Theoretical Population Biology 119 (2018) 48–56

Contents lists available at ScienceDirect

Theoretical Population Biology

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

Evolution of highly fecund haploid populations

Bjarki Eldon *, Wolfgang Stephan

Museum für Naturkunde, Invalidenstraße 43, 10115 Berlin, Germany

ARTICLE INFO

Article history:
Received 17 April 2017
Available online 31 October 2017

Keywords:
High fecundity
Sweepstakes reproduction
Viability selection
Fixation time
Fixation probability

ABSTRACT

We consider a model of viability selection in a highly fecund haploid population with sweepstakes reproduction. We use simulations to estimate the time until the allelic type with highest fitness has reached high frequency in a finite population. We compare the time between two reproduction modes of high and low fecundity. We also consider the probability that the allelic type with highest fitness is lost from the population *before* reaching high frequency. Our simulation results indicate that highly fecund populations can evolve faster (in some cases much faster) than populations of low fecundity. However, high fecundity and sweepstakes reproduction also confer much higher risk of losing the allelic type with highest fitness from the population by chance. The impact of selection on driving alleles to high frequency varies depending on the trait value conferring highest fitness; in some cases the effect of selection can hardly be detected.

1. Introduction

Adaptation by natural selection (Darwin, 1859) is a fundamental consequence of the inheritance of genetic variants which provide different fitness to individuals. Mathematical models of adaptation may be broadly classified into two main types. One is Fisher's geometric model (Fisher, 1930) which is among the first mathematical models of adaptation. In Fisher's model, selection acts on a continuous space of phenotypes and adaptation is driven by new beneficial mutations. Fisher's or related models have been extensively studied (Hartl, 1996; Hartl and Taubes, 1998; Barton, 1998, 2001; Orr, 1998, 2000, 2006). Gillespie considers a mutational landscape model (Gillespie, 1983, 1984, 1994) and Orr and Whitlock, 2002) a related model of selection acting on a discrete space of DNA sequences. Rokyta et al. (2005) test Orr's model of selection on DNA sequences. A key quantity in a model of adaptation is the probability that an advantageous type reaches high frequency, especially if the type is initially present in low frequency in the population. Efforts to quantify the fixation probability of a beneficial allele, and the time it takes to go to fixation, enjoy a long history. The results by Kimura (1957), Kimura (1962) and Kimura (1964) are used e.g. by Whitlock (2003) to consider the time and probability of fixation of a beneficial allele in a structured population. Greven et al. (2016) obtain rigorous limit results (in the limit of infinitely strong selection) on the fixation time of a beneficial allele in a structured population using a model of Wright–Fisher diffusion. Patwa and Wahl (2008) give an overview of work on the fixation probability of a beneficial allele.

Nature may act much more quickly than (Darwin, 1859) envisioned. The change in colour of the peppered moth during and after the industrial revolution, most probably caused by bird predation (Cook et al., 2012), is probably the best known example of a pacy evolution. A single gene has been implicated in regulating pigmentation in Lepidoptera (Nadeau et al., 2016). The rate at which bacteria develop antibiotic resistance (Beceiro et al., 2013; Reding-Roman et al., 2017), and changes in 50% age at maturity of the highly fecund Atlantic cod (Oosthuizen and Daan, 1974; May, 1967) in Gulf of St. Lawrence over roughly 20 years – possibly brought about by heavy fishing (Swain, 2011) – are other possible examples of rapid evolution.

Standard population genetic theory usually includes a model of reproduction, i.e. a law describing how to assign offspring numbers to individuals (or gene copies). A number of studies have shown that different offspring number laws (distributions) can predict drastically different patterns of neutral genetic variation (Birkner et al., 2013a, b; Blath et al., 2016; Sargsyan and Wakeley, 2008). More precisely, population models which admit high fecundity and sweepstakes reproduction (HFSR), possible characteristics of many marine populations (Hedgecock and Pudovkin, 2011), do so by introducing skewed, or heavy-tailed, offspring number laws (Schweinsberg, 2003; Eldon and Wakeley, 2006; Sargsyan and Wakeley, 2008; Huillet and Möhle, 2011; Möhle, 2011). Comparison with population genetic data of the highly fecund Atlantic cod (Birkner and Blath, 2008; Birkner et al., 2013c, b; Árnason and Halldórsdóttir, 2015; Blath et al., 2016) and Japanese sardines (Niwa et al., 2016) provide positive evidence for the applicability of models of HFSR to highly fecund natural populations.

Our main interest is to understand if and how high fecundity and sweepstakes reproduction facilitate (rapid) adaptation. To this

* Corresponding author.
E-mail address: bjarki.eldon@mfn-berlin.de (B. Eldon).

https://doi.org/10.1016/j.tpb.2017.10.002
0040-5809/

end we use simulations to estimate (*1*) the expected time it takes the allelic type with highest fitness under a model of viability selection to reach high frequency—conditional on the event it does so; (*2*) the probability that the fittest type is lost from the population *before* reaching high frequency. We compare our estimates of these 2 quantities between HFSR and non-HFSR populations. We model high fecundity and sweepstakes reproduction in which the effective size can be much smaller than the actual population size. Models of HFSR may also be applicable on a wider scale; to virus populations (Irwin et al., 2016), and more generally in conservation genetics and genomics (Montano, 2016). Models of HFSR introduce jumps in allele frequencies (Birkner and Blath, 2009). It is straightforward (Schweinsberg, 2003) to check that the ancestral process associated with our reproduction model (1) is an example of a Λ-coalescent which admits multiple mergers of ancestral lineages (Donnelly and Kurtz, 1999; Pitman, 1999; Sagitov, 1999). Therefore the classical diffusion approach of Kimura (1962), Kimura (1964) and Kimura and Ohta (1969) is not applicable to our framework (see for example Der et al., 2011).

2. The reproduction models

Now we describe the models we use for our simulations. We consider a haploid population of fixed size N. In each generation, individual i for $i \in [N] := \{1, 2, \ldots, N\}$ for $N \in \mathbb{N} := \{1, 2, \ldots\}$ independently contributes a random number X_i of *juveniles*, or *potential* offspring. If the total count of juveniles exceeds N random sampling of juveniles takes place in which N juveniles are sampled to form the new set of adults. In case of a highly fecund population with sweepstakes reproduction (HFSR population), the distribution of X_i for $i \in [N]$ (the X_i are i.i.d.) is heavy-tailed with parameters $\alpha, \gamma > 0$ and mass function

$$\mathbb{P}^{(\mathrm{HFSR})}(X_1 = k) := C\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right), \quad 1 \le k \le \gamma \tag{1}$$

and $\mathbb{P}^{(\mathrm{HFSR})}(X_1 > \gamma) = 0$ where C is a normalising constant. We remark that $\mathbb{P}^{(\mathrm{HFSR})}(X_1 \ge 1) = C\left(1 - (\gamma+1)^{-\alpha}\right)$. For our model to work, we need that $S_N := \sum_{i=1}^N X_i \ge N$ with high probability. One can choose C so that $\mathbb{P}(X_1 = 0) = 1 - \mathbb{P}(X_1 \ge 1) > 0$ and $\mathbb{E}[X_1] > 1$. If $\mathbb{E}[X_1] > 1$ then one can show (see e.g. Lemma 5 in Schweinsberg, 2003) that $S_N \ge N$ with high probability (in fact, $\mathbb{P}(S_N < N)$ decreases exponentially fast as N increases). For technical reasons (see Section 2.1) it would be more convenient to restrict the range of X_1 to $\{1, \ldots, \gamma\}$ since then $S_N \ge N$ almost surely (i.e. $\mathbb{P}(S_N \ge N) = 1$).

The truncation parameter γ limits the number of juveniles an individual can have. The parameter α determines how likely it is for an individual to have large numbers of juveniles (but no more than γ); the smaller the value of α the higher the stated probability. The model given by Eq. (1) is similar to the model by Schweinsberg (2003). The model by Schweinsberg is a limit model, i.e. $\mathbb{P}(X_1 \ge k)/k^\alpha \to C$ as $k \to \infty$. It can therefore only be used in theoretical derivations, and is biologically unreasonable; individuals even in highly fecund populations cannot produce an arbitrarily large number (even if always finite) of juveniles. The behaviour of our model (see Eq. (1)) should be intuitively clear. If $\alpha \in (1, 2)$ and $\gamma \ge N$ then the ancestral processes are multiple-merger coalescents and the allele frequencies can jump. On the other hand, if $\alpha \ge 2$ and/or γ is much less than N then we obtain the classical diffusion limits. We do not concern ourselves with the case $\alpha < 1$ as then, in the limit $N \to \infty$, we would be measuring time in discrete generations rather than in units on the order of (an appropriate power of) the population size (cf. Schweinsberg, 2003). The limit behaviour is treated in detail elsewhere. Our HFSR model (see Eq. (1)) can be readily extended to diploid populations.

