Theoretical Population Biology xx (xxxx) xxx–xxx

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 xxxx

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's 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's 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,

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