



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Theoretical Biology

journal homepage: [www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

## A monoecious and diploid Moran model of random mating

Ola Hössjer<sup>a,\*</sup>, Peder A. Tyvand<sup>b</sup><sup>a</sup> Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden<sup>b</sup> Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, N-1432 Ås, Norway

## H I G H L I G H T S

- 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.

## A R T I C L E I N F O

## 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

## A B S T R A C T

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.

© 2016 Elsevier Ltd. All rights reserved.

## 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](mailto:ola@math.su.se) (O. Hössjer), [peder.tyvand@nmbu.no](mailto:peder.tyvand@nmbu.no) (P.A. Tyvand).

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 ( $t = 1, 2, \dots$ ) 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  $\tilde{n}_1 + \tilde{n}_2 + \tilde{n}_3 = n$ .

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.

2.1. Offspring and elimination probabilities

In the next two subsections, we will look at the transition matrix of the Moran model. The probabilities for the different offspring genotypes  $AA, Aa$  and  $aa$  in the output population are denoted by  $P_1, P_2$  and  $P_3$ , respectively. Following Tyvand (1993), these probabilities are

$$\begin{aligned}
 P_1 &= \frac{(1-s)(n_1+n_2/2)^2+(sn-1)(n_1+n_2/4)}{n(n-1)}, \\
 P_2 &= \frac{(1-s)(n_1n_2+n_3n_2+2n_1n_3+n_2^2/2)+(sn-1)(n_2/2)}{n(n-1)}, \\
 P_3 &= \frac{(1-s)(n_3+n_2/2)^2+(sn-1)(n_3+n_2/4)}{n(n-1)}. \tag{1}
 \end{aligned}$$

We will see that there is one notable difference between our diploid Moran model and the diploid Wright–Fisher model in Tyvand (1993): the case of standard self-fertilization ( $s=1/n$ ) makes the diploid Wright–Fisher model degenerate into its conventional haploid version. In contrast, the case of standard self-fertilization makes our diploid Moran model degenerate into a haploid but modified Moran model. This modified haploid Moran model will have two eliminated gametes replaced by two newly formed gametes at each time step. In the conventional haploid Moran model, only one gamete is eliminated and replaced by a newly formed gamete.

Tyvand (1993) allowed the population size to be different from one time step to the next, involving rectangular transition matrices. The transition matrix is always a square matrix for the Moran model, since we assume that the number of individuals  $n$  in the parental population is the same as the number of individuals in the offspring population. Assuming that all possible states  $(n_2, n_3)$  are ordered as

$$\{(0, 0) = 1, (0, 1) = 2, (1, 0) = 3, (0, 2) = 4, \dots, (n, 0) = d(n)\}, \tag{2}$$

the running indices in the Markov chain algebra are the counting parameter  $i$  for all different possible input populations. It is defined as

$$i = 1 + n_3 + \frac{(n_2 + n_3)(n_2 + n_3 + 1)}{2}, \tag{3}$$

where  $i$  runs from 1 to a value  $d(n)$  that represents the number of different population compositions (where permutations are excluded). This number of different populations is given by

$$d(n) = \frac{(n+1)(n+2)}{2}. \tag{4}$$

The similar counting parameter  $j$  for all possible output populations is

$$j = 1 + \tilde{n}_3 + \frac{(\tilde{n}_2 + \tilde{n}_3)(\tilde{n}_2 + \tilde{n}_3 + 1)}{2}, \tag{5}$$

where  $j$  runs from 1 to  $d(n)$ . Even though  $n_1$  and  $\tilde{n}_1$  are not included in (3) and (5), these equations are essentially transformations from triplets to single natural numbers

$$(n_1, n_2, n_3) \rightarrow i, \quad (\tilde{n}_1, \tilde{n}_2, \tilde{n}_3) \rightarrow j. \tag{6}$$

It is desirable also to establish inverse transformations from natural numbers back to triplets

$$i \rightarrow (n_1, n_2, n_3), \quad j \rightarrow (\tilde{n}_1, \tilde{n}_2, \tilde{n}_3), \tag{7}$$

and for the input population these are given by the formulas

$$n_2 + n_3 = \left\lfloor \frac{\sqrt{8i-7}-1}{2} \right\rfloor, \quad n_3 = i - 1 - \frac{(n_2 + n_3)(n_2 + n_3 + 1)}{2}, \tag{8}$$

where  $\lfloor x \rfloor$  denotes the integer part of  $x$ , which is the largest integer that is smaller or equal to  $x$ . From the relationships (8) we first determine the sum  $n_2 + n_3$ , thereafter  $n_3$  and then  $n_2$  from the sum

that we already know. Once we know  $n_2$  and  $n_3$ , we find  $n_1 = n - n_2 - n_3$ , since the total number of diploid individuals  $n$  has a given constant value. The inverse transformation from the counting parameter  $j$  of Eq. (5) back to the genotype numbers  $(\tilde{n}_1, \tilde{n}_2, \tilde{n}_3)$  for the output population is analogous to (8).

We will now construct the transition matrix  $\mathbf{M} = (M_{ij})_{i,j=1}^{d(n)}$  between all the possible input populations  $i$  and the corresponding output populations  $j$ . We must give seven different expressions for seven different cases.

Define frequencies  $Q_1^{(t)} = Q_1, Q_2^{(t)} = Q_2$  and  $Q_3^{(t)} = Q_3$  at time  $t$ , for the three genotypes  $AA, Aa, aa$ , respectively. They are given by

$$(Q_1, Q_2, Q_3) = \frac{(n_1, n_2, n_3)}{n}, \tag{9}$$

and equal the elimination probabilities in Step (iii). We introduce them here in the interest of compacting formulas for the transition matrix.

There is elimination with full replacement, since the eliminated genotype is replaced by the newly formed offspring genotype. Thus the sum of offspring probabilities adds up to one ( $P_1 + P_2 + P_3 = 1$ ). The sum of elimination probabilities must also add up to one ( $Q_1 + Q_2 + Q_3 = 1$ ). Now we will combine these probabilities of offspring and elimination for each genotype to construct the elements in the transition matrix, where there are seven distinct cases.

(i) *Genotype Aa replaces genotype aa*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1, \quad \tilde{n}_2 = n_2 + 1, \quad \tilde{n}_3 = n_3 - 1. \tag{10}$$

The associated elements  $M_{ij}$  in the transition matrix are given by the product of the probability of reproducing one genotype  $Aa$  and the removal of one genotype  $aa$

$$M_{ij} = P_2 Q_3. \tag{11}$$

(ii) *Genotype aa replaces genotype AA*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1, \quad \tilde{n}_2 = n_2 - 1, \quad \tilde{n}_3 = n_3 + 1. \tag{12}$$

The associated elements in the transition matrix  $M_{ij}$  are given by the product of the probability of reproducing one genotype  $aa$  and the removal of one genotype  $AA$

$$M_{ij} = P_3 Q_2. \tag{13}$$

(iii) *Genotype aa replaces genotype Aa*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1 - 1, \quad \tilde{n}_2 = n_2, \quad \tilde{n}_3 = n_3 + 1. \tag{14}$$

The associated elements in the transition matrix  $M_{ij}$  are given by the product of the probability of reproducing one genotype  $aa$  and the removal of one genotype  $Aa$

$$M_{ij} = P_3 Q_1. \tag{15}$$

(iv) *Genotype AA replaces genotype aa*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1 + 1, \quad \tilde{n}_2 = n_2, \quad \tilde{n}_3 = n_3 - 1. \tag{16}$$

The associated elements in the transition matrix  $M_{ij}$  are given by the product of the probability of reproducing one genotype  $AA$  and the removal of one genotype  $aa$

$$M_{ij} = P_1 Q_3. \tag{17}$$

(v) *Genotype AA replaces genotype Aa*: This is the case where the relationships between the numbers of output and input genotypes

are

$$\tilde{n}_1 = n_1 + 1, \quad \tilde{n}_2 = n_2 - 1, \quad \tilde{n}_3 = n_3. \tag{18}$$

The associated elements in the transition matrix  $M_{ij}$  are given by the product of the probability of reproducing one genotype  $AA$  and the removal of one genotype  $Aa$

$$M_{ij} = P_1 Q_2. \tag{19}$$

(vi) *Genotype Aa replaces genotype AA*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1 - 1, \quad \tilde{n}_2 = n_2 + 1, \quad \tilde{n}_3 = n_3. \tag{20}$$

The associated elements in the transition matrix are given by the product of the probability of reproducing one genotype  $Aa$  and the removal of one genotype  $AA$

$$M_{ij} = P_2 Q_1. \tag{21}$$

(vii) *All genotype numbers AA, Aa, aa remain unchanged*: This is the case where the relationships between the numbers of output and input genotypes are

$$\tilde{n}_1 = n_1, \quad \tilde{n}_2 = n_2, \quad \tilde{n}_3 = n_3, \tag{22}$$

which means that the input and output population numbers are equal. The associated elements along the main diagonal in the transition matrix are given as

$$M_{ij} = P_1 Q_1 + P_2 Q_2 + P_3 Q_3. \tag{23}$$

The sum of all these seven expressions for  $M_{ij}$  is one, when  $i$  is fixed and  $j$  varies. This is a constraint that can be used as a check. All combinations of  $i$  and  $j$  that do not belong to one of these seven categories will give  $M_{ij} = 0$ .

### 2.2. The transition matrix

It is helpful to combine the seven formulas above into one common expression that is valid for all components in the transition matrix of the Markov chain. In order to do this, we introduce the following notation

$$\delta(\alpha, \beta) = \delta_{\alpha\beta} \tag{24}$$

for the Kronecker delta  $\delta_{\alpha\beta}$ , so that  $\delta(\alpha, \beta) = 1$  when  $\alpha = \beta$  and  $\delta(\alpha, \beta) = 0$  when  $\alpha \neq \beta$ . The transition matrix can then be expressed by one formula

$$M_{ij} = P_2 Q_3 \delta(\tilde{n}_1, n_1) \delta(\tilde{n}_2, n_2 + 1) \delta(\tilde{n}_3, n_3 - 1) + P_3 Q_2 \delta(\tilde{n}_1, n_1) \delta(\tilde{n}_2, n_2 - 1) \delta(\tilde{n}_3, n_3 + 1) + P_1 Q_3 \delta(\tilde{n}_1, n_1 + 1) \delta(\tilde{n}_2, n_2) \delta(\tilde{n}_3, n_3 - 1) + P_3 Q_1 \delta(\tilde{n}_1, n_1 - 1) \delta(\tilde{n}_2, n_2) \delta(\tilde{n}_3, n_3 + 1) + P_1 Q_2 \delta(\tilde{n}_1, n_1 + 1) \delta(\tilde{n}_2, n_2 - 1) \delta(\tilde{n}_3, n_3) + P_2 Q_1 \delta(\tilde{n}_1, n_1 - 1) \delta(\tilde{n}_2, n_2 + 1) \delta(\tilde{n}_3, n_3) + (P_1 Q_1 + P_2 Q_2 + P_3 Q_3) \delta(\tilde{n}_1, n_1) \delta(\tilde{n}_2, n_2) \delta(\tilde{n}_3, n_3). \tag{25}$$

A probability distribution vector over all possible population compositions of time steps  $t = 0, 1, \dots$  is given by the vector

$$\mathbf{V}^{(t)} = \left( V_1^{(t)}, \dots, V_{d(n)}^{(t)} \right), \tag{26}$$

where

$$V_i^{(t)} = P \left[ (n_2^{(t)}, n_3^{(t)}) = (n_2, n_3) \right], \quad i = 1, 2, \dots, d(n), \tag{27}$$

$d(n)$  is the population number of formula (4), and  $(n_2, n_3)$  is related to  $i$  as in (3). We may refer to (26) as the population distribution vector at time  $t$ . It represents the probabilities of each of the different populations after  $t$  steps of random mating. The Markov chain algebra produces the output population distribution vector  $\mathbf{V}^{(t+1)}$  from the input population distribution vector  $\mathbf{V}^{(t)}$  by the Kolmogorov–Chapman equation  $\mathbf{V}^{(t+1)} = \mathbf{V}^{(t)} \mathbf{M}$ . In component

form it is given by

$$V_j^{(t+1)} = \sum_{i=1}^{d(n)} V_i^{(t)} M_{ij}. \tag{28}$$

The above-mentioned equation (28) is to be applied recursively for  $t = 0, 1, 2, \dots$ , starting with the founder population  $t=0$ . Since we will choose a specific founder population, the initial vector  $\mathbf{V}^{(0)}$  will by definition only have one nonzero component, according to

$$V_i^{(0)} = \begin{cases} 1, & i = (n_2^{(0)}, n_3^{(0)}), \\ 0, & i \neq (n_2^{(0)}, n_3^{(0)}). \end{cases} \tag{29}$$

### 3. Statistics obtained from the Moran model

The exact stochastic process is represented in our Markov chain computations. We start with choosing a constant population size of  $n$  diploid individuals, and we choose the probability  $s$  of self-fertilization. Next we choose the founder population  $t=0$  to be composed of the numbers of  $(n_1, n_2, n_3)$  individuals of the genotypes  $(AA, Aa, aa)$ , respectively. Thereby an initial population distribution vector  $\mathbf{V}^{(0)}$  with only one nonzero component (29) is specified.