The random number of *offspring* (surviving juveniles) produced by an individual in a haploid Wright–Fisher population of size N is binomial with parameters N and $1/N$, which is approximated by a Poisson distribution with mean 1, as $N \to \infty$. In case of a non-HFSR population we model the distribution of the random number X_i of juveniles of individual i as Poisson with a fixed parameter $\lambda > 0$ and mass function

$$\mathbb{P}^{(\mathrm{Pois})}(X_1 = k) := \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \{0, 1, \ldots\}. \tag{2}$$

Therefore, as $\lambda \to \infty$, our non-HFSR model (see Eq. (2)) approaches the classical Wright–Fisher sampling. If we had modelled the number of *surviving* offspring as i.i.d. Poisson conditional on a fixed sum N our sampling would be exactly Wright–Fisher sampling regardless of the value of λ since (as is well known) the joint distribution of i.i.d. Poisson random variables conditional on a fixed sum is a multinomial.

In contrast to our HFSR model (see Eq. (1)) we do not (in most of our simulations) consider a truncated Poisson distribution. Corollary 1(ii) of Glynn (1987) gives, with λ fixed, writing $p_\lambda(k) = e^{-\lambda}\lambda^k/(k!)$,

$$\lim_{n\to\infty} \frac{\sum_{k\ge n} p_\lambda(k)}{p_\lambda(n)} = 1, \tag{3}$$

i.e. nearly all of the mass to the right of the point n sits at n. For example, $\sum_{k=0}^{100} p_{10}(k) \approx 1$ while the corresponding value for our HFSR model (see Eq. (1)) is approximately $1 - (1/101) \approx 0.99$ for $\alpha = 1$ and $\gamma \gg 100$ (so that $C \approx 1$).

One can compare the two distributions, (1) and (2), by considering the untruncated version $\tilde{\mathbb{P}}^{(\mathrm{HFSR})}(X = k)$ defined as in (1) but taking $\gamma \to \infty$. It is straightforward to check that, for fixed λ and α, where $\tilde{\mathbb{P}}^{(\mathrm{HFSR})}(X \ge k) = Ck^{-\alpha}$ for $k \ge 1$,

$$\lim_{k\to\infty} \frac{\sum_{j=k}^\infty p_\lambda(j)}{\tilde{\mathbb{P}}^{(\mathrm{HFSR})}(X \ge k)} = \lim_{k\to\infty} \frac{1}{C} k^\alpha \sum_{j=k}^\infty p_\lambda(j) = 0; \tag{4}$$

the law (1) admits a much heavier right-tail than the Poisson, even if $\lambda = \mathbb{E}^{(\mathrm{HFSR})}[X_1]$. Fig. A5 compares the relative mass for $\gamma = 10^3$. For $k \ge 20$, the mass of the Poisson is much smaller than the mass given by (1), even though the mean is the same.

Another reason for why we do not (except for $\gamma = 10$ in Figs. 1 and 2) consider a truncated Poisson distribution is that then the means do not match exactly (although they are still quite similar). One could obviously alter the parameter (λ, say) of the untruncated Poisson (2) so that the means of the truncated Poisson and the HFSR model (1) would match, but then the interpretation of λ would not be completely clear.

2.1. Small N_e/N

Let ν_i denote the number of *offspring*, i.e. the surviving juveniles, from parent i (arbitrarily labelled) that were sampled from the pool of juveniles. Since we assume constant population size $\mathbb{E}[\nu_1] = 1$. The effective size N_e is given by $N_e = 1/c_N$ where

$$c_N = \frac{\mathbb{E}[\nu_1(\nu_1 - 1)]}{N-1} = \frac{\mathrm{Var}(\nu_1)}{N-1} \tag{5}$$

and $\mathrm{Var}(\nu_1)$ denotes the variance of ν_1. Therefore,

$$\frac{N_e}{N} = \frac{N-1}{N\mathrm{Var}(\nu_1)}. \tag{6}$$

Schweinsberg (2003) obtains the relation for large N, where $\mathbb{1}(A) = 1$ if A holds, and is zero otherwise,

$$c_N \approx N\mathbb{E}\left[\frac{X_1(X_1-1)}{S_N(S_N-1)}\mathbb{1}(S_N \ge N)\right]. \tag{7}$$

Model (1) is a natural model of HFSR which admits high variance in the offspring distribution, and therefore low effective size, even when population size is large (Table A2).

Population models which admit high variance in the offspring distribution, even in a large population, therefore predict low N_e/N ratio, as also pointed out by Kimura and Crow (1963), Crow and Denniston (1988) and Hedrick (2005). Let $\{X_1, \ldots, X_N\}$ be the i.i.d. numbers of juveniles, and write $\underline{X} = (X_1, \ldots, X_N)$.

The surviving juveniles $\{\nu_1, \ldots, \nu_N\}$ are obtained by uniform random sampling without replacement from the pool of juveniles, i.e. ν_1 is hypergeometrically distributed given $\underline{X}$. This, together with the law of total variance

$$\mathrm{Var}(\nu_1) = \mathbb{E}\left[\mathrm{Var}(\nu_1|\underline{X})\right] + \mathrm{Var}(\mathbb{E}\left[\nu_1|\underline{X}\right])$$

gives, recalling that with $X_1 \geq 1$ almost surely then $S_N \geq N$ a.s.,

$$\begin{aligned}\mathrm{Var}(\nu_1) &= \mathbb{E}\left[N(N-1)\frac{X_1(X_1-1)}{S_N(S_N-1)} + N\frac{X_1}{S_N} - \left(N\frac{X_1}{S_N}\right)^2\right] \\ &\quad + \mathrm{Var}\left(N\frac{X_1}{S_N}\right) \\ &\approx N^2\mathrm{Var}\left(\frac{X_1}{S_N}\right) + N\mathbb{E}\left[\frac{X_1}{S_N}\right]\end{aligned}$$

for large N. Suppose the X_i have distribution (1). One can adapt the calculations by Schweinsberg (2003) to show that $\mathbb{E}\left[X_1^2/S_N^2\right] = \mathcal{O}(N^{-\alpha})$ for $1 < \alpha < 2$ and $\gamma \geq N$. Hence, $\mathrm{Var}(\nu_1) = \mathcal{O}(N^{2-\alpha})$ and it follows from Eq. (6) that $N_e/N = \mathcal{O}(N^{\alpha-2})$ for $1 < \alpha < 2$ and as $N \to \infty$. Our model (1) can predict tiny N_e/N ratios (see also Eq. (16) in Section 4.1). Models that can predict small N_e/N ratios were also considered by Hedrick (2005) and Waples (2016).

Now assume X_i for $i \in \{1, \ldots, N\}$ are i.i.d. Poisson with rate λ; and consider the approximation (7). Then one can approximate $\mathbb{E}\left[X_1(X_1-1)/(S_N(S_N-1))\right]$ with a Taylor approximation in 2 variables. Write $R_N = S_N(S_N-1)$. Since now $\mathbb{P}(X_1 = 0) > 0$ we have $\mathbb{P}(S_N = 0) > 0$. Write Π for the untruncated Poisson (see Eq. (2)) and Π^* for the Poisson distribution restricted to $\mathbb{N}$, i.e. $\mathbb{P}^{(\Pi^*)}(X_1 \geq 1) = 1$. Let f denote a (central) moment function. Then, with $p_\lambda(k)$ as in (2),

$$\begin{aligned}\mathbb{E}^{(\Pi^*)}[f(X_1)] &= \frac{1}{1-e^{-\lambda}}\sum_{k\geq 1} f(k)p_\lambda(k) \\ &= \frac{1}{1-e^{-\lambda}}\left(\sum_{k\geq 0} f(k)p_\lambda(k) - f(0)p_\lambda(0)\right) \\ &= \frac{1}{1-e^{-\lambda}}\mathbb{E}^{(\Pi)}[f(X_1)] - \frac{f(0)p_\lambda(0)}{1-e^{-\lambda}},\end{aligned}$$

or $\mathcal{O}(\mathbb{E}^{(\Pi^*)}[f(X_1)]) = \mathcal{O}(\mathbb{E}^{(\Pi)}[f(X_1)])$ which is what we care about in the following calculation. Let $\mathrm{Var}^{(\Pi)}$ resp. $\mathrm{Cov}^{(\Pi)}$ denote the variance resp. covariance under Π. Then we have (recall Eq. (7)), since $\mathbb{P}^{(\Pi^*)}\{S_N \geq N\} = 1$ so that $c_N \approx N\mathbb{E}^{(\Pi^*)}[X_1(X_1-1)/R_N]$,

