



First-order effect of frequency-dependent selection on fixation probability in an age-structured population with application to a public goods game

Cíntia Dalila Soares, Sabin Lessard*

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7



ARTICLE INFO

Article history:

Received 16 October 2018

Available online 23 May 2019

MSC:

primary 92D25

secondary 60J70

Keywords:

Fixation-fitness coefficient

Ancestry coefficient

Population-structure coefficient

Genealogical process

Reproductive value

ABSTRACT

In this paper, we deduce the first-order effect of frequency-dependent viability and fertility selection on the probability of fixation of a mutant in a large finite haploid population with a fixed age structure by applying a direct small perturbation method to the neutral two-timescale genealogical process. This effect is expressed in terms of fixation-fitness coefficients times ancestry coefficients that are related to the effective population size. In the case of constant selection, the fixation-fitness coefficients are functions of the coefficients of viability and fertility selection weighted by reproductive values and population-structure coefficients for the different age classes. This explains the difference between the effects of viability selection and fertility selection on fixation probability in age-structured populations. With frequency-dependent selection in the form of a public goods game, the fixation-fitness coefficients depend also on the public good allocation strategy of the population and the resource allocation strategies of the individuals. In this case, the results show that weak selection may favor the fixation of cooperation if the cooperators allocate more resources to reproduction versus survival than the defectors do.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

In a neutral finite haploid population with a fixed age structure in discrete time, the number of individuals in each age class being kept constant from one time step to the next as in [Felsenstein \(1971\)](#), the probability of fixation of a mutant type is given by a weighted average of its frequencies in the different age classes ([Emigh, 1979a](#), see also [Emigh and Pollak, 1979](#), for a diploid population). The weights in this average correspond to the elements of the stationary probability distribution of a backward transition matrix for the ancestry of the population. What are the effects of small differences in survival and reproduction between the types of individuals on this fixation probability? This is the first question that will be addressed in this paper.

Let us mention that, incorporating constant viability and fertility parameters in [Felsenstein's \(1971\)](#) model for an age-structured population, [Emigh \(1979b\)](#) obtained the fixation probability for a mutant type from a diffusion approximation. This approximation was based on numerical results showing that the frequency of each type tends rapidly to the same value in each age class like in the neutral model, and that this value changes slowly under the effects of selection. This two-timescale argument was

proved in [Soares and Lessard \(2019\)](#), but the proof involves the verification of very technical conditions.

In this paper, we will consider frequency-dependent viability and fertility selection in a finite population undergoing discrete, overlapping generations but keeping a fixed age structure in the limit of a large population size. In order to get the first-order effect of selection on the fixation probability for a mutant, we will use a direct approach proposed in [Rousset \(2003\)](#) and asserted in [Lessard and Ladret \(2007\)](#). It is based on a first-order approximation of the expected change of a weighted average frequency from one time step to the next over all future time steps following the introduction of the mutant type in the population. This approach has already been used to study exchangeable selection models extending the neutral ([Canning's 1974](#)) model in a finite well-mixed population ([Lessard and Ladret, 2007](#)) as well as frequency-dependent selection models in finite structured populations, including populations with a two-deme structure ([Ladret and Lessard, 2008](#)) and [Wright's \(1931\)](#) island structure with a finite number of demes of finite sizes ([Ladret and Lessard, 2007](#)) in the limit of a large number of demes ([Lessard, 2011a](#)) and in the limit of large deme sizes ([Kroumi and Lessard, 2014, 2015a](#)).

In the context of a population with a fixed age structure, our approximation of the fixation probability for a mutant under

* Corresponding author.

E-mail address: lessards@dms.umontreal.ca (S. Lessard).

weak selection will rely on a study of the genealogical process and its different timescales in the limit of a large neutral population. Previous studies in geographically structured populations with non-overlapping generations have already given strong-migration limits (Notohara, 1993; Kroumi and Lessard, 2014, 2015a). An analogous limit is expected in an age-structured population which is subdivided into age classes.

Our approximation of the fixation probability will allow us to study exact conditions for weak selection to favor a mutant type in such a population. Following Nowak et al. (2004), this will be the case if the fixation probability is larger under weak selection than under neutrality. Moreover, the differences between the effects of selection on survival and selection on reproduction on the fixation probability will be addressed. Such differences were revealed recently in Li et al. (2016) by numerical iterations and simulations in the case of constant selection parameters in small populations. This study has shown the importance of considering life histories in selection models. A review of studies on fertility measures and survival probabilities in nature in relation with the age of the individuals can be found, e.g., in Jones et al. (2014). Our theoretical results will complement these studies by exhibiting the exact factors that come into play in the fixation probability in the case of weak selection.

Finally, by letting the viability and fertility parameters depend on type frequencies in different age classes, we will be able to get conditions for weak selection to favor the evolution of cooperation in a public goods game with a cost paid by each cooperator and a benefit shared by all individuals, cooperators and defectors (see, e.g., Archetti and Scheuring, 2012 and references therein). Of particular interest for a public goods game in the framework of a finite age-structured population is the effect of the resource allocation strategy between survival and reproduction (see, e.g., Baudisch and Vaupel, 2012, for resource allocation theory). This is the second question that will be addressed in this paper.

It may be useful to recall here that a public goods game corresponds to a N -person prisoner's dilemma with payoffs for cooperation and defection depending on the number of cooperators. This is an extension of the classical prisoner's dilemma with two players, which was the first game to be considered to study the evolution of cooperation in a finite population. It is in the context of an iterated prisoner's dilemma that the one-third law for weak selection to favor cooperation in the limit of a large population was deduced. This was done first with an update of strategies according to a Moran model (Nowak et al., 2004) and a Wright–Fisher model (Lessard, 2005; Imhof and Nowak, 2006). It was shown later on to hold more generally for exchangeable models in the domain of a diffusion approximation in a large population, which corresponds to the domain of Kingman's (1982) coalescent in the limit of neutrality, and extended beyond this domain (Lessard and Ladret, 2007). The one-third law of evolution has also been extended to a N -person prisoner's dilemma, first for a Moran model (Kurokawa and Ihara, 2009; Gokhale and Traulsen, 2010) and later on to the same domain as for the one-third law (Lessard, 2011b). Interactions within groups of random size have been considered too (Hilbe, 2011; McAvoy et al., 2018), as well as group-structured populations with hard or soft selection and complete or partial dispersal (Lessard, 2009, 2011a).

The present paper on evolution in a finite age-structured population is organized as follows. In Section 2, the selection model with a fixed age structure is presented. In Section 3, the first-order effect of selection on the fixation probability of a given type among a finite number of types is deduced in the general case of frequency-dependent parameters for survival and reproduction, and developed further in the case of two types with

constant or linearly frequency-dependent parameters. Ancestry coefficients in the limit of a large neutral population are obtained in Section 4, and used to approximate the fixation probability in a large population under weak selection in Section 5. This approximation is applied to a population with two age classes with constant selection on survival and reproduction in Section 6, and with frequency-dependent selection according to a public goods game in Sections 7 and 8. This is followed by a discussion in Section 9. Technical results on the ancestral process under neutrality are relegated to Appendices.

2. Selection model with fixed age structure

Consider a finite population of size N with $n \geq 1$ types of individuals distributed in $d \geq 1$ age classes. Time is assumed to be discrete. The number of individuals in each age class stays constant from one time step to the next. Let the size of age class k be $N_k = f_k N$ for $k = 1, \dots, d$ with $\sum_{k=1}^d f_k = 1$. The population is assumed to be haploid so that each new individual produced from one time step to the next comes from only one parent.

From one time step to the next, N_1 offspring are produced independently from one another to form the next cohort of individuals in age class 1, while N_{k+1} individuals are chosen without replacement among the N_k individuals in age class k to form the next cohort of individuals in age class $k+1$ for $k = 1, \dots, d-1$.

Let the number of individuals of type $i = 1, \dots, n$ in age class $k = 1, \dots, d$ at time step $\tau \geq 0$ be $c_{i,k}(\tau)$. The array of type numbers $\mathbf{c}(\tau) = (c_{i,k}(\tau))$ determines the population state at time step $\tau \geq 0$. Let

$$x_{i,k}(\tau) = \frac{c_{i,k}(\tau)}{N_k} \quad (1)$$

be the frequency of type $i = 1, \dots, n$ in age class $k = 1, \dots, d$ at time step $\tau \geq 0$, so that the population state at this time step can be represented by the array of type frequencies $\mathbf{x}(\tau) = (x_{i,k}(\tau))$. Moreover, from this time step to the next, an individual of type $j = 1, \dots, n$ in age class $k = 1, \dots, d$ has relative fertility and viability given by

$$r_{j,k}(\mathbf{x}(\tau)) = 1 + s\rho_{j,k}(\mathbf{x}(\tau)) \quad (2)$$

and

$$s_{j,k}(\mathbf{x}(\tau)) = 1 + s\sigma_{j,k}(\mathbf{x}(\tau)), \quad (3)$$

respectively, for some intensity of selection $s > 0$ with continuous functions $\rho_{j,k}(\mathbf{x}(\tau))$ and $\sigma_{j,k}(\mathbf{x}(\tau))$ as coefficients. It is assumed that $s_{j,d}(\mathbf{x}(\tau)) = 0$ so that the last age class is d . Moreover, weak selection is modeled by assuming s small enough.

The cohort of individuals in age class $k+1$ at time step $\tau+1$ for $k = 1, \dots, d-1$ is obtained by sampling without replacement N_{k+1} individuals among the N_k individuals in age class k at time step τ with weight $1 + s\sigma_{i,k}(\mathbf{x}(\tau))$ given to individuals of type $i = 1, \dots, n$, respectively. The state in age class k at time step τ is given by the column vector $\mathbf{c}_k(\tau) = (c_{1,k}(\tau), \dots, c_{n,k}(\tau))^T$, where T stands for transpose. Then, the state in the next age class at the next time step, $\mathbf{c}_{k+1}(\tau+1)$, follows a multivariate Wallenius' noncentral hypergeometric distribution (Wallenius, 1963; Chesson, 1976). Actually, given $\mathbf{c}(\tau) = (\mathbf{c}_k(\tau))$, we have

$$\mathbf{c}_{k+1}(\tau+1) \mid \mathbf{c}(\tau) \sim mwnchypg(N_{k+1}, N_k, \mathbf{c}_k(\tau), \mathbf{1} + s\boldsymbol{\sigma}_k(\mathbf{x}(\tau))),$$

where $\mathbf{1}$ denotes the n -dimensional vector with all components equal to 1, while $\boldsymbol{\sigma}_k(\mathbf{x}(\tau)) = (\sigma_{1,k}(\mathbf{x}(\tau)), \dots, \sigma_{n,k}(\mathbf{x}(\tau)))^T$, for $k = 1, \dots, d-1$.

Using (1), it can be shown (as in Appendix B in Soares and Lessard, 2019 with $Y_i = c_{i,k+1}(\tau+1)$, $m_i = c_{i,k}(\tau)$, $r = N_{k+1}$,

$R = N_k$ and $\omega_i = 1 + s\sigma_{i,k}(\mathbf{x}(\tau))$ that

$$E_s \left[(x_{i,k+1}(\tau + 1) \mid \mathbf{x}(\tau)) \right] = x_{i,k}(\tau) + s x_{i,k}(\tau) D_k \left(\sigma_{i,k}(\mathbf{x}(\tau)) - \sum_{j=1}^n \sigma_{j,k}(\mathbf{x}(\tau)) x_{j,k}(\tau) \right) + o(s) \tag{4}$$

for $i = 1, \dots, n$, where

$$D_k = \frac{N_k(N_k - N_{k+1})}{(N_k - 1)N_{k+1}} S_{N_{k+1}, N_k} \tag{5}$$

with

$$S_{N_{k+1}, N_k} = \sum_{m=N_k - N_{k+1} + 1}^{N_k} \frac{1}{m}$$

for $k = 1, \dots, d - 1$. Here, E_s denotes an expectation when the intensity of selection is s , and $o(s)$ a function such that $o(s)/s \rightarrow 0$ as $s \rightarrow 0$. The coefficient D_k depends on the population age structure, more precisely on the sizes of the age classes k and $k + 1$. It modulates the strength of viability selection from one age class to the next. Note that

$$D_k = C_k + O(N^{-1}),$$

where

$$C_k = \lim_{N \rightarrow +\infty} D_k = \frac{f_k - f_{k+1}}{f_{k+1}} \ln \left(\frac{f_k}{f_k - f_{k+1}} \right) \tag{6}$$

is a population-structure coefficient for age class $k = 1, \dots, d - 1$ in the limit of a large population.

As for reproduction, selection can be soft with competition within each age class (and each age class contributing a fixed expected proportion of offspring) or hard with competition in the whole population (see, e.g., Karlin, 1982). With soft selection taking place and p_k being the expected proportion of offspring coming from age class k for $k = 1, \dots, d$, the probability for an offspring produced in a population in state $\mathbf{x}(\tau)$ to be of type i is given by

$$P_i(\mathbf{x}(\tau)) = \sum_{k=1}^d p_k \left(\frac{x_{i,k}(\tau)(1 + s\rho_{i,k}(\mathbf{x}(\tau)))}{1 + s \sum_{j=1}^n x_{j,k}(\tau)\rho_{j,k}(\mathbf{x}(\tau))} \right) \tag{7}$$

for $i = 1, \dots, n$. This probability can be written as

$$\sum_{k=1}^d p_k x_{i,k}(\tau) + s \sum_{k=1}^d p_k x_{i,k}(\tau) \left(\rho_{i,k}(\mathbf{x}(\tau)) - \sum_{j=1}^n x_{j,k}(\tau)\rho_{j,k}(\mathbf{x}(\tau)) \right) + o(s). \tag{8}$$

Assuming N_1 independent trials to form the next cohort of individuals in age class 1, we have

$$\mathbf{c}_1(\tau + 1) \mid \mathbf{c}(\tau) \sim \text{multinomial}(N_1, P_i(\mathbf{x}(\tau))),$$

from which

$$E_s \left[(x_{i,1}(\tau + 1) \mid \mathbf{x}(\tau)) \right] = P_i(\mathbf{x}(\tau)) \tag{9}$$

for $i = 1, \dots, n$.