At each time step  $t$  we calculate recursively the inherited population distribution vector  $\mathbf{V}^{(t+1)}$  from its parental distribution vector  $\mathbf{V}^{(t)}$ . The value of  $t$  does not represent a generation, as in the Wright–Fisher model, but a much smaller time unit, since only one individual out of  $n$  individuals is exchanged at each step. Actually there are no distinct generations in the Moran model, since there is only a slow gradual replacement of parental individuals with their offspring. Through the common diffusion approximation we can link a Moran model to its corresponding Wright–Fisher model, see below.

The population distribution vector  $\mathbf{V}^{(t)}$  contains information which may be summarized in a number of ways. We are interested in the average genotype frequencies  $\mathbf{f}^{(t)} = (f_1^{(t)}, f_2^{(t)}, f_3^{(t)})$  at time  $t$  for the three genotypes  $(AA, Aa, aa)$ , respectively. These average genotype frequencies at a given time step  $t$  are given by

$$\mathbf{f}^{(t)} = \sum_{i=1}^{d(n)} \frac{(n_1, n_2, n_3)_i V_i^{(t)}}{n}, \tag{30}$$

where the values of  $n_1, n_2$  and  $n_3$  have to vary with  $i$  in this sum, in a way specified by the relationship (3) or its inverse relationship (8). The frequency  $f_2^{(t)}$  of the heterozygous genotype  $Aa$  deserves special attention, since it is a distinguishing feature of the diploid model. Only when  $s = 1/n$ , the diploid model degenerates to having a haploid gamete pool, where heterozygosity is no longer a property of the model. It is shown in the appendix that

$$f_2^{(t)} = 2x^{(0)}(1-x^{(0)}) \cdot \frac{1-s}{1-\frac{s}{2}} \cdot e^{-t/(nn_e)} + \left[ Q_2^{(0)} - 2x^{(0)}(1-x^{(0)}) \cdot \frac{1-s}{1-\frac{s}{2}} \right] e^{-(1-s/2)t/n} + \frac{1}{n} \cdot x^{(0)}(1-x^{(0)}) \frac{(1-s)\left(3-\frac{s}{2}\right)}{\left(1-\frac{s}{2}\right)^2 \left(1-\frac{s}{2}-\frac{1}{n_e}\right)} \left[ e^{-t/(nn_e)} - e^{-(1-\frac{s}{2})t/n} \right] + o(n^{-1}), \tag{31}$$

where

$$x^{(t)} = \frac{1}{2} Q_2^{(t)} + Q_3^{(t)}$$

is the frequency of allele  $a$  at time  $t$ ,

$$n_e = n \left(1 - \frac{s}{2}\right) \tag{32}$$

is the effective population size for selfing, see Pollak (1997), Nordborg and Donnelly (1987) and Hössjer et al. (2015), and  $o(n^{-1})$  a remainder term of smaller order than  $n^{-1}$ . The first term on the right-hand side of (31) will dominate when  $t$  is large, so that  $\exp[-1/(nn_e)]$  is the multiplicative rate by which the expected heterozygosity decreases per time step. We will interpret  $n$  as an average generation length, since only one randomly chosen individual is replaced at each time point, and therefore the expected life length of each individual is  $n$  time steps. The more selfing there is, the smaller is  $n_e$ , and the faster is the decay rate  $\exp(-1/n_e)$  of heterozygosity per generation,

The standard deviation for the frequency of genotype  $\alpha = 1, 2, 3$  is given by

$$\sigma_\alpha^{(t)} = \sqrt{\sum_{i=1}^{d(m)} \left(\frac{n_\alpha}{n}\right)^2 V_i^{(t)} - (f_\alpha^{(t)})^2} \tag{33}$$

at time point  $t$ . We also introduce the probabilities  $\phi_A^{(t)}$  and  $\phi_a^{(t)}$  for fixation at time  $t$  in the gametes  $A$  and  $a$  respectively. These are given by

$$\begin{aligned} \phi_A^{(t)} &= V_1^{(t)}, \\ \phi_a^{(t)} &= V_{d(m)}^{(t)}. \end{aligned} \tag{34}$$

Under Hardy–Weinberg (HW) equilibrium the frequency of  $Aa$  would be  $2x^{(t)}(1-x^{(t)})$ . In order to quantify departure from random mating in terms of HW-equilibrium, we use the fixation index

$$f_{IS}^{(t)} = \begin{cases} \frac{2x^{(t)}(1-x^{(t)}) - Q_2^{(t)}}{2x^{(t)}(1-x^{(t)})}, & 0 < x^{(t)} < 1, \\ \text{NaN}, & x^{(t)} \in \{0, 1\}, \end{cases} \tag{35}$$

where NaN (Not a Number) means that  $f_{IS}^{(t)}$  is undefined.

#### 4. Approximations for large populations

##### 4.1. Genotype frequency equilibrium curve

Recall that  $n_\alpha^{(t)} = n_\alpha$  and  $n_\alpha^{(t+1)} = \tilde{n}_\alpha$  for  $\alpha = 1, 2, 3$  denote the three genotype counts at time points  $t$  and  $t+1$ . Since  $\tilde{n}_\alpha$  can at most differ from  $n_\alpha$  by one unit, the genotype counts will have a systematic drift

$$E(\tilde{n}_\alpha | n_1, n_2, n_3) - n_\alpha = P(\tilde{n}_\alpha = n_\alpha + 1) - P(\tilde{n}_\alpha = n_\alpha - 1) = P_\alpha - Q_\alpha.$$

A genotype frequency fixed point is obtained by putting  $P_\alpha = Q_\alpha$  for  $\alpha = 1, 2, 3$ . We will derive the large population limit of this fixed point, assuming that the two parents, in the case of no selfing, are drawn with replacement in step (i) of the reproduction cycle between time points  $t$  and  $t+1$ . This corresponds to a  $n = \infty$  limit in (1), with simpler formulas

$$P_1^n \stackrel{\infty}{=} (1-s)(Q_1 + Q_2/2)^2 + s(Q_1 + Q_2/4) = (1-s)(1-x^{(t)})^2 + s(1-x^{(t)}) - \frac{1}{4}sQ_2,$$

$$P_2^n \stackrel{\infty}{=} 1 - Q_1 - Q_3 = 2(1-s)x^{(t)}(1-x^{(t)}) + \frac{1}{2}sQ_2^{(t)},$$

$$P_3^n \stackrel{\infty}{=} (1-s)(Q_3 + Q_2/2)^2 + s(Q_3 + Q_2/4) = (1-s)(x^{(t)})^2 + sx^{(t)} - \frac{1}{4}sQ_2 \tag{36}$$

for all three offspring probabilities. Let  $Q_\alpha(x) = Q_\alpha = P_\alpha$  be the large population fixed points of the frequencies of genotypes  $\alpha = 1, 2, 3$  when the allele frequency is  $x^{(t)} = x$ . By solving (36), we find

that

$$\begin{aligned} Q_1(x) &= (1-s)(1-x)^2 + s(1-x) - x(1-x)\frac{s(1-s)}{2-s}, \\ Q_2(x) &= 2x(1-x)\frac{1-s}{1-s/2}, \\ Q_3(x) &= (1-s)x^2 + sx - x(1-x)\frac{s(1-s)}{2-s}. \end{aligned} \tag{37}$$

It follows from (37) that one single parameter  $x$  suffices to describe the system when the population is large, and  $\{(Q_1(x), Q_2(x), Q_3(x)); 0 < x < 1\}$  constitutes an equilibrium curve of genotype frequencies that equal Hardy Weinberg proportions in the absence of selfing ( $s=0$ ). Inserting (37) into (35) we get a large population fixed point

$$f_{IS}^{fix} = \frac{2x(1-x) - Q_2(x)}{2x(1-x)} = \frac{s}{2-s} \tag{38}$$

of the fixation index if  $0 < x < 1$ . It only depends on the selfing rate, not on the allele frequency  $x$ .

##### 4.2. Diffusion approximations

In this section we derive a large population approximation of the genotype probabilities  $V_i^{(t)}$  that relies on a diffusion fluctuation around (37). Since  $\sum_{\alpha=1}^3 Q_\alpha^{(t)} = 1$ , it suffices to consider the last two genotypes  $\alpha = 2, 3$ . We first rewrite (27) in terms of the frequencies of these two genotypes;

$$V_i^{(t)} = P(Q_2^{(t)} = q_2, Q_3^{(t)} = q_3 | Q_2^{(0)} = \pi_2, Q_3^{(0)} = \pi_3), \tag{39}$$

where  $i$  and  $(q_2, q_3) = (n_2, n_3)/n$  are related through (3). Consequently, the genotype frequencies are defined on a lattice

$$q_2, q_3 \in \left\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\},$$

and satisfy  $0 \leq q_2 + q_3 \leq 1$ . We will treat two cases separately, depending on whether  $(q_2, q_3)$  corresponds to an absorbing state or not. That is, we give separate approximations of  $V_i^{(t)}$  depending on whether any of the two alleles  $A$  or  $a$  has been fixed in the whole population at time  $t$  or not.

##### 4.2.1. Non-absorbing states

If  $(q_2, q_3) \notin \{(0, 0), (0, 1)\}$  is a nonabsorbing state we approximate (39) by

$$V_i^{(t)} \approx n^{-2} f_{\mathbf{Q}^{(t)} | \mathbf{Q}^{(0)}}(q_2, q_3 | \pi_2, \pi_3).$$

This is accurate for large  $n$ , viewing  $\mathbf{Q}^{(t)} = (Q_2^{(t)}, Q_3^{(t)})$  as a two-dimensional absolutely continuous random vector with conditional density function  $f_{\mathbf{Q}^{(t)} | \mathbf{Q}^{(0)}}$  for  $t > 0$ , given the starting genotype configuration  $\mathbf{Q}^{(0)} = (\pi_2, \pi_3)$ . In order to find this density, we need to look a two time scales simultaneously. The first one

$$T = \frac{t}{n} \tag{40}$$

is a generation counter, and the second one

$$\tau = \frac{t}{nn_e} \tag{41}$$

is the time scale (31) for genetic drift. It turns out that allele frequencies change according to the slower time scale  $\tau$ , and it is therefore convenient to introduce the time transformed allele frequency process

$$X(\tau) = x^{(nn_e\tau)}. \tag{42}$$

A process that changes more quickly, on time scale  $T$ , is the excess fraction  $Q_2^{(t)} - Q_2(x^{(t)})$  of heterozygotes compared to an infinitely large population with a given selfing rate  $s$ . This excess fraction gets smaller as  $n$  increases, at rate  $1/\sqrt{n}$ . We introduce the

standardized heterozygosity excess process

$$Y(T) = \sqrt{n}[Q_2^{(nT)} - Q_2(x^{(nT)})], \tag{43}$$

which has a non-vanishing limit for large populations.

With these definitions, we can rewrite the genotype frequencies as

$$\begin{aligned} Q_1^{(t)} &= Q_1(X(\tau)) - Y(T)/(2\sqrt{n}), \\ Q_2^{(t)} &= Q_2(X(\tau)) + Y(T)/\sqrt{n}, \\ Q_3^{(t)} &= Q_3(X(\tau)) - Y(T)/(2\sqrt{n}), \end{aligned} \tag{44}$$

with  $T$  and  $\tau$  as in (40) and (41). Denote the last two components of (44) as  $\mathbf{Q}^{(t)} = \mathbf{G}(X(\tau), Y(T))$ , where  $\mathbf{G} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is nonlinear in  $X$  but linear in  $Y$ . Using transformation rules for probability densities, we verify

$$f_{\mathbf{Q}^{(t)}|\mathbf{Q}^{(0)}}(q_2, q_3 | \pi_2, \pi_3) \approx \sqrt{n} f_{X(\tau)|X(0)}(x|p) f_{Y(T)|X(\tau)}(y|x) \tag{45}$$

in the appendix, where  $p = \pi_2/2 + \pi_3$  is the initial frequency of allele  $a$ , and

$$(x, y) = \mathbf{G}^{-1}(q_2, q_3) = (q_2/2 + q_3, \sqrt{n}(q_2 - Q_2(x)))$$

the inverse transformation from genotype frequencies to allele frequency and heterozygosity excess. It is also shown in the appendix that

$$\begin{aligned} E[X(\tau + \delta) | X(\tau) = x] &= x, \\ \text{Var}[X(\tau + \delta) | X(\tau) = x] &= x(1-x)\delta + o(\delta), \end{aligned} \tag{46}$$

as the time increment  $\delta \rightarrow 0$ , so that  $X(\tau)$  satisfies the same stochastic differential equation as the haploid Wright–Fisher and Moran models. The Kolmogorov forward equation gives a partial differential equation associated with (46), for the allele frequency density. Kimura (1955) derived the solution of this equation in terms of an infinite series

$$f_{X(\tau)|X(0)}(x|p) = p(1-p) \sum_{m=1}^{\infty} m(m+1)(2m+1)e^{-\lambda_m\tau/2} \cdot {}_2F_1(1-m, m+2; 2; p) \cdot {}_2F_1(1-m, m+2; 2; x), \tag{47}$$

see also Section 8.4 of Crow and Kimura (1970). Here  $\lambda_m = m(m+1)$  is the  $m$ th eigenvalue of the partial differential operator that transforms  $f_{X(\tau)|X(0)}(x|p)$  to  $-2\partial f_{X(\tau)|X(0)}(x|p)/\partial\tau$ , and  ${}_2F_1$  is the hypergeometric function, defined in detail for instance in Abramowitz and Stegun (1964).