$$\begin{aligned}&\mathbb{E}^{(\Pi^*)}\left[\frac{X_1(X_1-1)}{R_N}\right] \\ &\approx \frac{\mathbb{E}^{(\Pi)}[X_1(X_1-1)]}{\mathbb{E}^{(\Pi)}[R_N]} \\ &\quad + \mathcal{O}\left(\frac{\mathrm{Cov}^{(\Pi)}(X_1(X_1-1), R_N)}{(\mathbb{E}^{(\Pi)}[R_N])^2} + \frac{\mathrm{Var}^{(\Pi)}(R_N)\mathbb{E}^{(\Pi)}[X_1(X_1-1)]}{(\mathbb{E}^{(\Pi)}[R_N])^3}\right) \\ &= \frac{1}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)\end{aligned}$$

by linearity of the covariance operator, writing $S_N = X_1 + S_{N-1}$ where X_1 and S_{N-1} are independent, by approximating $\mathrm{Var}(R_N)$ with $\mathrm{Var}(S_N^2) = \mathcal{O}(N^3)$, and by noting that it suffices to consider leading terms in the numerator. Therefore, when X_i are Poisson with fixed mean λ, then $c_N \approx 1/N$ and the effective size $N_e(N, \lambda) \approx N$ for large N. The Poisson model (2) cannot predict small N_e/N unlike the HFSR model (1) even if the means ($\mathbb{E}[X_1]$) are matched. Notice that our approximation of the effective size under the Poisson model is independent of the mean λ. However, as $\lambda \to \infty$, the Poisson approximates a unit mass at infinity [i.e. $\lim_{\lambda\to\infty}\sum_{k=0}^{n} p_\lambda(k)/p_\lambda(n) = 1$ by Cor. 1(i) (Glynn, 1987)] and the sampling therefore becomes equivalent to Wright–Fisher sampling.

3. Modelling viability selection

One can model selection in a myriad of ways. Since we are interested in the effect of HFSR on time to high frequency and probability of losing the fittest type, we consider only viability selection. In viability selection, individuals contribute juveniles irrespective of allelic type; the juveniles inherit the type of their parent. We exclude mutation from our models. The viability of a juvenile is influenced by the type it carries.

We consider separately the *trait* function and the *fitness* function. The trait function maps the genetic types of individuals to the trait value (phenotype). We set the type space $E = \{0, 1, \ldots, n-1\}$ where n denotes the number of allelic types. Each type will be associated with a unique trait value and hence a unique fitness value. We will assume the trait function (recall we are modelling a haploid population)

$$z(a) := \frac{a}{a+1}, \quad a \in E. \tag{8}$$

We can assign highest fitness to any allelic type. Let z_0 denote the optimum trait value. By way of an example, if $z_0 = 0$ then allelic type $a = 0$ will have highest fitness, and the next fittest type will have trait value $z(1) = 1/2$, which could result in considerable difference in fitness. In contrast, if we choose $z_0 = 1$, then the fittest type will be type $a = n-1$ with trait value $(n-1)/n$, and the next-fittest type $(n-2)$ will have trait value $(n-2)/(n-1)$; and for most fitness functions and large n these two types will have almost the same fitness (see Fig. A1). The trait function (8) will be used throughout in our simulations to assign trait value to a given genetic type.

The strength of selection is given by the constant $s > 0$. We consider two fitness functions, an algebraic and an exponential fitness function. The fitness $w(z)$ of an individual with trait value z in case of the exponential fitness function is

$$w(z) = \exp\left(-s(z-z_0)^2\right). \tag{9}$$

The Gaussian fitness function (9) has been widely employed (e.g. Lande, 1976; Bürger, 2000; Matuszewski et al., 2015; Martin and Lenormand, 2008), apparently mostly for mathematical convenience.

For comparison we also consider an algebraic fitness function. In case of the algebraic fitness function the fitness is determined by

$$w(z) = \frac{1}{1+s(z-z_0)^2}. \tag{10}$$

In both cases highest fitness (at most 1) is obtained when the trait value is as close as possible to z_0. The difference between these functions becomes clear when $s(z-z_0)^2$ is large; if s is not very small, being far away from the optimum induces a stronger selection pressure under the exponential function (9) than under the algebraic function (10).

To model viability selection we proceed as follows. In each generation individuals independently contribute a random number of juveniles according to either the skewed model (1) or the Poisson distribution. If the realised number $\sum_{i=1}^{N} x_i$ of juveniles in any

given generation exceeds the population size (N), then viability selection acts as follows to 'extract' the surviving juveniles from the total pool of juveniles. We draw $\sum_{i=1}^{N} x_i$ independent exponentials each with rate w_j for $j \in \{1, \ldots, \sum_{i=1}^{N} x_i\}$ where w_j is the fitness of juvenile j. The juveniles associated with the N lowest realised exponentials then form the new set of adults. This selection procedure is equivalent to drawing N surviving juveniles without replacement, with probabilities proportional to w_j, from the total pool of $\sum_{i=1}^{N} x_i$ juveniles. One may also consider the exponential times as waiting times until the juveniles reach reproductive maturity. In case the total number of juveniles is less than N, we draw a new set of juveniles until the total number is at least N. If $\sum_{i=1}^{N} x_i = N$ then all the juveniles survive.

3.1. Measuring the time to high frequency

One way to measure the time until the optimum trait value (z_0) has been 'reached' is simply to count the generations, starting from some arbitrary point, until $|\overline{Z}_r - z_0| < \varepsilon$ for some $\varepsilon > 0$ where $\overline{Z}_r$ is the population mean trait value at generation r. More precisely,

$$T := \inf\{r \in \mathbb{N} : |\overline{Z}_r - z_0| < \varepsilon\}.$$

One can also consider the time until the fittest allelic type has reached a given frequency. This is the option we will take. Let Y_r denote the count of the fittest allelic type at time r, and $y \in [0, 1]$ some fixed threshold frequency (0.95, say). Then we consider the time

$$T := \inf\{r \in \mathbb{N} : Y_r \geq \lfloor yN \rfloor\}. \tag{11}$$

In addition, we define $T_0 := \inf\{r \in \mathbb{N} : Y_r = 0\}$. In our simulations (see Section 4) we will be concerned with the quantities

$$\begin{aligned} \tau^{(\Pi,s)} &:= \mathbb{E}^{(\Pi,s)}\{T | T < T_0\}, \\ p_0^{(\Pi,s)} &:= \mathbb{P}^{(\Pi,s)}(T_0 < T). \end{aligned} \tag{12}$$

The conditional expected time $\tau^{(\Pi,s)}$ is the expected value of T (11) conditional on reaching high frequency. In the absence of mutation, the fittest allelic type may be lost from the population before reaching the given frequency (y); in that case $T = \infty$ a.s. We therefore only consider the time T (11) conditional on the event $\{T < T_0\}$ that Y_r reaches the given count ($\lfloor yN \rfloor$). Since the number of states is finite, 2 of which are absorbing states, $\tau^{(\Pi,s)}$ exists (i.e. is finite). The probability $p_0^{(\Pi,s)}$ is the probability of losing the fittest type *without* the type having reached high count.

Let $\Pi^{(\text{Pois})}$ denote a Poisson distribution of number of juveniles with mean λ (see Eq. (2)), and $\Pi^{(\text{HFSR})}$ a sweepstakes reproduction mechanism (e.g. (1)). Assume also that there is a selection mechanism acting on the types independently of the reproduction mechanism. A meaningful comparison of $\Pi^{(\text{Pois})}$ and $\Pi^{(\text{HFSR})}$ in our context would be to compare $\tau^{(\Pi,s)}$ and $p_0^{(\Pi,s)}$ (see Eq. (12)) between $\Pi^{(\text{Pois})}$ and $\Pi^{(\text{HFSR})}$ so that $\lambda = \mathbb{E}^{(\text{HFSR})}[X_1]$. One would further assume that the population size N would be the same under the two reproduction schemes. By matching the mean and the population size between $\Pi^{(\text{Pois})}$ and $\Pi^{(\text{HFSR})}$ one can more clearly see the effect of the skewness of the offspring distribution on $\tau^{(\Pi,s)}$ and other statistics one might wish to consider.