We conclude that the vector of the frequencies of type i in the different age classes at time step $\tau + 1$, that is,

$$\mathbf{x}_i(\tau + 1) = (x_{i,1}(\tau + 1), \dots, x_{i,d}(\tau + 1))^T,$$

has conditional expected value

$$E_s[\mathbf{x}_i(\tau + 1) \mid \mathbf{x}(\tau)] = M\mathbf{x}_i(\tau) + s\phi_i(\mathbf{x}(\tau)) + o(s), \tag{10}$$

where

$$M = \begin{pmatrix} p_1 & p_2 & p_3 & \dots & p_d \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}, \tag{11}$$

while

$$\phi_i(\mathbf{x}(\tau)) = (\phi_{i,1}(\mathbf{x}(\tau)), \phi_{i,2}(\mathbf{x}(\tau)), \dots, \phi_{i,d}(\mathbf{x}(\tau)))^T$$

with

$$\phi_{i,1}(\mathbf{x}(\tau)) = \sum_{k=1}^d p_k x_{i,k}(\tau) \left(\rho_{i,k}(\mathbf{x}(\tau)) - \sum_{j=1}^n x_{j,k}(\tau)\rho_{j,k}(\mathbf{x}(\tau)) \right) \tag{12}$$

according to (9), and

$$\phi_{i,k}(\mathbf{x}(\tau)) = x_{i,k-1}(\tau) D_{k-1} \times \left(\sigma_{i,k-1}(\mathbf{x}(\tau)) - \sum_{j=1}^n \sigma_{j,k-1}(\mathbf{x}(\tau)) x_{j,k-1}(\tau) \right) \tag{13}$$

for $k = 2, \dots, d$ according to (4), for $i = 1, \dots, n$.

The stochastic matrix $M = (m_{kl})$ is the backward transition matrix under neutrality, with m_{kl} being the expected proportion of individuals in age class k coming from age class l one time step back in the absence of selection for $k, l = 1, \dots, d$. This matrix is assumed to be ergodic, that is, irreducible and aperiodic. This is the case, for instance, if $p_{d-1}p_d > 0$. Then, the matrix M possesses a stationary probability distribution given by

$$\mathbf{w}^T = (w_1, \dots, w_d) = \mathbf{w}^T M, \tag{14}$$

that is,

$$w_k = p_k w_1 + w_{k+1} \tag{15}$$

for $k = 1, \dots, d$, with $\sum_{k=1}^d w_k = 1$ and $w_{d+1} = 0$. The proportion w_k represents the fraction of time that the lineage of an individual backward in time spends in the long run in age class k in the neutral model.

Note that

$$w_k = \frac{1}{\bar{k}} \sum_{l=k}^d p_l \tag{16}$$

for $k = 1, \dots, d$, with

$$\bar{k} = \sum_{k=1}^d k p_k = \frac{1}{w_1}. \tag{17}$$

Then, w_k represents the reproductive value of age class k and \bar{k} , the mean age of reproduction, the generation time in the neutral model.

Table 1 summarizes the notation used in this section and the following ones.

3. First-order effect of selection on fixation probability

In this section, we shall use an approach suggested in Rousset (2003) and more rigorously established in Lessard and Ladret (2007) to obtain an approximation of the fixation probability for a mutant type under weak selection.

Table 1
Notation.

d	Number of age classes
n	Number of types
N_k	Number of individuals in age class k
N	Total population size
f_k	Proportion of individuals in age class k
f	Proportion of individuals in age class 1
$c_{i,k}(\tau)$	Number of individuals of type i in age class k at time step τ
$x_{i,k}(\tau)$	Frequency of type i in age class k at time step τ
$\mathbf{c}(\tau)$	Array of type numbers at time step τ
$\mathbf{x}(\tau)$	Array of type frequencies at time step τ
$\rho_{i,k}(\mathbf{x})$	Coefficient of fertility selection for type i in age class k
$\sigma_{i,k}(\mathbf{x})$	Coefficient of viability selection for type i in age class k
M	Backward transition matrix (m_{ki}) in one time step under neutrality
p_k	Expected proportion of offspring from age class k in the neutral model
\bar{k}	Generation time in the neutral model
\mathbf{w}	Stationary probability distribution of M
w_k	Reproductive value of age class k under neutrality
$z_i(\tau)$	Weighted average frequency of type i at time step τ
$\mathbf{z}(\tau)$	Array of weighted averages of type frequencies at time step τ
D_k	Population-structure coefficient of age class k
C_k	Population-structure coefficient of age class k in a large population
α	Rate of coalescence under neutrality with N/\bar{k} time steps as unit of time
E_s	Expectation with intensity of selection s
E_0	Expectation under neutrality
$u_i(s)$	Probability of fixation of type i with intensity of selection s
$u_i(0)$	Probability of fixation of type i under neutrality
c_k	Cost for cooperation in age class k
b_k	Public good for cooperation in age class k
h_k	Proportion of total public good allocated to age class k
r	Reward factor for public good
$\beta_{i,k}(\mathbf{x})$	Resources allocated by type i individual in age class k
$q_{i,k}$	Fraction of resources allocated to reproduction by type i individual in age class k
$1 - q_{i,k}$	Fraction of resources allocated to survival by type i in age class k
$\pi_{i,k}^*$	Fixation-fitness coefficient for type i in age class k in a large population
$v_{i,k}^*$	Fixation-fitness coefficient for type i in age class k per unit of resources in a large population

The array of type frequencies $\mathbf{x}(\tau) = (x_{i,k}(\tau))$ for $\tau \geq 0$ is a discrete Markov chain on the state space

$$S = \left\{ \mathbf{x} = (x_{i,k}) : x_{i,k} = \frac{c_{i,k}}{N_k}, c_{i,k} \in \mathbb{N}, \sum_{i=1}^n c_{i,k} = N_k, \right. \\ \left. \text{for } i = 1, \dots, n \text{ and } k = 1, \dots, d \right\}.$$

Given any initial population state $\mathbf{x}(0) = \mathbf{x}$, this chain has n absorbing states $\mathbf{e}_1, \dots, \mathbf{e}_n$, with \mathbf{e}_i denoting a $n \times d$ matrix with all entries on the i th line equal to 1 and all other entries equal to 0. All other states are transient. According to the ergodic theorem (see, e.g., Karlin and Taylor, 1975, p.85), the Markov chain will reach the absorbing state \mathbf{e}_i in a finite time with some probability $u_i(s)$. This represents the probability of fixation of type i as a function of the intensity of selection for $i = 1, \dots, n$ with $\sum_{i=1}^n u_i(s) = 1$. This guarantees that $\mathbf{x}(\tau)$ converges in probability as $\tau \rightarrow +\infty$ to a random array $\mathbf{x}(\infty)$ that takes the value \mathbf{e}_i with probability $u_i(s)$ for $i = 1, \dots, n$. Being uniformly bounded, the chain also converges in mean. This is also the case for $\mathbf{z}(\tau) = (z_i(\tau))$ with $z_i(\tau)$ being the weighted frequency of type i given

by

$$z_i(\tau) = \mathbf{w}^T \mathbf{x}_i(\tau) = \sum_{k=1}^d w_k x_{i,k}(\tau) \tag{18}$$

for $i = 1, \dots, n$. Consequently, we have

$$\begin{aligned} u_i(s) &= E_s[z_i(\infty)] \\ &= \lim_{T \rightarrow +\infty} E_s[z_i(T)] \\ &= \lim_{T \rightarrow +\infty} E_s \left[z_i(0) + \sum_{\tau=0}^{T-1} (z_i(\tau+1) - z_i(\tau)) \right] \\ &= z_i(0) + \sum_{\tau \geq 0} E_s[z_i(\tau+1) - z_i(\tau)] \\ &= z_i(0) + \sum_{\tau \geq 0} E_s[E_s[z_i(\tau+1) - z_i(\tau) | \mathbf{x}(\tau)]]. \end{aligned}$$

Let us recall that E_s denotes an expectation when the intensity of selection is s . The neutral case corresponds to $s = 0$.

Using (18) and (10), we have

$$\begin{aligned} E_s[z_i(\tau+1) | \mathbf{x}(\tau)] &= \mathbf{w}^T E_s[\mathbf{x}_i(\tau+1) | \mathbf{x}(\tau)] \\ &= \mathbf{w}^T \mathbf{x}_i(\tau) + s \mathbf{w}^T \boldsymbol{\phi}_i(\mathbf{x}(\tau)) + o(s) \\ &= z_i(\tau) + s \mathbf{w}^T \boldsymbol{\phi}_i(\mathbf{x}(\tau)) + o(s). \end{aligned}$$

Then, according to Result 1 in Lessard and Ladret (2007), the probability of fixation of type i is approximated as

$$u_i(s) = z_i(0) + s \sum_{\tau \geq 0} E_0[\mathbf{w}^T \boldsymbol{\phi}_i(\mathbf{x}(\tau))] + o(s). \tag{19}$$

Therefore, the fixation probability under neutrality is

$$u_i(0) = z_i(0) = \mathbf{w}^T \mathbf{x}_i(0) = \sum_{k=1}^d w_k x_{i,k}(0), \tag{20}$$

while the derivative of the fixation probability with respect to the intensity of selection evaluated at $s = 0$ is

$$u_i'(0) = \sum_{\tau \geq 0} E_0[\mathbf{w}^T \boldsymbol{\phi}_i(\mathbf{x}(\tau))]. \tag{21}$$

This derivative represents the first-order effect of selection on the fixation probability.

Note that (12) and (13) lead to

$$\begin{aligned} \mathbf{w}^T \boldsymbol{\phi}_i(\mathbf{x}(\tau)) &= w_1 \sum_{k=1}^d p_k x_{i,k}(\tau) \left(\rho_{i,k}(\mathbf{x}(\tau)) - \sum_{j=1}^n x_{j,k}(\tau) \rho_{j,k}(\mathbf{x}(\tau)) \right) \\ &\quad + \sum_{k=2}^d w_k x_{i,k-1}(\tau) D_k \\ &\quad \times \left(\sigma_{i,k-1}(\mathbf{x}(\tau)) - \sum_{j=1}^n \sigma_{j,k-1}(\mathbf{x}(\tau)) x_{j,k-1}(\tau) \right) \\ &= \sum_{k=1}^d \left(\pi_{i,k}(\mathbf{x}(\tau)) - \sum_{j=1}^n x_{j,k}(\tau) \pi_{j,k}(\mathbf{x}(\tau)) \right) x_{i,k}(\tau), \end{aligned} \tag{22}$$

where

$$\pi_{i,k}(\mathbf{x}(\tau)) = \rho_{i,k}(\mathbf{x}(\tau)) p_k w_1 + \sigma_{i,k}(\mathbf{x}(\tau)) D_k w_{k+1} \tag{23}$$

for $k = 1, \dots, d$ with D_k given in (5) and $w_{d+1} = 0$.

In the particular case of two types ($n = 2$), we have

$$\mathbf{w}^T \boldsymbol{\phi}_1(\mathbf{x}(\tau)) = \sum_{k=1}^d x_{1,k}(\tau) (1 - x_{1,k}(\tau)) (\pi_{1,k}(\mathbf{x}(\tau)) - \pi_{2,k}(\mathbf{x}(\tau))). \tag{24}$$

In this case, the first-order effect of selection on the probability of fixation of type 1 is given by

$$u'_1(0) = \sum_{\tau \geq 0} \sum_{k=1}^d E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau)) (\pi_{1,k}(\mathbf{x}(\tau)) - \pi_{2,k}(\mathbf{x}(\tau)))]. \tag{25}$$

3.1. Two types with constant selection

Let us consider first the case of two types with fertility and viability selection coefficients that do not depend on the population state. Then we have

$$\rho_{i,k}(\mathbf{x}(\tau)) = \rho_{i,k}$$

and

$$\sigma_{i,k}(\mathbf{x}(\tau)) = \sigma_{i,k}$$

for $i = 1, 2$ and $k = 1, \dots, d$ at every time step $\tau \geq 0$. In this case, the first-order effect of selection on the fixation probability for type 1 takes the form

$$u'_1(0) = \sum_{k=1}^d (\pi_{1,k} - \pi_{2,k}) \sum_{\tau \geq 0} E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau))] \tag{26}$$

with

$$\pi_{i,k} = \rho_{i,k}p_k w_1 + \sigma_{i,k}D_k w_{k+1}$$

for $i = 1, 2$ and $k = 1, \dots, d$.

3.2. Two types with frequency-dependent selection

Consider next the case of two types with fertility and viability selection coefficients that depend linearly on the population state. Suppose that the total resources for reproduction and survival available to an individual of type i in age class $k \geq 1$ when the population state is $\mathbf{x}(\tau)$ is given by the linear function

$$\beta_{i,k}(\mathbf{x}(\tau)) = \sum_{l=1}^d (\beta_{i,k,1,l}x_{1,l}(\tau) + \beta_{i,k,2,l}(1 - x_{1,l}(\tau))) \tag{27}$$

for $i = 1, 2$ and $k = 1, \dots, d$. The parameter $\beta_{i,k,j,l}$ is such that $N_i^{-1}\beta_{i,k,j,l}$ represents the resources available to an individual of type i in age class k per individual of type j in age class l , for $i, j = 1, 2$ and $k, l = 1, \dots, d$. Moreover, let these resources be allocated to reproduction and survival with probabilities $q_{i,k}$ and $1 - q_{i,k}$, respectively, which define a resource allocation strategy (see, e.g., Baudisch and Vaupel, 2012; Lessard and Soares, 2018). More precisely, let the selection coefficients be given by

$$\rho_{i,k}(\mathbf{x}(\tau)) = q_{i,k}\beta_{i,k}(\mathbf{x}(\tau))$$

and

$$\sigma_{i,k}(\mathbf{x}(\tau)) = (1 - q_{i,k})\beta_{i,k}(\mathbf{x}(\tau)),$$

where $0 \leq q_{i,k} \leq 1$ for $i = 1, 2$ and $k = 1, \dots, d$, with $q_{i,d} = 1$ for $i = 1, 2$ so that all resources of the individuals in the last age class are allocated to reproduction. In this case, (23) becomes

$$\pi_{i,k}(\mathbf{x}(\tau)) = \beta_{i,k}(\mathbf{x}(\tau))v_{i,k} \tag{28}$$

with

$$v_{i,k} = q_{i,k}p_k w_1 + (1 - q_{i,k})D_k w_{k+1} \tag{29}$$

for $k = 1, \dots, d - 1$ and $v_{i,d} = p_d w_1$, for $i = 1, 2$. This yields

$$\begin{aligned} \pi_{1,k}(\mathbf{x}(\tau)) - \pi_{2,k}(\mathbf{x}(\tau)) &= \beta_{1,k}(\mathbf{x}(\tau))v_{1,k} - \beta_{2,k}(\mathbf{x}(\tau))v_{2,k} \\ &= \sum_{l=1}^d (a_{k,l}x_{1,l}(\tau) + d_{k,l}), \end{aligned} \tag{30}$$

where

$$a_{k,l} = (\beta_{1,k,1,l} - \beta_{1,k,2,l})v_{1,k} - (\beta_{2,k,1,l} - \beta_{2,k,2,l})v_{2,k} \tag{31}$$

and

$$d_{k,l} = \beta_{1,k,2,l}v_{1,k} - \beta_{2,k,2,l}v_{2,k} \tag{32}$$

for $l, k = 1, \dots, d$. Then, (24) can be written as

$$\begin{aligned} \mathbf{w}^T \phi_1(\mathbf{x}(\tau)) &= \sum_{k,l=1}^d a_{k,l}x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,l}(\tau) \\ &+ \sum_{k,l=1}^d d_{k,l}x_{1,k}(\tau)(1 - x_{1,k}(\tau)). \end{aligned} \tag{33}$$

This leads to

$$\begin{aligned} u'_1(0) &= \sum_{k,l=1}^d a_{k,l} \sum_{\tau \geq 0} E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,l}(\tau)] \\ &+ \sum_{k,l=1}^d d_{k,l} \sum_{\tau \geq 0} E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau))] \end{aligned} \tag{34}$$

for the first-order effect of selection on the fixation probability for type 1.