The diffusion representation of the  $Y$ -process is conditional on the allele frequency process  $X$ . It is proved in the appendix that

$$\begin{aligned} E[Y(T+\delta) - Y(T) | X(\tau) = x, Y(T) = y] &= -\left(1 - \frac{s}{2}\right) \left[ y - \frac{c(x)}{\sqrt{n}} \right] \delta + o(\delta), \\ \text{Var}[Y(T+\delta) | X(\tau) = x, Y(T) = y] &= b(x)x(1-x)\delta + o(\delta) \end{aligned} \tag{48}$$

as  $\delta \rightarrow 0$ , with

$$b(x) = \frac{2(1-s)}{\left(1 - \frac{1}{2}s\right)^2} \left[ 1 + (2x-1)^2 \frac{(1-s)\left(\frac{3}{2}s-1\right)}{1 - \frac{1}{2}s} \right], \tag{49}$$

and

$$c(x) = x(1-x) \frac{(1-s)\left(3 - \frac{s}{2}\right)}{\left(1 - \frac{s}{2}\right)^3}. \tag{50}$$

We notice from (48) that the  $Y$ -process is mean-reverting, with a systematic drift towards  $c(x)/\sqrt{n}$ . Since this number converges to 0 as  $n$  grows, a large population drifts  $Y$  back to 0 and hence tries to maintain heterozygosity balance (37). For large populations, the  $X$ -process changes much more slowly than the  $Y$ -process, so that  $x$  becomes a time invariant constant in (48). This corresponds to a stationary Ornstein–Uhlenbeck process, whose marginal density

function

$$f_{Y(T)|X(\tau)}(y|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{\left[y - \frac{c(x)}{\sqrt{n}}\right]^2}{2\sigma^2(x)}\right) \tag{51}$$

is normal with mean  $c(x)/\sqrt{n}$  and variance

$$\sigma^2(x) = \frac{b(x)x(1-x)}{2-s}. \tag{52}$$

In the appendix we use weak convergence results of Norman (1975), Ethier and Nagylaki (1980, 1988) and Ethier and Kurz (1986) in order to motivate that the joint distribution of the two processes (42) and (43) converge to (47) and (51) on the local time scale as  $n \rightarrow \infty$ .

It follows from (38), (44) and (51) that the fixation index (35) is approximately normally distributed

$$\begin{aligned} f_{IS}^{(t)} | X(\tau) = x &\sim N\left(\frac{s}{2-s} - \frac{c(x)}{2x(1-x)} \cdot \frac{1}{n}, \frac{\sigma(x)^2}{(2x(1-x))^2} \cdot \frac{1}{n}\right) \\ &= N\left(\frac{s}{2-s} - \frac{(1-s)\left(3 - \frac{s}{2}\right)}{2\left(1 - \frac{s}{2}\right)^3} \cdot \frac{1}{n}, \frac{b(x)}{4(2-s)x(1-x)} \cdot \frac{1}{n}\right), \end{aligned} \tag{53}$$

conditionally on an allele frequencies before fixation ( $0 < x < 1$ ). Of particular interest is the case of no selfing  $s=0$ . Then  $b(x) = 8x(1-x)$ , so that the fixation index, to a first approximation, has a distribution

$$f_{IS}^{(t)} | X(\tau) = x \sim N\left(-\frac{3}{2n}, \frac{1}{n}\right) \tag{54}$$

independently of the allele frequency  $x$ .

#### 4.2.2. Absorbing states

The probabilities (39) for the two absorbing states  $(q_2, q_3) \in \{(0, 0), (0, 1)\}$  correspond fixation of  $A$  and  $a$  respectively, as defined in (34). Reaching any of these two states is equivalent to the allele frequency process  $x^{(t)}$  hitting any of its two absorbing barriers 0 or 1. In Section 4.2.1 we found that the allele frequency process has the same diffusion limit as the haploid Wright Fisher and Moran models, on a time scale (41). The fixation probability for this process at the upper boundary 1 is

$$\begin{aligned} \Phi(\tau | p) &= P(X(\tau) = 1 | X(0) = p) \\ &= p + p(1-p) \sum_{m=1}^{\infty} (2m+1)(-1)^m e^{-\lambda_m\tau/2} \cdot {}_2F_1(1-m, m+2; 2; p), \end{aligned} \tag{55}$$

see for instance Sections 8.4 and 8.8 of Crow and Kimura (1970), or McKane and Waxman (2007). Since the two alleles are treated symmetrically, this yields approximations

$$\begin{aligned} V_1^{(t)} &= \phi_A^{(t)} \approx \Phi(\tau | 1-p), \\ V_{d(n)}^{(t)} &= \phi_a^{(t)} \approx \Phi(\tau | p) \end{aligned} \tag{56}$$

of the exact fixation probabilities at both boundaries at any time point  $t$ .

### 5. Continuous time Moran model

It is possible to define a continuous time version of the diploid Moran model with time parameter  $T \geq 0$ . Individuals are born as in Section 2. But this does not happen at equidistant time points, but rather according to a Poisson process with rate  $n$ , at time points  $0 < T^{(1)} < T^{(2)} < \dots$ . This implies in particular that  $T^{(t)} - T^{(t-1)}$  are independent and exponentially distributed random variables with

mean  $1/n$ . Since the expected number of generations  $t$  before an individual dies is  $n$ , the expected life time is  $n \cdot 1/n = 1$ . The continuous time Moran model can be represented as a Markov process

$$(\bar{n}_2^{(T)}, \bar{n}_3^{(T)}) = n(\bar{Q}_2^{(T)}, \bar{Q}_3^{(T)})$$

for  $T \geq 0$ , where  $\bar{n}_2^{(T)} = n_2^{(t)}$  is the genotype frequency of  $Aa$  for  $T^{(t)} \leq T < T^{(t+1)}$ , with  $T^{(0)} = 0$ . Similarly,  $\bar{n}_3^{(T)} = n_3^{(t)}$  is the genotype frequency of  $aa$  during the same time interval. This implies in particular that the discrete time Moran model  $(n_2^{(t)}, n_3^{(t)})$ ,  $t = 0, 1, 2, \dots$ , is the embedded Markov chain of the continuous time Moran model, obtained by recording values of the latter process after each jump.

Since the number of births from 0 to  $T$  is Poisson distributed with mean  $nT$ , it follows from the theory of Markov processes (see for instance [Grimmett and Stirzaker, 2001](#)) that the transition matrix over a time interval of length  $T$  has entries

$$P_{ij}(T) = P\left((\bar{n}_2^{(T)}, \bar{n}_3^{(T)}) = j \mid (\bar{n}_2^{(0)}, \bar{n}_3^{(0)}) = i\right) = \sum_{k=0}^{\infty} e^{-nT} \frac{(nT)^k}{k!} M_{ij}^{(k)} = \exp(t\mathbf{A})_{ij},$$

for  $1 \leq i, j \leq d(n)$ . Here  $M_{ij}^{(k)}$  denotes element  $(ij)$  of  $\mathbf{M}^k$ , and  $\mathbf{A} = (A_{ij})_{i,j=1}^{d(n)}$  is the intensity matrix, with elements

$$A_{ij} = \begin{cases} nM_{ij}, & j \neq i, \\ -\sum_{k:k \neq i} A_{ik} = -n, & j = i, \end{cases}$$

that equal the rates between all pairs of different states, and minus the rate to leave states along the diagonal.

By the law of large numbers for the Poisson distribution, the number of births up to time  $T$  is  $nT(1 + o(1))$  when  $n$  is large. The triplet  $(\bar{Q}_1^{(T)}, \bar{Q}_2^{(T)}, \bar{Q}_3^{(T)})$  of genotype frequencies for the continuous time Moran model will therefore be well approximated by the right-hand side of (44), with  $X(\tau)$  and  $Y(T)$  as in (47) and (51), and  $\tau = T/n_e$ . Moreover, the joint distribution of the frequencies  $(\bar{Q}_2^{(T)}, \bar{Q}_3^{(T)})$  of genotypes  $Aa$  and  $aa$ , is well approximated by the right-hand side of (45).

### 6. Monoecious diploid Wright–Fisher model

For the monoecious diploid Wright–Fisher model we let

$$(n_0^{(r)}, n_1^{(r)}, n_2^{(r)}) = n(Q_0^{(r)}, Q_1^{(r)}, Q_2^{(r)})$$

be the number of  $AA$ ,  $Aa$  and  $aa$  genotypes in generation  $r = 0, 1, 2, \dots$ . This is a discrete time Markov chain with the same state space (2) as for the Moran model. If coding (3)–(5) is used for these states, we get multinomial transition probabilities

$$M_{ij} = \frac{n!}{\tilde{n}_1! \tilde{n}_2! \tilde{n}_3!} P_1^{\tilde{n}_1} P_2^{\tilde{n}_2} P_3^{\tilde{n}_3}, \tag{57}$$

with  $P_\alpha$  as in (1),  $n_\alpha = n_\alpha^{(r)}$  and  $\tilde{n}_\alpha = n_\alpha^{(r+1)}$ , see [Tyvand \(1993\)](#). In the haploid case, it is well-known that one WF generation corresponds to half a Moran generation. We will prove that the same is true in the diploid case. To this end, we introduce a time scale

$$\tau = \frac{r}{2n_e}$$

of genetic drift, let  $x^{(r)} = Q_2^{(r)}/2 + Q_3^{(r)}$  be the frequency of allele  $a$  in WF generation  $r$ , and

$$X(\tau) = x^{(2n_e\tau)}$$

the corresponding time transformed process. In order to model oscillations around the equilibrium curve (37) of genotype

frequencies, we define the Moran model generation number as

$$T = \frac{r}{2},$$

and let

$$Y(T) = \sqrt{n} [Q_2^{(2T)} - Q_2(x^{(2T)})] \tag{58}$$

be the time transformed and normalized excess of heterozygotes process at time  $T$ . An important distinction from the Moran model is that  $T \in \{0, 1/2, 1, 3/2, \dots\}$  is defined on the same lattice for each  $n$ , and it is *not* a continuous parameter in the limit of large populations.

Notice that the  $Aa$  and  $aa$  genotype frequencies  $(Q_2^{(r)}, Q_3^{(r)})$  of generation  $r$  still satisfy (44), provided we change the Moran time step index  $t$  to  $r$  on the left-hand side of this equation. Likewise, the joint distribution of these two genotypes is approximated by (45). In order to find this approximate distribution, we prove in the appendix that  $X(\tau)$  satisfies the same diffusion equation (46) as for the Moran model, with an asymptotic distribution given by (47). The distribution of (58) is on the other hand different from (51). We prove in the appendix that asymptotically for large population, (58) is an autoregressive process

$$Y\left(T + \frac{1}{2}\right) - \frac{\tilde{c}(x)}{\sqrt{n}} = \frac{s}{2} \left[ Y(T) - \frac{\tilde{c}(x)}{\sqrt{n}} \right] + \varepsilon(T), \tag{59}$$

of order 1, with a bias term that involves

$$\tilde{c}(x) = x(1-x) \frac{(1-s)(2-\frac{s}{2})}{(1-\frac{s}{2})^3}, \tag{60}$$

and an innovation term  $\varepsilon(T) \sim N(0, b(x)x(1-x)/2)$  that is normally distributed conditionally on  $X(\tau) = x$ , with  $b(x)$  as in (49). Combining (59) with time theory analysis (see for instance [Brockwell and Davis, 1991](#)), we find that the normal marginal distribution

$$f_{Y(T)|X(\tau)=x}(y|x) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2(x)}} \exp\left(-\frac{\left[y - \frac{\tilde{c}(x)}{\sqrt{n}}\right]^2}{2\tilde{\sigma}^2(x)}\right)$$

of  $Y(T)$  conditionally on the allele frequency process, with variance

$$\tilde{\sigma}^2(x) = \frac{\text{Var}[\varepsilon(T)|X(\tau)=x]}{1 - (\frac{s}{2})^2} = \frac{b(x)x(1-x)}{2 - \frac{s^2}{2}}. \tag{61}$$

Comparing (60) and (61) with (50) and (52), we notice that  $\tilde{c}(x) < c(x)$  and  $\tilde{\sigma}^2(x) \leq \sigma^2(x)$ , with equality if and only if  $s=0$ . This implies that the frequency of heterozygotes  $Aa$  on average departs less from the stable fixed point (37) for the WF model than for the Moran model.