Let $\mathbb{E}^{(\Pi,s)}$ denote expectation with regard to reproduction mechanism Π and selection. One would like to find conditions under which

$$\mathbb{E}^{(\Pi_1,s)}[T \mid T < T_0] > \mathbb{E}^{(\Pi_2,s)}[T \mid T < T_0], \tag{13}$$

ideally under general reproduction mechanisms denoted by Π_1 and Π_2 with $\mathbb{E}^{(\Pi_1)}[X_1] = \mathbb{E}^{(\Pi_2)}[X_1]$ and such that

$$\lim_{k\to\infty} \frac{\mathbb{P}^{(\Pi_1)}(X_1 \geq k)}{\mathbb{P}^{(\Pi_2)}(X_1 \geq k)} = 0. \tag{14}$$

For an example, see Eq. (4).

In light of our results on the probability of extinction (being lost from the population) of the fittest allelic type, we would like to find conditions under which

$$\mathbb{P}^{(\Pi_1,s)}(Y_r = 0) < \mathbb{P}^{(\Pi_2,s)}(Y_r = 0), \tag{15}$$

or $p_0^{(\Pi_1,s)} < p_0^{(\Pi_2,s)}$ (see Eq. (12)). However, we postpone a rigorous mathematical treatment of (13) and (15) for further study. The jumps in the type frequencies, the requirement to compare the models on the same timescale, and ideally to keep track of the rate of convergence to a diffusion approximation, make rigorous mathematical derivations non-trivial. However, a small insight into the mechanism of our model may be gained by considering the transition probabilities of $\{Y_r, r \geq 0\}$ (see Sec. A11).

4. Simulation results

In our simulations we consider the time it takes the fittest type to reach high frequency ($\tau^{(\Pi,s)}$, see (12)), and the probability that the fittest type is lost from the population *before* reaching high frequency ($p_0^{(\Pi,s)}$, see (12)); Π refers to a given reproduction mechanism and s denotes dependence on the selection mechanism. Once high frequency is reached, the fittest type should reach fixation with high probability (see Sec. A11). Given that we need to condition on reaching high frequency, it is natural to consider the probability $p_0^{(\Pi,s)}$ (see Eq. (12)) that the fittest type is lost from the population *before* reaching high frequency. Therefore one can view the probability of fixation of the fittest type as well approximated by $1 - p_0^{(\Pi,s)}$. In all our simulation results we use C in model (1) (and in (A1)) as a normalising constant so that $\mathbb{P}^{(\text{HFSR})}(X_1 \geq 1) = 1$ (and therefore $\mathbb{E}[X_1] > 1$).

We obtain and compare results for two reproduction modes; the HFSR model (1) (and a slightly altered version (A1)) and one in which the number of juveniles is drawn from a Poisson (see Eq. (2)) with the same mean as the corresponding HFSR model. The CWEB (Knuth and Levy, 1994) code written for the simulations is available at https://github.com/eldonb/evolution.

4.1. Measuring time in the HFSR model

How one measures time in population models is essential, and determines how to 'calibrate' the strength of additional forces such as mutation and selection. The natural way to measure time, in the limit $N \to \infty$, is in units of c_N, where c_N is the probability that 2 distinct individuals sampled at the same time from the population have the same parent (Möhle, 2001; Der et al., 2011). One unit of time in the diffusion timescale (as $N \to \infty$) then corresponds to $1/c_N$ generations. Equivalently, 1 generation corresponds to c_N units of time in the diffusion timescale. In the classical haploid Wright–Fisher model, $c_N = 1/N$. Lemma 13 in Schweinsberg (2003) gives, for $1 < \alpha < 2$ and $\gamma \geq N$,

$$\lim_{N\to\infty} C_\gamma c_N N^{\alpha-1} = 1 \tag{16}$$

where C_γ is a normalising constant whose exact form depends on γ. This means that for this parameter range, i.e. with $1 < \alpha < 2$ and $\gamma \geq N$ we have a chance of observing very large families. Indeed, fix $x \in (0, 1]$ such that $xN \in \mathbb{N}$. Then $\mathbb{P}^{(\text{HFSR})}(X_1 \geq xN)$ is of order $\mathcal{O}(N^{-\alpha})$ (with $\gamma = N$). The probability of observing at least 1 large family (of order $\mathcal{O}(N)$) at a given timepoint is therefore of order $\mathcal{O}(N^{1-\alpha})$, the same order as c_N (see Eq. (16)). The implication is that one should, on average, observe large families before fixation of a given type occurs, and therefore that the occurrence of large families of the given type should 'speed up' the process relative to models which do not admit large families (or only do so with negligible probability).

One can further deduce from Schweinsberg (2003) that in the case $\alpha = 2$ one obtains units of time of order $\mathcal{O}(\log(N)/N)$ (as $N \to \infty$). Since, for $\alpha = 2$ and $\gamma \geq N$, one obtains $\mathbb{P}^{(\mathrm{HFSR})}(X_1 \geq xN) \in \mathcal{O}\left(1/N^2\right)$ the probability of observing large families is of a larger order than the unit of time (but only just). For $\alpha \geq 2$ or $\gamma \in o(N)$ we do not have a 'HFSR' model in the sense of the model capturing high fecundity and skewed offspring distribution, but we elect to refer to models given by offspring laws (1) or (A1) as HFSR models for consistency.

Fig. 2 (which contains information from Tables A3, A4, A5) shows estimates of $\tau^{(\Pi,s)}$ (see Eq. (12)) as a function of the truncation γ. The estimates are calibrated by the maximum estimate for each combination of α and selection strength s. For $\alpha = 1$ ($\circ$ and $\square$ in Fig. 2) the effect of the truncation is clear, especially for $s = 0.1$ (shown as $\circ$) and so the truncation is clearly affecting the timescale in that case. For $\alpha = 3$ ($\diamond$ and $+$) there is no effect of truncation (for $\gamma \geq 10$) and hence α is determining the timescale. The estimates in Fig. 2 were obtained for optimal trait value $z_0 = 0$ and initially with the 2 genetic types of equal count. The similarly calibrated estimates of $\tau^{(\Pi,s)}$ shown in Fig. A4 (which plots information in Table A6) were obtained for $z_0 = 1$ and 100 genetic types with type 0 in highest frequency. The interpretation of the results in Fig. A4 is the same as for Fig. 2. There is, however, one noticeable difference between Fig. 2A and Fig. A4 in the estimates for $\alpha = 1$ ($\circ$ and $\square$); the effect of selection is weaker in Fig. A4 since the circles ($\circ$, representing $s = 0.1$ or 'weak' selection) are nearly identical to the squares ($\square$, representing $s = 1$ or 'strong' selection). We will return to this point in Section 4.4.

4.2. Expected time to high frequency ($\tau^{(\Pi,s)}$)

In Fig. 1 (which contains information in Tables A3, A4, A5, and A7), Fig. A2 (which contains information in Tables A3 and A7), and Fig. A3 (which contains information in Tables A8 and A9) we compare estimates of $\tau^{(\Pi,s)}$ (see Eq. (12)) between the HFSR model (1) (a slightly different HFSR model (A1) in Fig. A3) and the Poisson model (2) of the number of juveniles calibrated to have the same mean, for a given value of α and truncation γ, as the corresponding HFSR model. In the case the truncation $\gamma = 10^1$ (shown as '+' in Fig. 1) the Poisson distribution is truncated at γ. For clarity the graphs on the left in Fig. 1 are for selection strength $s = 0.1$ and on the right for $s = 1.0$; the vertical scale of the graphs differ. As one might expect, the smallest times are obtained for $\alpha = 1$ and high truncation ($\gamma = 10^6$, $\square$; $\gamma = 10^3$, $\diamond$) under the HFSR model, especially in the case of weak selection ($s = 0.1$). In the case of strong selection ($s = 1.0$) the difference in the time to high frequency between the HFSR and the Poisson models is much less. One also notices that the curves for the estimates of $\tau^{(\Pi,s)}$ are quite different for the two models; the curve for the Poisson model (slashed lines in Fig. 1) is approximately linear as a function of α while the curve for the HFSR model (solid lines in Fig. 1) is convex as a function of α. The estimate of $\tau^{(\Pi,s)}$ for the HFSR model is therefore much larger than the estimate for the Poisson model for large (≥ 2) α. We suggest that the main reason for why the Poisson model has much smaller estimates of $\tau^{(\Pi,s)}$ than HFSR model (1) for $\alpha = 3$ is the effect of α on the effective size. The effective size of the neutral HFSR model, and by implication also when selection is included, depends much more strongly on α (see Table A2) than the effective size associated with the Poisson model. Recall that, as we calculate in Section 3.1, the effective size associated with the neutral Poisson model is independent of the mean.

In Figs. 1, A2, and A3 the optimum trait value is at $z_0 = 0$, i.e. type 0 is the fittest type.