4. Ancestry coefficients in a large neutral population

In order to obtain the first-order effect of selection on the probability of fixation of one of two types, say type 1 over type 2, when the selection coefficients are either constant or linear with respect to the type frequencies, we need to compute

$$\begin{aligned} &\sum_{\tau \geq 0} E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau))] \\ \text{and} \\ &\sum_{\tau \geq 0} E_0 [x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,l}(\tau)] \end{aligned} \tag{35}$$

for $k, l = 1, \dots, d$. These are coefficients that depend on the ancestry of sampled individuals in the population. The computation will be done for individuals in a large population under neutrality and the assumption that, at time step 0, there is only one individual of type 1 in age class 1, while all other individuals are of type 2.

The method used below for the computation comes from Ladret and Lessard (2007, 2008) and Lessard (2011a). Ordering arbitrarily the N_k individuals in age class $k = 1, \dots, d$, we have

$$x_{1,k}(\tau) = \frac{1}{N_k} \sum_{m=1}^{N_k} \xi_{k,m}(\tau),$$

where

$$\xi_{k,m}(\tau) = \begin{cases} 1 & \text{if individual } m \text{ in age class } k \text{ is of type 1 at time step } \tau, \\ 0 & \text{if individual } m \text{ in age class } k \text{ is of type 2 at time step } \tau. \end{cases}$$

Then,

$$\begin{aligned} E_0[x_{1,k}(\tau)(1 - x_{1,k}(\tau))] &= \frac{1}{N_k^2} \sum_{m=1}^{N_k} \sum_{n=1}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))] \\ &= \frac{1}{N_k^2} \sum_{m=1}^{N_k} \sum_{n=1, n \neq m}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))] \\ &= \frac{(N_k - 1)}{N_k} E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))]. \end{aligned}$$

Note that $E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))]$ is the probability for two individuals chosen at random without replacement at time step τ to be of types 1 and 2 in this order. On the other hand, going backward in time from time step τ to time step 0 and using the assumption of a single individual of type 1 in age class 1 of size N_1 at time step 0, we have

$$E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))] = \sum_{l=1}^d p_{kk,1l}(\tau) \frac{1}{N_1},$$

where $p_{kk,1l}(\tau)$ is the probability for two individuals chosen at random in age class k at time step τ to have two distinct ancestors in age classes 1 and l , respectively, at time step 0, while $1/N_1$ is the probability for an ancestor in age class 1 at time step 0 to be of type 1. With $f_1 = N_1/N$ and N/α time steps as unit of time where

$$\alpha = \frac{1}{f_1} w_1^2 + \sum_{i=1}^{d-1} w_{i+1}^2 \left(\frac{1}{f_{i+1}} - \frac{1}{f_i} \right), \tag{36}$$

we have

$$\sum_{\tau \geq 0} p_{kk,1l}(\tau) \frac{1}{N_1} = \frac{1}{\alpha f_1} \int_0^{+\infty} p_{kk,1l}(\lfloor Nt/\alpha \rfloor) dt.$$

Here, $\lfloor Nt/\alpha \rfloor$ denotes the floor value of Nt/α and gives the number of time steps corresponding to time t . Moreover, owing to a convergence result on the genealogical process based on a lemma for Markov chains with two timescales due to Möhle (1998) (see Appendix A), we have

$$\lim_{N \rightarrow +\infty} \int_0^{+\infty} p_{kk,1l}(\lfloor Nt/\alpha \rfloor) dt = w_1 w_l \int_0^{+\infty} e^{-t} dt = w_1 w_l.$$

This is the expected time for the lineages backward in time of two distinct individuals in age class k to be in age classes 1 and l , respectively, in the limit of a large population size. This means that the lineages forget their starting age class over a short timescale, while they spend time in the different age classes according to their stationary distribution independently of one another over a long timescale. Finally, using the fact that $\sum_{l=1}^d w_l = 1$, we get

$$\lim_{N \rightarrow +\infty} \sum_{\tau \geq 0} E_0[x_{1,k}(\tau)(1 - x_{1,k}(\tau))] = \frac{w_1}{f_1 \alpha} \tag{37}$$

for $k = 1, \dots, d$.

In a similar way, we have

$$\begin{aligned} E_0[x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,k}(\tau)] &= \frac{1}{N_k^3} \sum_{m=1}^{N_k} \sum_{n=1}^{N_k} \sum_{l=1}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))\xi_{k,l}(\tau)] \\ &= \frac{1}{N_k^3} \sum_{m=1}^{N_k} \sum_{n=1, n \neq m}^{N_k} \sum_{l=1, l \neq m, n}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))\xi_{k,l}(\tau)] \\ &+ \frac{1}{N_k^3} \sum_{m=1}^{N_k} \sum_{n=1, n \neq m}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))\xi_{k,m}(\tau)] \end{aligned}$$

$$\begin{aligned} &= \frac{(N_k - 1)(N_k - 2)}{N_k^2} E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))\xi_{k,3}(\tau)] \\ &+ \frac{(N_k - 1)}{N_k^2} E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))] \end{aligned}$$

and

$$\begin{aligned} E_0[x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,j}(\tau)] &= \frac{1}{N_k^2 N_j} \sum_{m=1}^{N_k} \sum_{n=1}^{N_k} \sum_{l=1}^{N_j} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))\xi_{j,l}(\tau)] \\ &= \frac{1}{N_k^2 N_j} \sum_{l=1}^{N_j} \sum_{m=1}^{N_k} \sum_{n=1, n \neq m}^{N_k} E_0[\xi_{k,m}(\tau)(1 - \xi_{k,n}(\tau))\xi_{j,l}(\tau)] \\ &= \frac{(N_k - 1)}{N_k} E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))\xi_{j,3}(\tau)] \end{aligned}$$

for $j \neq k$. Going backward in time from time step τ to time step 0, we have in this case

$$E_0[\xi_{k,1}(\tau)(1 - \xi_{k,2}(\tau))\xi_{j,3}(\tau)] = \sum_{l=1}^d p_{kkj,1l}(\tau) \frac{1}{3N_1},$$

where $p_{kkj,1l}(\tau)$ is the probability for three individuals at time step τ , two in age class k and one in age class j , to have exactly two ancestors at time step 0, one ancestor of two individuals in age class 1 and the other ancestor in age class l , while $1/3$ is the probability for two individuals in particular to be the ones having a common ancestor in age class 1 at time step 0 and $1/N_1$ is the probability for this ancestor to be the only one of type 1 in age class 1. As previously, we have

$$\sum_{\tau \geq 0} p_{kkj,1l}(\tau) \frac{1}{N_1} = \frac{1}{\alpha f_1} \int_0^{+\infty} p_{kkj,1l}(\lfloor Nt/\alpha \rfloor) dt$$

with $f_1 = N_1/N$, the change of variable $t = \alpha\tau/N$ and α given in (36). The convergence result in Appendix A yields

$$\begin{aligned} \lim_{N \rightarrow +\infty} \int_0^{+\infty} p_{kkj,1l}(\lfloor Nt/\alpha \rfloor) dt &= w_1 w_l \int_0^{+\infty} \frac{3}{2} (e^{-t} - e^{-3t}) dt \\ &= w_1 w_l. \end{aligned}$$

This is the same limit as previously since only the time spent with two lineages matters for the event considered. Using the fact that $\sum_{l=1}^d w_l = 1$, we have in this case

$$\lim_{N \rightarrow +\infty} \sum_{\tau \geq 0} E_0[x_{1,k}(\tau)(1 - x_{1,k}(\tau))x_{1,j}(\tau)] = \frac{w_1}{3f_1 \alpha}$$

for $k, j = 1, \dots, d$.

5. Effect of weak selection on fixation probability in a large population

Following Nowak et al. (2004), a given type is said to be favored by selection if its fixation probability is larger under selection than that it would be under neutrality. In a context of two types, 1 and 2, in a large population and ignoring degenerate cases, type 1 is favored by weak selection if and only if

$$\lim_{N \rightarrow +\infty} u'_1(0) > 0. \tag{38}$$

Here, $u'_1(0)$ is the derivative of the fixation probability with respect to the intensity of selection evaluated at 0 and represents the first-order effect of selection on the fixation probability. In the framework of our age-structured population model, we assume throughout that there is initially only one individual of type 1 and that this individual is in age class 1.

5.1. Constant selection

Owing to the previous sections, the first-order effect of constant fertility and viability selection on the fixation probability for type 1 in a large population is given by

$$\lim_{N \rightarrow +\infty} u'_1(0) = \frac{w_1}{f_1 \alpha} \sum_{k=1}^d (\pi_{1,k}^* - \pi_{2,k}^*), \tag{39}$$

where α is given in (36), while

$$\pi_{i,k}^* = \lim_{N \rightarrow +\infty} \pi_{i,k} = \rho_{i,k} p_k w_1 + \sigma_{i,k} C_k w_{k+1} \tag{40}$$

for $i = 1, 2$ and $k = 1, \dots, d$ with $w_{d+1} = 0$ and C_k given in (6) for $k = 1, \dots, d - 1$. Owing to (38), we conclude that type 1 is favored by weak selection in the limit of a large population if and only if

$$\sum_{k=1}^d \pi_{1,k}^* > \sum_{k=1}^d \pi_{2,k}^*. \tag{41}$$

The quantity $\pi_{i,k}^*$ can be viewed as a fixation-fitness coefficient for type i in age class k and its sum over $k = 1, \dots, d$ as a fixation-fitness coefficient for type i in the whole population.

5.2. Frequency-dependent selection

In the case of frequency-dependent viability and fertility selection, we have

$$\lim_{N \rightarrow +\infty} u'_1(0) = \frac{w_1}{3f_1 \alpha} \sum_{k,l=1}^d (a_{k,l}^* + 3d_{k,l}^*), \tag{42}$$

where α is given in (36), while

$$a_{k,l}^* = \lim_{N \rightarrow +\infty} a_{k,l} = (\beta_{1,k,1,l} - \beta_{1,k,2,l}) v_{1,k}^* - (\beta_{2,k,1,l} - \beta_{2,k,2,l}) v_{2,k}^* \tag{43}$$

and

$$d_{k,l}^* = \lim_{N \rightarrow +\infty} d_{k,l} = \beta_{1,k,2,l} v_{1,k}^* - \beta_{2,k,2,l} v_{2,k}^* \tag{44}$$

for $k, l = 1, \dots, d$ with

$$v_{i,k}^* = \lim_{N \rightarrow +\infty} v_{i,k} = q_{i,k} p_k w_1 + (1 - q_{i,k}) C_k w_{k+1} \tag{45}$$

for $i = 1, 2$ and $k = 1, \dots, d$. Owing to (38), this leads to the necessary and sufficient condition

$$\sum_{k,l=1}^d (a_{k,l}^* + 3d_{k,l}^*) > 0, \tag{46}$$

where

$$a_{k,l}^* + 3d_{k,l}^* = (q_{1,k} p_k w_1 + (1 - q_{1,k}) C_k w_{k+1}) (\beta_{1,k,1,l} + 2\beta_{1,k,2,l}) - (q_{2,k} p_k w_1 + (1 - q_{2,k}) C_k w_{k+1}) (\beta_{2,k,1,l} - 2\beta_{2,k,2,l})$$

for $k, l = 1, \dots, d$, for type 1 to be favored by weak selection in the limit of a large population.

6. Fixation probability with constant selection in two age classes

In this section, we are interested in the fixation probability in a population with two types in two age classes. As previously, type 1 is represented only once at time step 0 in age class 1.

Let the age classes produce offspring in the same proportions as their sizes so that

$$p_1 = f_1 = \frac{N_1}{N} = f$$

and

$$p_2 = f_2 = \frac{N_2}{N} = 1 - f.$$

Note that $1/2 \leq f \leq 1$, since $N_2 \leq N_1 = N - N_2$. Moreover, according to (16), the reproductive values of age classes 1 and 2 under neutrality are given by

$$w_1 = \frac{1}{2 - f}$$

and

$$w_2 = 1 - w_1 = \frac{1 - f}{2 - f},$$

respectively. Then, the parameter α in (36) for the time change in the neutral genealogical process becomes

$$\alpha = \frac{1}{f_1} w_1^2 + w_2^2 \left(\frac{1}{f_2} - \frac{1}{f_1} \right) = \frac{3 - 2f}{(2 - f)^2}, \tag{47}$$

while (6) gives

$$C_1 = \frac{f_1 - f_2}{f_2} \ln \left(\frac{f_1}{f_1 - f_2} \right) = \frac{2f - 1}{1 - f} \ln \left(\frac{f}{2f - 1} \right) \tag{48}$$

as population-structure coefficient for age class 1. Note that α and C_1 as well as w_1 , the reproductive value of age class 1 under neutrality, are increasing functions with respect the frequency f of age class 1.

