Journal of Theoretical Biology 394 (2016) 182–196

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

A monoecious and diploid Moran model of random mating

Ola Hössjer [a,*], Peder A. Tyvand [b]

[a] Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
[b] Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, N-1432 Ås, Norway

HIGHLIGHTS

- A diploid monoecious Moran model is proposed for random mating with possible selfing.
- The Moran model is compared with a diploid and monoecious Wright–Fisher model.
- Diffusion approximations are derived on two time scales.
- Genotype frequencies oscillate as an Ornstein–Uhlenbeck process on the local time scale.
- Fixation index f_{IS} oscillates as an Ornstein–Uhlenbeck process around a fixed point.

ARTICLE INFO

Article history:
Received 17 May 2015
Received in revised form
23 November 2015
Accepted 20 December 2015
Available online 22 January 2016

Keywords:
Diploid
Diffusion approximation
Markov chain
Moran model
Random mating

ABSTRACT

An exact Markov chain is developed for a Moran model of random mating for monoecious diploid individuals with a given probability of self-fertilization. The model captures the dynamics of genetic variation at a biallelic locus. We compare the model with the corresponding diploid Wright–Fisher (WF) model. We also develop a novel diffusion approximation of both models, where the genotype frequency distribution dynamics is described by two partial differential equations, on different time scales. The first equation captures the more slowly varying allele frequencies, and it is the same for the Moran and WF models. The other equation captures departures of the fraction of heterozygous genotypes from a large population equilibrium curve that equals Hardy–Weinberg proportions in the absence of selfing. It is the distribution of a continuous time Ornstein–Uhlenbeck process for the Moran model and a discrete time autoregressive process for the WF model. One application of our results is to capture dynamics of the degree of non-random mating of both models, in terms of the fixation index f_{IS}. Although f_{IS} has a stable fixed point that only depends on the degree of selfing, the normally distributed oscillations around this fixed point are stochastically larger for the Moran than for the WF model.

1. Introduction

The classical haploid Wright–Fisher model of population genetics describes random mating in discrete generations without overlap (Fisher, 1922; Wright, 1931). As a contrast the haploid Moran model (Moran, 1958a) of population genetics has a higher degree of continuity between consecutive generations, as it exchanges one individual at discrete or continuous time points. A diploid and dioecious (two-sex) version of the discrete time Moran model was introduced by Moran (1958b). It has a more complicated state space, with genotype frequencies for males and females, so that exact computation becomes unfeasible for all but very small populations. In this paper we develop an exact Markov process for a simpler monoecious (one-sex) diploid Moran model with possible selfing, at a biallelic locus. It has a simpler state space, where only two genotype frequencies are needed to analyze the dynamics of the population, in discrete or continuous time.

There is also a monoecious and diploid Wright–Fisher (WF) model, see for instance Moran (1958c), Crow and Denniston (1988), Tyvand (1993) and references therein. We will compare the exact Markov chains for the monoecious and diploid Moran and WF models, and in particular check the validity of the standard equivalence-time scaling that connects the two models. In order to facilitate this comparison, we develop diffusion approximations for both models that are increasingly accurate for large populations. It is well-known (Watterson, 1964, Ethier and Nagylaki, 1980, 1988) that such approximations work on two different time scales in terms of a system of two partial differential equations instead of one. The first equation is the same for the Moran and

* Corresponding author.
E-mail addresses: ola@math.su.se (O. Hössjer), peder.tyvand@nmbu.no (P.A. Tyvand).

http://dx.doi.org/10.1016/j.jtbi.2015.12.028
0022-5193/

WF models. It describes the slower allele frequencies dynamics in the same way as for the haploid Moran and WF models (Kimura, 1955). The second equation operates on a more local time scale, and it is different for the Moran and WF models. It captures the more rapidly varying departure of the fraction of heterozygous genotypes from a large population equilibrium curve that corresponds to Hardy–Weinberg (HW) proportions when there is no selfing.

Previous work for diploid models have shown that the allele frequency process dominates for large populations, and that the oscillations around the equilibrium curve are asymptotically negligible. Here we rescale the latter and obtain a nondegenerate limit that has the more slowly varying allele frequency process as a fixed parameter. This makes the frequency oscillations of the heterozygous genotypes locally time invariant and normally distributed around the equilibrium curve. It is a stationary Gaussian (Ornstein–Uhlenbeck, OU) process in continuous time for the Moran model, and an autoregressive process in discrete time for the WF model.

