The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation

H. Allen Orr*

Department of Biology, University of Rochester, Rochester, NY 14627, USA

Abstract

Recent models of adaptation at the DNA sequence level assume that the fitness effects of new mutations show certain statistical properties. In particular, these models assume that the distribution of fitness effects among new mutations is in the domain of attraction of the so-called Gumbel-type extreme value distribution. This assumption has not, however, been justified on any biological or theoretical grounds. In this note, I study random mutation in one of the simplest models of mutation and adaptation—Fisher's geometric model. I show that random mutation in this model yields a distribution of mutational effects that belongs to the Gumbel type. I also show that the distribution of fitness effects among rare beneficial mutations in Fisher's model is asymptotically exponential. I confirm these analytic findings with exact computer simulations. These results provide some support for the use of Gumbel-type extreme value theory in studies of adaptation and point to a surprising connection between recent phenotypic- and sequence-based models of adaptation: in both, the distribution of fitness effects among rare beneficial mutations is approximately exponential.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Adaptation; Extreme value theory; Geometric model; Mutational landscape

1. Introduction

Recent work in the theory of adaptation has focused on DNA sequence models. Real adaptation in real organisms must, after all, occur in a space of alternative DNA sequences. These recent efforts build on pioneering work by John Maynard Smith (1962, 1970) and John Gillespie (1983, 1984, 1991), who emphasized that, with realistically low mutation rates, natural selection can only "see" mutant sequences that differ from wild-type by a single base-pair change: double and triple, etc. mutants are too rare to be of much significance to molecular evolution. Gillespie further emphasized that, because the wild-type allele typically enjoys high fitness and adaptation involves the substitution of sequences

having yet higher fitness, almost all adaptive evolution occurs among the fittest few alleles locally available at a locus or small genome. Put differently, almost all adaptation occurs within the right-hand tail of the distribution of allelic fitnesses (Gillespie, 1991; Orr, 2003, 2005). As Gillespie further argued, this means that we can import extreme value theory—a body of probability theory that characterizes extreme draws from distributions (Gumbel, 1958; Leadbetter et al., 1983; Embrechts et al., 1997)—into the study of adaptation.

Gillespie (1983, 1984, 1991) used extreme value theory to characterize the statistical properties of molecular evolution in his "mutational landscape model," a model of adaptation over rugged fitness landscapes. More recent work has used extreme value theory to study the genetics of adaptation in this model. Orr (2002), for instance, showed that if the wild-type allele represents the ith fittest allele (more precisely, single base-pair

*Tel.: +585 275 3838; fax: +585 275 2070.
E-mail address: aorr@mail.rochester.edu.

changes to the wild-type yield $i - 1$ beneficial mutations), natural selection will on average substitute a mutant allele having fitness rank $(i + 2)/4$ at the next step in adaptation. It has also been shown that the mean selection coefficient, s, fixed at subsequent steps in adaptation falls off as an approximate geometric sequence (Orr, 2002), and that parallel evolution should be common at the DNA sequence level; indeed parallel evolution should occur about twice as often under positive selection as under neutrality (Orr, 2004a). Finally, it has been shown that new beneficial mutations have exponentially distributed fitness effects (Orr, 2003).

These results, like most that depend on extreme value theory, are robust to many biological details. Most important, these results hold for many possible distributions of allelic fitnesses—a distribution that is almost always unknown. Studies of the mutational landscape model do, however, depend on certain assumptions about the tail behavior of the distribution of allelic fitnesses.

In particular, all studies of the mutational landscape model assume that the right tail of the distribution of allelic fitnesses falls within the domain of attraction of the so-called Gumbel-type extreme value distribution (EVD), which has cumulative distribution function $\Lambda(x) = \exp[-\exp(-x)]$. An EVD describes the distribution of maxima (or a linear transformation of maxima) drawn from a distribution. In reality, there are three different types of extreme value distribution (Gumbel, 1958; Leadbetter et al., 1983; Embrechts et al., 1997). The Gumbel type holds for almost all "ordinary" distributions, including the exponential, gamma, normal, lognormal, and logistic. The Frechet type holds for very heavy-tailed distributions, like the Cauchy, that lack all or higher moments. The Weibull type holds for many (though not all) distributions that are truncated on the right.

There are good reasons why the theory of adaptation has, so far, assumed that the distribution of allelic fitnesses is of the Gumbel type. For one thing, the Gumbel type was the focus of classical extreme value distribution and arguably is better understood than the alternatives; indeed the Gumbel type is often referred to as *the* EVD. More important, the Gumbel type holds for a wider range of distributions than the Frechet and Weibull types (Embrechts et al., 1997). Although it is sometimes claimed that the Gumbel EVD holds only for exponential-like distributions, this is misleading. In reality, distributions having infinite *or* finite (truncated) right end-points can belong to the Gumbel type (Leadbetter et al., 1983; Embrechts et al., 1997). Moreover, distributions whose tails are lighter than exponential ("subexponential," like the lognormal), or whose tails are heavier than exponential ("superexpoential," like the normal) can belong to the Gumbel type (Embrechts et al., 1997, pp. 138, 145, 277).

The other EVD types may also be inappropriate biologically. The Weibull type, for instance, appears inappropriate as it is hard to see why there should, in principle, be a ceiling on the highest fitness possible at a gene. (In any given case, i.e. given a particular wild-type allele, there *is* a best possible mutant allele, but that is a different matter; see Section 3.) The situation may be worse for the Frechet type, which does not easily allow weak selection (extreme draws from heavy-tailed distributions are separated by large spacings). Also, because the Frechet type holds for distributions lacking all or higher moments, we would have no guarantee that mean fitness at a gene could even be defined.

These arguments are, however, obviously not decisive. In this note, I present some support for the Gumbel assumption. In particular, I show that random mutation in Fisher's (1930) geometric model of adaptation gives rise to a distribution of mutational fitness effects of the Gumbel type. Fisher's geometric model represents one of the simplest and best studied models of mutation and adaptation. The model pictures a population as a point in a high-dimensional phenotypic space, in which each axis represents a trait. The population is assumed to be presently off the (local) phenotypic optimum and moves closer to it by producing random mutations. These mutations are represented by vectors having some magnitude and random direction in phenotypic space. Mutations that fall closer to the optimum are beneficial, while those that fall farther away from the optimum are deleterious; because fitness declines monotonically with distance from the optimum (i.e. the landscape is locally smooth), one can calculate the fitness effect of any mutation.

Fisher (1930) used this geometric model to calculate the probability that a mutation of some phenotypic size will be beneficial. He showed that this probability falls off very rapidly with the size of a mutation; Fisher interpreted this to mean that mutations of very small phenotypic effect must be the stuff of adaptation. Kimura (1983) showed, however, that, when taking into account the stochastic loss of beneficial mutations, the distribution of phenotypic effects among mutations fixed at the first step in adaptation is bell-shaped, with mutations of intermediate effect getting substituted most often (also see Otto and Jones, 2000). Finally, Orr (1998) showed that, when integrating over entire adaptive walks (which may involve many substitutions), the distribution of phenotypic effects among mutations fixed during adaptation is nearly exponential.

Here I show that random mutation in Fisher's model gives rise to a distribution of mutational fitness effects of the Gumbel type. I also show that the distribution of fitness effects among *beneficial* mutations in Fisher's model is approximately exponential.

My approach is mostly analytic. But because this work involves several approximations, I check the accuracy of all results with exact computer simulations.