According to (20), the fixation probability for type 1 under neutrality is given by

$$u_1(0) = \frac{w_1}{N_1} = \frac{1}{Nf(2 - f)}, \tag{49}$$

which is a decreasing function with respect to f . On the other hand, according to (39), the first-order effect of constant fertility and viability selection on the fixation probability in the limit of a large population is given by

$$\begin{aligned} & \lim_{N \rightarrow +\infty} u'_1(0) \\ &= \frac{w_1}{f \alpha} \sum_{k=1}^2 (\pi_{1,k}^* - \pi_{2,k}^*) \\ &= \frac{1}{f(3 - 2f)} ((\rho_{1,1} - \rho_{2,1})f + (\sigma_{1,1} - \sigma_{2,1})(1 - f)C_1 \\ & \quad + (\rho_{1,2} - \rho_{2,2})(1 - f)). \end{aligned} \tag{50}$$

Taking $s = N^{-1}$ as the intensity of selection, we have

$$\begin{aligned} & \lim_{N \rightarrow +\infty} N u_1(N^{-1}) \\ &= \frac{1}{f(2 - f)} + \frac{1}{f(3 - 2f)} ((\rho_{1,1} - \rho_{2,1})f + (\sigma_{1,1} - \sigma_{2,1})(1 - f)C_1 \\ & \quad + (\rho_{1,2} - \rho_{2,2})(1 - f)). \end{aligned} \tag{51}$$

This gives an approximation of the fixation probability for type 1 as a function of f , the proportion of age class 1, in the case of a large population size. Relative to the fixation probability for a

neutral mutant, the fixation probability under selection satisfies

$$\begin{aligned} & \lim_{N \rightarrow +\infty} \frac{u_1(N^{-1})}{u_1(0)} \\ &= 1 + \frac{1}{\alpha} \left((\rho_{1,1} - \rho_{2,1})fw_1 + (\sigma_{1,1} - \sigma_{2,1})(1-f)C_1w_2 \right. \\ & \quad \left. + (\rho_{1,2} - \rho_{2,2})(1-f)w_1 \right) \end{aligned}$$

with α given in (47).

6.1. Selection on reproduction only

Consider the case where

$$\rho_{1,1} = \rho_{1,2} = 1, \rho_{2,1} = \rho_{2,2} = 0 \text{ and } \sigma_{1,1} = \sigma_{2,1} = 0.$$

This means that type 1 is favored with respect to reproduction and this evenly in both age classes, 1 and 2, in comparison to type 2. On the other hand, both types are undistinguishable with respect to survival from age class 1 to age class 2.

In this case, we have

$$\lim_{N \rightarrow +\infty} Nu_1(N^{-1}) = \frac{1}{f(2-f)} + \frac{1}{f(3-2f)} = \frac{5-3f}{f(2-f)(3-2f)}.$$

This approximation of the fixation probability for type 1 in a large population for $1/2 \leq f \leq 1$ is represented in Fig. 1a.

This recovers the classical fixation probability for an advantageous mutant times the population size when $f = 1$, which is given by 2. Note that $f = 1$ corresponds to the Wright–Fisher model. Moreover, relative to the fixation probability for a neutral mutant, we have

$$\lim_{N \rightarrow +\infty} \frac{u_1(N^{-1})}{u_1(0)} = 1 + \frac{w_1}{\alpha} = 1 + \frac{2-f}{3-2f}.$$

This approximation, which is an increasing function with respect to $1/2 \leq f \leq 1$, is represented in Fig. 2a.

As f increases from $1/2$ to 1 , the fixation probability for a mutant type 1 with selection on reproduction only starts decreasing with the decrease of the frequency of the single mutant in age class 1, but ends up increasing with the increase of the reproductive value of age class 1. The value of f that minimizes the fixation probability is approximately equal to 0.85.

6.2. Selection on survival only

In the case where

$$\sigma_{1,1} = 1, \sigma_{2,1} = 0 \text{ and } \rho_{1,1} = \rho_{1,2} = \rho_{2,1} = \rho_{2,2} = 0,$$

type 1 is favored with respect to survival from age class 1 to age class 2 in comparison to type 2, but both types reproduce equally in both age classes.

In this case, we have

$$\lim_{N \rightarrow +\infty} Nu_1(N^{-1}) = \frac{1}{f(2-f)} + \frac{(2f-1)}{(3-2f)f} \ln\left(\frac{f}{2f-1}\right),$$

which is represented in Fig. 1b for $1/2 \leq f \leq 1$.

The approximation of the fixation probability for type 1 relative to the fixation probability for a neutral mutant is given in this case by

$$\begin{aligned} \lim_{N \rightarrow +\infty} \frac{u_1(N^{-1})}{u_1(0)} &= 1 + \frac{(1-f)C_1w_2}{\alpha} \\ &= 1 + \frac{(2f-1)(2-f)}{3-2f} \ln\left(\frac{f}{2f-1}\right). \end{aligned}$$

This is a concave function with respect to $1/2 \leq f \leq 1$ (see Fig. 2b).

With selection on survival only, the fixation probability for type 1 as a function of $1/2 \leq f \leq 1$ starts increasing with a strong increase of the population-structure coefficient of age class 1, but ends up decreasing with the decrease of the reproductive value of age class 2. This leads to a maximum fixation probability when f is approximately equal to 0.55.

6.3. Selection on both reproduction and survival

Finally, in the case where

$$\rho_{1,1} = \rho_{1,2} = \sigma_{1,1} = 1 \text{ and } \rho_{2,1} = \rho_{2,2} = \sigma_{2,1} = 0,$$

type 1 is favored with respect to both reproduction and survival in comparison to type 2.

In this case, we have

$$\begin{aligned} \lim_{N \rightarrow +\infty} Nu_1(N^{-1}) &= \frac{1}{f(2-f)} + \frac{1}{f(3-2f)} \\ & \quad \times \left(1 + (2f-1) \ln\left(\frac{f}{2f-1}\right) \right). \end{aligned}$$

The curve of this function for $1/2 \leq f \leq 1$ is represented in Fig. 1c. Moreover, we have

$$\begin{aligned} \lim_{N \rightarrow +\infty} \frac{u_1(N^{-1})}{u_1(0)} &= 1 + \frac{1}{\alpha} (w_1 + (1-f)Cw_2) \\ &= 1 + \frac{2-f}{3-2f} \left(1 + (2f-1) \ln\left(\frac{f}{2f-1}\right) \right) \end{aligned}$$

as illustrated in Fig. 2c. As a function of $1/2 \leq f \leq 1$, the fixation probability for type 1 is between the fixation probabilities with fertility selection only and with viability selection only. There is a maximum value when f is approximately equal to 0.53.

7. Public goods game

In this section, we consider a population with two types of individuals ($n = 2$), cooperators (type 1) and defectors (type 2). Each cooperator pays a cost to produce a public good that is multiplied by some multiplicative reward factor and distributed evenly among all individuals in the same group, cooperators and defectors. Defectors do not pay anything. With groups of size N , this is a public goods game known as the N -person prisoner's dilemma (see, e.g., Archetti and Scheuring, 2012, for a review in the case of a well-mixed population without age structure).

In an age-structured population with d age classes, let a cooperator in age class k pay a price c_k to produce a public good b_k , with $0 < c_k \leq 1$ and $0 < b_k$, for $k = 1, \dots, d$. The total public good produced if the frequency of cooperators in age class k of size N_k is $x_{1,k}(\tau)$ at time step $\tau \geq 0$ for $k = 1, \dots, d$ is

$$\sum_{k=1}^d b_k N_k x_{1,k}(\tau).$$

The total public good is multiplied by a reward factor $r > 0$ and then distributed among the individuals in the population according to their age classes. It is assumed that a fraction h_k of the total public good produced is allocated equally to all individuals in age class k with $0 \leq h_k \leq 1$ for $k = 1, \dots, d$ and

$$\sum_{k=1}^d h_k = 1.$$

As a result, at time step $\tau \geq 0$, a defector (type 2) in age class k receives a payoff

$$\pi_{2,k}(\mathbf{x}(\tau)) = \frac{h_k}{N_k} r \sum_{l=1}^d b_l N_l x_{1,l}(\tau) = \sum_{l=1}^d r b_l h_k \frac{f_l}{f_k} x_{1,l}(\tau), \tag{52}$$

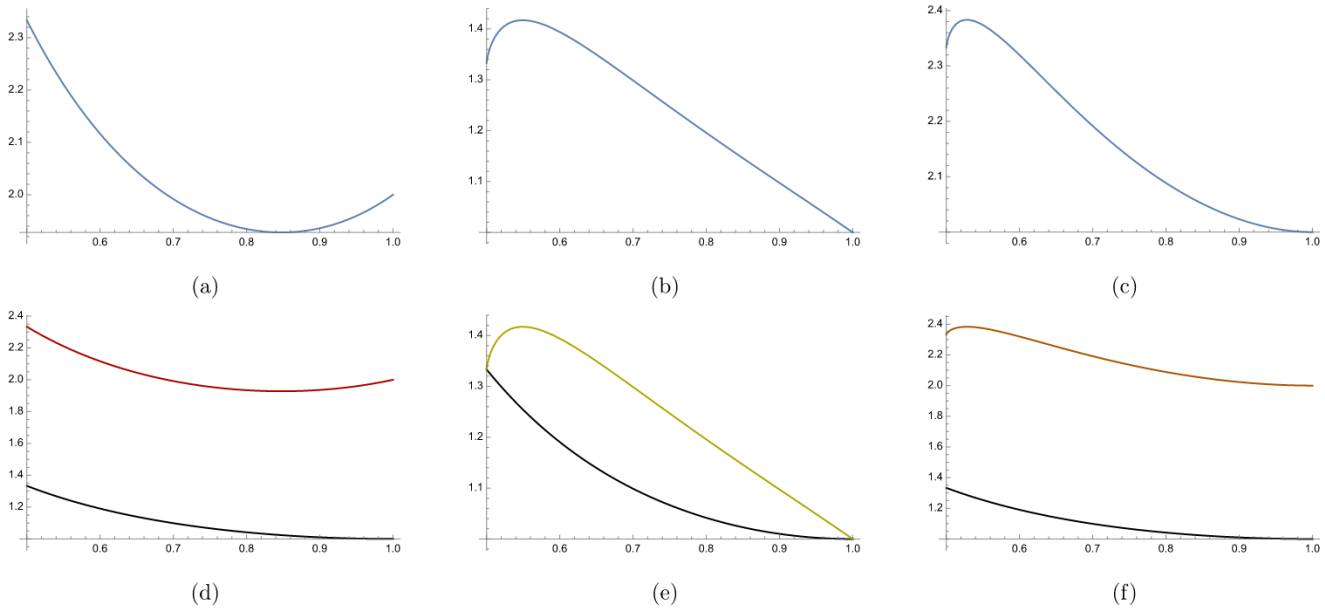


Fig. 1. Fixation probability for type 1 times the population size in the limit of a large two-age-class population with respect to the proportion of age class 1: (a) selection on reproduction only; (b) selection on survival only; (c) selection on both reproduction and survival; (d) selection on reproduction only (upper curve) and neutral model (lower curve); (e) selection on survival only (upper curve) and neutral model (lower curve); and (f) selection on both reproduction and survival (upper curve) and neutral model (lower curve), with the inverse of the population size as the intensity of selection.

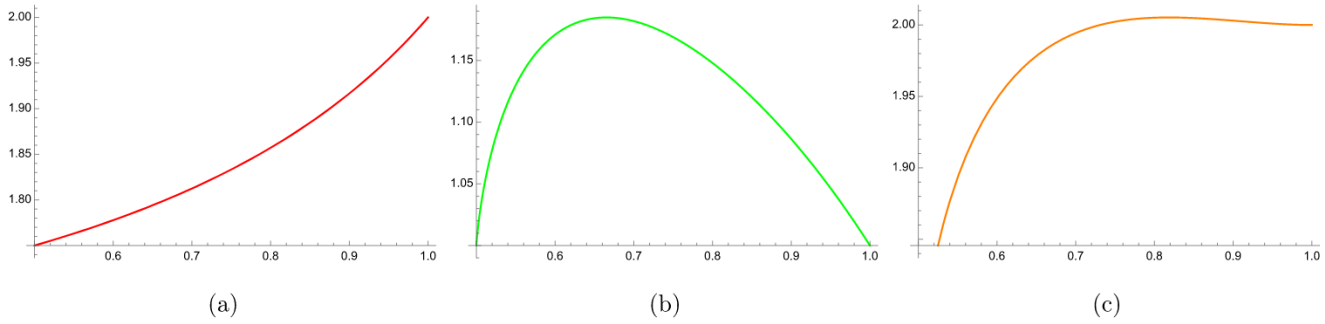


Fig. 2. Fixation probability for type 1 relative to fixation probability for a neutral mutant in the limit of a large two-age-class population with respect to the proportion of age class 1: (a) selection on survival only (b) selection on reproduction only; and (c) selection on both reproduction and survival, with the inverse of the population size as the intensity of selection.

while a cooperator (type 1) in the same age class receives a payoff

$$\pi_{1,k}(\mathbf{x}(\tau)) = \pi_{2,k}(\mathbf{x}(\tau)) - c_k, \tag{53}$$

for $k = 1, \dots, d$.

Suppose that the payoff received by an individual affects its capacity to survive and reproduce. More precisely, let

$$\beta_{i,k}(\mathbf{x}(\tau)) = 1 + \pi_{i,k}(\mathbf{x}(\tau))$$

be the total resources allocated to reproduction and survival by an individual of type i in age class k at time step $\tau \geq 0$ for $i = 1, 2$ and $k = 1, \dots, d$. Then, the parameters in the linear function (27) are given by

$$\beta_{2,k,1,l} = rb_l h_k \frac{f_l}{f_k} + \frac{1}{d}, \quad \beta_{1,k,1,l} = rb_l h_k \frac{f_l}{f_k} + \frac{1 - c_k}{d}$$

and

$$\beta_{2,k,2,l} = \frac{1}{d}, \quad \beta_{1,k,2,l} = \frac{1 - c_k}{d}$$

for $k, l = 1, \dots, d$. With these parameters, (43) and (44) become

$$a_{k,l}^* = rb_l h_k \frac{f_l}{f_k} (v_{1,k}^* - v_{2,k}^*)$$

and

$$d_{k,l}^* = \frac{1 - c_k}{d} v_{1,k}^* - \frac{1}{d} v_{2,k}^*$$

with $v_{i,k}^*$ given in (45) for $i = 1, 2$ and $k = 1, \dots, d$. Then, condition (46) for cooperation (type 1) to be favored by weak selection in the limit of a large population reduces to

$$\sum_{k=1}^d v_{1,k}^* \left(1 - c_k + r\bar{b} \frac{h_k}{3f_k} \right) > \sum_{k=1}^d v_{2,k}^* \left(1 + r\bar{b} \frac{h_k}{3f_k} \right),$$

that is,

$$r\bar{b} \sum_{k=1}^d (v_{1,k}^* - v_{2,k}^*) \frac{h_k}{3f_k} > S \tag{54}$$

with

$$S = \sum_{k=1}^d (v_{2,k}^* - v_{1,k}^* (1 - c_k)), \tag{55}$$

where

$$\bar{b} = \sum_{l=1}^d f_l b_l.$$

The quantity $v_{i,k}^*$ can be viewed as a fixation-fitness coefficient for type i in age class k per unit of resources. Note that the types differ not only by their behavior (cooperation or defection) but also by their allocation of resources to survival or reproduction.