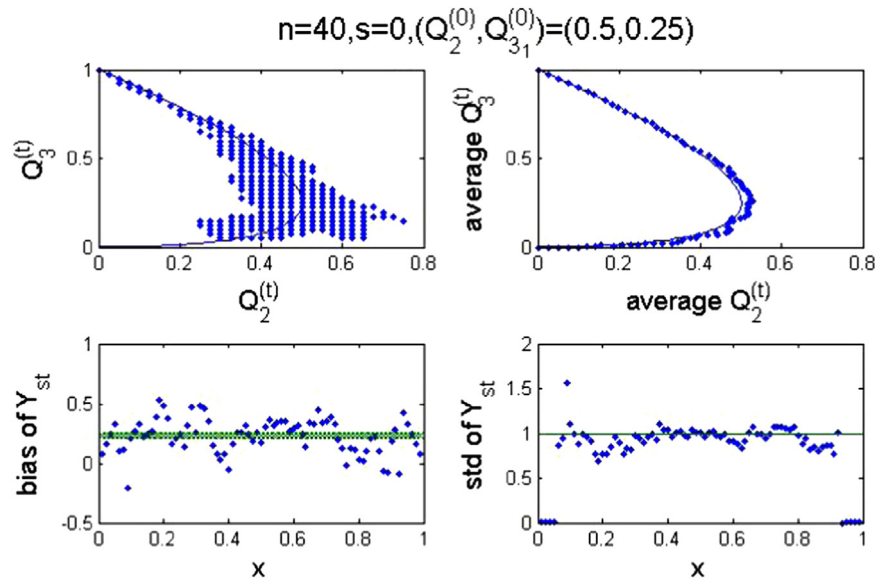
For the fixation index we obtain an approximate normal distribution

$$f_{IS}^{(r)} | X(\tau) = x \sim N\left(\frac{s}{2-s} - \frac{(1-s)(2-\frac{s}{2})}{2(1-\frac{s}{2})^3} \cdot \frac{1}{n}, \frac{b(x)}{4(2-\frac{s^2}{2})x(1-x)} \cdot \frac{1}{n}\right),$$

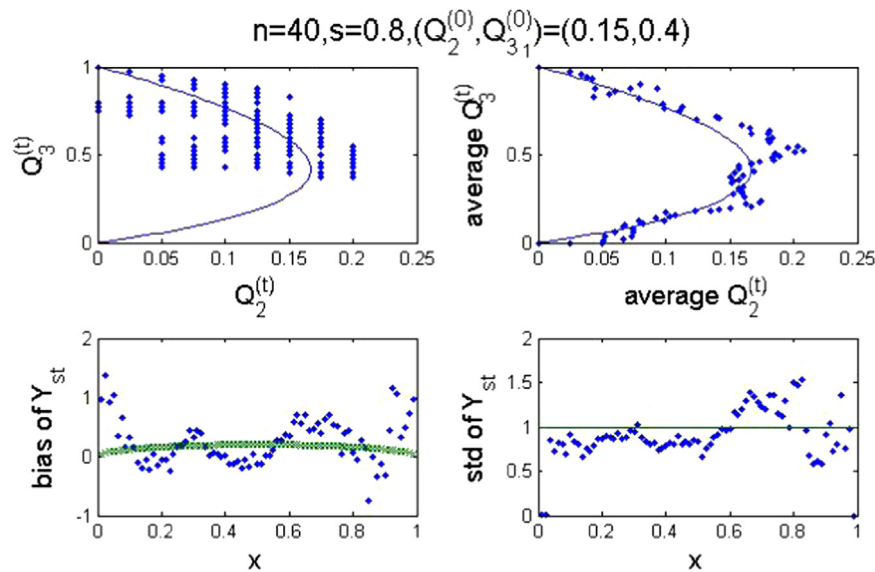
similarly as in (53). This expression simplifies a lot when there is no selfing ( $s=0$ ). Then the fixation index process has independent and identically distributed components

$$f_{IS}^{(r)} | X(\tau) = x \sim N\left(-\frac{1}{n}, \frac{1}{n}\right) \tag{62}$$

for  $r = 1, 2, \dots$ , independently of the allele frequency  $x$ . This Levene effect can more easily be obtained by a direct argument, see for instance [Levene \(1949\)](#) and Section 2.10 of [Crow and Kimura \(1970\)](#). It confirms a well-known fact that the fraction of



**Fig. 1.** Simulation results for a diploid Moran population of size  $n=40$  with no selfing ( $s=0$ ) that starts with genotype frequencies  $(Q_2^{(0)}, Q_3^{(0)}) = (0.5, 0.25)$ . Upper left is a scatterplot from one time series  $(Q_2^{(t)}, Q_3^{(t)})$ ,  $t = 1, \dots, N = 10,000$ , together with the limit (solid) curve  $\{(Q_2(x), Q_3(x)); 0 \leq x \leq 1\}$  for large populations. The remaining three subplots give aggregated results from ten time series of length  $N$ , conditionally on the allele frequency process  $x^{(t)} = Q_2^{(t)}/2 + Q_3^{(t)} = x$ . The upper right subplot depicts the average of  $(Q_2^{(t)}, Q_3^{(t)})$  for each  $x$ . The two lower subplots show for each  $x$  estimates of the mean and standard deviation of the standardized heterozygosity excess process  $Y_{st} = Y(T)/\sigma(x^{(t)})$ , with  $Y(T)$  as in (43). The corresponding predictions from the diffusion approximation in (51) are  $c(x)/(\sqrt{\pi}\sigma(x))$  (thick line, left) and 1 (thin line, right).



**Fig. 2.** Simulation results for a diploid Moran population of size  $n=40$  with selfing probability  $s=0.8$  that starts with genotype frequencies  $(Q_2^{(0)}, Q_3^{(0)}) = (0.15, 0.4)$ . See Fig. 1 for a description of the four subplots.

heterozygotes of a randomly mating WF population has a slight upward bias compared to HW equilibrium.

### 7. Numerical results

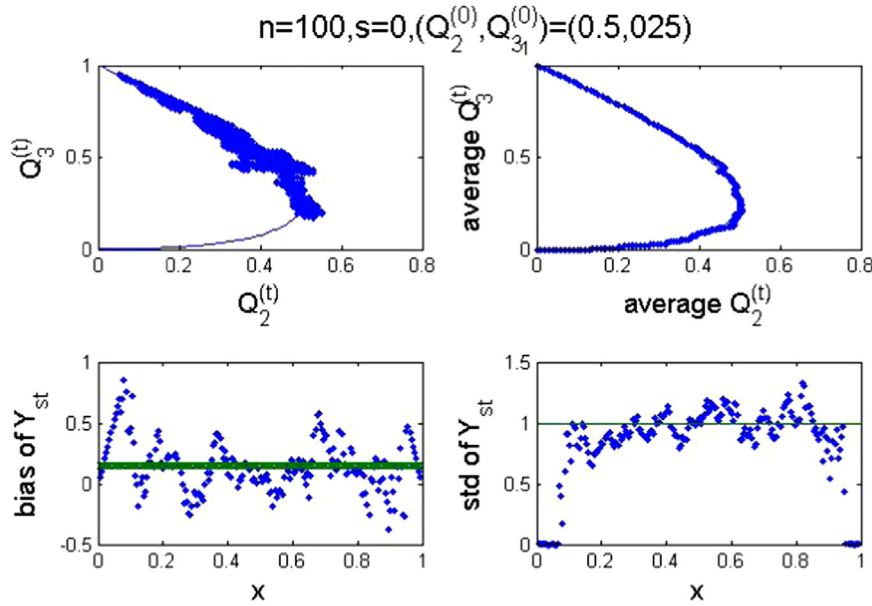
We will show some numerical results from computing the exact Markov chain of the diploid discrete time Moran model, as well as the diffusion approximation.

Scatterplots of  $(Q_2^{(t)}, Q_3^{(t)})$  are displayed in the upper left graphs of Figs. 1–3, for two population sizes  $n=40$  and  $n=100$ , and two selfing rates  $s=0$  and  $s=0.8$ . The distance between the realized genotype frequencies and the equilibrium curve (37) is sometimes quite substantial for the smaller population, although the average genotype

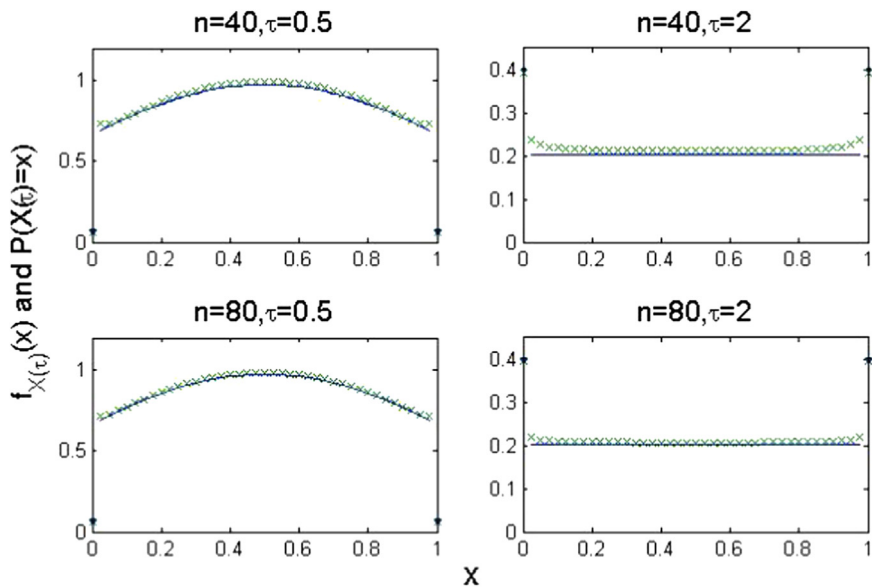
frequencies over 10 realizations follow the curve closely. The bias and standard deviation of this distance are estimated in the lower subplots, and they agree well with the mean and standard deviation of the normal distribution in (51).

Fig. 4 illustrates the accuracy of the diffusion approximation for the allele frequency distribution at interior points (47) and at the two boundaries (56) when the initial allele frequency  $p=0.5$ . It can be seen that the approximation is good already for  $n=40$ , and even better for  $n=80$ . Notice also that the conditional distribution of the allele frequency distribution conditional on non-fixation is close to uniform when  $\tau=2$ , but not for  $\tau=0.5$ . The reason is that the first eigenfunction  $x \rightarrow {}_2F_1(0, 3; 2; x)$  in (47) is uniform, and its coefficient  $6e^{-\tau/2} {}_2F_1(0, 3; 2; 0.5)$  will dominate the infinite series when  $\tau=2$ .





**Fig. 3.** Simulation results for a diploid Moran population of size  $n=100$  with no selfing ( $s=0$ ) that starts with genotype frequencies  $(Q_2^{(0)}, Q_3^{(0)}) = (0.15, 0.4)$ . See Fig. 1 for a description of the four subplots.



**Fig. 4.** Exact (crosses) and approximate (dots) distribution of allele frequency  $x$  for a diploid Moran model of size  $n$  with no selfing ( $s=0$ ) at standardized time  $\tau$ , defined in (41). It is assumed that  $x = v/(2n)$ , with  $v = 0, \dots, 2n$ . The density  $f_{X(\tau)}(x)$  in (47) that solves the diffusion equation, and the standardized probability  $2nP(x^{(2n\tau)} = x)$ , are plotted for interior allele frequencies, whereas the approximate and exact fixation probabilities (56) are displayed at the two boundary points  $x \in \{0, 1\}$ .

