Theoretical Population Biology 80 (2011) 197–207

Contents lists available at SciVerse ScienceDirect

Theoretical Population Biology

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

The rate of multi-step evolution in Moran and Wright–Fisher populations

Stephen R. Proulx

Ecology, Evolution and Marine Biology Department, UC Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 27 January 2011
Available online 23 July 2011

Keywords:
Population genetics
Stochastic process
Stochastic tunneling
Fixation probability

ABSTRACT

Several groups have recently modeled evolutionary transitions from an ancestral allele to a beneficial allele separated by one or more intervening mutants. The beneficial allele can become fixed if a succession of intermediate mutants are fixed or alternatively if successive mutants arise while the previous intermediate mutant is still segregating. This latter process has been termed stochastic tunneling. Previous work has focused on the Moran model of population genetics. I use elementary methods of analyzing stochastic processes to derive the probability of tunneling in the limit of large population size for both Moran and Wright–Fisher populations. I also show how to efficiently obtain numerical results for finite populations. These results show that the probability of stochastic tunneling is twice as large under the Wright–Fisher model as it is under the Moran model.

1. Introduction

Evolutionary biologists have long understood that transitions between adaptive sets of traits may involve multiple substitutions separated by neutral or maladaptive intermediate states (Wright, 1932). There has been a resurgence of interest in these ideas, in part because of advances in methods to measure epistatic interactions (e.g. Tong et al., 2001; Tong and Lesage et al., 2004) and the ability to observe evolutionary trajectories (Weinreich and Chao, 2005). Several researchers have modeled evolutionary processes when epistatic interactions allow for multiple genotypes to have the same direct effect on fitness but experience different evolutionary dynamics because of differences in their genetic robustness (van Nimwegen et al., 1999; de Visser et al., 2003; Proulx and Phillips, 2005; Draghi et al., 2010) or the local mutational landscape (Wilke et al., 2001; O'Fallon et al., 2007). These scenarios can be called circum-neutral because alternative genotypes differ in their long-term evolutionary dynamics only because of the genomic circumstances in which they are found (Proulx and Adler, 2010).

Several groups have extended the theory to describe the rates and probability of transition along a multi-step evolutionary pathway. Weinreich and Chao (2005) took the approach of calculating the total waiting time along various pathways and comparing the relative waiting times to reach a final state. Hermisson and Pennings (2005) considered a scenario where previously accumulated genetic variation may become adaptive following an environmental shift. In this scenario the population genetic dynamics of standing variation plays an important role in determining how evolution proceeds at the next step in the process (see also Kopp and Hermisson, 2009). Iwasa et al. (2004) derived approximate results on the waiting time and probability of a two-step sequence of mutational transitions using the Moran model, while Iwasa et al. (2003) derived results in a Wright–Fisher model for a scenario where multiple mutations are required to escape the immune response. These results have been utilized by several other groups to study the rate of multi-step evolutionary processes (Durrett and Schmidt, 2008; Lynch, 2010; Lynch and Abegg, 2010). Several other works have explored the probability and timing of multi-step processes, as well as exploring the validity of approximations (Schweinsberg, 2008; Weissman et al., 2009; Durrett et al., 2009). Both Schweinsberg (2008) and Weissman et al. (2009) have presented branching process approximations for large populations that are equivalent to the large population size limit results for the Moran model presented here.

The goal of this paper is to show how the finite population processes for both the Moran model and the Wright–Fisher model can be written and solved using the method of first step analysis. This helps to clarify some of the terms described by Iwasa et al. (2004) and gives an algorithm for efficiently solving the finite population Moran model. The Moran tunneling probabilities have previously been applied to Wright–Fisher populations without verifying that these results still hold. I show that the Wright–Fisher tunneling probabilities differ from the Moran probabilities by a factor of 2. This correction will allow stochastic tunneling results to be applied to a wider range of scenarios. I also compare the large population size approximations for the rate of tunneling with simulations and exact calculations for small population size.

E-mail addresses: stephen.proulx@gmail.com, proulx@proulxresearch.org.