We are interested in the public good allocation strategy $\mathbf{h} = (h_1, \dots, h_d)^T$ that makes it easier for cooperation to be favored by weak selection. Let us define

$$L(\mathbf{h}) = \frac{\bar{b}}{3S} \sum_{k=1}^d (v_{1,k}^* - v_{2,k}^*) \frac{h_k}{f_k}. \tag{56}$$

If $S > 0$, then condition (54) is equivalent to

$$\frac{1}{r} < L(\mathbf{h}).$$

The linear function $L(\mathbf{h})$ with $0 \leq h_k \leq 1$ for $k = 1, \dots, d$ and $\sum_{k=1}^d h_k = 1$ has d critical points, namely, \mathbf{e}_k with

$$L(\mathbf{e}_k) = \frac{\bar{b}(v_{1,k}^* - v_{2,k}^*)}{3Sf_k}$$

for $k = 1, \dots, d$, where \mathbf{e}_k is the d -dimensional vector with 1 in the k th entry and 0 elsewhere. Therefore, the most advantageous public good allocation strategy for cooperation to be favored by weak selection is $\mathbf{h} = \mathbf{e}_{\hat{k}}$ where \hat{k} is such that

$$\frac{v_{1,\hat{k}}^* - v_{2,\hat{k}}^*}{f_{\hat{k}}} = \max_{1 \leq k \leq d} \frac{(v_{1,k}^* - v_{2,k}^*)}{f_k} > 0.$$

Note that, if $v_{1,k}^* - v_{2,k}^* \leq 0$ for all $k = 1, \dots, d$, then cooperation can never be favored by weak selection. On the other hand, if $S < 0$, then there exists \hat{k} such that

$$v_{2,\hat{k}}^* - v_{1,\hat{k}}^* \leq v_{2,\hat{k}}^* - v_{1,\hat{k}}^*(1 - c_{\hat{k}}) < 0,$$

while condition (54) is equivalent to

$$\frac{1}{r} > L(\mathbf{h}).$$

This condition holds for $\mathbf{h} = \mathbf{e}_{\hat{k}}$ since $L(\mathbf{e}_{\hat{k}}) < 0$ and $r > 0$. Finally, if $S = 0$, then

$$\sum_{k=1}^d v_{2,k}^* = \sum_{k=1}^d v_{1,k}^*(1 - c_k) < \sum_{k=1}^d v_{1,k}^*,$$

while condition (54) is equivalent to

$$\sum_{k=1}^d (v_{1,k}^* - v_{2,k}^*) \frac{h_k}{f_k} > 0. \tag{57}$$

This is the case, for instance, if the public good is allocated according to $\mathbf{h} = (h_1, \dots, h_d)^T$ with $h_k = f_k$ for $k = 1, \dots, d$.

8. Public goods game in two age classes

In this section, we analyze in more detail the public goods game in two age classes ($d = 2$), age class 1 for young and age class 2 for adults. Using $p_1 = f_1 = f$ and $p_2 = f_2 = 1 - f$ for $1/2 \leq f \leq 1$, and recalling that $q_{1,2} = q_{2,2} = 1$, Eq. (45) yields

$$v_{i,2}^* = \frac{1-f}{2-f}$$

and

$$v_{i,1}^* = q_{i,1} \frac{f}{2-f} + (1 - q_{i,1}) C_1 \frac{1-f}{2-f} \tag{58}$$

Table 2

Conditions for weak selection to favor cooperation in the case of a public goods game in two age classes.

Case	Condition for weak selection to favor cooperation
$q_{1,1} \leq q_{2,1}$	Never
$q_{1,1} > q_{2,1}$ and $S > 0$	$r > \frac{1}{L((1,0))}$
$q_{1,1} > q_{2,1}$ and $S < 0$	$r > 0$
$q_{1,1} > q_{2,1}$ and $S = 0$	$r > 0$ if $0 < h \leq 1$

with C_1 given in (48). Hence,

$$v_{1,2}^* - v_{2,2}^* = 0$$

and

$$v_{1,1}^* - v_{2,1}^* = (q_{1,1} - q_{2,1}) \left(\frac{f}{2-f} - C_1 \frac{1-f}{2-f} \right),$$

where

$$\frac{f}{2-f} - C_1 \frac{1-f}{2-f} > 0$$

for $1/2 \leq f \leq 1$. Moreover, (56) and (55) reduce to

$$L(\mathbf{h}) = \frac{\bar{b}h}{3Sf} (v_{1,1}^* - v_{2,1}^*)$$

and

$$S = v_{2,1}^* - v_{1,1}^*(1 - c_1) + v_{1,2}^* c_2,$$

respectively.

If $S > 0$ and $q_{1,1} > q_{2,1}$, then $L(\mathbf{h}) > 0$ and weak selection favors cooperation if

$$\frac{1}{r} < L(\mathbf{h}) \leq L((1,0)),$$

which implies that the most advantageous public good allocation strategy for this to occur is $\mathbf{h} = (1, 0)^T$. If $S > 0$ and $q_{1,1} \leq q_{2,1}$, then $L(\mathbf{h}) \leq 0$ and it is not possible to have

$$\frac{1}{r} < L(\mathbf{h})$$

for every $r > 0$, which means that cooperation can never be favored by weak selection. On the other hand, if $S < 0$, then necessarily $q_{2,1} < q_{1,1}$ and $L(\mathbf{h}) < 0$, which implies that the condition

$$\frac{1}{r} > L(\mathbf{h})$$

holds for every $r > 0$, in which case cooperation is always favored by weak selection. Finally, if $S = 0$, then necessarily $q_{1,1} > q_{2,1}$, which implies that

$$(v_{1,1}^* - v_{2,1}^*) \frac{h}{f} > 0$$

for $0 < h \leq 1$, in which case cooperation is favored by weak selection for all reward factor $r > 0$ according to condition (57).

These results are summarized in Table 2.

In the case of two age classes, the fixation-fitness coefficients of adults are equal for both types, cooperators and defectors, since they both allocate all their resources to reproduction. The only possibility for cooperators to increase their fixation probability under weak selection is to allocate more resources to reproduction than defectors do when they are young. The reason is that the fixation-fitness coefficient in age class 1 increases with the proportion of resources allocated to reproduction (see Eq. (58)).

9. Discussion

In this paper, we have studied weak frequency-dependent viability and fertility selection in a finite haploid population in discrete time with a fixed age class structure, where the number of individuals in each age class is kept constant from one time step to the next one.

We have been interested in the first-order effect of selection on the fixation probability for a mutant type. We have obtained this effect in terms of selection coefficients for the different types, reproductive values of age classes and ancestry coefficients for sampled individuals by applying a small perturbation method to the neutral process. In the limit of a large population size, a two-timescale argument based on a lemma due to Möhle (1998) as in Kroumi and Lessard (2014, 2015a) for a strong migration limit in a subdivided population with non-overlapping generations leads to ancestry coefficients in a well-mixed population of effective size $N_e = N/(\bar{k}\alpha)$ as in Felsenstein (1971), where N is the total size of the neutral age-structured population and \bar{k} the generation time in this population. The parameter α corresponds to the rate of coalescence of two lineages in the limit of a large neutral age-structured population with N/\bar{k} time steps as unit of time. As for the reproductive values of the age classes in this population, they are given by the elements of the stationary probability distribution of a backward transition matrix under neutrality. These correspond to the contributions of the different age classes to all future generations in a neutral population. Finally, and not the least, the age structure of the population introduces a bias in the strength of viability selection. The coefficients of viability selection are weighted not only by reproductive values but also by population-structure coefficients for the age classes that depend on their relative sizes.

We have applied our results to the case of constant viability and fertility parameters and to a public goods game with the total public good distributed among the age classes according to some population allocation strategy and the resources allocated to survival and reproduction according to the type of the individuals (cooperator or defector). Considering that a mutant type is favored by weak selection if the first-order effect of selection on its fixation probability is positive (Nowak et al., 2004), we have shown that this is the case with constant selection parameters if some fixation-fitness coefficient of the mutant type exceeds that of the resident type. This coefficient is a sum of fixation-fitness coefficients for the different age classes that depend on the reproductive values, the coefficients of selection and the population-structure coefficients.

We have shown that the reproductive values and the population-structure coefficients affect differently the coefficients of viability selection and fertility selection. With two age classes and a single mutant introduced in the first age class, the fixation probability as a function of the frequency of the first age class is decreasing under neutrality. The first-order effect of a selective advantage of the mutant is increasing with respect to this frequency because this is the case for the reproductive value of the first age class with a selective advantage in reproduction, and for the population-structure coefficient from the first age class to the second one with a selective advantage in survival. This explains the different patterns found in Li et al. (2016) by numeral iterations and simulations in small populations with two age classes.

In the case of the public goods game with some public good produced at a fixed cost and multiplied by some reward factor, we have shown that cooperation is favored by weak selection if a sum of coefficients of fixation fitness for cooperation in the different age classes exceeds that for defection. These depend not only on the reproductive values, the selection coefficients and

the population-structure coefficients, but also on the public good allocation strategy of the population and the resource allocation strategies of the individuals. At least in a population with two age classes, it becomes possible for cooperation to be favored by weak selection if cooperators allocate more resources to reproduction versus survival than defectors do. This provides a way for cooperation to evolve in age-structured populations.

It may be useful to recall that the payoff to cooperation in a prisoner's dilemma with random pairwise interactions in a well-mixed population is always smaller than the payoff to defection whatever the population state is, in which case cooperation cannot be favored by selection. This is also the case for a public goods game. It is possible for cooperation in a prisoner's dilemma to become favored by selection if the dilemma is repeated enough time so that the payoff to the tit-for-tat strategy starting with cooperation gets larger than the payoff to defection as an effect of reciprocity if the frequency of tit-for-tat exceeds some threshold value. The threshold value $1/3$ in the limit of a large population with an update of strategies according to a Moran or Wright-Fisher model gives the one-third law of evolution (Nowak et al., 2004; Lessard, 2005; Imhof and Nowak, 2006), but it is less than $1/3$ with more skewed reproduction schemes (Lessard and Ladret, 2007). The one-third law can be explained by the fact that the fixation probability is determined by a kind of average over the population state up to fixation (Ohtsuki et al., 2007; Lessard and Lahaie, 2009; Lessard, 2011a). In family-structured populations, and more generally group-structured populations, the benefit of cooperation for relatives or group members can compensate for its cost (see, e.g., Lessard and Rocheleau, 2004; Nowak, 2006; Lessard, 2009, 2011a,c). Similarly, in graph-structured populations or social networks, viscosity of the environment can create conditions favorable to the evolution of cooperation (see, e.g., Ohtsuki et al., 2006; Nowak et al., 2010; Kroumi and Lessard, 2014, 2015a,b, for rules based on relative abundance in the presence of recurrent mutation instead of fixation probability in the absence of mutation). We have seen that other factors come into play in age-structured populations with age classes contributing differently to fixation and the possibility for individuals to invest in reproduction or survival.

Some simplifying assumptions have been made in our analysis of selection in age-structured populations. We have considered soft selection, which means competition within the age classes. Hard selection would introduce competition between the age classes. Apart more complex systems of equations, this would lead to consider ancestry coefficients for individuals sampled in different age classes. In the strong migration limit, however, these are the same as in a well-mixed population. Therefore, the analysis is not out of reach. This would also be the case in principle for a non-linear frequency-dependent selection model, e.g., a model for the evolution of cooperation in which the payoffs are general functions of the number of cooperators, but in practice ancestry coefficients for more than three individuals would have to be considered, and these become rapidly cumbersome. As for the assumption of a fixed age structure, it does not seem possible to get rid of it without considerably increasing the difficulty of the analysis. Without this assumption, we would have to prove fast convergence to a stable age distribution and slow changes in type frequencies in the limit of weak selection, and this is not obvious even for a deterministic model (Lessard and Soares, 2018).

Finally, note that the fixation probability in an age-structured population can be obtained from a diffusion approximation of a Markov chain with two timescales in the limit of a large population by checking conditions given in Ethier and Nagylaki (1980). The proof relies on rigorous approximations of the moments of the multivariate Wallenius' noncentral hypergeometric distribution by the moments of a multinomial distribution. Such an

approximation is more general and it can be used not only to obtain the fixation probability but also the stationary distribution in the case of recurrent mutation. The conditions for the validation of such an approximation, however, are more difficult to ascertain (Soares and Lessard, 2019).

Acknowledgments

This research was supported in part by NSERC of Canada, Grant no. 8833, and CAPES, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil, Finance Code 0968-13-7. We also appreciate constructive comments by two anonymous reviewers.

Appendix A. Limit genealogical process in the neutral model

In this section, the genealogical process in the neutral model is described. Considering the age classes of the ancestors for up to three individuals, we have a Markov chain on a discrete state space S from one time step to the previous one. The state space S can be partitioned into three subsets according to the number of ancestors. These are

$$S_1 = \{i \mid i = 1, \dots, d\}, \quad (59)$$

$$S_2 = \{ij \mid i, j = 1, \dots, d\} \quad (60)$$

and

$$S_3 = \{ijk \mid i, j, k = 1, \dots, d\}. \quad (61)$$

The order of the elements in S_1 is the natural order of the integers. The order of the elements in S_2 is given by

$$ij < i'j' \text{ if } [i < i'] \text{ or } [i = i' \text{ with } j < j'].$$

Finally, the order of the elements in S_3 is given by

$$ijk < i'j'k' \text{ if } [i < i'] \text{ or}$$

$$[i = i' \text{ with } j < j'] \text{ or } [i = i' \text{ and } j = j' \text{ with } k < k'].$$

As shown in Appendix B, the transition matrix of the Markov chain on S takes the form

$$P = R + N^{-1}A + o(N^{-1}). \quad (62)$$

Here, R is the transition matrix that would be obtained in a population of infinite size in which there is no coalescence event, while the elements of A/N are transition probabilities for events involving only one coalescence backward in time. Moreover,

$$R = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_3 \end{pmatrix}, \quad (63)$$

where R_1 , R_2 and R_3 are transition matrices for 1, 2 and 3 ancestors, respectively. Actually, we have $(R_2)_{ij,i'j'} = m_{i'i'}m_{j'j}$ and $(R_3)_{ijk,i'j'k'} = m_{i'i'}m_{j'j}m_{k'k}$, where $m_{i'i'} = (R_1)_{i,i'}$. It is assumed that $M = (m_{i'i'})$ is ergodic, in which case R_l is ergodic for $l = 1, 2, 3$. Then, the ergodic theorem (see, e.g., Karlin and Taylor, 1975, p.85) guarantees that

$$\lim_{n \rightarrow +\infty} R_l^n = \mathbf{1}_l \Pi_l^T,$$

where Π_l is the unique probability distribution satisfying $\Pi_l^T R_l = \Pi_l^T$ and $\mathbf{1}_l = (1, \dots, 1)^T$ is the d^l -dimensional vector of all ones, for $l = 1, 2, 3$. Since $R_1 = M$, we have $\Pi_1 = \mathbf{w}$ with $\mathbf{w} = (w_1, \dots, w_d)^T$ being the stationary probability distribution of the stochastic matrix M . Moreover, we have

$$(\Pi_2)_{ij} = w_i w_j$$

and

$$(\Pi_3)_{ijk} = w_i w_j w_k$$

for $i, j, k = 1, \dots, d$. On the other hand, the matrix A is in the block form

$$A = \begin{pmatrix} 0 & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & A_{32} & A_{33} \end{pmatrix}. \quad (64)$$

See Appendix B for more details.