In order to check the accuracy of the diffusion approximation of the distribution of  $Y(T)|X(\tau)$ , we plot the ratio of

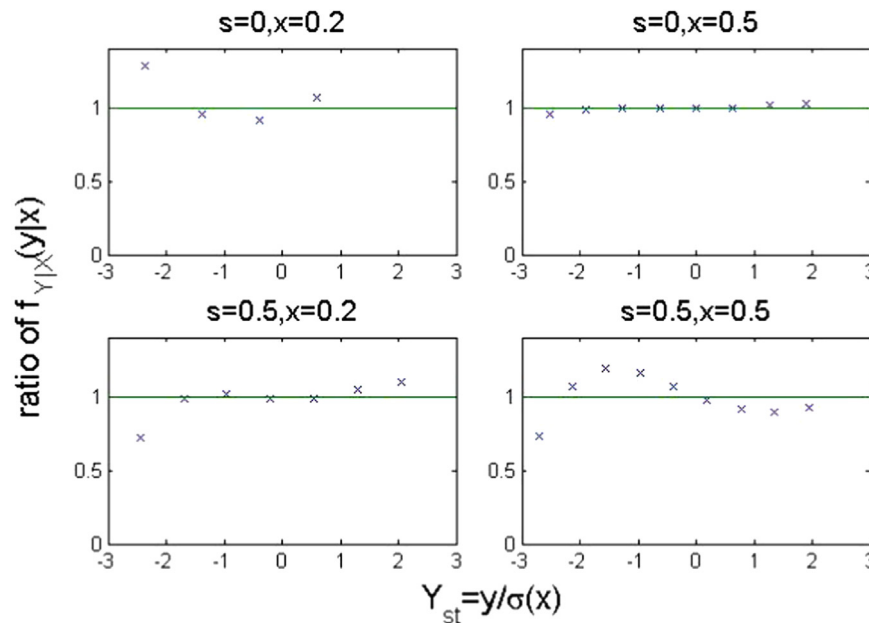
Without this correction, the ratio deviates much more from unity (results not shown).

$$\hat{f}_{Y(T)|X(\tau)}(y|x) = \frac{\sqrt{n}}{2} \frac{P\left(\left(n_2^{(t)}, n_3^{(t)}\right) = n \left[ Q_2(x) + \frac{y}{\sqrt{n}}, Q_3(x) - \frac{y}{\sqrt{2n}} \right] \right)}{P\left(n_2^{(t)} + 2n_3^{(t)} = 2nx\right)}, \quad (63)$$

**8. Conclusions**

and (51) in Fig. 5, for different values of  $x$  and  $s$ , when  $n=40$ . The first term on the right-hand side of (63) takes into account that  $Y|X(\tau) = x$  varies on a lattice of width  $2/\sqrt{n}$  when  $x$  is fixed. It is seen from the figure that the ratio of the two densities is quite close to 1, indicating that the diffusion approximation is accurate. It is crucial in this context to include in (51) the bias term  $c(x)/\sqrt{n}$ .

In this paper we introduced an exact monoecious diploid Moran model and compared it with the monoecious diploid Wright–Fisher model. We also derived novel diffusion approximations of both models on two time scales, where the more slowly varying allele frequency process is asymptotically the same for both models, whereas the more rapid oscillations of genotype frequencies around an equilibrium curve, are described by different stochastic processes for the two models.



**Fig. 5.** Ratio of exact and asymptotic approximations, (63) and  $f_{Y(T)|X(\tau)}(y|x)$ , of the conditional distribution of  $Y(T)$  given  $X(\tau)$ . It is plotted as a function of  $Y_{st} = Y(T)/\sigma(X(\tau))$  when  $X(\tau) = x$  is fixed. The population size is  $n = 40$  in all four subplots, the standardized time in (41) is  $\tau = 2$ , and the starting configuration  $(Q_2^{(0)}, Q_3^{(0)})$  of the genotype frequencies of  $Aa$  and  $aa$ , is  $(0.5, 0.25)$  in the upper subplots ( $s = 0$ ) and  $(0.325, 0.325)$  in the lower subplots ( $s = 0.5$ ).

The results of this paper can be extended in several ways. First, it is possible to define a more general hybrid model that includes the Moran model and the Wright–Fisher model as special cases. Such a hybrid model can have a given number of  $n$  diploid individuals, and in each generation it lets a number  $m$  of offsprings be generated. At the next time step,  $m$  diploid individuals are eliminated and replaced by the  $m$  newly formed offspring. The present Moran model represents  $m = 1$ , and if all offspring choose their parents independently, the diploid Wright–Fisher model corresponds to  $m = n$ . But it is also possible with more skewed offspring distributions, where, for instance, one single couple is the parent of all  $m$  offspring.

When  $s = 1/n$ , we get a haploid hybrid model where  $2m$  gametes are replaced in each generation. If the new gametes are chosen randomly with replacement from the previous generation and  $m = n$ , we get an ordinary haploid Wright Fisher model of size  $2n$ . A proper rescaling of the number of  $a$ -alleles  $n_2^{(t)} + 2n_3^{(t)}$  at time  $t = 0, 1, 2, \dots$  for this haploid model converges to the same asymptotic diffusion limits (47) and (55) as for the diploid allele frequency process, for any value of  $m$ . On the other hand, if all  $2m$  gametes have the same parents, we obtain the model of Eldon and Wakeley (2006). These haploid reproduction scenarios are all instances of the exchangeable Cannings model of a homogeneous population (Cannings, 1974).

Second, it is not immediately clear how the Moran model should be generalized to take a varying population size into account. The most natural extension is to allow the number of individuals  $m^{(t)}$  and  $q^{(t)}$  that are born or die, to vary with time. If  $q^{(t)}$  differs from  $m^{(t)}$ , the population size will change from  $n^{(t)}$  to  $n^{(t)} + m^{(t)} - q^{(t)}$  between time points  $t$  and  $t + 1$ .

Third, we believe the diffusion approximation approach with two time scales is even more important for higher dimensional processes. Whereas our genotype process is two-dimensional, exact Markov chain computation becomes infeasible for all but very small populations in higher dimensions. This includes the diploid two-sex Moran and WF models of a biallelic gene. It requires four-dimensional processes, with separate genotype

frequencies of males and females. It is well-known that this process has small oscillations around a one-dimensional curve of Hardy–Weinberg proportions when there is no selfing (Ethier and Nagylaki, 1980, 1988). The position along this curve depends on the allele frequency, and the allele frequency process is asymptotically the same as for a haploid model (46)–(47). We conjecture that the genotype frequency oscillations around this curve follow a multivariate OU-process for the Moran model and multivariate autoregressive process for the WF model, on a local time scale.

Fourth, it is of interest to work out the dynamics of a biallelic gene in a haploid subdivided population with non-overlapping generations and strong migration between the  $S$  subpopulations. Nagylaki (1980) has shown that a reproductively weighted (Fisher, 1958) average of the subunits allele frequencies gives a global allele frequency that determines genetic drift according to (46)–(47), on a slowly varying time scale determined by the variance effective population size. We conjecture that the  $S$ -dimensional vector of local allele frequencies for all subpopulations will oscillate around the line  $\{(x, \dots, x); 0 < x < 1\}$  on a local time scale, as a multivariate autoregressive process of order 1. An implication of this result would be that the fixation index  $f_{ST}$  of Wright (1943), which quantifies the amount of subpopulation differentiation, will exhibit oscillations around a quasi-equilibrium fixed point obtained explicitly in Hössjer et al. (2013) for the island model, and by Hössjer and Ryman (2014) for more general models.

### Acknowledgements

Ola Hössjer was financially supported by the Swedish Research Council, Contract no. 621-2013-4633. The authors want to thank one anonymous reviewer for helpful comments that considerably improved our paper, in particular the reference to Norman (1975).

**Appendix A. Mathematical proofs and motivations**

**A.1. Motivation of (45)**

It follows from transformation rules of densities that

$$\begin{aligned}
 f_{\mathbf{Q}^{(t)}|\mathbf{Q}^{(0)}}(q_2, q_3 | \pi_2, \pi_3) &= |J(x, y)|^{-1} f_{X(\tau)|X(0), Y(0)}(x | p, y_0) \\
 &\quad \cdot f_{Y(T)|X(\tau), X(0), Y(0)}(y | x, p, y_0) \\
 &= \sqrt{nf} f_{X(\tau)|X(0), Y(0)}(x | p, y_0) \\
 &\quad \cdot f_{Y(T)|X(\tau), X(0), Y(0)}(y | x, p, y_0) \\
 &\approx \sqrt{nf} f_{X(\tau)|X(0)}(x | p) f_{Y(T)|X(\tau)}(y | x), \tag{A.1}
 \end{aligned}$$

where  $y_0 = \sqrt{n}[\pi_2 - Q_2(p)]$  is the value of the heterozygous excess process at time 0. of (A.1) we used that the determinant of the Jacobian equals

$$|J(x, y)| = \left| \frac{d\mathbf{G}(x, y)}{d(x, y)} \right| = \begin{vmatrix} (1-2x)\frac{2(1-s)}{1-s/2} & 2(1-s)x + s - \frac{1}{2}(1-2x)\frac{s(1-s)}{1-s/2} \\ \frac{1}{\sqrt{n}} & -\frac{1}{2\sqrt{n}} \end{vmatrix} = \frac{1}{\sqrt{n}},$$

and in the last step that asymptotically for large time points and  $n$ , it is only the allele frequency  $p$  at time 0 that influences the joint distribution of  $X(\tau)$  and  $Y(T)$ . This will be motivated further in the weak convergence proof below.

**A.2. Proof of (46)**

Let

$$v^{(t)} = n_2^{(t)} + 2n_3^{(t)} = 2nx^{(t)} = 2nX\left(\frac{t}{nn_e}\right) \tag{A.2}$$

and

$$w^{(t)} = n_2^{(t)} - nQ_2(x^{(t)}) = \sqrt{n}Y\left(\frac{t}{n}\right) \tag{A.3}$$

refer to the number of  $a$ -alleles and excess number of  $Aa$  genotypes (compared to fixed point (37)) at time point  $t$ .

It follows from (32) and (42) that Eq. (46) is equivalent to

$$E(v^{(t+1)} | v^{(t)} = v) = v,$$

$$\text{Var}(v^{(t+1)} | v^{(t)} = v = 2nx) = \frac{(2n)^2}{nn_e} x(1-x) + o(nn_e^{-1}) = \frac{4x(1-x)}{1-s/2} + o(1) \tag{A.4}$$

as  $n \rightarrow \infty$ . In order to prove (A.4), define  $Q_\alpha = Q_\alpha^{(t)}$  and  $P_\alpha = P_\alpha^{(t)}$  for  $\alpha = 1, 2, 3$ , and recall that the genotype frequency transition between time points  $t$  and  $t+1$  was divided into seven cases (i)–(vii) in Section 2.1. Only the first six of these, (i)–(vi), give a non-zero value of  $v^{(t+1)} - v^{(t)}$  that equals either of  $-2, -1, 1, 2$ . Summing over all possible  $v^{(t+1)} - v^{(t)}$ , weighted by their probabilities to occur, we find that

$$\begin{aligned}
 E(v^{(t+1)} - v^{(t)} | v^{(t)} = v, w^{(t)} = w) &= -P_2Q_3 + P_3Q_2 + 2P_3Q_1 - 2P_1Q_3 - P_1Q_2 + P_2Q_1 \\
 &= P_3 - P_1 - (Q_3 - Q_1) \\
 &= \frac{(1-s)(n_1 + n_2 + n_3)(n_3 - n_1) + (sn - 1)(n_3 - n_1)}{n(n-1)} - \frac{n_3 - n_1}{n} = 0, \tag{A.5}
 \end{aligned}$$

where in the second step we made some rearrangements of terms and used

$$P_1 + P_2 + P_3 = Q_1 + Q_2 + Q_3 = 1, \tag{A.6}$$

and in the third step we inserted the definitions of  $P_1$  and  $P_3$  in (1). Averaging over  $w^{(t)}$  we then find that

$$E(v^{(t+1)} - v^{(t)} | v^{(t)} = v) = E[E(v^{(t+1)} - v^{(t)} | v^{(t)} = v, w^{(t)} = w)] = 0. \tag{A.7}$$

In order to verify the second part of (A.4), we notice that the genotype frequencies for large populations are close to the fixed

point (37), so that

$$\begin{aligned}
 P_\alpha &= Q_\alpha(x) + o(1), \\
 Q_\alpha &= Q_\alpha(x) + o(1), \tag{A.8}
 \end{aligned}$$

for  $\alpha = 1, 2, 3$  as  $n \rightarrow \infty$ . This gives

$$\begin{aligned}
 \text{Var}(v^{(t+1)} | v^{(t)} = 2nx, w^{(t)} = w) &= E[(v^{(t+1)} - v^{(t)})^2 | v^{(t)} = 2nx, w^{(t)} = w] \\
 &= P_2Q_3 + P_3Q_2 + 2^2P_3Q_1 + 2^2P_1Q_3 + P_1Q_2 + P_2Q_1 \\
 &= 2Q_2(x)(1 - Q_2(x)) + 8Q_1(x)Q_3(x) + o(1) \\
 &= \frac{4x(1-x)}{1-\frac{s}{2}} + o(1), \tag{A.9}
 \end{aligned}$$

as was to be proved. The last step of (A.9) follows after some calculations from the definitions of  $Q_\alpha(x)$  in (37). We finally deduce the second part of (A.4) by averaging over  $w^{(t)}$ , similarly as in (A.7).  $\square$

