The drift component of evolution by selection
and the role of rare alleles in evolution

R. B. Campbell   University of Northern Iowa   campbell@math.uni.edu
June 2006   Stony Brook, NY

Introduction

Evolution is the result of the combination of selection and drift. Although both forces act at all frequencies, their relative import varies with the frequency of a selected allele. It is possible to interpret selection either as more important when an allele is rare, or more important when it is frequent. In the following, s will be the selection coefficient (fitness equals $1 + s$), but it will just be a parameter in the neutral case which is juxtaposed to selection. The population size is N (haploid), and frequency $1/s$ (relative frequency $1/(Ns)$) is used to demarcate rare from common alleles. Only strong selection ($Ns \gg 1$) is considered.

Fate of an advantageous mutation

The canonical model considers the fate of a new mutation (frequency 1, relative frequency $1/N$) which has selective advantage s. The time until fixation can be approximately parsed into the time for a neutral mutation to increase from 1 to $1/s$, which is $1/(2s)$, plus the time for exponential increase from frequency $1/s$ to N, which is $\ln(Ns)/s$. Thus rate of increase suggests that drift dominates while a mutation is rare, but selection dominates while it is common.

The probability of increasing in frequency from 1 to $1/s$ is approximately $2s$, which is only twice s, the probability that a neutral mutation increases to that frequency. But the probability of increasing in frequency from $1/s$ to N is almost 1, which is manifold $1/(Ns)$, the probability that a neutral mutation increases from $1/s$ to N. Thus the probability of increase also provides that selection is more important when a mutation is frequent.

Model of environmental shifts

An alternative model of evolution with selection posits that drift governs neutral mutations until an environmental shift occurs which favors one of them. Pure drift indeed occurs before selection, and the formula

$$\theta[1 + \int_{1/N}^1 x^{-1}(1-x)^{\theta-1}dx]$$

(and its restriction to intervals) is used to determine the number of alleles at various frequencies at the time selection manifests. It is assumed that all allelic types (not weighted by their frequency) are equally likely to become the selected allele.

The probability of fixation of a selected rare (frequency less than $1/s$) allele ranges from $2s$ to more than .86 (over the frequency range 1 to $1/s$), hence the probability per copy ranges from $2s$ to $.86s$. Subsequent calculations employ s as the per copy probability of fixation of a rare allele. The probability of fixation of a selected common (frequency greater than $1/s$) allele ranges between .86 and 1, subsequent calculations employ 1 as the per allele (not per copy) probability of fixation of common alleles. Given that a selected allele became fixed, the probability that it was rare when selection manifested is ($s \times$ [the cumulative frequency of rare alleles])/([the number of common alleles] $+ s \times$ [the cumulative frequency of rare alleles]). Numerical evaluations employ $u = 10^{-5}$ or 10^{-6} (u is the mutation rate) and $\theta = 2Nu$ ranging from 1 to 10 to study the impact of selection under this model.

Strong selection $Ns = 5$

The number of alleles with relative frequency less than $1/Ns = .2$ ranged from 10.21 (with $u = 10^{-5}$ and $N = 50\,000$) to 135.55 (with $u = 10^{-6}$ and $N = 5\,000\,000$); the number of alleles with relative frequency greater than .2 ranged from 1.61 to .41 for those values. Weighted by their frequencies, the cumulative relative frequency of the rare alleles ranged from .2 to .89, which is the probability that one of them would have become fixed under neutrality. However, if selection manifested (with each allele, not weighted by its frequency, having equal probability of becoming selected) and the selected allele became fixed, the probability that it was a rare allele that became fixed ranges from 38% (versus 20% under neutrality) to 92% (versus 89% under neutrality). Thus selection increases the probability of fixation of rare alleles up to two-fold. *(over)*