Then, applying a lemma due to Möhle (1998) to the transition matrix from time 0 to time t in the past with N time steps as unit of time, we obtain

$$\lim_{N \rightarrow +\infty} P^{\lfloor Nt \rfloor} = He^{tG},$$

where

$$H = \begin{pmatrix} \mathbf{1}_1 \Pi_1^T & 0 & 0 \\ 0 & \mathbf{1}_2 \Pi_2^T & 0 \\ 0 & 0 & \mathbf{1}_3 \Pi_3^T \end{pmatrix}$$

and

$$G = HAH = \begin{pmatrix} 0 & 0 & 0 \\ (\Pi_2^T A_{21} \mathbf{1}_1) \mathbf{1}_2 \Pi_1^T & (\Pi_2^T A_{22} \mathbf{1}_2) \mathbf{1}_2 \Pi_2^T & 0 \\ 0 & (\Pi_3^T A_{32} \mathbf{1}_2) \mathbf{1}_3 \Pi_2^T & (\Pi_3^T A_{33} \mathbf{1}_3) \mathbf{1}_3 \Pi_3^T \end{pmatrix}. \quad (65)$$

Moreover, as shown in Appendix B, we have

$$\Pi_2^T A_{21} \mathbf{1}_1 = -\Pi_2^T A_{22} \mathbf{1}_2 = \alpha_2$$

and

$$\Pi_3^T A_{32} \mathbf{1}_2 = -\Pi_3^T A_{33} \mathbf{1}_3 = \alpha_3.$$

This leads to the result

$$G^k = (-1)^k \begin{pmatrix} 0 & 0 & 0 \\ -\alpha_2^k \mathbf{1}_2 \Pi_1^T & \alpha_2^k \mathbf{1}_2 \Pi_2^T & 0 \\ \gamma_k \mathbf{1}_3 \Pi_1^T & \delta_k \mathbf{1}_3 \Pi_2^T & \alpha_3^k \mathbf{1}_3 \Pi_3^T \end{pmatrix},$$

where $\gamma_1 = 0$, $\delta_1 = \alpha_3$, while, for $k \geq 2$,

$$\delta_k = -\sum_{n=0}^{k-1} \alpha_2^n \alpha_3^{k-n} = -\alpha_3 \frac{\alpha_3^k - \alpha_2^k}{\alpha_3 - \alpha_2}$$

and

$$\gamma_k = \sum_{n=1}^{k-1} \alpha_2^n \alpha_3^{k-n} = \frac{\alpha_2 \alpha_3^k - \alpha_3 \alpha_2^k}{\alpha_3 - \alpha_2}.$$

This result can be shown by induction. As a matter of fact, using

$$\Pi_1^T \mathbf{1}_1 = \Pi_2^T \mathbf{1}_2 = 1,$$

the result for $k = 2$ ensues, since

$$G^2 = \begin{pmatrix} 0 & 0 & 0 \\ \alpha_2 \mathbf{1}_2 \Pi_1^T & -\alpha_2 \mathbf{1}_2 \Pi_2^T & 0 \\ 0 & \alpha_3 \mathbf{1}_3 \Pi_2^T & -\alpha_3 \mathbf{1}_3 \Pi_3^T \end{pmatrix} \\ \times \begin{pmatrix} 0 & 0 & 0 \\ \alpha_2 \mathbf{1}_2 \Pi_1^T & -\alpha_2 \mathbf{1}_2 \Pi_2^T & 0 \\ 0 & \alpha_3 \mathbf{1}_3 \Pi_2^T & -\alpha_3 \mathbf{1}_3 \Pi_3^T \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ -\alpha_2^2 \mathbf{1}_2 \Pi_1^T & \alpha_2^2 \mathbf{1}_2 \Pi_2^T & 0 \\ \alpha_2 \alpha_3 & -\alpha_3(\alpha_3 + \alpha_2) \mathbf{1}_3 \Pi_2^T & \alpha_3^2 \mathbf{1}_3 \Pi_3^T \end{pmatrix}$$

with $\gamma_2 = \alpha_2 \alpha_3$ and $\delta_2 = -\alpha_3(\alpha_2 + \alpha_3)$. Suppose next that the result is true for some $k \geq 2$. Then, we find

$$\begin{aligned} G^{k+1} &= GG^k \\ &= (-1)^k \begin{pmatrix} 0 & 0 & 0 \\ \alpha_2 \mathbf{1}_2 \Pi_1^T & -\alpha_2 \mathbf{1}_2 \Pi_2^T & 0 \\ 0 & \alpha_3 \mathbf{1}_3 \Pi_2^T & -\alpha_3 \mathbf{1}_3 \Pi_3^T \end{pmatrix} \\ &\quad \times \begin{pmatrix} 0 & 0 & 0 \\ -\alpha_2^k \mathbf{1}_2 \Pi_1^T & \alpha_2^k \mathbf{1}_2 \Pi_2^T & 0 \\ \gamma_k \mathbf{1}_3 \Pi_1^T & \delta_k \mathbf{1}_3 \Pi_2^T & \alpha_3^k \mathbf{1}_3 \Pi_3^T \end{pmatrix} \\ &= (-1)^k \begin{pmatrix} 0 & 0 & 0 \\ \alpha_2^{k+1} \mathbf{1}_2 \Pi_1^T & -\alpha_2^{k+1} \mathbf{1}_2 \Pi_2^T & 0 \\ (-\alpha_3 \alpha_2^k - \alpha_3 \gamma_k) \mathbf{1}_3 \Pi_1^T & (\alpha_2^k \alpha_3 - \alpha_3 \delta_k) \mathbf{1}_3 \Pi_2^T & -\alpha_3^{k+1} \mathbf{1}_3 \Pi_3^T \end{pmatrix}, \end{aligned}$$

where

$$\alpha_3 \alpha_2^k + \alpha_3 \gamma_k = \alpha_3 \alpha_2^k + \alpha_3 \frac{\alpha_2 \alpha_3^k - \alpha_3 \alpha_2^k}{\alpha_3 - \alpha_2} = \frac{\alpha_2 \alpha_3^{k+1} - \alpha_3 \alpha_2^{k+1}}{\alpha_3 - \alpha_2} = \gamma_{k+1}$$

and

$$\begin{aligned} \alpha_2^k \alpha_3 - \alpha_3 \delta_k &= \alpha_2^k \alpha_3 - \alpha_3 \left(-\alpha_3 \frac{\alpha_3^k - \alpha_2^k}{\alpha_3 - \alpha_2} \right) \\ &= \alpha_3 \frac{\alpha_3^{k+1} - \alpha_2^{k+1}}{\alpha_3 - \alpha_2} = -\delta_{k+1}. \end{aligned}$$

Therefore,

$$G^{k+1} = (-1)^{k+1} \begin{pmatrix} 0 & 0 & 0 \\ -\alpha_2^{k+1} \mathbf{1}_2 \Pi_1^T & \alpha_2^{k+1} \mathbf{1}_2 \Pi_2^T & 0 \\ \gamma_{k+1} \mathbf{1}_3 \Pi_1^T & \delta_{k+1} \mathbf{1}_3 \Pi_2^T & \alpha_3^{k+1} \mathbf{1}_3 \Pi_3^T \end{pmatrix},$$

which is the result for $k + 1$.

Note that

$$\begin{aligned} \sum_{k=1}^{+\infty} \frac{(-t)^k \delta_k}{k!} &= \frac{\alpha_3}{\alpha_2 - \alpha_3} \sum_{k=1}^{+\infty} \frac{(-t\alpha_3)^k - (-t\alpha_2)^k}{k!} \\ &= \frac{\alpha_3}{\alpha_2 - \alpha_3} (e^{-t\alpha_3} - e^{-t\alpha_2}), \end{aligned}$$

and similarly that

$$\begin{aligned} \sum_{k=1}^{+\infty} \frac{(-t)^k \gamma_k}{k!} &= \frac{1}{\alpha_3 - \alpha_2} \sum_{k=1}^{+\infty} \frac{\alpha_3(t\alpha_2)^k - \alpha_2(t\alpha_3)^k}{k!} \\ &= 1 - \frac{\alpha_2 e^{-t\alpha_3} - \alpha_3 e^{-t\alpha_2}}{\alpha_2 - \alpha_3}. \end{aligned}$$

Therefore,

$$e^{tG} = \sum_{k=0}^{+\infty} \frac{t^k G^k}{k!}$$

is given by the equation given in **Box I** from which we get He^{tG} given in **Box II**. Moreover, as shown in **Appendix B**, we have

$$\alpha_2 = \alpha_3/3.$$

Finally, redefining time t in number of N/α time steps as unit of time with $\alpha = \alpha_2$ and using the notation $\lfloor Nt/\alpha \rfloor$ for the floor

value of Nt/α , we get

$$Q(t) = \lim_{N \rightarrow +\infty} P^{\lfloor Nt/\alpha \rfloor} = \begin{pmatrix} \mathbf{1}_1 \Pi_1^T & 0 & 0 \\ (1 - e^{-t}) \mathbf{1}_2 \Pi_1^T & e^{-t} \mathbf{1}_2 \Pi_2^T & 0 \\ \left(1 - \frac{3e^{-t} - e^{-3t}}{2}\right) \mathbf{1}_3 \Pi_1^T & \frac{3}{2} (e^{-t} - e^{-3t}) \mathbf{1}_3 \Pi_2^T & e^{-3t} \mathbf{1}_3 \Pi_3^T \end{pmatrix}. \tag{66}$$

This is a transition matrix from time 0 to time t for a continuous time Markov chain with generator G . Transitions occur according to two different timescales. After an initial phase during which instantaneous transitions take place, there is a phase during which transitions within S occur according to **Kingman's (1982)** coalescent with rate 1 with N/α time steps as unit of time as $N \rightarrow \infty$.

Appendix B. Backward transition matrix under neutrality

We are interested in the elements of the transition matrix $P = (p_{s,s'})$ in (62) for the age classes of the ancestors of one, two or three individuals from one time step to the previous one in a neutral population with d age classes of fixed finite sizes. More precisely, we want to obtain the elements of the matrices $R = (r_{s,s'})$ and $A = (a_{s,s'})$ such that

$$p_{s,s'} = r_{s,s'} + N^{-1} a_{s,s'} + o(N^{-1}), \tag{67}$$

where $No(N^{-1}) \rightarrow 0$ as the total population size $N \rightarrow +\infty$, for s and s' in the state space $S = S_1 \cup S_2 \cup S_3$ with S_1, S_2 and S_3 being the sets of states with 1, 2 and 3 ancestors given in (59), (60) and (61), respectively. The possible transitions are summarized in **Fig. 3**.

Let the stochastic matrix $M = (m_{ij})$, where m_{ij} represents the probability that an individual in age class i comes from an individual in age class j one time step back for i and j in S_1 , be in the form

$$M = F + T, \tag{68}$$

where $F = (f_{ij})$ and $T = (t_{ij})$ with

$$f_{ij} = \begin{cases} p_j & \text{if } i = 1, \\ 0 & \text{if } i = 2, \dots, d, \end{cases} \tag{69}$$

and

$$t_{ij} = \begin{cases} 1 & \text{if } i = j + 1, \\ 0 & \text{if } i \neq j + 1, \end{cases} \tag{70}$$

for $i, j = 1, \dots, d$. Moreover, consider the Kronecker delta functions

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\delta_{ijk} = \begin{cases} 1 & \text{if } i = j = k, \\ 0 & \text{otherwise,} \end{cases}$$

for $i, j, k = 1, \dots, d$. Then, from state $s = ijk$ in S_3 , there is transition to state $s' = i'j'k'$ in S_3 with probability

$$\begin{aligned} P_{ijk,i'j'k'} &= (f_{i'j'j'} f_{kk'} + f_{i'j'j'} t_{kk'} + f_{i'j'j'} f_{kk'} + t_{i'j'j'} f_{kk'}) \\ &\quad \times \left(1 - \frac{\delta_{i'j'}}{N_{i'}} - \frac{\delta_{i'k'}}{N_{i'}} - \frac{\delta_{j'k'}}{N_{j'}} + \frac{\delta_{i'j'k'}}{N_{i'}^2} \right) \\ &\quad + f_{i'j'j'} t_{kk'} \left(1 - \frac{\delta_{i'j'}}{N_{i'}} - \frac{\delta_{i'k'}}{N_{i'}} \right) + t_{i'j'j'} f_{kk'} \left(1 - \frac{\delta_{k'j'}}{N_{k'}} - \frac{\delta_{k'i'}}{N_{k'}} \right) \end{aligned}$$

$$\begin{pmatrix} I_1 & 0 & 0 \\ (1 - e^{-\alpha_2 t}) \mathbf{1}_2 \Pi_1^T & I_2 + (e^{-\alpha_2 t} - 1) \mathbf{1}_2 \Pi_2^T & 0 \\ \left(1 - \frac{\alpha_2 e^{-\alpha_3 t} - \alpha_3 e^{-\alpha_2 t}}{\alpha_2 - \alpha_3}\right) \mathbf{1}_3 \Pi_1^T & \frac{\alpha_3}{\alpha_2 - \alpha_3} (e^{-\alpha_3 t} - e^{-\alpha_2 t}) \mathbf{1}_3 \Pi_2^T & I_3 + (e^{-\alpha_3 t} - 1) \mathbf{1}_3 \Pi_3^T \end{pmatrix}$$

Box I.

$$He^{tG} = \begin{pmatrix} \mathbf{1}_1 \Pi_1^T & 0 & 0 \\ (1 - e^{-\alpha_2 t}) \mathbf{1}_2 \Pi_1^T & e^{-\alpha_2 t} \mathbf{1}_2 \Pi_2^T & 0 \\ \left(1 - \frac{\alpha_2 e^{-\alpha_3 t} - \alpha_3 e^{-\alpha_2 t}}{\alpha_2 - \alpha_3}\right) \mathbf{1}_3 \Pi_1^T & \frac{\alpha_3}{\alpha_2 - \alpha_3} (e^{-\alpha_3 t} - e^{-\alpha_2 t}) \mathbf{1}_3 \Pi_2^T & e^{-\alpha_3 t} \mathbf{1}_3 \Pi_3^T \end{pmatrix}$$

Box II.