**A.3. Proof of (48)**

We will prove that

$$\begin{aligned}
 E(w^{(t+1)} - w^{(t)} | x^{(t)} = x, w^{(t)} = w) &= -\left(1 - \frac{s}{2}\right)\frac{w}{n} + \frac{C}{n} + o(n^{-1}), \\
 \text{Var}(w^{(t+1)} | x^{(t)} = x, w^{(t)} = w) &= b(x)x(1-x) + o(1) \tag{A.10}
 \end{aligned}$$

as  $n \rightarrow \infty$ , with

$$C = 2x(1-x)\frac{1-s}{2-s} - Q_2(x)\frac{x(1-x)}{2(1-\frac{s}{2})} = x(1-x)\frac{1-s}{1-\frac{s}{2}} + \frac{2x(1-x)(1-s)}{(1-\frac{s}{2})^2} =: C_1 + C_2 = c(x)\left(1 - \frac{s}{2}\right) \tag{A.11}$$

a constant,  $Q_2'(x)$  refers to the second derivative of the middle equation in (37) and  $c(x)$  is the function defined in (50). In order to verify the first part of (A.10), we Taylor expand  $Q_2(\cdot)$  around  $x$  and use (A.2)–(A.4) to deduce that

$$\begin{aligned}
 E(w^{(t+1)} - w^{(t)} | x^{(t)} = x, w^{(t)} = w) &= E\left(n_2^{(t+1)} - n_2^{(t)} | x^{(t)} = x, w^{(t)} = w\right) \\
 &= -n \cdot \frac{1}{2n} Q_2'(x) E(v^{(t+1)} - v^{(t)} | x^{(t)} = x, w^{(t)} = w) \\
 &= -n \cdot \frac{1}{2(2n)^2} Q_2''(x) E[(v^{(t+1)} - v^{(t)})^2 | x^{(t)} = x, w^{(t)} = w] \\
 &= E\left(n_2^{(t+1)} - n_2^{(t)} | x^{(t)} = x, w^{(t)} = w\right) + \frac{C_2}{n} + o(n^{-1}) \\
 &= P_2Q_3 - P_3Q_2 + P_2Q_1 - P_1Q_2 + \frac{C_2}{n} + o(n^{-1}) \\
 &= P_2 - Q_2 + \frac{C_2}{n} + o(n^{-1}) = \left[(1-s)2x(1-x)\frac{n}{n-1} + \frac{1}{2}Q_2s\frac{n-1/s}{n-1}\right] - Q_2 + \frac{C_2}{n} + o(n^{-1}) \\
 &= (1-s)2x(1-x)\frac{n}{n-1} + \left[2x(1-x)\frac{1-s}{1-s/2} + \frac{w}{n}\right] \left[\frac{1}{2}s - 1 + \frac{s-1}{2(n-1)}\right] + \frac{C_2}{n} + o(n^{-1}) \\
 &= -\left(1 - \frac{s}{2}\right)\frac{w}{n} + \frac{C_1}{n} + \frac{C_2}{n} + o(n^{-1}) \\
 &= -\left(1 - \frac{s}{2}\right)\frac{w}{n} + \frac{C}{n} + o(n^{-1})
 \end{aligned}$$

where in the fourth step we used (A.6), in the fifth step the definition of  $P_2$  in (1) and in the sixth step we combined (37) and (A.3) to deduce that

$$Q_2 = 2x(1-x)\frac{1-s}{1-\frac{s}{2}} + \frac{w}{n}.$$

In order to verify the second part of (A.10), we rewrite the conditional variance of  $w^{(t+1)} - w^{(t)}$  as

$$\begin{aligned}
 \text{Var}(w^{(t+1)} - w^{(t)} | x^{(t)} = x, w^{(t)} = w) &= E[(w^{(t+1)} - w^{(t)})^2 | x^{(t)} = x, w^{(t)} = w] + o(1) \\
 &= E\left\{\left[n_2^{(t+1)} - n_2^{(t)} - \frac{1}{2}Q_2'(x)(v^{(t+1)} - v^{(t)})\right]^2 | x^{(t)} = x, w^{(t)} = w\right\} + o(1) \\
 &= (P_1Q_2 + P_2Q_1)\left(1 - \frac{1}{2}Q_2'(x)\right)^2 + (P_1Q_3 + P_3Q_1)\left(0 + 2 \cdot \frac{1}{2}Q_2'(x)\right)^2
 \end{aligned}$$

$$\begin{aligned}
 & + (P_2 Q_3 + P_3 Q_2) \left( 1 + \frac{1}{2} Q_2'(x) \right)^2 + o(1) \\
 = & P_1 Q_2 + P_2 Q_1 + P_2 Q_3 + P_3 Q_2 + (P_2 Q_3 + P_3 Q_2 - P_1 Q_2 - P_2 Q_1) Q_2'(x) \\
 & + (P_1 Q_2 + P_2 Q_1 + P_2 Q_3 + P_3 Q_2 + 4P_1 Q_3 + 4P_3 Q_1) \frac{Q_2'(x)^2}{4} + o(1) \\
 = & Q_1(1 - P_2) + P_2(1 - Q_2) + [Q_2(P_3 - P_1) + P_2(Q_3 - Q_1)] Q_2'(x) + [2Q_2(x)(1 - Q_2(x)) \\
 & + 8Q_1(x)Q_3(x)] \frac{Q_2'(x)^2}{4} + o(1) \\
 = & 2Q_2(x)(1 - Q_2(x)) + 2Q_2(x)(Q_3(x) - Q_1(x)) Q_2'(x) + \frac{4x(1-x)}{1-\frac{s}{2}} \cdot \frac{Q_2'(x)^2}{4} + o(1). \\
 = & b(x)x(1-x) + o(1),
 \end{aligned}$$

where in the first step we used the upper part of (A.10), and in the second step the definition of  $w^{(t)}$  in (A.3) and a first order Taylor expansion of  $Q_2(\cdot)$ . In the fifth step we used (A.6), in the sixth step (A.8) and (A.9), and the last step, finally, follows after some computations from the definition of  $b(x)$  in (49), by inserting formulas (37) for  $Q_\alpha(x)$  and its derivative

$$Q_2'(x) = \frac{2(1-2x)(1-s)}{1-\frac{s}{2}}$$

for  $\alpha = 2$ .

To finalize the proof of (48), we notice that its first part follows by combining (42), (A.3), the upper part of equation of (A.10), and (A.11);

$$\begin{aligned}
 & E[Y(T + \delta) - Y(T) | X(\tau) = x, Y(T) = y] \\
 = & \delta n \cdot \frac{1}{\sqrt{n}} E[w^{(t+1)} - w^{(t)} | x^{(t)} = x, w^{(t)} = \sqrt{ny}] + o(1) \\
 = & \delta n \cdot \frac{1}{\sqrt{n}} \left[ -\left(1 - \frac{s}{2}\right) \frac{\sqrt{ny}}{n} + \frac{c(x)(1-\frac{s}{2})}{n} + o(n^{-1}) \right] + o(1) \\
 = & -\left(1 - \frac{s}{2}\right) \left( y - \frac{c(x)}{\sqrt{n}} \right) \delta + o(1). \tag{A.12}
 \end{aligned}$$

The second part of (48) follows similarly from the lower equation of (A.10). □

**A.4. Proof of weak convergence of discrete time Moran process on local time scale**

We start by indexing the  $a$  allele frequency process  $X(\tau) = X_n(\tau)$  in (42), and the heterozygous excess process  $Y(T) = Y_n(T)$  in (43), by population size  $n$ . We will prove weak convergence

$$(X_n(\tau), Y_n(T)) | X_n(0) = p, Y_n(0) = y(0) \xrightarrow{L} (X_\infty(\tau), Y_\infty) \tag{A.13}$$

when  $\tau$  is fixed and  $n \rightarrow \infty$ , and hence  $T = \tau n_e \rightarrow \infty$ . The first component  $X_\infty(\tau)$  on the right-hand side of (A.13) has a distribution (47) that depends on  $p$  but not on  $y(0)$ . The second component has a normal distribution

$$Y_\infty | X_\infty(\tau) = x \sim N(0, \sigma^2(x)) \tag{A.14}$$

conditionally on the first, with a variance (52) that depends on  $x$ , but on neither  $p$  nor  $y(0)$ . This distribution corresponds to the limit of (51) as  $n \rightarrow \infty$ . We will actually prove a result more general than (A.13); functional weak convergence

$$\begin{aligned}
 & \left\{ \left( X_n \left( \tau + \frac{U}{n_e} \right), Y_n(T + U) \right); U_1 \leq U \leq U_2 \right\} | X_n(0) = p, Y_n(0) = y(0) \\
 & \xrightarrow{L} \{ (X_\infty(\tau), OU[U; \sigma^2(X_\infty(\tau))]); U_1 \leq U \leq U_2 \} \tag{A.15}
 \end{aligned}$$

of a stochastic process defined over an interval with fixed end points  $U_1 < 0 < U_2$ , using a standardized local time scale  $U$  such that  $U=0$  corresponds to (40). The second component  $OU$  on the right-hand side of (A.15) is a stationary Ornstein–Uhlenbeck process conditionally on  $X_\infty(\tau)$ . This Ornstein–Uhlenbeck process has

a marginal density (A.14), and a transition density

$$OU(U'; \sigma^2) | OU(U'; \sigma^2) = y \sim N(yg(U' - U), \sigma^2 [1 - g^2(U' - U)]) \tag{A.16}$$

for any pair  $U' < U''$  of time points, with  $g(u) = \exp(-(1-\frac{s}{2})u)$ . It corresponds to a diffusion solution (48) with  $n = \infty$ .

In order to prove (A.15), we introduce a second time point  $\tau_0 = \tau + U_0/n_e$  on the slowly varying time scale (41), with  $U_0 < U_1 < 0$  fixed. Then we introduce the process

$$\mathbf{Z}_n(U') = \left( \sqrt{n} \left[ X_n \left( \tau_0 + \frac{U'}{n_e} \right) - X_n(\tau_0) \right], Y_n(T_0 + U') \right) \tag{A.17}$$

of allele frequency oscillations and heterozygous excess around  $\tau_0$ , on a time scale  $U' = U - U_0 \geq 0$ , with  $T_0 = \tau_0 n_e = T + U_0$ . We will prove functional weak convergence

$$\begin{aligned}
 & \{ \mathbf{Z}_n(U'); U' \geq 0 \} | X_n(\tau_0) = x_0, Y_n(T_0) = y_0 \xrightarrow{L} \{ \mathbf{Z}(U'); U' \geq 0 \} \\
 = & \left\{ \left[ \sqrt{\frac{x_0(1-x_0)}{1-\frac{s}{2}}} B(U'), OU(U'; \sigma^2(x_0)) \right]; U' \geq 0 \right\} | OU(0; \sigma^2(x_0)) = y_0 \tag{A.18}
 \end{aligned}$$

of this process as  $n \rightarrow \infty$ , towards a limit process  $\mathbf{Z}$  whose first component involves a standard Brownian motion  $B$  and whose second component is an Ornstein–Uhlenbeck process (A.16), defined on the non-negative real line.

Before proving (A.18), we will first show how to use it in order to verify (A.15). To this end, conditionally on  $X_n(0) = p$  and  $Y_n(0) = y(0)$  we rewrite the left-hand side of (A.15) as

$$\begin{aligned}
 & \left( X_n \left( \tau + \frac{U}{n_e} \right), Y_n(T + U) \right) = \left( X_n(\tau) + \left[ X_n \left( \tau + \frac{U}{n_e} \right) - X_n(\tau) \right], Y_n(T + U) \right) \\
 = & \left( X_n(\tau) + \left[ X_n \left( \tau_0 + \frac{U - U_0}{n_e} \right) - X_n(\tau_0) \right], Y_n(T + U) \right) \\
 = & \left( X_n(\tau) + \left[ X_n \left( \tau_0 + \frac{U - U_0}{n_e} \right) - X_n(\tau_0) \right], Y_n(T_0 + U - U_0) \right) \\
 \stackrel{L}{\approx} & \left( X_n(\tau) + \frac{1}{\sqrt{n}} \sqrt{\frac{X_n(\tau_0)(1-X_n(\tau_0))}{1-\frac{s}{2}}} [B(U - U_0) - B(-U_0)], \right. \\
 & OU[U - U_0; \sigma^2(X_n(\tau_0))] | OU[0; \sigma^2(X_n(\tau_0))] = Y_n(T_0) \\
 \stackrel{L}{\approx} & (X_n(\tau), OU[U - U_0; \sigma^2(X_n(\tau_0))]) | OU[0; \sigma^2(X_n(\tau_0))] = Y_n(T_0) \\
 \stackrel{L}{\approx} & (X_\infty(\tau), OU[U - U_0; \sigma^2(X_\infty(\tau))]) | OU[0; \sigma^2(X_\infty(\tau))] = Y_n(T_0) \\
 \stackrel{L}{\approx} & (X_\infty(\tau), OU[U - U_0; \sigma^2(X_\infty(\tau))]) \\
 \stackrel{L}{\approx} & (X_\infty(\tau), OU[U; \sigma^2(X_\infty(\tau))]), \tag{A.19}
 \end{aligned}$$

for all  $U_1 \leq U \leq U_2$ . In the third step of (A.19) we introduced the notation  $A_n \stackrel{L}{\approx} B_n$ , which refers to two sequences  $A_n$  and  $B_n$  of random elements that have the same weak limit  $A$ , i.e.  $A_n \xrightarrow{L} A$  and  $B_n \xrightarrow{L} A$ . In the fifth step of (A.19) we used that the allele frequency process converges weakly

$$X_n(\tau) | X_n(0) = p, Y_n(0) = y(0) \xrightarrow{L} X_\infty(\tau) \tag{A.20}$$

at rescaled time point  $\tau$ , as  $n \rightarrow \infty$ . This follows from the weak convergence theory of Ethier and Nagylaki (1980, 1988) and Ethier and Kurtz (1986). In the sixth step of (A.19) we used (A.16) and chose  $-U_0 > 0$  so large that conditioning on the value of  $OU$  at  $U_0$  has no impact on its behaviour over the interval  $[U_1, U_2]$ . Finally, in the last step of (A.19) we used that  $U \rightarrow OU[U; \sigma^2(X_\infty(\tau))]$  is a mixture of stationary Ornstein–Uhlenbeck processes, and hence stationary itself. Formula (A.15) follows since its left- and right-hand sides coincides with the left- and right-hand sides of (A.19).