The effect of the truncation (γ) on $\tau^{(\Pi,s)}$ can be understood in light of the relation of the truncation to the population size N. Only in the case $\gamma \geq N$ and $\alpha < 2$ are the estimates of $\tau^{(\Pi,s)}$ smaller for the HFSR models than for the Poisson model. The estimates of the effective size under HFSR model (1) in Table A2 also show how the effect of γ on N_e depends on the relation of γ to N.

In Figs. A2 and A3 we exclude the estimates of $\tau^{(\Pi,s)}$ at $\alpha = 3$, and for $\gamma = 10^1$, but include estimates for $\gamma = 10^5$, and for $n = 100$ genetic types. The simulation results shown in Figs. 1, A2 and A3 indicate that $\tau^{(\Pi,s)}$ increases as α increases for both reproduction modes. This is as expected, since both the probability of contributing large numbers of juveniles, and the mean $\mathbb{E}^{(\mathrm{HFSR})}[X_1]$, decrease as α increases.

In the case $z_0 = 0$ (Figs. 1, A2 and A3) the difference between the HFSR models (1) and (A1) and the corresponding Poisson models is only noticeable when population size is 'small' ($N = 10^3$) and selection is 'weak' ($s = 0.1$). Here the estimates of $\tau^{(\Pi,s)}$ roughly agree when $\alpha = 2$, even though the variance of X_1 ($\mathrm{Var}(X_1)$) and the effective size (N_e) is much higher, resp. smaller, when associated with the HFSR model than with the Poisson model (Table A2). Similarly, the estimates of $\tau^{(\Pi,s)}$ also roughly agree in a 'large' ($N = 10^5$) population, again regardless of the considerable difference in $\mathrm{Var}(X_1)$ and N_e between the reproduction modes (Table A2). For 'strong' selection ($s = 1.0$) and $z_0 = 0$ the difference between the two reproduction modes is negligible. In Table 1 we compare estimates of $\tau^{(\Pi,s)}$ between the trait optima $z_0 = 0$ and $z_0 = 1$ for $n = 100$ allelic types. There the estimates for the Poisson mode when $z_0 = 1$, and type $n - 1$ is the fittest type (see Fig. A1), are very similar across the parameter values considered, and much higher than the estimates under the HFSR model (A1), even for strong selection. In the case $z_0 = 1$ there are many types with very similar fitness (see Fig. A1). The population therefore likely becomes concentrated on a subset of the type space in which the types have very similar fitness, and then evolves approximately neutrally. Number of allelic types ($n = 2$ or $n = 100$) seems to matter little for $\tau^{(\Pi,s)}$ in case of $z_0 = 0$ (see Figs. A2 and A3).

Fig. A3 shows results from Tables A9 and A8 and tells the same story as Fig. A2. Note the similarity in the estimates of $\tau^{(\Pi,s)}$ for corresponding values of N and s between the different truncations when associated with the Poisson distribution; this comes about since the expected values $\mathbb{E}^{(\mathrm{HFSR})}[X_1]$ are very similar for $\alpha \geq 1.5$ (see Table A2). The results shown in Figs. A2 and A3 were obtained in the case of all the types initially being present in equal frequency. One could argue for the initial configuration of equal frequencies with a modest change in the environment. Starting with type $n-1$ in $N - n + 1$ copies, and all the other $n - 1$ types in 1 copy each (and $z_0 = 0$) roughly doubles the time to high frequency (results not shown). One could argue for an initial configuration where type 0 (resp. $n - 1$) is in $N - n + 1$ copy, and $z_0 = 1$ (resp. $z_0 = 0$) and all other types in 1 copy each, with a significant change in the environment.

One can compare the variance in the time to high frequency (T) by comparing the coefficient of variation $c_v = \left(\sqrt{\mathrm{Var}(Y)}\right)/\mathbb{E}[Y]$, i.e. the standard deviation relative to the mean of a positive random variable (Y). The general pattern is that when associated with the HFSR model (1), the c_v decreases as α increases (Tables A3 and A9). The exception is for two types, weak selection ($s = 0.1$), and small population size ($N = 10^3$). In contrast, the c_v increases with α – almost without exception – when associated with the Poisson distribution (Tables A7 and A8). This indicates that the distribution of T changes with α in different ways for the two reproduction modes.

4.3. Probability of extinction ($p_0^{(\Pi,s)}$)

Since we exclude mutation, and only consider the action of drift and selection on variation already present in the population, we must consider the time (T) to high frequency conditional on the event that the fittest allelic type will reach high frequency. If the

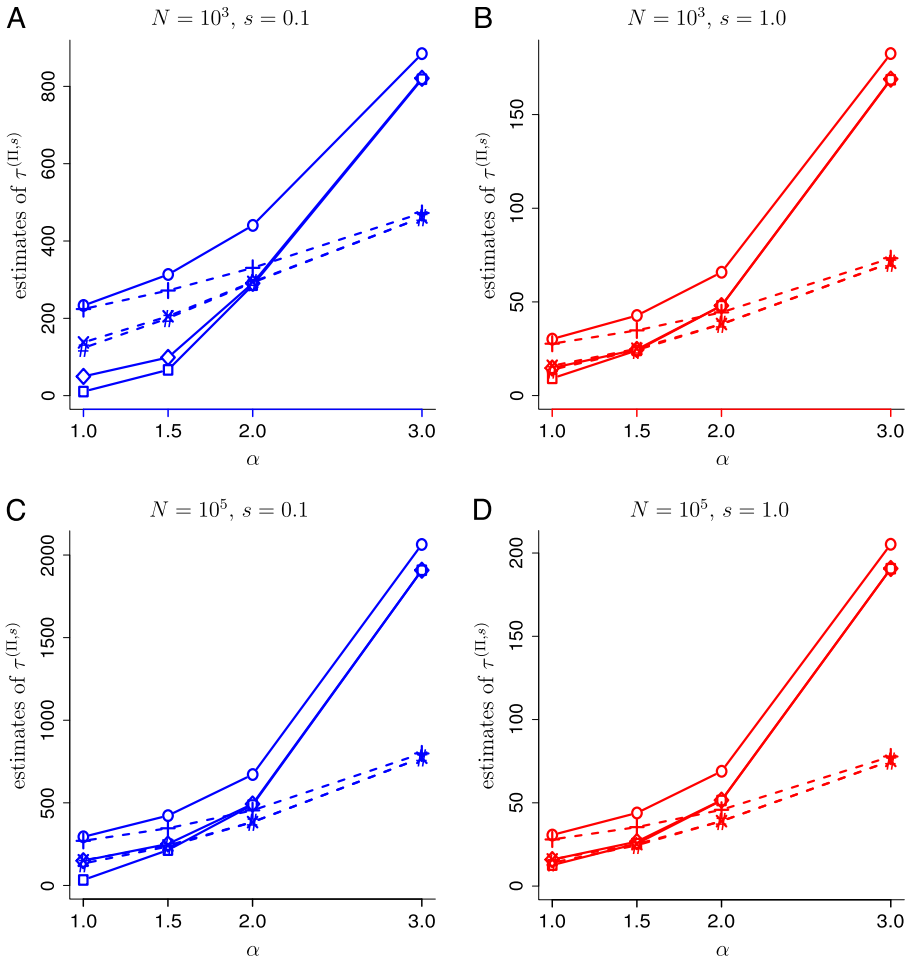

Fig. 1. Estimates of expected time to high frequency ($\tau^{(\Pi,s)}$, see Eq. (12)) as a function of skewness parameter α for population size N and selection strength s as shown. The number of allelic types is $n = 2$. Optimal trait value is $z_0 = 0$ and trait values were mapped to fitness values via the exponential fitness function (Eq. (9)). Initially all types are in equal frequency. Solid lines connect estimates (shown as $\circ$, $\square$, $\diamond$) obtained under HFSR model (1) while dashed lines connect estimates (shown as $+$, $\times$, #) obtained under the Poisson model with same mean number of juveniles ($\mathbb{E}[X_1]$) as the HFSR model. The truncation parameter takes values $\gamma = 10^1$ ($\circ$, $+$), 10^3 ($\diamond$, $\times$), 10^6 ($\square$, #). Results from Tables A3, A7, A4, and A5 and obtained for 10^4 replicates.