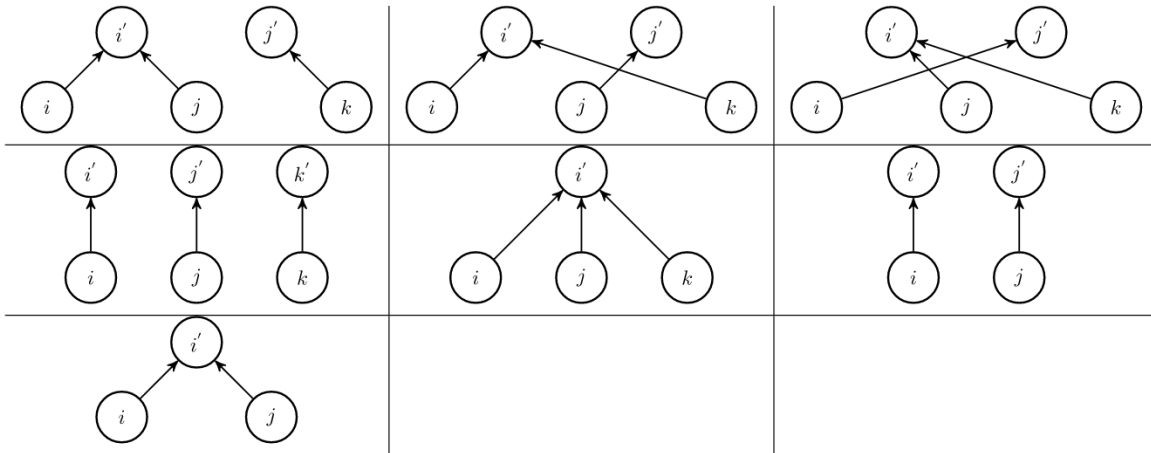


Fig. 3. Backward transitions for the age classes of the ancestors for up to three individuals.

$$\begin{aligned} &+ t_{ii'} f_{jj'} t_{kk'} \left(1 - \frac{\delta_{i'j'}}{N_{j'}} - \frac{\delta_{j'k'}}{N_{j'}}\right) + t_{ii'} t_{jj'} t_{kk'} \\ &= (f_{ii'} f_{jj'} f_{kk'} + f_{ii'} f_{jj'} t_{kk'} + f_{ii'} t_{jj'} f_{kk'} + t_{ii'} f_{jj'} f_{kk'}) \\ &\times \left(1 - \frac{\delta_{i'j'}}{N_{j'}} - \frac{\delta_{j'k'}}{N_{j'}} - \frac{\delta_{j'k'}}{N_{j'}}\right) \\ &+ f_{ii'} t_{jj'} t_{kk'} \left(1 - \frac{\delta_{i'j'}}{N_{j'}} - \frac{\delta_{j'k'}}{N_{j'}}\right) + t_{ii'} t_{jj'} f_{kk'} \left(1 - \frac{\delta_{k'j'}}{N_{k'}} - \frac{\delta_{k'i'}}{N_{k'}}\right) \\ &+ t_{ii'} f_{jj'} t_{kk'} \left(1 - \frac{\delta_{i'j'}}{N_{j'}} - \frac{\delta_{j'k'}}{N_{j'}}\right) + t_{ii'} t_{jj'} t_{kk'} + o(N^{-1}) \\ &= m_{ii'} m_{jj'} m_{kk'} \\ &- (f_{ii'} f_{jj'} f_{kk'} + f_{ii'} f_{jj'} t_{kk'} + f_{ii'} t_{jj'} f_{kk'} + t_{ii'} f_{jj'} f_{kk'}) \left(\frac{\delta_{i'j'}}{N_{j'}} + \frac{\delta_{j'k'}}{N_{j'}} + \frac{\delta_{j'k'}}{N_{j'}}\right) \\ &- f_{ii'} t_{jj'} t_{kk'} \left(\frac{\delta_{i'j'}}{N_{j'}} + \frac{\delta_{j'k'}}{N_{j'}}\right) - t_{ii'} t_{jj'} f_{kk'} \left(\frac{\delta_{k'j'}}{N_{k'}} + \frac{\delta_{k'i'}}{N_{k'}}\right) \\ &- t_{ii'} f_{jj'} t_{kk'} \left(\frac{\delta_{i'j'}}{N_{j'}} + \frac{\delta_{j'k'}}{N_{j'}}\right) + o(N^{-1}); \end{aligned} \tag{71}$$

transition to state $s' = i'j'$ in S_2 (with coalescence in age class i') with probability

$$\begin{aligned} p_{ijk, i'j'} &= (f_{ii'} f_{jj'} f_{kk'} + f_{ii'} f_{jj'} t_{kk'} + f_{ii'} t_{jj'} f_{kk'} + t_{ii'} f_{jj'} f_{kk'}) \frac{1}{N_{j'}} \left(1 - \frac{\delta_{i'j'}}{N_{j'}}\right) \\ &+ (f_{ii'} t_{jj'} t_{kk'} + t_{ii'} f_{jj'} t_{kk'}) \frac{1}{N_{j'}} \\ &+ (f_{ii'} f_{kk'} f_{jj'} + f_{ii'} f_{kk'} t_{jj'} + f_{ii'} t_{kk'} f_{jj'} + t_{ii'} f_{kk'} f_{jj'}) \frac{1}{N_{j'}} \left(1 - \frac{\delta_{i'j'}}{N_{j'}}\right) \\ &+ (f_{ii'} t_{kk'} t_{jj'} + t_{ii'} f_{kk'} t_{jj'}) \frac{1}{N_{j'}} \\ &+ (f_{jj'} f_{kk'} f_{ij'} + f_{jj'} f_{kk'} t_{ij'} + f_{jj'} t_{kk'} f_{ij'} + t_{jj'} f_{kk'} f_{ij'}) \frac{1}{N_{j'}} \left(1 - \frac{\delta_{i'j'}}{N_{j'}}\right) \\ &+ (f_{jj'} t_{kk'} t_{ij'} + t_{jj'} f_{kk'} t_{ij'}) \frac{1}{N_{j'}} \\ &= (f_{ii'} f_{jj'} f_{kk'} + f_{ii'} f_{jj'} t_{kk'} + f_{ii'} t_{jj'} f_{kk'} + t_{ii'} f_{jj'} f_{kk'}) \frac{1}{N_{j'}} \\ &+ (f_{ii'} t_{jj'} t_{kk'} + t_{ii'} f_{jj'} t_{kk'}) \frac{1}{N_{j'}} \\ &+ (f_{ii'} f_{kk'} f_{jj'} + f_{ii'} f_{kk'} t_{jj'} + f_{ii'} t_{kk'} f_{jj'} + t_{ii'} f_{kk'} f_{jj'}) \frac{1}{N_{j'}} \end{aligned}$$

$$\begin{aligned}
 &+ (f_{i'v}t_{ki'}t_{j'v} + t_{i'v}f_{ki'}t_{j'v}) \frac{1}{N_{i'}} \\
 &+ (f_{j'v}f_{ki'}f_{i'v} + f_{j'v}f_{ki'}t_{i'v} + f_{j'v}t_{ki'}f_{i'v} + t_{j'v}f_{ki'}f_{i'v}) \frac{1}{N_{i'}} \\
 &+ (f_{j'v}t_{ki'}t_{i'v} + t_{j'v}f_{ki'}t_{i'v}) \frac{1}{N_{i'}} + o(N^{-1}) \\
 &= (m_{i'v}m_{j'v}m_{ki'} + m_{i'v}m_{ki'}m_{j'v} + m_{j'v}m_{ki'}m_{i'v}) \\
 &- t_{i'v}t_{j'v}f_{ki'} - t_{i'v}t_{j'v}t_{ki'} - t_{i'v}t_{ki'}f_{j'v} - t_{i'v}t_{ki'}t_{j'v} - t_{j'v}t_{ki'}f_{i'v} - t_{j'v}t_{ki'}t_{i'v}) \\
 &\times \frac{1}{N_{i'}} + o(N^{-1}) \\
 &= (m_{i'v}m_{j'v}m_{ki'} + m_{i'v}m_{ki'}m_{j'v} + m_{j'v}m_{ki'}m_{i'v}) \\
 &- t_{i'v}t_{j'v}(f_{ki'} + t_{ki'}) - t_{i'v}t_{ki'}(f_{j'v} + t_{j'v}) - t_{j'v}t_{ki'}(f_{i'v} + t_{i'v})) \frac{1}{N_{i'}} + o(N^{-1}) \\
 &= (m_{i'v}m_{j'v}m_{ki'} + m_{i'v}m_{ki'}m_{j'v} + m_{j'v}m_{ki'}m_{i'v}) \\
 &- t_{i'v}t_{j'v}m_{ki'} - t_{i'v}t_{ki'}m_{j'v} - t_{j'v}t_{ki'}m_{i'v}) \frac{1}{N_{i'}} + o(N^{-1}); \tag{72}
 \end{aligned}$$

and transition to state $s' = i'$ in S_1 with probability

$$p_{ijk,i'} = (f_{i'v}f_{j'v}f_{ki'} + f_{i'v}f_{j'v}t_{ki'} + f_{i'v}t_{j'v}f_{ki'} + t_{i'v}f_{j'v}f_{ki'}) \left(\frac{1}{N_{i'}}\right)^2 = o(N^{-1}).$$

On the other hand, from state $s = ij$ in S_2 , there is transition to state $s' = i'j'$ in S_2 with probability

$$\begin{aligned}
 p_{ij,i'j'} &= (f_{i'v}f_{j'v} + f_{i'v}t_{j'v} + t_{i'v}f_{j'v}) \left(1 - \delta_{i'j'} \frac{1}{N_{i'}}\right) + t_{i'v}t_{j'v} \\
 &= m_{i'v}m_{j'v} - (f_{i'v}f_{j'v} + f_{i'v}t_{j'v} + t_{i'v}f_{j'v}) \frac{\delta_{i'j'}}{N_{i'}}; \tag{73}
 \end{aligned}$$

and transition to state $s' = i'$ in S_1 with probability

$$p_{ij,i'} = (f_{i'v}f_{j'v} + f_{i'v}t_{j'v} + t_{i'v}f_{j'v}) \frac{1}{N_{i'}}. \tag{74}$$

Finally, from state $s = i$ in S_1 , there is transition to state $s' = i'$ in S_1 with probability

$$p_{i,i'} = m_{i'v}. \tag{75}$$

B.1. Backward transition matrix in an infinite population

Consider $R_l = (r_{s,s'})$ for s and s' in S_l , which corresponds to the transition matrix for l lineages backward in time in an infinite neutral population with d age classes, for $l = 1, 2, 3$. Owing to (67) and (75), we have $R_1 = M = (m_{i'v})$ for i and i' in $S_1 = \{1, \dots, d\}$. Moreover, owing to (73) and (71), we have

$$(R_2)_{ij,i'j'} = m_{i'v}m_{j'v}$$

and

$$(R_3)_{ijk,i'j'k'} = m_{i'v}m_{j'v}m_{k'v}$$

for $i, j, k, i', j', k' = 1, \dots, d$.

We want to determine the stationary probability distribution of R_l defined as a probability vector Π_l that satisfies $\Pi_l^T = \Pi_l^T R_l$, for $l = 1, 2, 3$. Under the assumption that $R_1 = M$ is ergodic, its stationary probability distribution $\Pi_1 = \mathbf{w} = (w_1, \dots, w_d)^T$ is uniquely determined by

$$\sum_{i=1}^d w_i m_{i'v} = w_{i'v}$$

for $i' = 1, \dots, d$ with $\sum_{i=1}^d w_i = 1$. Then, we have

$$\sum_{i,j=1}^d w_i w_j (R_2)_{ij,i'j'} = \sum_{i=1}^d w_i m_{i'v} \sum_{j=1}^d w_j m_{j'v} = w_{i'v} w_{j'v}$$

and

$$\begin{aligned}
 \sum_{i,j,k=1}^d w_i w_j w_k (R_3)_{ijk,i'j'k'} &= \sum_{i=1}^d w_i m_{i'v} \sum_{j=1}^d w_j m_{j'v} \sum_{k=1}^d w_k m_{k'v} \\
 &= w_{i'v} w_{j'v} w_{k'v}
 \end{aligned}$$

for $i', j', k' = 1, \dots, d$. We conclude that

$$(\Pi_2)_{ij} = w_i w_j$$

for ij in $S_2 = \{ij \mid i, j = 1, \dots, d\}$, and

$$(\Pi_3)_{ijk} = w_i w_j w_k$$

for ijk in $S_3 = \{ijk \mid i, j, k = 1, \dots, d\}$.

