Journal of Theoretical Biology 391 (2016) 51–64

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

The influence of genetic drift on the formation and stability of polymorphisms arising from negative frequency-dependent selection

Lei Zhao, David Waxman *

Centre for Computational Systems Biology, Fudan University, 220 Handan Road, Shanghai 200433, People's Republic of China

HIGHLIGHTS

- We determine the effects of genetic drift on negative frequency-dependent selection.
- The stable polymorphism of an infinite population is destabilised by genetic drift.
- We find the timescales of mutant loss, achievement and decay of the polymorphism.
- We clarify the different influences of census and effective population sizes.
- This work bears on maintenance of variation in finite populations.

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form
13 November 2015
Accepted 17 November 2015
Available online 30 November 2015

Keywords:
Allele loss
Experimental evolution
Branching process
Diffusion approximation

ABSTRACT

We consider the simplest form of negative frequency-dependent selection in a biallelic haploid population, where the selection coefficient of a mutant allele is a linear function of the allele's frequency, and changes from positive to negative as the frequency is increased. In an effectively infinite population this behaviour leads to a stable polymorphism. We present a theoretical investigation of what occurs in a finite population, where a long-lived polymorphism may be formed, but which fluctuates and ultimately disappears due to random genetic drift. We model the dynamics as a branching process and explicitly take into account differences between the census population size and the effective population size, which play different roles in the dynamics. We characterise the behaviour of the population in terms of three distinct timescales associated with: (i) early loss of mutant alleles, (ii) achievement of the long-lived polymorphism, (iii) disappearance of the polymorphism. Timescales (i) and (iii) depend on the effective population size and are, as a consequence, affected by random genetic drift, while timescale (ii) depends primarily on the census size and is relatively insensitive to genetic drift. Analysis and simulations of the branching process clarify the different influences of the census and effective population sizes. One substantial quantitative difference, between populations where the effective and census population sizes coincide and where they differ, lies in the number of mutant alleles in the long-lived polymorphism. This number is approximately proportional to the census size. Thus assuming the census size equals a much smaller effective population size predicts a much smaller number of mutants in the long-lived polymorphism.

1. Introduction

In this work we provide an analysis motivated by some phenomena which have been studied in laboratory populations (Chelo et al., 2013). These studies come under the general topic of "experimental evolution", which is defined as "… research in which populations are studied across multiple generations under defined and reproducible conditions, whether in the laboratory or in nature" (Garland and Rose, 2009). Such studies "… require most if not all of the following fundamental design elements: maintenance of control populations, simultaneous replication, observation over multiple generations, and the prospect of detailed genetic analysis" (Garland and Rose, 2009). Experimental evolution allows the active control and testing of phenomena that, previously, could only be inferred from historical observations.

The investigations of the present work are specifically concerned with the behaviour of alleles in finite populations where negative frequency-dependent selection is acting. The finiteness of the population size means that, in addition to selection, the alleles are also influenced by random genetic drift (see e.g., Ewens, 2004;

* Corresponding author.
E-mail address: davidwaxman@fudan.edu.cn (D. Waxman).

http://dx.doi.org/10.1016/j.jtbi.2015.11.011
0022-5193/

Barton et al., 2007). In such a situation frequency-dependent selection may allow establishment of a long-lived polymorphism, but as we shall see, the polymorphism cannot persist indefinitely, due to effects of random genetic drift.

Negative frequency-dependent selection in diploid populations is a particular case of *balancing selection*, which generally maintains variation of multiple alleles in very large populations. Genes experiencing multi-allelic balancing selection, such as MHC genes and S-genes (at self-incompatibility loci in plants) have been theoretically and empirically studied for a long time (see, e.g., Wright, 1939; Takahata et al., 1991; Apanius et al., 1997; Schierup et al., 1997; Yokoyama and Nei, 1979; Vekemans and Slatkin, 1994; Ohashi and Tokunaga, 2000) as has biallelic overdominant selection (see, e.g., Robertson, 1962; Nei and Roychoudhury, 1973). Recently, biallelic balancing selection in primates has become of interest (Segurel et al., 2012; Leffler et al., 2013) and explicit statistical methods have been developed to detect this form of selection (De Giorgio et al., 2014).

Some diploid models with weak frequency-dependent selection are mathematically equivalent to models with overdominant selection (Takahata and Nei, 1990) and this equivalence applies, in particular, to biallelic loci. The present work analyses frequency-dependent selection at a biallelic locus in a *haploid* population, which is the simplest situation to study frequency-dependent selection and, effectively, is the nature of the population investigated in experiments by Chelo et al. (2013). While there is no notion of dominance in a haploid model, it is straightforward to construct a diploid model without dominance (based on genic selection) with very similar or identical behaviour to the haploid model.[1] While this diploid model might be thought to be equivalent to (or derivable from) a model with overdominant selection (cf. Takahata and Nei, 1990), such an overdominant model will generally need parameters lying outside allowed ranges (e.g., involving negative fitnesses), as we consider in the Discussion (Section 5). Such overdominant models have thus not been the subject of previous analyses, and there is no previous study on what is equivalent to a haploid population with frequency-dependent selection (the topic considered here).