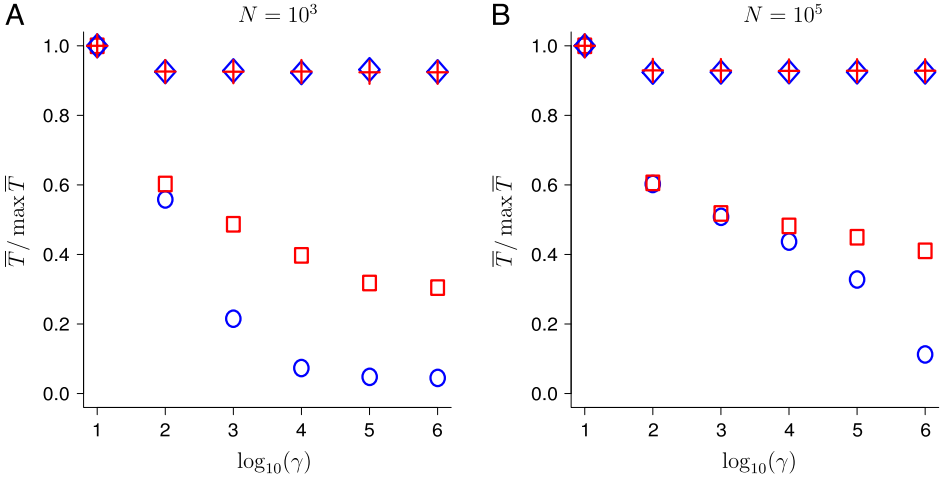

Fig. 2. Estimates ($\overline{T}$) of the expected time to high frequency ($\tau^{(\Pi,s)}$, see Eq. (12)) as a function of the truncation γ on $\log_{10}$ scale of the HFSR model (1). The estimates are normalised by the maximum estimate for each combination of α and selection strength s. The symbols are $\circ$ which represents $(\alpha, s) = (1.0, 0.1)$, $\diamond$ $(3.0, 0.1)$, $\square$ $(1.0, 1.0)$, $+$ $(3.0, 1.0)$. Number of genetic types $n = 2$ and trait values were mapped to fitness values using the exponential fitness function (9) with optimal trait value $z_0 = 0$. The results are from Tables A3, A4, and A5 and were obtained from 10^4 replicates.

fittest allelic type is lost from the population *before* reaching high frequency, then $T = \infty$ a.s. Recall that $p_0^{(\Pi,s)}$ (see Eq. (12)) denotes the probability of losing the fittest type *before* it reaches some given frequency (y). Table A10 holds estimates of $p_0^{(\Pi,s)}$ with $y = 0.95$ in

Tables A10 and A11. The general pattern is that $p_0^{(\Pi,s)}$ in a small HFSR population ($N = 10^3$) can be much higher than in a Poisson population with the same mean (Table A10). Especially noticeable is the difference in case of strong selection ($s = 1.0$). Recall that the estimates of the mean time to high frequency of the fittest type (conditional on reaching high frequency) are very similar between the two reproduction modes in case of strong selection (Figs. A2 and A3). In case of a large population ($N = 10^5$) the general pattern is that the fittest type does reach high frequency. An exception is in case of weak selection ($s = 0.1$) and strong drift ($\alpha = 1.0$, $\gamma = 10^5$). Thus, even in a large population and starting from anything but a small initial frequency, $p_0^{(\Pi,s)}$ can be high in a HFSR population.

Table A11 holds estimates of $p_0^{(\Pi,s)}$ when initially the fittest type (type 0) is in 1 copy—with type $n-1$ in $N-n+1$ copies. One could argue for this framework with a sudden large shift in the environment. In all cases is $p_0^{(\Pi,s)}$ higher when associated with the HFSR model; the difference is up to roughly ninefold in some cases. Even in a large population under 'modest' sweepstakes effect ($\alpha = 1.5$) and under strong selection ($s = 1.0$), in only approximately half the cases will the beneficial type reach high frequency when starting from a single copy. Selection must therefore be very strong, even in a large population, if a beneficial type, starting from a single copy, is to have a good chance of reaching high frequency in a HFSR population.

Given our trait function (8) one can compare at least two frameworks; where the population mostly consists of type 0 (resp. $n-1$), and type $n-1$ (resp. 0) is the fittest type. Table A11 holds results for the case when most of the population is of type $n-1$, and selection acts to push it to type 0. It seems much harder for selection to drive a population in the other direction, from type 0 to type $n-1$ (results not shown); i.e. one may say, given our trait function (8) (see also Fig. A1), that it seems harder for selection to push the population 'uphill' than 'downhill'. We conclude that the probability of success depends strongly on the trait function.

4.4. The impact of selection

The effect of selection on $\tau^{(\Pi,s)}$ appears to depend on which type confers highest fitness. In Table 1 we compare estimates of $\tau^{(\Pi,s)}$ for different optimal trait values, $z_0 = 0$ vs. $z_0 = 1$. The difference in estimates of $\tau^{(\Pi,s)}$ between weak and strong selection in case $z_0 = 1$ is very small in comparison to the case $z_0 = 0$. This may be understood in view of the shape (see Fig. A1) of our trait function (8). When $z_0 = 0$ the difference in fitness between the fittest type (type 0) and all other types can be large for large s. Since $w(0) = 1$ then for the exponential fitness function (9) we have $\frac{w(z(i))}{w(0)} = e^{-si^2/(i+1)^2}$, and for the algebraic function (10) $w(1)/w(0) = (i+1)^2/((i+1)^2 + si^2)$. Therefore, one should see an impact of selection on $\tau^{(\Pi,s)}$. In case $z_0 = 1$ we have, since $w(n-1) \approx 1$ for large n, in case of the exponential fitness function $w(z(i))/w(n-1) \approx \exp\left(-s/(i+1)^2\right)$, and $w(z(i))/w(n-1) \approx (i+1)^2/((i+1)^2 + s)$ for the algebraic fitness function (10); i.e. the effect of selection on the fitness function is much less in case $z_0 = 1$ in comparison with the case $z_0 = 0$. What likely happens in the case $z_0 = 1$ is that the population (possibly quite quickly) becomes concentrated on a subset of the type space, say on $A = \{\lfloor n/2 \rfloor, \ldots, n-1\}$, and then evolves effectively neutrally until the fittest type has reached high frequency, since on A the different types confer almost identical fitness.

Table 1
Estimates ($\overline{T}$) of the expected time for the fittest type to reach high frequency ($\tau^{(\Pi,s)}$, see Eq. (12)) in the case Π is the HFSR model given by (A1). The algebraic fitness function (10) maps the trait value of $n = 100$ types to fitness values. The fittest type is 0 in the case when the optimal trait value $z_0 = 0$, type $n-1$ in the case $z_0 = 1$. Initially all types were present in equal frequency. The expected values $\mathbb{E}^{(\mathrm{HFSR})}[X_1]$ for $\alpha \geq 1.5$ are very similar for the values of γ shown. Values for population size $N = 10^3$ only shown, since estimates for $N = 10^5$ and $z_0 = 1$ take very long to obtain. Results were obtained from 10^4 replicates. For $\alpha \geq 1.5$ the mean $\mathbb{E}^{(\mathrm{HFSR})}[X_1]$ is very similar between $\gamma = 10^3$ and $\gamma = 10^5$.

N	γ	α	s	HFSR		$\mathrm{Pois}\left(\mathbb{E}^{(\mathrm{HFSR})}[X_1]\right)$	
				$z_0 = 1.0$ $\overline{T}$	$z_0 = 0$ $\overline{T}$	$z_0 = 1.0$ $\overline{T}$	$z_0 = 0$ $\overline{T}$
10^3	10^3	1.0	0.1	54.1	46.1	1668.9	148.0
			1.0	53.1	17.3	1636.3	18.8
		1.5	0.1	128.1	109.7	1685.7	208.3
			1.0	127.6	35.0	1649.4	30.9
		2.0	0.1	496.4	314.1	1702.4	281.5
			1.0	490.0	69.1	1663.2	50.3
	10^5	1.0	0.1	14.6	15.4	1682.8	138.6
			1.0	14.5	14.4	1636.0	17.3
		1.5	0.1	98.8	98.3		
			1.0	97.2	34.6		
		2.0	0.1	482.6	311.3		
			1.0	464.1	68.3		

5. Discussion

Understanding if and how natural populations can adapt to environmental changes is of utmost importance. In order to resolve this for any given population, it might help to know how the different allelic types (or genotypes) affect a phenotype (the trait function), and how selection acts on the phenotype (the fitness function). Our results suggest that one also needs to consider how the different genetic types affect a trait (the trait function), as well as the offspring law.

Our results indicate that (i) $\tau^{(\Pi,s)}$ (see Eq. (12)) can be shorter for a HFSR population than a non-HFSR population, in some cases much shorter; (ii) $\tau^{(\Pi,s)}$ can be very robust to changes in selection strength in some cases, which means that the impact of selection varies depending on the trait function; (iii) for large α (≥ 3) estimates of $\tau^{(\Pi,s)}$ are higher for the HFSR model (1) than for the Poisson model (2); (iv) $p_0^{(\Pi,s)}$ (see Eq. (12)) can be much higher in a HFSR than in a non-HFSR population; (v) the probability of the fittest type reaching high frequency seems to depend strongly on the trait function.