B.2. First-order effects of population size on backward transition matrix

The first-order effects of the population size N on the backward transition matrix $P = (p_{s,s'})$ in (67) are given by the elements of the matrix $A = (a_{s,s'})$ for s and s' in S . This is a block matrix with blocks given by $A_{ll'} = (a_{s,s'})$ for s in S_l and s' in $S_{l'}$ for $l, l' = 1, 2, 3$. Owing to (71), (72), (73) and (74), the only non null blocks are given by

$$(A_{21})_{ij,i'} = (f_{i'v}f_{j'v} + f_{i'v}t_{j'v} + t_{i'v}f_{j'v}) \frac{1}{f_{i'v}},$$

$$(A_{22})_{ij,i'j'} = -(f_{i'v}f_{j'v} + f_{i'v}t_{j'v} + t_{i'v}f_{j'v}) \frac{\delta_{i'j'}}{f_{i'v}},$$

$$\begin{aligned}
 (A_{32})_{ijk,i'j'k'} &= -(t_{i'v}t_{j'v}f_{k'v} - t_{i'v}t_{j'v}t_{k'v} - t_{i'v}t_{k'v}f_{j'v} \\
 &- t_{i'v}t_{k'v}t_{j'v} - t_{j'v}t_{k'v}f_{i'v} - t_{j'v}t_{k'v}t_{i'v}) \frac{1}{f_{i'v}}
 \end{aligned}$$

and

$$\begin{aligned}
 (A_{33})_{ijk,i'j'k'} &= -(f_{i'v}f_{j'v}f_{k'v} + f_{i'v}f_{j'v}t_{k'v} + f_{i'v}t_{j'v}f_{k'v} + t_{i'v}f_{j'v}f_{k'v}) \\
 &\times \left(\frac{\delta_{i'j'}}{f_{i'v}} + \frac{\delta_{i'k'}}{f_{i'v}} + \frac{\delta_{j'k'}}{f_{j'v}}\right) \\
 &- f_{i'v}t_{j'v}t_{k'v} \left(\frac{\delta_{i'j'}}{f_{i'v}} + \frac{\delta_{i'k'}}{f_{i'v}}\right) - t_{i'v}t_{j'v}f_{k'v} \left(\frac{\delta_{k'j'}}{f_{k'v}} + \frac{\delta_{k'i'}}{f_{k'v}}\right) \\
 &- t_{i'v}f_{j'v}t_{k'v} \left(\frac{\delta_{i'j'}}{f_{j'v}} + \frac{\delta_{j'k'}}{f_{j'v}}\right)
 \end{aligned}$$

with $N_i = f_i N$ for $i, j, k, i', j', k' = 1, \dots, d$. Denoting by $\mathbf{1}_l$ the d^l -dimensional vector of all ones for $l = 1, 2, 3$, we have

$$(A_{22} \mathbf{1}_2)_{ij} = - \sum_{i'=1}^d (m_{i'v}m_{j'v} - t_{i'v}t_{j'v}) \frac{1}{f_{i'v}} = -(A_{21} \mathbf{1}_1)_{ij}$$

and

$$\begin{aligned}
 (A_{33} \mathbf{1}_3)_{ijk} &= - \sum_{j',k'=1}^d (f_{i'v}f_{j'v}f_{k'v} + f_{i'v}f_{j'v}t_{k'v} + f_{i'v}t_{j'v}f_{k'v} \\
 &+ t_{i'v}f_{j'v}f_{k'v} + f_{i'v}t_{j'v}t_{k'v} + t_{j'v}f_{i'v}f_{k'v} + t_{j'v}f_{i'v}t_{k'v}) \frac{1}{f_{i'v}} \\
 &- \sum_{i',j'=1}^d (f_{i'v}f_{j'v}f_{k'v} + f_{i'v}f_{j'v}t_{k'v} + f_{i'v}t_{j'v}f_{k'v} + t_{i'v}f_{j'v}f_{k'v} \\
 &+ f_{i'v}t_{j'v}t_{k'v} + t_{i'v}t_{j'v}f_{k'v}) \frac{1}{f_{i'v}} \\
 &- \sum_{i',k'=1}^d (f_{i'v}f_{j'v}f_{k'v} + f_{i'v}f_{j'v}t_{k'v} + f_{i'v}t_{j'v}f_{k'v}
 \end{aligned}$$

$$\begin{aligned}
 & +t_{i'j'k'}f_{kk'} + t_{i'j'k'}f_{kk'} + t_{i'j'k'}f_{kk'}) \frac{1}{f_{k'}} \\
 & = - \sum_{j',k'=1}^d (f_{i'j'j'}m_{kk'} + f_{i'j'j'}m_{kk'} + t_{i'j'j'}m_{kk'}) \frac{1}{f_{j'}} \\
 & - \sum_{i',j'=1}^d (f_{i'i'}m_{jj'}f_{k'i'} + f_{i'i'}m_{jj'}t_{k'i'} + t_{i'i'}m_{jj'}f_{k'i'}) \frac{1}{f_{i'}} \\
 & - \sum_{i',k'=1}^d (m_{i'i'}f_{j'k'}f_{kk'} + m_{i'i'}f_{j'k'}t_{kk'} + m_{i'i'}t_{j'k'}f_{kk'}) \frac{1}{f_{k'}} \\
 & = - \sum_{j'=1}^d (f_{i'j'}f_{j'j'} + f_{i'j'}t_{j'j'} + t_{i'j'}f_{j'j'}) \frac{1}{f_{j'}} \\
 & - \sum_{i'=1}^d (f_{i'i'}f_{k'i'} + f_{i'i'}t_{k'i'} + t_{i'i'}f_{k'i'}) \frac{1}{f_{i'}} \\
 & - \sum_{k'=1}^d (f_{j'k'}f_{kk'} + f_{j'k'}t_{kk'} + t_{j'k'}f_{kk'}) \frac{1}{f_{k'}} \\
 & = - \sum_{i'=1}^d (f_{i'i'}f_{j'i'} + f_{i'i'}t_{j'i'} + t_{i'i'}f_{j'i'} + f_{i'i'}f_{k'i'} + f_{i'i'}t_{k'i'}) \\
 & + t_{i'i'}f_{k'i'} + f_{j'i'}f_{k'i'} + f_{j'i'}t_{k'i'} + t_{j'i'}f_{k'i'}) \frac{1}{f_{i'}} \\
 & = - \sum_{i'=1}^d (f_{i'i'}m_{j'i'} + t_{i'i'}f_{j'i'} + f_{i'i'}m_{k'i'} + t_{i'i'}f_{k'i'} + f_{j'i'}m_{k'i'} + t_{j'i'}f_{k'i'}) \frac{1}{f_{i'}} \\
 & = - \sum_{i'=1}^d (m_{i'i'}m_{j'i'} - t_{i'i'}t_{j'i'} + m_{i'i'}m_{k'i'} - t_{i'i'}t_{k'i'} + m_{j'i'}m_{k'i'} - t_{j'i'}t_{k'i'}) \frac{1}{f_{i'}} \\
 & = -(A_{32}\mathbf{1}_2)_{ijk}.
 \end{aligned}$$

Therefore, using the facts that $\mathbf{w}^T M = \mathbf{w}^T$ and $\mathbf{w}^T \mathbf{1}_1 = 1$, we find

$$\begin{aligned}
 \alpha_3 & = -\Pi_3^T A_{33} \mathbf{1}_3 = \Pi_3^T A_{32} \mathbf{1}_2 \\
 & = \sum_{i,j,k=1}^d w_i w_j w_k \sum_{i'=1}^d \frac{1}{f_{i'}} (m_{i'i'}m_{j'i'} + m_{i'i'}m_{k'i'} + m_{j'i'}m_{k'i'} - t_{i'i'}t_{j'i'} \\
 & \quad - t_{i'i'}t_{k'i'} - t_{j'i'}t_{k'i'}) \\
 & = 3 \sum_{i'=1}^d \frac{1}{f_{i'}} \left(w_{i'} w_{i'} - \sum_{i,j} w_i w_j t_{i'i'} t_{j'i'} \right)
 \end{aligned}$$

and

$$\begin{aligned}
 \alpha_2 & = -\Pi_2^T A_{22} \mathbf{1}_2 = \Pi_2^T A_{21} \mathbf{1}_1 \\
 & = \sum_{i,j=1}^d w_i w_j \sum_{i'=1}^d \frac{1}{f_{i'}} (m_{i'i'}m_{j'i'} - t_{i'i'}t_{j'i'}) \\
 & = \sum_{i'=1}^d \frac{1}{f_{i'}} \left(w_{i'}^2 - \sum_{i,j=1}^d w_i w_j t_{i'i'} t_{j'i'} \right).
 \end{aligned}$$

Finally, we get

$$\alpha = \frac{\alpha_3}{3} = \alpha_2 = \sum_{i'=1}^d \frac{1}{f_{i'}} \left(w_{i'}^2 - \sum_{i,j=1}^d w_i w_j t_{i'i'} t_{j'i'} \right).$$

Note that in Felsenstein's (1971) model with $F = (f_{ij})$ and $T = (t_{ij})$ given in (69) and (70), respectively, we have

$$\alpha = \frac{1}{f_1} w_1^2 + \sum_{i=1}^{d-1} w_{i+1}^2 \left(\frac{1}{f_{i+1}} - \frac{1}{f_i} \right). \tag{76}$$

Then, according to Felsenstein (1971), the inbreeding effective number in this population is given by $N_e = N/(k\alpha)$ where k represents the generation time.

References

Archetti, M., Scheuring, I., 2012. Game theory of public goods in one-shot social dilemmas without assortment. *J. Theoret. Biol.* 299, 9–20.

Baudisch, A., Vaupel, J.W., 2012. Getting to the root of aging. *Science* 338, 618–619.

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.

Chesson, J., 1976. A non-central multivariate hypergeometric distribution arising from biased sampling with application to selective predation. *J. Appl. Probab.* 13, 795–797.

Emigh, T.H., 1979a. The dynamics of finite haploid populations with overlapping generations. I. Moments, fixation probabilities and stationary distributions. *Genetics* 92, 323–337.

Emigh, T.H., 1979b. The dynamics of finite haploid populations with overlapping generations. II. The diffusion approximation. *Genetics* 92, 339–351.

Emigh, T.H., Pollak, E., 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theor. Popul. Biol.* 15, 86–107.

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.

Felsenstein, J., 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68, 581–597.

Gokhale, C., Traulsen, A., 2010. Evolutionary games in the multiverse. *Proc. Natl. Acad. Sci. USA* 107, 5500–5504.

Hilbe, C., 2011. Local replicator dynamics: a simple link between deterministic and stochastic models of evolutionary game theory. *Bull. Math. Biol.* 73, 2068–2087.

Imhof, L., Nowak, M., 2006. Evolutionary game dynamics in a Wright-Fisher process. *J. Math. Biol.* 52, 667–681.

Jones, O.R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C.G., Schaible, R., Casper, B.B., Dahlgren, J.P., Ehrlén, J., García, M.B., Menges, E.S., Quintana-Ascencio, P.F., Caswell, H., Baudisch, A., Vaupel, J.W., 2014. Diversity of ageing across the tree of life. *Nature* 505, 169–173.

Karlin, S., 1982. Classification of selection-migration structures and conditions for a protected polymorphism. In: Hecht, M.K., Wallace, B., Prance, G.T. (Eds.), *Evolutionary Biology*. Plenum, New York, NY, pp. 61–204.

Karlin, S., Taylor, H.M., 1975. *A First Course in Stochastic Processes*. Academic Press, New York.

Kingman, J.F.C., 1982. The coalescent. *Stochastic Process. Appl.* 13, 235–248.

Kroumi, D., Lessard, S., 2014. Conditions for cooperation to be more abundant than defection in a hierarchically structured population. *Dynam. Games Appl.* 5, 239–262.

Kroumi, D., Lessard, S., 2015a. Strong migration limit for games in structured populations: applications to dominance hierarchy and set structure. *Games* 6, 318–346.

Kroumi, D., Lessard, S., 2015b. Evolution of cooperation in a multidimensional phenotype space. *Theor. Popul. Biol.* 102, 60–75.

Kurokawa, S., Ihara, Y., 2009. Emergence of cooperation in public goods games. *Proc. R. Soc. B* 276, 1379–1384.

Ladret, V., Lessard, S., 2007. Fixation probability for a beneficial allele and a mutant strategy in a linear game under weak selection in a finite island model. *Theor. Popul. Biol.* 72, 409–425.

Ladret, V., Lessard, S., 2008. Evolutionary game dynamics in a finite asymmetric two-deme population and emergence of cooperation. *J. Theor. Biol.* 255, 137–151.

Lessard, S., 2005. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. *Theor. Popul. Biol.* 68, 19–27.

Lessard, S., 2009. Diffusion approximations for one-locus multi-allele kin selection, mutation and random drift in group-structured populations: a unifying approach to selection models in population genetics. *J. Math. Biol.* 59, 659–696.

Lessard, S., 2011a. Evolution of cooperation in finite populations. In: Sigmund, K. (Ed.), *Evolutionary Game Dynamics*. American Mathematical Society, Providence, Rhode Island, pp. 143–171.

Lessard, S., 2011b. On the robustness of the extension of the one-third law of evolution to the multi-player game. *Dynam. Games Appl.* 1, 408–418.

- Lessard, S., 2011c. Effective game matrix and inclusive payoff in group-structured populations. *Dynam. Games Appl.* 1, 301–318.
- Lessard, S., Ladret, V., 2007. The probability of fixation of a single mutant in an exchangeable selection model. *J. Math. Biol.* 54, 721–744.
- Lessard, S., Lahaie, P., 2009. Fixation probability with multiple alleles and projected average allelic effect on selection. *Theor. Popul. Biol.* 75, 266–277.
- Lessard, S., Rocheleau, G., 2004. Kin selection and coefficients of relatedness in family-structured populations with inbreeding. *Theor. Popul. Biol.* 66, 287–306.
- Lessard, S., Soares, C.D., 2018. Frequency-dependent growth in class-structured populations: continuous dynamics in the limit of weak selection. *J. Math. Biol.* 77, 229–259.
- Li, X.-Y., Kurokawa, S., Giaimo, S., Traulsen, A., 2016. How life history can sway the fixation probability of mutants. *Genetics* 203, 1297–1313.
- McAvoy, A., Fraiman, N., Hauert, C., Wakeley, J., Nowak, M.A., 2018. Public goods games in populations with fluctuating size. *Theor. Popul. Biol.* 121, 72–84.
- Möhle, M., 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Probab.* 30, 493–512.
- Notohara, M., 1993. The strong-migration limit for the genealogical process in geographically structured populations. *J. Math. Biol.* 31, 115–122.
- Nowak, M.A., 2006. Five rules for the evolution of cooperation. *Science* 314, 1560–1563.
- Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650.
- Nowak, M.A., Tarnita, C., Antal, T., 2010. Evolutionary dynamics in structured populations. *Philos. Trans. R. Soc. B* 365, 19–30.
- Ohtsuki, H., Bordalo, P., Nowak, M.A., 2007. The one-third law of evolutionary dynamics. *J. Theoret. Biol.* 249, 289–295.
- Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M.A., 2006. A simple rule for evolution of cooperation on graphs and social networks. *Nature* 441, 502–505.
- Rousset, F., 2003. A minimal derivation of convergence stability measures. *J. Theoret. Biol.* 221, 665–668.
- Soares, C.D., Lessard, S., 2019. Diffusion approximation for an age-class-structured population under viability and fertility selection with application to fixation probability of an advantageous mutant. *J. Math. Biol.* revision under review.
- Wallenius, K.T., 1963. *Biased Sampling: The Non-Central Hypergeometric Probability Distribution* (Ph.D.). Stanford University.
- Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16, 97–159.