Hence it remains to prove (A.18). We will do this by applying a weak convergence result of Norman (1975), henceforth denoted as

N75. Theorem 3 of N75 provides functional weak convergence for a one-dimensional process towards a Gaussian diffusion limit, but we will need the multivariate extension of this result, as outlined in Section 9 of N75. To this end, we will first show that (A.17) is equivalent to

$$\mathbf{Z}_n(U') = \sqrt{n} \left[ (\mathbf{R}_n^{(t_0+nU')} - \boldsymbol{\gamma}_n^{(t_0+nU')}) + (\mathbf{0}, \gamma_{n2}^{(t_0+nU')}) \right], \quad (\text{A.21})$$

where  $t_0 = \tau_0/(nn_e)$  is the discrete time step corresponding to  $\tau_0$ , and

$$\mathbf{R}_n^{(t)} = (\mathbf{R}_{n1}^{(t)}, \mathbf{R}_{n2}^{(t)}) = (\mathbf{x}^{(t)}, \mathbf{Q}_2^{(t)} - \mathbf{Q}_2(\mathbf{x}^{(t)})) = \left( \frac{v^{(t)}}{2n}, \frac{w^{(t)}}{n} \right), \quad (\text{A.22})$$

is a vector that for each time step  $t \geq t_0$  has one  $a$  allele frequency component  $\mathbf{x}^{(t)}$ , and one heterozygous excess component  $\mathbf{Q}_2^{(t)} - \mathbf{Q}_2(\mathbf{x}^{(t)})$ , with  $v^{(t)}$  and  $w^{(t)}$  as defined in (A.2) and (A.3). The function  $\boldsymbol{\gamma}_n^{(t)} = (\gamma_{n1}^{(t)}, \gamma_{n2}^{(t)})$  corresponds to the mean drift of  $\mathbf{R}_n^{(t)}$ . It is defined for  $t \geq t_0$ , with initial value  $\boldsymbol{\gamma}_n^{(t_0)} = (x_0, y_0/\sqrt{n})$  obtained from the conditioning in (A.18). The recursion for the subsequent time steps  $t = t_0 + 1, t_0 + 2, \dots$ , is

$$\boldsymbol{\gamma}_n^{(t)} = E(\mathbf{R}_n^{(t)} | \mathbf{R}_n^{(t-1)}) = \boldsymbol{\gamma}_n^{(t-1)}.$$

It follows from (A.4) and (A.10) that

$$\begin{aligned} \gamma_{n1}^{(t_0+nU')} &= x_0, \\ \gamma_{n2}^{(t_0+nU')} &= \frac{y_0}{\sqrt{n}} \prod_{u=1}^{nU'} \left( 1 - \left( 1 - \frac{s}{2} \right) \frac{1}{n} \right) (1 + o(1)) \\ &= \frac{y_0}{\sqrt{n}} \exp \left( - \left( 1 - \frac{s}{2} \right) U' \right) (1 + o(1)) \\ &= \frac{y_0}{\sqrt{n}} g(U') (1 + o(1)) \end{aligned} \quad (\text{A.23})$$

as  $n \rightarrow \infty$ . In the second equation of (A.23) we used (A.22) and the upper part of (A.10) to conclude that  $\gamma_{n2}^{(t+1)} = \gamma_{n2}^{(t)} (1 - (1 - s/2)/n) + O(Cn^{-2}) + o(n^{-2})$ . From (42), (43) and the upper part of (A.23) we deduce that (A.17) and (A.21) are equivalent.

In order to prove (A.18), we need to establish certain regularity conditions for the increment

$$\Delta \mathbf{R}_n^{(t)} = \mathbf{R}_n^{(t+1)} - \mathbf{R}_n^{(t)} \quad (\text{A.24})$$

of (A.22) over one time step. These regularity conditions can be phrased in terms of its first two moments

$$E(\Delta \mathbf{R}_n^{(t)} | \mathbf{R}_n^{(t)}) = \frac{1}{n} \mathbf{W}(\mathbf{R}_n^{(t)}) + \mathbf{e}_{1n}^{(t)}, \quad (\text{A.25})$$

and

$$\text{Var}(\Delta \mathbf{R}_n^{(t)} | \mathbf{R}_n^{(t)}) = \frac{1}{n^2} \mathbf{S}(\mathbf{R}_n^{(t)}) + \mathbf{e}_{2n}^{(t)}, \quad (\text{A.26})$$

and the martingale difference

$$\mathbf{e}_{3n}^{(t)} = \Delta \mathbf{R}_n^{(t)} - E(\Delta \mathbf{R}_n^{(t)} | \mathbf{R}_n^{(t)}).$$

We regard  $\mathbf{e}_{1n}^{(t)}$ ,  $\mathbf{e}_{2n}^{(t)}$  and  $\mathbf{e}_{3n}^{(t)}$  as remainder terms, whereas the leading terms of the expected value and variance of  $\Delta \mathbf{R}_n^{(t)}$  include

$$\mathbf{W}(r_1, r_2) = \left( 0, - \left( 1 - \frac{s}{2} \right) r_2 \right) \quad (\text{A.27})$$

and

$$\mathbf{S}(r_1, r_2) = r_1 (1 - r_1) \begin{pmatrix} \frac{1}{1-s} & \tilde{b}(r_1) \\ \tilde{b}(r_1) & b(r_1) \end{pmatrix}, \quad (\text{A.28})$$

where  $b(r_1)$  is the function defined in (49), and

$$\tilde{b}(r_1) = - \frac{(1-s) \left( 1 + \frac{s}{2} \right)}{\left( 1 - \frac{s}{2} \right)^2} (1 - 2r_1).$$

Eqs. (A.27)–(A.28) follow from formulas (A.22), (A.4) and (A.10) for the mean and variance of the increments of  $v^{(t)}$  and  $w^{(t)}$  in (A.2) and (A.3), and the corresponding formula

$$\begin{aligned} E[(v^{(t+1)} - v^{(t)})(w^{(t+1)} - w^{(t)}) | v^{(t)} = 2nx, w^{(t)} = w] &= E \left\{ (v^{(t+1)} - v^{(t)}) \right. \\ &\left. \left[ n_2^{(t+1)} - n_2^{(t)} - \frac{1}{2} Q_2'(x)(v^{(t+1)} - v^{(t)}) \right] \middle| v^{(t)} = 2nx, w^{(t)} = w \right\} + o(1) \\ &= P_2 Q_3 \cdot (-1) \cdot 1 + P_3 Q_2 \cdot 1 \cdot (-1) + P_3 Q_1 \cdot 2 \cdot 0 \\ &\quad + P_1 Q_3 \cdot (-2) \cdot 0 + P_1 Q_2 \cdot (-1) \cdot (-1) + P_2 Q_1 \cdot 1 \cdot 1 \\ &\quad - \frac{1}{2} Q_2'(x) E[(v^{(t+1)} - v^{(t)})^2 | v^{(t)} = 2nx, w^{(t)} = w] + o(1) \\ &= 2[Q_1(x)Q_2(x) - Q_3(x)Q_2(x)] - \frac{1}{2} Q_2'(x) \frac{4x(1-x)}{1-\frac{s}{2}} + o(1) \\ &= 2\tilde{b}(x)x(1-x) + o(1) \end{aligned} \quad (\text{A.29})$$

for the covariance of the increments of these two processes. In the second step of (A.29), the first six terms correspond to cases (i)–(vi) of Section 2.1, with probabilities determined by the Markov transition matrix, with  $v^{(t+1)} - v^{(t)}$  attaining one of the values  $-2, -1, 1, 2$ , and  $n_2^{(t+1)} - n_2^{(t)}$  one of the values  $-1, 0, 1$ . In the third step of (A.29) we used (A.8) and (A.9), and in the last step the definition of  $Q_\alpha(x)$  in (37).

The process  $\mathbf{Z}_n(U')$  in (A.21) corresponds to the random polygonal line process on p. 229 of N75, except that we added a term  $\sqrt{n} \gamma_{n2}^{(t_0+nU')} \rightarrow y_0 g(U')$  to the second component. In order to see this, we notice that the normalizing factor

$$\sqrt{n} = \left( \frac{1/n}{1/n^2} \right)^{1/2}$$

of (A.21) is obtained from the two multiplicative factors  $1/n$  and  $1/n^2$  in (A.25) and (A.26), and the factor  $1/n$  of the time scale  $t = t_0 + U'/n$  is also derived from (A.25), just as in N75. The limiting process  $\mathbf{Z}$  in (A.18) is a Gaussian diffusion, whose distribution is fully specified by its mean and covariance functions. These are obtained by inserting the two functions  $\mathbf{W}$  and  $\mathbf{S}$  in (A.27) and (A.28) into formulas (2.15)–(2.17) of N75, and then adding the asymptotic limit  $y_0 g(U')$  of the second term of (A.21). After some elementary calculations, it follows that the resulting Gaussian diffusion  $\mathbf{Z}$  corresponds to the process on the right-hand side of (A.18), with one Brownian motion and one Ornstein–Uhlenbeck component. Since  $\tilde{b}(r_1) \neq 0$ , these two processes will not be independent.

In order to complete the proof of (A.18), some regularity conditions of Theorem 3 in N75 need to be checked. It can easily be shown that the differentiability conditions (2.3)–(2.7) on  $\mathbf{W}(r_1, r_2)$  and  $\mathbf{S}(r_1, r_2)$  hold. The remaining regularity conditions (2.9'), (2.10) and (2.12') of N75 involve the three remainder terms  $\mathbf{e}_{in}^{(t)}$ , for  $i = 1, 2, 3$ . They can be written as

$$\left( E(|\mathbf{e}_{1n}^{(t)}|^3) \right)^{1/3} = o(n^{-3/2}), \quad (\text{A.30})$$

$$E(|\mathbf{e}_{2n}^{(t)}|) = o(n^{-2}) \quad (\text{A.31})$$

and

$$\left( E(|\mathbf{e}_{3n}^{(t)}|^3) \right)^{1/3} = O(n^{-1}). \quad (\text{A.32})$$

Formulas (A.30)–(A.31) follow by looking more closely at the remainder terms of (A.4), (A.10) and (A.29), with third moment estimates for the remainder terms of expected values of increments, and first moment estimates for remainder terms of variance and covariance of increments. Finally, (A.32) follows from that  $|\Delta \mathbf{x}^{(t)}|$  and  $|\Delta \mathbf{Q}_2^{(t)}|$  are less or equal to  $1/n$ , and the fact that  $Q_2'$  is bounded. □

A.5. Proof of (31)

Formulas (37) and (44) imply

$$f_2^{(t)} = \frac{2(1-s)}{1-\frac{s}{2}} E[x^{(t)}(1-x^{(t)})] + \mu(T), \tag{A.33}$$

with  $\mu(T) = \sqrt{n}E(Y(T))$ . Starting with the first term, we condition on what happens at time step  $t$ , and find from (32) and (A.2)–(A.4) that

$$\begin{aligned} E[x^{(t+1)}(1-x^{(t+1)}) | x^{(t)} = x] &= x(1-x) - E[(x^{(t+1)} - x^{(t)})^2 | x^{(t)} = x] \\ &= x(1-x) - \frac{1}{(2n)^2} \cdot \frac{4x(1-x)}{1-\frac{s}{2}} + o(n^{-2}) \\ &= x(1-x) \left[ 1 - \frac{1}{nn_e} + o\left(\frac{1}{nn_e}\right) \right]. \end{aligned}$$

Taking expectations on both sides of the last displayed equation, we obtain

$$E[x^{(t+1)}(1-x^{(t+1)})] = E[x^{(t)}(1-x^{(t)})] \cdot \left[ 1 - \frac{1}{nn_e} + o\left(\frac{1}{nn_e}\right) \right].$$

Iterating this last relation  $t$  times, we get

$$\begin{aligned} E[x^{(t)}(1-x^{(t)})] &= x^{(0)}(1-x^{(0)}) \cdot \left[ 1 - \frac{1}{nn_e} + o\left(\frac{1}{nn_e}\right) \right]^t \\ &= x^{(0)}(1-x^{(0)}) \cdot \left[ 1 - \frac{1}{nn_e} + o\left(\frac{1}{nn_e}\right) \right]^{nn_e \cdot t / (nn_e)} \\ &= x^{(0)}(1-x^{(0)}) \exp\left(-\frac{t}{nn_e}\right) + o(1). \end{aligned} \tag{A.34}$$

The first term of (31) is obtained by inserting (A.34) into (A.33).