HFSR populations can exhibit very small ratio of effective to actual population size (N_e/N) (Waples, 2016; Hedrick, 2005). The model of HFSR we consider also admits very small N_e/N. Importantly for us, the corresponding forward-in-time process of gene frequencies admits jumps (Birkner and Blath, 2009). Classical diffusion methods as used by Kimura (1957, 1962) are therefore not applicable to our case. An additional complicating factor is that we compare the time to high frequency between different reproduction modes; we therefore need to ensure that we are comparing time on the same scale. Models of HFSR often have a timescale associated with them which can be of much smaller order than $\mathcal{O}(N)$ (see Section 4.1), the timescale most often assumed.

Since we are comparing results between different reproduction modes, a natural question is which comparison is meaningful in our case. Our interest lies primarily in comparing distributions of numbers of juveniles which compare in the way given by (14). Our reproduction models distinguish between juveniles and the surviving offspring. The offspring law (1), or the Poisson distribution, refers to the numbers of juveniles contributed by each individual. In classical Wright–Fisher sampling, the marginal random number of offspring contributed by each individual is binomial with parameters N and $1/N$ which, for large N, is well approximated by a Poisson with mean 1. We therefore compare our HFSR model (1) with a Poisson calibrated to have the same mean as the HFSR model. See also Der et al. (2011, 2012) and Der and Plotkin (2014) who compare population models calibrated to have the same mean number of offspring (so-called 'Generalised Wright–Fisher'

models). Since the population size is the same between the two models, we cannot calibrate the Poisson population to have the same (inbreeding) effective size as the HFSR one, since the only way to change the effective size of the Poisson is by changing the population size.

Most models of selection in population genetics simply assign differential selective advantage to allelic types. We consider a trait function which assigns different phenotypic value to different types. The trait function we choose (8) allows us to consider at least two scenarios in which the difference in fitness between the fittest type and (most of) the other types is either 'large' (type 0 the fittest type for our function (8)) or 'small' (type $n-1$ the fittest for our function (8)). In addition, one can view our trait function (8) as a model of beneficial or deleterious mutations depending on the trait optimum (i.e. the trait value with highest fitness). If the trait optimum is at $z_0 = 0$ one may consider all other types as carrying different numbers of deleterious mutations. If $z_0 = 1$ one may view the types as carrying different amount of beneficial mutations.

Our simulation results (Figs. A2 and A3 and Table 1) suggest conditions exist in which $\tau^{(\Pi,s)}$ is lower in a HFSR population than in a Poisson population with the same mean. These conditions are, in the case of $z_0 = 0$ (and type 0 is the fittest type), small actual and effective population size and weak selection. The very small effective size required (see Table A2) results from the highly skewed reproduction law which for our model translates to small α and large γ. The difference is much less in a large population, probably since the effective size is larger, even though the ratio N_e/N is smaller in a larger population (for $\alpha < 2$). If selection is strong enough in the case $z_0 = 0$, there is no difference in $\tau^{(\Pi,s)}$ between the two reproduction modes, even in a small ($N = 10^3$) population. In the case $z_0 = 1$ (type $n-1$ is the fittest type) then many types (if n is large) have very similar high fitness and it appears selection would have to be very strong to have any impact on $\tau^{(\Pi,s)}$. Most theoretical results in population genetics are obtained in the limit of an infinite population. One exception is the work by Molina and Earn (2017) who consider conditions for evolutionary stability in finite populations. Our simulation results indicate that it is important to consider a finite population size when deriving theoretical results.

Another example of why it is important to consider finite population size in mathematical derivations is the probability that the fittest type reaches high frequency (or is lost before reaching high frequency). Der et al. (2012) consider this problem in the context of the HFSR model of Eldon and Wakeley (2006). In the model proposed by Eldon and Wakeley (2006), a single individual contributes 1 offspring with probability $1-\varepsilon_N$, and $\lfloor \psi N \rfloor$ with probability ε_N where $\psi \in (0, 1]$ is fixed. Der et al. (2012) consider the forward-in-time limit dynamics, as $N \to \infty$, of allele frequencies under a model of selection. They show that the fittest type can be assured to fix if selection is strong enough for any initial frequency of the fittest type. However, the results of Der et al. (2012) are obtained in the limit $N \to \infty$. Our simulation results with a finite population size indicate that fixation is all but assured in our model of HFSR and selection even when $N = 10^5$ (Tables A10 and A11). Foucart (2013) considers the fate of a *deleterious* allelic type in a Lambda–Wright–Fisher process with selection and finds conditions on the strength of selection under which the deleterious allele will be lost from the population with probability 1. The asymptotic (as $N \to \infty$) critical strength of selection is shown by both Der et al. (2012) and Foucart (2013) to depend on the particulars of the HFSR model. Our simulation results suggest that either population size needs to be extremely large for the continuous-time approximations to be a good approximation of the discrete-time process, and/or that how genetic types are mapped to fitness values matter. By way of an example, consider the following framework in which initially the least fit type is present in $N-n+1$ copies ($n = 100$) and all other types, including the fittest type, in 1 copy each. In Table A11 the estimates of $p_0^{(\Pi,s)}$ range from 0.99 to 0.23 for $N = 10^3$ and HFSR model (1). In Table A11 the optimal trait value $z_0 = 0$ so type 0 is the fittest type. If we set $z_0 = 1$ so that type 0 is now the least fit type, and type $n-1$ the fittest type, we estimate $p_0^{(\Pi,s)} \geq 0.99$ regardless of α and s.

Our model differs from most standard models of selection in population genetics in that we include a function (8) of how the allelic type space is mapped to the trait space. Our results (Table 1) show that in case of many allelic types the type of the fittest type changes the impact of selection. The parameterised model by Fisher (1918) of three genotypes in a diploid population is not comparable to ours since we are modelling many allelic types in a haploid population. Obviously the actual way different types determine a phenotype will depend on the biology of each population; our very simple parameter-free model has the advantage that it models many types with very similar phenotype, and hence very similar fitness, and a subset of types further apart.

Our work has focussed on the pace and opportunity for evolution (the progression of the fittest type to high frequency) in HFSR populations. The work by Der et al. (2012) and Foucart (2013) is almost the only published work on the subject. We introduce a new model of high fecundity and sweepstakes reproduction and use it to investigate the evolution in a finite HFSR population under a simple form of viability selection. Our results highlight the importance of deriving results for finite population size, and modelling the way genetic types map to phenotypes. Many open questions remain regarding the pace of evolution in HFSR populations. One concerns the scenario of polygenic, or multi-loci, selection (Jain and Stephan, 2017). One may also wish to model epistasis (Lessard and Kermany, 2012; Kermany and Lessard, 2012). Our model of HFSR could also be useful to model virus dynamics (Irwin et al., 2016).

Can our model explain the observation of Swain (2011) regarding the apparent sudden and significant drop in age at maturity of the highly fecund Atlantic cod? Assuming that the onset of heavy fishing results in a significant shift in optimal age at maturity, that there were genetic types present in low frequency in the population that confer early age at maturity, and we exclude mutation, then selection pressure must have been quite strong since the probability that the fittest type goes from low to high frequency even in a large HFSR population can be quite low (Table A11). In this example, one may also need to include significant decline in population size, an important remaining open problem.

Acknowledgements

We are grateful for the many and detailed comments of anonymous reviewers, which helped to improve the manuscript. BE was supported by DFG grant STE 325/17-1 to Wolfgang Stephan through the DFG Priority Programme SPP 1819: Rapid Evolutionary Adaptation.

A Conflict of Interest Statement: The authors declare no conflict of interests.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.tpb.2017.10.002.

References

Árnason, E., Halldórsdóttir, K., 2015. Nucleotide variation and balancing selection at the Ckma gene in Atlantic cod: analysis with multiple merger coalescent models. PeerJ 3, e786. URL http://dx.doi.org/10.7717/peerj.786.

Barton, N., 1998. Evolutionary biology: The geometry of adaptation. Nature 395 (6704), 751–752.

Barton, N., 2001. The role of hybridization in evolution. Mol. Ecol. 10 (3), 551–568.

Beceiro, A., Tomas, M., Bou, G., 2013. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? Clin. Microbiol. Rev. 26 (2), 185–230. URL https://doi.org/10.1128%2Fcmr.00059-12.

Birkner, M., Blath, J., 2008. Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. J. Math. Biol. 57, 435–465.

Birkner, M., Blath, J., 2009. Measure-valued diffusions, general coalescents and population genetic inference. In: Blath, J., Mörters, P., Scheutzow, M. (Eds.), Trends in Stochastic Analysis. Cambridge University Press, pp. 329–363.