The present work is mainly concerned with *timescales* of the dynamics of a population, where a timescale is a time-interval over which quantities, such as mean values and probability distributions, change appreciably. This work has the particular aim of producing a theoretical analysis of all of the relevant timescales in a haploid population where both negative frequency-dependent selection and random genetic drift act. We shall assume that the negative frequency-dependent selection acts on a biallelic locus, and manifests itself as the selection coefficient of one of the two alleles *changing sign*, from positive to negative, as its frequency is increased, while the other allele has a selection coefficient of zero.

A key aspect of the present work is that we distinguish between the census and effective population sizes. These two measures of the population size do not coincide in the experiments of Chelo et al. (2013) and this is the general case. For an analysis that is not oversimplified, it is necessary that equality of these two parameters is not assumed; they play different roles in the dynamics, as we show in the analysis of this work.

The results we present have a broader applicability than the experimental work that motivated this analysis (Chelo et al., 2013). Negative frequency-dependent selection occurs in a number of biologically important situations. Three examples (the list is not exhaustive) are: (i) host–pathogen coevolution (see e.g., May and Anderson, 1983); (ii) Batesian mimicry (Anderson and Johnson, 2006); (iii) mitochondrial inheritance (Kazancioğlu and Arnqvist, 2014). More generally, negative frequency-dependent selection is viewed as being ubiquitous in situations where local density regulation in heterogeneous environments occurs (Bell, 1997, p. 498).

2. Model

The theoretical analysis of the present work considers a model based on the following.

1. Haploid asexual organisms.
   These have a single locus with two alleles, denoted A and B.
   We shall often refer to the A allele as the *mutant allele* or simply the *mutant*.
2. Finite population size.
   The census population size is N and the effective population size is N_e.
   We shall not assume that N_e coincides with N, rather we assume that N_e is generally smaller than N.
3. Negative frequency-dependent selection.
   When the frequency of the A allele is x, the relative fitnesses of carriers of the A and B alleles are $1+s(x)$ and 1, respectively. The selection coefficient $s(x)$ associated with the A allele is frequency dependent and in the low frequency regime ($x \ll 1$) we model it by the linear function
   $$s(x) = a - bx \tag{1}$$
   where both a and b are positive and $a < b$.
   The form of negative frequency-dependent selection in Eq. (1) is the maximally simplest form compatible with a selection coefficient that changes from positive to negative values.
   We shall assume that $a/b \ll 1$ so that the main phenomena associated with the form of frequency-dependent selection adopted occur at low frequencies.[2]
4. Initial number of A alleles. We assume the initial number of A alleles in the population is known and we denote this number by n.

The above model, while having a fundamental interest in its own right, also captures important features of the experiments reported by Chelo et al. (2013) on marked lines of the nematode *Caenorhabditis elegans*, in the regime of low frequencies[3]; see Chelo et al. (2013) and references therein for more details on this experimental system.

When we present numerical cases, we shall take, as a concrete example, the parameters a, b, N, N_e and n to have the values inferred/used in the experiments of Chelo et al. (2013), namely

$$a = 0.12,$$

[1] We formulate such a diploid model, without dominance, from a biallelic haploid model with frequency-dependent selection as follows. Let A and B denote the two alleles, and x the frequency of the A allele. In the haploid model assume relative fitnesses of the A and B alleles of $1+s(x)$ and 1, respectively (selection is frequency dependent when $s(x)$ depends on x). The corresponding diploid model has relative fitnesses of the AA, AB and BB genotypes of $[1+s(x)]^2$, $1+s(x)$ and 1, respectively. This model has frequency dependent fitnesses but no dominance. In an effectively infinite population, the dynamics of both the haploid and diploid models are identical. In a finite population, under a diffusion approximation, the two models have identical descriptions – apart from a factor of 2 in the population size.

[2] The behaviour of $s(x)$ at large frequencies is irrelevant when low frequency phenomena dominate the problem. However, in simulations (carried out later) we need a complete specification of $s(x)$. We then take $s(x)$ to equal $a-bx$ for $x < (1+a)/b$ and -1 for $x \geq (1+a)/b$.

[3] Chelo et al. (2013) assumed that under their experimental conditions, *C. elegans* reproduces primarily by selfing. This assumption is valid providing the population size is not too large, otherwise a small degree of outcrossing may occur (Chelo et al., 2013; Teotónio et al., 2006).