As for the last two terms of (31), we first take expectation with respect to  $X(\tau)$  in the upper equation of (48) and then let  $\delta \rightarrow 0$ . This gives a linear and inhomogeneous first order ordinary differential equation

$$\begin{aligned} \mu'(T) &= -\left(1 - \frac{s}{2}\right)\mu(T) + \frac{1}{n} \cdot \left(1 - \frac{s}{2}\right) c E[X(\tau)(1-X(\tau))] \\ &= -\left(1 - \frac{s}{2}\right)\mu(T) + \frac{1}{n} \cdot \left(1 - \frac{s}{2}\right) c x^{(0)}(1-x^{(0)}) e^{-T/n_e} \end{aligned} \tag{A.35}$$

for  $\mu(T)$ , with

$$c = \frac{(1-s)\left(3 - \frac{s}{2}\right)}{\left(1 - \frac{s}{2}\right)^3},$$

and in the last step of (A.35) we made use of (A.34). The solution of (A.35) is

$$\begin{aligned} \mu(T) &= \mu(0)e^{-(1-\frac{s}{2})T} \\ &\quad + \int_0^T \frac{1}{n} \cdot \left(1 - \frac{s}{2}\right) c x^{(0)}(1-x^{(0)}) e^{-T'/n_e} e^{-(1-\frac{s}{2})(T-T')} dT'. \end{aligned}$$

After evaluating the integral and making some other rearrangements, it can be seen that this expression equals the sum of the last two terms of (31), since

$$\mu(0) = Q_2^{(0)} - 2x^{(0)}(1-x^{(0)}) \frac{1-s}{1-\frac{s}{2}}. \square$$

A.6. Proof of (46) for the Wright–Fisher model

Since the argument is analogous to the proof of (46) for the Moran model, we will be more brief. It suffices to prove

$$\begin{aligned} E(x^{(r+1)} | x^{(r)} = x, w^{(r)} = w) &= x, \\ \text{Var}(x^{(r+1)} | x^{(r)} = x, w^{(r)} = w) &= \frac{x(1-x)}{2\left(1 - \frac{s}{2}\right)} \cdot \frac{1}{n} + o(n^{-1}), \end{aligned} \tag{A.36}$$

for all  $0 < x < 1$  and  $w$ , with

$$w^{(r)} = n_2^{(r)} - Q_2(x^{(r)}) = -\frac{Y\left(\frac{r}{2}\right)}{\sqrt{n}} \tag{A.37}$$

the excess number of heterozygots. In order to prove the upper part of (A.36), we use that the expected value of the multinomial distribution in (57) is

$$E(\tilde{n}_1, \tilde{n}_2, \tilde{n}_3 | n_1, n_2, n_3) = n(P_1, P_2, P_3).$$

Since  $x^{(r+1)} = (\tilde{n}_2 + 2\tilde{n}_3)/(2n)$ , it follows that

$$\begin{aligned} E(x^{(r+1)} | x^{(r)} = x, w^{(r)} = w) - x &= \frac{1}{2}P_2 + P_3 - \left(\frac{1}{2}Q_2 + Q_3\right) \\ &= \frac{1}{2}[(P_3 - P_1) - (Q_3 - Q_1)] = 0, \end{aligned} \tag{A.38}$$

where the last equality is a consequence of (A.5). For the lower part of (A.36) we use that the covariance matrix of the multinomial distribution in (57) has entries

$$\text{Cov}(\tilde{n}_\alpha, \tilde{n}_\beta | n_1, n_2, n_3) = n(P_\alpha \delta_{\alpha\beta} - P_\alpha P_\beta),$$

for  $1 \leq \alpha, \beta \leq 3$ . This yields

$$\begin{aligned} \text{Var}(x^{(r+1)} | x^{(r)} = x, w^{(r)} = w) &= \frac{1}{n} \left[ \frac{1}{4}P_2(1-P_2) + P_3(1-P_3) - P_2P_3 \right] \\ &= \frac{1}{n} \left[ \frac{1}{4}P_2(1-P_2) + P_1P_3 \right] \\ &= \frac{1}{n} \left[ \frac{1}{4}Q_2(x)(1-Q_2(x)) + Q_1(x)Q_3(x) \right] + o(n^{-1}) \\ &= \frac{1}{n} \cdot \frac{x(1-x)}{2\left(1 - \frac{s}{2}\right)} + o(n^{-1}), \end{aligned} \tag{A.39}$$

where in the last two steps we used (A.8) and (A.9). $\square$

A.7. Proof of (58)

We will prove that

$$\begin{aligned} E\left[Y\left(T + \frac{1}{2}\right) - y | X(\tau) = x, Y(T) = y\right] &= -\left(1 - \frac{s}{2}\right) \left(y - \frac{\tilde{c}(x)}{\sqrt{n}}\right) + o(n^{-1/2}), \\ \text{Var}\left[Y\left(T + \frac{1}{2}\right) | X(\tau) = x, Y(T) = y\right] &= \frac{1}{2}b(x)x(1-x). \end{aligned} \tag{A.40}$$

It can be seen that this is equivalent to (58), if  $\varepsilon(T)$  has a normal distribution. This follows by applying a Central Limit Theorem to the multinomial distribution in (57) when  $n$  is large.

Let  $w^{(r)}$  be the excess number of heterozygots of generation  $r$ , defined in (A.37), and  $w = \sqrt{n}y$ . We have that

$$\begin{aligned} \sqrt{n}E\left[Y\left(T + \frac{1}{2}\right) - y | X(\tau) = x, Y(T) = y\right] &= E(w^{(r+1)} - w^{(r)} | x^{(r)} = x, w^{(r)} = w) \\ &= E(n_2^{(r+1)} - n_2^{(r)} | x^{(r)} = x, w^{(r)} = w) \\ &\quad - \frac{n}{2}Q_2(x)E\left[(x_2^{(r+1)} - x^{(r)})^2 | x^{(r)} = x, w^{(r)} = w\right] \\ &= n(P_2 - Q_2) - \frac{n}{2} \left( -\frac{4(1-s)}{1-\frac{s}{2}} \right) \cdot \frac{x(1-x)}{2n\left(1 - \frac{s}{2}\right)} + o(1) \\ &= -\left(1 - \frac{s}{2}\right)w + C_1 + \frac{C_2}{2} + o(1), \end{aligned} \tag{A.41}$$

where the third step follows from (A.38) and a Taylor expansion of  $Q_2(\cdot)$ , in the fourth step we used (A.39) and the expected value of the multinomial distribution in (57). In the last step we defined  $C_1$  and  $C_2$  as in (A.11) and used an expression for  $P_2 - Q_2$  from the proof of (48). But (A.41) is equivalent to the upper part of (A.40), since  $C_1 + C_2/2 = (1-s/2)\tilde{c}(x)$ .

For the second part of (A.40) we use that

$$\text{Var}\left[Y\left(T + \frac{1}{2}\right) | X(\tau) = x, Y(T) = y\right]$$

$$\begin{aligned}
&= \frac{1}{n} E \left[ (n_2^{(r+1)} - n_2^{(r)})^2 \mid x^{(r)} = x, w^{(r)} = w \right] \\
&\quad - 2Q_2'(x) E \left[ (n_2^{(r+1)} - n_2^{(r)}) (x^{(r+1)} - x^{(r)}) \mid x^{(r)} = x, w^{(r)} = w \right] \\
&\quad + nQ_2'(x)^2 E \left[ (x^{(r+1)} - x^{(r)})^2 \mid x^{(r)} = x, w^{(r)} = w \right] + o(1) \\
&= P_2(1 - P_2) - Q_2'(x)[P_2(1 - P_2) - 2P_2P_3] \\
&\quad + \frac{1}{4} Q_2'(x)^2 [P_2(1 - P_2) - 4P_2P_3 + 4P_3(1 - P_3)] + o(1) \\
&= Q_2(x)(1 - Q_2(x)) \\
&\quad + Q_2'(x)Q_2(x)[Q_3(x) - Q_1(x)] \\
&\quad + \frac{1}{4} Q_2'(x)^2 [Q_2(x)(1 - Q_2(x)) + 4Q_1(x)Q_3(x)] + o(1) \\
&= \frac{1}{2} b(x)x(1 - x) + o(1),
\end{aligned}$$

where the last step follows as in the proof of (48). $\square$

## References

- Abramowitz, M., Stegun, I.A. 1964. Handbook of Mathematical Functions. Nat. Bureau Stand.
- Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, 2nd ed. Springer-Verlag, New York.
- Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Adv. Appl. Probab.* 6, 260–290.
- Cavalli-Sforza, L.L., Bodmer, W.F., 1971. The Genetics of Human Populations. Freeman, San Francisco.
- Coad, R.W., 2000. Diffusion approximation of the Wright–Fisher model of population genetics: single-locus two alleles. *Ukr. Math. J.* 3 (52), 388–399.
- 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, Caldwell, New Jersey.
- Eldon, B., Wakeley, J., 2006. Coalescent process when the distribution of number of offspring is highly skewed. *Genetics* 172, 2621–2633.
- Ethier, S.N., Nagylaki, T., 1980. Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Probab.* 12, 14–49.
- Ethier, S.N., Nagylaki, T., 1988. Diffusion approximations of Markov chains with two time scales and applications to population genetics, II. *Adv. Appl. Probab.* 20, 525–545.
- Ethier, S.N., Kurtz, T.G., 1986. Markov Processes, Characterization and Convergence. John Wiley & Sons, Hoboken, New Jersey.
- Fisher, R.A., 1922. On the dominance ratio. *Proc. R. Soc. Edinb.* 42, 321–431.
- Fisher, R.A., 1958. The General Theory of Natural Selection, 2nd rev. ed. Dover, New York.
- Grimmett, G., Stirzaker, D., 2001. Probability and Random Processes, 3rd ed. Oxford University Press, Oxford.
- Hössjer, O., Jorde, P.E., Ryman, N., 2013. Quasi equilibrium approximations of the fixation index under neutrality: The finite and infinite island model. *Theor. Popul. Biol.* 84, 9–24.
- 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>.
- Hössjer, O., Olsson, F., Laikre, L., Ryman, N., 2015. Metapopulation inbreeding dynamics, effective size and subpopulation differentiation—a general analytical approach for diploid organisms. *Theor. Popul. Biol.* 102, 40–59. <http://dx.doi.org/10.1016/j.tpb.2015.03.006>.
- Kimura, M., 1955. Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. U.S.A.* 41, 141–150.
- Korolyuk, V.S., Korolyuk, D., 1995. Diffusion approximation of stochastic Markov models with persistent regression. *Ukr. Math. J.* 47 (7), 1065–1073.
- McKane, A.J., Waxman, D., 2007. Singular solutions of the diffusion equation of population genetics. *J. Theor. Biol.* 247, 849.
- Levene, H., 1949. On a matching problem arising in genetics. *Ann. Math. Stat.* 20, 91–94.
- Moran, P.A.P., 1958a. Random processes in genetics. *Proc. Camb. Philos. Soc.* 54, 60.
- Moran, P.A.P., 1958b. A general theory of the distribution of gene frequencies. I. Overlapping generations. *Proc. Camb. Philos. Soc.* B149, 102–112.
- Moran, P.A.P., 1958c. A general theory of the distribution of gene frequencies. II. Non-overlapping generations. *Proc. Camb. Philos. Soc.* B149, 113–116.
- Nagylaki, T., 1980. The strong migration limit in geographically structured populations. *J. Math. Biol.* 9, 101–114.
- Nordborg, M., Donnelly, P., 1987. The coalescent process with selfing. *Genetics* 146, 1185–1195.
- Norman, F., 1975. Approximation of stochastic processes by Gaussian diffusion and application to Wright–Fisher genetic models. *SIAM J. Appl. Math.* 29 (2), 225–242.
- Pollak, E., 1997. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117, 353–360.
- Tyvand, P.A., 1993. An exact algebraic theory of genetic drift in finite diploid populations with random mating. *J. Theor. Biol.* 163, 315–331.
- Watterson, G.A., 1964. The application of diffusion theory to two population genetic models of Moran. *J. Appl. Probab.* 1, 233–246.
- Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16, 97–159.
- Wright, S., 1943. Isolation by distance. *Genetics* 16, 114–138.