Birkner, M., Blath, J., Eldon, B., 2013a. An ancestral recombination graph for diploid populations with skewed offspring distribution. Genetics 193, 255–290.

Birkner, M., Blath, J., Eldon, B., 2013b. Statistical properties of the site-frequency spectrum associated with Λ-coalescents. Genetics 195, 1037–1053.

Birkner, M., Blath, J., Steinrücken, M., 2013c. Analysis of DNA sequence variation within marine species using Beta-coalescents. Theor. Popul. Biol. 87, 15–24.

Blath, J., Cronjäger, M.C., Eldon, B., Hammer, M., 2016. The site-frequency spectrum associated with Ξ-coalescents. Theor. Popul. Biol. 110, 36–50. URL http://dx.doi.org/10.1016/j.tpb.2016.04.002.

Bürger, R., 2000. The Mathematical Theory of Selection, Recombination, and Mutation, Vol. 228. Wiley Chichester.

Cook, L.M., Grant, B.S., Saccheri, I.J., Mallet, J., 2012. Selective bird predation on the peppered moth: the last experiment of Michael Majerus. Biol. Lett. 8 (4), 609–612. URL https://doi.org/10.1098%2Frsbl.2011.1136.

Crow, J.F., Denniston, C., 1988. Inbreeding and variance effective population numbers. Evolution 482–495.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection. John Murray, London.

Der, R., Epstein, C.L., Plotkin, J.B., 2011. Generalized population models and the nature of genetic drift. Theor. Popul. Biol. 80 (2), 80–99. URL http://dx.doi.org/10.1016/j.tpb.2011.06.004.

Der, R., Epstein, C., Plotkin, J.B., 2012. Dynamics of neutral and selected alleles when the offspring distribution is skewed. Genetics 191 (4), 1331–1344.

Der, R., Plotkin, J.B., 2014. The equilibrium allele frequency distribution for a population with reproductive skew. Genetics 196 (4), 1199–1216. URL http://dx.doi.org/10.1534/genetics.114.161422.

Donnelly, P., Kurtz, T.G., 1999. Particle representations for measure-valued population models. Ann. Probab. 27, 166–205.

Eldon, B., Wakeley, J., 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172, 2621–2633.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52 (02), 399–433.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection: A Complete Variorum Edition. Oxford University Press.

Foucart, C., 2013. The impact of selection in the Λ-Wright–Fisher model. Electron. Commun. Probab. 18, 1–10.

Gillespie, J.H., 1983. A simple stochastic gene substitution model. Theor. Popul. Biol. 23 (2), 202–215.

Gillespie, J.H., 1984. Molecular evolution over the mutational landscape. Evolution 1116–1129.

Gillespie, J.H., 1994. The Causes of Molecular Evolution. Oxford University Press.

Glynn, P.W., 1987. Upper bounds on Poisson tail probabilities. Oper. Res. Lett. 6 (1), 9–14.

Greven, A., Pfaffelhuber, P., Pokalyuk, C., Wakolbinger, A., 2016. The fixation time of a strongly beneficial allele in a structured population. Electron. J. Probab. 21, 1–42.

Hartl, D., 1996. Compensatory Nearly neutral mutations: Selection without adaptation. J. Theoret. Biol. 182 (3), 303–309.

Hartl, D.L., Taubes, C.H., 1998. Towards a theory of evolutionary adaptation. Genetica 102/103, 525–533.

Hedgecock, D., Pudovkin, A.I., 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull. Mar. Sci. 87, 971–1002.

Hedrick, P., 2005. Large variance in reproductive success and the N_e/N ratio. Evolution 59 (7), 1596.

Huillet, T., Möhle, M., 2011. Population genetics models with skewed fertilities: forward and backward analysis. Stoch. Models 27, 521–554.

Irwin, K.K., Laurent, S., Matuszewski, S., Vuilleumier, S., Ormond, L., Shim, H., Bank, C., Jensen, J.D., 2016. On the importance of skewed offspring distributions and background selection in virus population genetics. Heredity 117, 393–399.

Jain, K., Stephan, W., 2017. Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics 206, 389–406.

Kermany, A.R., Lessard, S., 2012. Effect of epistasis and linkage on fixation probability in three-locus models: An ancestral recombination–selection graph approach. Theor. Popul. Biol. 82 (2), 131–145.

Kimura, M., 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28 (4), 882–901.

Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47 (6), 713.

Kimura, M., 1964. Diffusion models in population genetics. J. Appl. Probab. 1 (2), 177–232.

Kimura, M., Crow, J.F., 1963. The measurement of effective population number. Evolution 279–288.

Kimura, M., Ohta, T., 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61 (3), 763.

Knuth, D.E., Levy, S., 1994. The CWEB System of Structured Documentation: Version 3.0. Addison-Wesley Longman Publishing Co., Inc., Reading, Massachusetts.

Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30 (2), 314. URL https://doi.org/10.2307%2F2407703.

Lessard, S., Kermany, A.R., 2012. Fixation probability in a two-locus model by the ancestral recombination–selection graph. Genetics 190 (2), 691–707.

Martin, G., Lenormand, T., 2008. The distribution of beneficial and fixed mutation fitness effects close to an optimum. Genetics 179 (2), 907–916.

Matuszewski, S., Hermisson, J., Kopp, M., 2015. Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. Genetics 200 (4), 1255–1274. URL https://doi.org/10.1534%2Fgenetics.115.178574.

May, A.W., 1967. Fecundity of atlantic cod. J. Fish. Res. Brd. Can. 24, 1531–1551.

Möhle, M., 2001. Forward and backward diffusion approximations for haploid exchangeable population models. Stochastic Process. Appl. 95 (1), 133–149.

Möhle, M., 2011. Coalescent processes derived from some compound Poisson population models. Electron. Commun. Probab. 16, 567–582.

Molina, C., Earn, D.J.D., 2017. On selection in finite populations. J. Math. Biol.. http://dx.doi.org/10.1007/s00285-017-1151-4.

Montano, V., 2016. Coalescent inferences in conservation genetics: should the exception become the rule? Biol. Lett. 12 (6), 20160211.

Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W.R., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., Hines, H., Salazar, C., Merrill, R.M., Dowling, A.J., ffrench Constant, R.H., Llaurens, V., Joron, M., McMillan, W.O., Jiggins, C.D., 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. Nature 534 (7605), 106–110. URL https://doi.org/10.1038%2Fnature17961.

Niwa, H.-S., Nashida, K., Yanagimoto, T., 2016. Reproductive skew in Japanese sardine inferred from DNA sequences. ICES J. Mar. Sci. 73 (9), 2181–2189. URL http://dx.doi.org/10.1093/icesjms/fsw070.

Oosthuizen, E., Daan, N., 1974. Egg fecundity and maturity of North Sea cod, Gadus morhua. Netherlands J. Sea Res. 8 (4), 378–397.

Orr, H.A., 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 935–949.

Orr, H.A., 2000. Adaptation and the cost of complexity. Evolution 54 (1), 13–20.

Orr, H.A., 2006. The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J. Theoret. Biol. 238 (2), 279–285.

Orr, H.A., Whitlock, M., 2002. The population genetics of adaptation: the adaptation of DNA sequences. Evolution 56 (7), 1317–1330.

Patwa, Z., Wahl, L.M., 2008. The fixation probability of beneficial mutations. J. R. Soc. Interface 5 (28), 1279–1289.

Pitman, J., 1999. Coalescents with multiple collisions. Ann. Probab. 27, 1870–1902.

Reding-Roman, C., Hewlett, M., Duxbury, S., Gori, F., Gudelj, I., Beardmore, R., 2017. The unconstrained evolution of fast and efficient antibiotic-resistant bacterial genomes. Nature Ecol. Evol. 1, 0050.

Rokyta, D.R., Joyce, P., Caudle, S.B., Wichman, H.A., 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nature Genet. 37 (4), 441–444.

Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Probab. 36, 1116–1125.

Sargsyan, O., Wakeley, J., 2008. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. Theor. Popul. Biol. 74, 104–114.

Schweinsberg, J., 2003. Coalescent processes obtained from supercritical Galton-Watson processes. Stochastic Process. Appl. 106, 107–139.

Swain, D.P., 2011. Life-history evolution and elevated natural mortality in a population of Atlantic cod (Gadus morhua). Evolut. Appl. 4 (1), 18–29.

Waples, R.S., 2016. Tiny estimates of the N_e/N ratio in marine fishes: Are they real? J. Fish Biol. 89 (6), 2479–2504.

Whitlock, M.C., 2003. Fixation probability and time in subdivided populations. Genetics 164 (2), 767–779.