Strong selection $Ns = 20$

The number of alleles with relative frequency less than $1/Ns = .05$ ranged from 8.82 (with $u = 10^{-5}$ and $N = 50\,000$) to 130.19 (with $u = 10^{-6}$ and $N = 5\,000\,000$); the number of alleles with relative frequency greater than .05 ranged from 3.00 to 5.76 for those values. Weighted by their frequencies, the cumulative relative frequency of the rare alleles ranged from .05 to .40, which is the probability that one of them would have become fixed under neutrality. However, if selection manifested (with each allele, not weighted by its frequency, having equal probability of becoming selected) and the selected allele became fixed, the probability that it was a rare allele that became fixed ranges from 25% (versus 5% under neutrality) to 58% (versus 40% under neutrality). Thus selection increases the probability of fixation of rare alleles up to five-fold.

Stronger selection $Ns = 100$

The number of alleles with relative frequency less than $1/Ns = .01$ ranged from 7.21 (with $u = 10^{-5}$ and $N = 50\,000$) to 117.32 (with $u = 10^{-6}$ and $N = 5\,000\,000$); the number of alleles with relative frequency greater than .01 ranged from 4.61 to 18.63 for those values. Weighted by their frequencies, the cumulative relative frequency of the rare alleles ranged from .01 to .10, which is the probability that one of them would have become fixed under neutrality. However, if selection manifested (with each allele, not weighted by its frequency, having equal probability of becoming selected) and the selected allele became fixed, the probability that it was a rare allele that became fixed ranges from 18% (versus 1% under neutrality) to 35% (versus 10% under neutrality). Thus selection increases the probability of fixation of rare alleles up to eighteen-fold.

Discussion

The law of large numbers provides an explanation for this paradox of selection favoring both rare and common alleles. Because many copies of an allele are necessary for the deterministic force of selection to overwhelm random fluctuation, selection does not have a significant impact on a new mutation until it is common. However, under the model of environmental shifts, the effects of many rare alleles are cumulated so that the impact of selection is realized although individual alleles are rare.

The paradox can also be viewed as the difference between immediate and ultimate results. The classical result that the probability of fixation of a new selected mutation is $2s$ extends to higher frequencies by increasing the probability of fixation, but decreasing the probability of fixation per copy of that mutation. The results for the fate of a single mutation which are assessed on two intervals reflect the convexity of the increase in the total probability of fixation. Over the interval from 1 to $1/s$, the probability of fixation increases more rapidly than later, providing a lower probability of traversing that interval; the interval from $1/s$ to N has less change in probability of fixation, thereby requiring greater probability of traversing that interval. The relative frequency $1/(Ns)$ demarcates where fixation probability is increasing from where fixation is essentially assured due to selection. Under the model of environmental shifts, only the probability of fixation (not attainment of an intermediate frequency) is considered. This provides an advantage for rare alleles because the ratio of the probability of fixation under selection to the neutral probaility decreases from $2Ns$ when there is a single mutant allele to near one near fixation.

Rare alleles are evolutionary potential. Although they will probably go extinct due to drift even if they acquire a selective advantage, their prospects for fixation are much greater than their frequency suggests.

Table 1: Impact of selection on fixation of rare (relative frequency $< 1/(Ns)$) alleles

NS	u	N	s	number rare alleles	number common alleles	cum rel freq rare alleles	cum rel freq common alleles	prob rare fixed	sel prob/ neut prob
5	10^{-5}	5×10^4	10^{-4}	10.21	1.61	.2	.8	.38	1.9
	10^{-6}	5×10^6	10^{-6}	135.55	.41	.89	.11	.92	1.03
20	10^{-5}	5×10^4	4×10^{-4}	8.82	3.00	.05	.95	.25	5
	10^{-6}	5×10^6	4×10^{-6}	130.19	5.76	.40	.60	.58	1.45
100	10^{-5}	5×10^4	2×10^{-3}	7.21	4.61	.01	.99	.18	18
	10^{-6}	5×10^6	2×10^{-5}	117.32	18.63	.10	.90	.35	3.5