A similar autoregressive limit process has previously been obtained by Korolyuk and Korolyuk (1995), and Coad (2000) in a different context, to model allele frequency fluctuations of a diploid Wright–Fisher model with balancing selection. However, these oscillations are one-dimensional and do not surround an equilibrium curve, but rather a stable fixed point. Norman (1975) also obtains normally distributed fluctuations of allele frequencies, but in the context of a monoecious diploid WF model or a dioecious diploid Moran model, for which the deterministic forces of selection and/or mutation are stronger than the stochastic genetic drift. In more detail, the allele frequency dynamics in Norman's paper is dominated by a deterministic function that solves an ordinary differential equation according to Haldane's theory (see for instance Chapter 4 of Cavalli-Sforza and Bodmer, 1971). The random fluctuations around this deterministic allele frequency curve are smaller, but in contrast to our results, they occur on the same time scale as the variations of the deterministic curve. These stochastic fluctuations are described by a Gaussian diffusion in the limit of large populations. Norman's results are formulated mathematically in a more general setting though, and we will apply them to the genotype frequency process of the monoecious and diploid Moran model, in order to prove its weak convergence on both time scales simultaneously.

An implication of our results is that the fixation index f_{IS} of Wright (1943) oscillates around a stable fixed point that is only a function of the selfing probability. These oscillations constitute an OU process for the Moran model and an autoregressive process for the WF model. Both these processes have a marginal normal distribution, whose variance and bias are larger for the Moran than for the WF model.

Our paper is organized as follows. We first define the diploid Moran model in Section 2 in terms of a Markov chain. In Section 3 we define some important statistics, such as the expected heterozygosity, effective population size and fixation index. The diffusion approximation of the Moran model is introduced in Section 4, and its continuous time version in Section 5. In Section 6 we introduce more briefly the diploid WF model, and derive its diffusion approximation. Simulation results are presented in Section 7, and extensions are discussed in Section 8. The mathematical derivations have been collected in the appendix.

2. Formulation of monoecious diploid Moran model

We consider one locus and assume two versions or alleles A and a of the gamete. Moran (1958a) noted that a model with overlapping generations, where only one individual is replaced at each instant can be formulated in discrete or continuous time. We will mainly focus on the discrete version, and briefly mention the continuous time extension in Section 5. Time $t=0$ represents the founder generation, and the composition of the founder population is given. The constant number of diploid and monoecious individuals is n. We take into account the probability s of self-fertilization.

The Markov chain for the discrete time diploid Moran model starts with a given founder population. At each of the following time steps ($t=1,2,..$) we work with a probability distribution over all possible populations. In an input (parental) population (time step t) there are $n_1^{(t)}=n_1$ individuals of genotype AA, $n_2^{(t)}=n_2$ genotypes of type Aa, and $n_3^{(t)}=n_3$ individuals of genotype aa. The total number of input individuals is $n=n_1+n_2+n_3$. In each output population there are $\tilde{n}_1=n_1^{(t+1)}$ individuals of genotype AA, $\tilde{n}_2=n_2^{(t+1)}$ individuals of genotype Aa, and $\tilde{n}_3=n_3^{(t+1)}$ individuals of genotype aa. We will not call the output population an offspring population, because there is just one new member appearing at each time step t. It is assumed that the total number of diploid individuals n remains constant for all t, although extensions to a variable population size are discussed in Section 8. This implies in particular that $\tilde{n}_1+\tilde{n}_2+\tilde{n}_2=n$.

At each time step t we first pick one mating individual at random, for reproduction. After that, we pick one more mating individual, with probability s for self-fertilization. These two mating individuals produce an offspring individual, which is put aside for a moment. We can say that the newly formed offspring individual is in exile, temporarily. The next step is that we select one individual at random for removal from the genotype pool. The last thing we do, is to bring the offspring individual in from its exile and put it into the genotype pool to replace the removed individual.

Each time step t thus comprises four substeps: (i) the random picking of two diploid parental individuals, one after the other, with a given probability for self-mating. These parental individuals are put back into the genotype pool after their mating. (ii) The formation of one offspring individual, by combining two haploid gametes at random from the parental individuals. This single newly formed diploid offspring individual is put temporarily into exile. (iii) The random removal of one diploid individual from the genotype pool. (iv) The waiting offspring individual is put into the genotype pool, where it replaces the individual that has been eliminated.

The net effect during one time step is that one diploid individual dies and is replaced by one newly composed individual. This new genotype is composed through random mating in the full genotype pool, including the one that will soon be picked to die and be replaced by the newly composed one. This is an important distinction. The formation of the offspring takes place before an individual is removed from the genotype pool. A different type of Moran model could be constructed if one assumed the reverse order of removal and replication. A stronger genetic drift would result if the dying individual had been excluded from having offspring.

The present model will not take mutation and selection into account, but it could be modified to do so. A probability s of self-fertilization is taken into account though. The case $s=1/n$ is of particular interest because it lets the mating take place in a gamete pool instead of in a genuine genotype pool. It is thus a reference case which can be called standard self-fertilization, where the diploid model effectively degenerates into a haploid model where the heterozygosity is no longer an essential property, since no heterozygosity can be identified in a haploid gamete pool.