, 4) a generalized unified theory of different types of effective size N_e , and 5) a novel method of computational reduction for populations with symmetries.

We briefly describe these five contributions. First, we consider a class of diploid, monoecious populations evolving in discrete generations under selective neutrality, with mutations, migration and genetic drift (due to random sampling of genes when a finite population reproduces) as evolutionary forces affecting the amount of genetic variation. Whereas formulas for N_e and G_{ST} are often developed under a variety of assumptions, such as a large population, a small migration rate between subpopulations, or a long time frame, we will rather use matrix analytic methods [10] in order to define exact linear recursions for a number of quantities, including identity by descent and coalescence probabilities, standardized variances of allele frequency change, and predicted (expected) gene identities/gene diversities. All these recursions are very similar, with matrices that have rows and columns indexed by pairs of subpopulations. Although several authors have considered such recursions, starting with the seminal work in [56], our setup is more general in that we allow the demographics, in terms of migration patterns, local census and local effective sizes to vary in an arbitrary way, including global and local bottlenecks, subpopulation extinction and recolonization, cyclic changes, or exponential growth. From these recursions we get novel and exact expressions of predictions of G_{ST} and various types of N_e (inbreeding, variance, nucleotide diversity, eigenvalue).

Second, it is essential for protection of genetic characteristics of populations and species to know the rate of loss of genetic variability and subpopulation differentiation over short and long time intervals. For a subdivided population it is typically not possible to summarize this information with just a few parameters, not even when the subpopulation census sizes are constant over time. When a population is isolated its degree of inbreeding, for example, will increase at a fixed rate so that N_e is constant from one generation to the next. In contrast, if a subpopulation of a population system receives migrants from the rest of the system, then the rate of inbreeding, and thereby N_e , will vary over time. As a consequence, N_e of the whole system will fluctuate as well, as we will see below. For this reason we generalize a new approach initiated in [73] for age structured models and variance effective sizes, and report N_e as a function of the time interval under which genetic loss takes place. Similarly, the predicted G_{ST} depends on when the forecast is made, and it can therefore be computed as a function of the distance between the present and the time point of prediction. This enables researchers to investigate the predicted genetic effects of various demographic scenarios and management schemes that include population systems rather than single, isolated populations.

Third, we consider a large and novel class of schemes of weighting subpopulations and show how they influence N_e and predictions of G_{ST} . Of particular interest are weights that are uniform (all

subpopulations weighted equally), proportional to subpopulation sizes (each individual weighted equally) or reproductive (each individual weighted proportionally to its predicted or expected long term number of descendants). We also consider local schemes for which only subsets of subpopulations are assigned positive weights. This could be of interest in practical management when the population managed in a particular area is genetically connected to one or several other populations, which act as more or less known ‘ghost populations’ [4,91]. For such local weights, it is possible to quantify exactly how various migration scenarios between the population of interest and the other subpopulations affect G_{ST} and N_e .

Fourth, in an influential paper Whitlock and Barton [110] showed that several notions of effective size are closely related for subdivided populations, and here we extend their results by considering time intervals of arbitrary length, and a larger class of effective sizes. To this end, we utilize that each type of effective size involves a quantity that is either defined backwards (identity by descent and coalescence probabilities) or forwards (predicted gene identity, standardized variance of allele frequency change) in time, and the matrices of the corresponding linear recursions can be described in terms of pairs of genes, drawn with or without replacement from the population. Although the latter distinction has a negligible effect for a population with a size of order, say, 100 or larger, it makes it possible to put all notions of effective size into a unified framework, expressing each one of them as a very explicit function of the initial conditions and matrices of its linear recursion and of the subpopulation weights.

Fifth, we define a general way of exploiting invariance between subpopulations, so that whenever certain symmetry conditions hold, the size of the state space can be reduced from s^2 , where s is the number of subpopulations.

In Section 2 we define the population dynamics and specify in particular how migration, genetic drift and mutations enter into the model. Subpopulation differentiation is treated in Section 3 and the various types of effective sizes in Section 4. In Section 5 we consider the special case when local census sizes and migration rates are time independent. This is illustrated with several examples that highlight the importance of reporting N_e and predictions of G_{ST} as functions of time. State space reduction is defined and exemplified in Section 6, a summary and discussion are provided in Section 7, some extensions of the theory and proofs can be found in Appendix A, and finally, Table 1 provides a list of notations for some of the most important quantities of the paper.

2. Model for demographics, reproduction, and mutations

Consider a diploid and monoecious population with a random amount of selfing that is subdivided, with s subpopulations. It evolves in discrete generations $t = \dots, -1, 0, 1, 2, \dots$ of which $t < 0$ represents the past, $t = 0$ the present and $t > 0$ the future. Let $N_{ti} \geq 0$ be the local census size of subpopulation i in generation t , with $N_{ti} = 0$ corresponding to extinction. Each individual carries two copies of a gene, so that the total number of genes in subpopulation i and generation t is $2N_{ti}$. The total census size $N_t = \sum_{i=1}^s N_{ti}$ in generation t is assumed to be positive, so that at least one subpopulation is non-extinct.

The local effective size N_{eti} in generation t of subpopulation i is usually (but not necessarily) smaller than or equal to its local census size N_{ti} . The more variable reproduction between individuals in i is, the smaller is N_{eti} .

Occasionally, migration between subpopulations takes place, as quantified by the forward and backward migration rates $M_{t-1,ki}$ and B_{tik} from subpopulation k of generation $t-1$ to subpopulation i of generation t . More precisely, each gene of subpopulation k and generation $t-1$ has an expected number $M_{t-1,ki}$ of offspring, which, in the next generation t , live in subpopulation i , whereas B_{tik} is the probability that a parent of a gene in subpopulation i and generation t originates from subpopulation k in the previous generation $t-1$. The

Download English Version:

<https://daneshyari.com/en/article/6372029>

Download Persian Version:

<https://daneshyari.com/article/6372029>

[Daneshyari.com](https://daneshyari.com)