0040-5809/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2011.07.003

Notations

μ_1	probability that a wild-type allele mutates to produce a primary mutation
μ_2	probability that a primary mutant allele mutates to produce a secondary mutation
N	number of haploid genomes in the population
$U(x)$	Probability an allele with relative fitness x becomes fixed when initially present as a single copy
U_∞	probability that a tunnel of infinitely many steps will open.
r	fitness of primary mutants relative to the wild-type
a	fitness of secondary mutants relative to the wild-type
S_1	probability that a primary mutation destined to become fixed arises in a given generation
S_2	probability that a secondary mutation destined to become fixed arises in a population composed entirely of primary mutants
T	probability, in a population of wild-type alleles, of a primary mutant destined to (before the primary mutant becomes fixed) give rise to a secondary mutant that then becomes fixed
π_i	probability of eventual extinction of a lineage descending from i primary mutants
ω	composite parameter equal to $\mu_2 U(a)$
v_i	probability that no successful secondary mutations are produced from a lineage descending from i primary mutants, conditional on the non-fixation of the primary mutation
$\boldsymbol{v}$	vector of the probability that no successful secondary mutations are produced
$\tilde{v}_i$	unconditional probability that no successful secondary mutations are produced from a lineage descending from i primary mutants
α	probability that no successful secondary mutants are spawned from a lineage descending from a single primary mutant
α_{IMN}	approximate α derived by Iwasa et al. (2004)
T_{M}	for the Moran model, the probability that a single primary mutant will produce a lineage that produces a successful secondary mutant
T_{WF}	for the Wright–Fisher model, the probability that a single primary mutant will produce a lineage that produces a successful secondary mutant

1.1. Preliminary definitions and results

By considering the population level evolutionary process as a series of transitions between populations fixed for a single genotype we can calculate the waiting time for the population to become fixed for secondary mutations. So long as $N\mu \ll 1$ we will seldom have multiple mutants arising in the same generation. This approach also assumes that each attempt at tunneling, if unsuccessful, is over before another primary mutant arises. Determining when this condition actually holds is more difficult because the sojourn time of the primary mutant goes up as its selective disadvantage decreases. In the case of circum-neutral primary mutants, the sojourn times are characterized by large variances that become undefined as population size approaches infinity. A rigorous analysis of the parameter combinations that allow this approximation to be applied is provided in Schweinsberg (2008) and Weissman et al. (2009).

The first mutational step (the primary mutant) is assumed to have relative fitness $r \leq 1$, while the second mutational step (secondary mutant) is assumed to have fitness $a > 1$ relative to the ancestral allele. In the case where r is exactly one, the first mutational step has no direct effect on fitness and the primary mutants can be considered circum-neutral (Proulx and Adler, 2010). Such circum-neutral substitutions do not directly affect reproductive fitness but do alter the long-term evolutionary trajectory of the population. The ancestral population can evolve to be fixed for the secondary mutant either through a sequential mutational pathway or because a lineage of primary mutants destined for extinction produces a secondary mutant which is destined for fixation, a process termed stochastic tunneling by Komarova et al. (2003).

The waiting time until a secondary mutation becomes fixed can be expressed in terms of the waiting times for the sequential and tunneling paths. I define the per generation probability of successful sequential substitutions S_1 and S_2 and the per generation probability of the opening of a successful tunnel as T. The waiting time for the transition between population states is well described by an exponential waiting time so long as population size is not too small (Iwasa et al., 2005). This means that the process is characterized by a race between waiting for a primary mutation to arise and fix and the start of a tunneling pathway. The expectation of the total waiting time until a secondary mutation is given by

$$E[t] = \frac{T}{(T+S_1)^2} + \frac{S_1(S_1+S_2+T)}{S_2(T+S_1)^2}, \tag{1}$$

where the first term represents the contribution to the expected waiting time from tunneling pathways and the second term represents the contribution from sequential pathways. If $T = 0$ this is simply the sum of the waiting times for primary and secondary mutations to sequentially fix. This approximation ignores the time that it takes for beneficial mutations to spread through the population and the amount of time that primary mutants are segregating before a secondary mutation arises. The time required for alleles destined to fix to spread to fixation is typically much smaller than the waiting times for them to arise, and in any case it can be simply added to the total waiting time (see Lynch and Abegg, 2010).

The per generation probabilities of sequential fixation are

$$S_1 = N\mu_1 U(r), \tag{2}$$

$$S_2 = N\mu_2 U(a/r), \tag{3}$$

where N is the haploid population size (for simplicity I assume this is approximately the effective population size as well), μ_1 is the probability that an ancestral allele will mutate into a primary mutant, μ_2 is the probability that the primary mutant will mutate into a secondary mutant, and $U(x)$ is the fixation probability of a mutation with relative fitness x when initially present as a single copy. Because this follows sequential fixation of mutants, the secondary mutant is invading into a population fixed for the primary mutant, giving it a relative fitness of a/r.

Following Iwasa et al. (2004), the probability of tunneling can be written as

$$T = (1-U(r))(1-E[\text{no secondary substitution}|\text{extinction}]), \tag{4}$$

where $E[\text{no secondary substitution}|\text{extinction}]$ represents the probability that no successful secondary mutations arise while the primary mutant is segregating conditioned on the eventual extinction of the lineage of primary mutants. This can be related to the unconditional expectation by

$$E[\text{no secondary substitution}|\text{extinction}] = E[\text{no secondary substitution}](1-U(r)) \tag{5}$$

(Iwasa et al., 2004). This provides a simple relationship between calculations made using the conditioned trajectory of primary mutations and the unconditioned trajectory of primary mutations.