Fixation Probabilities and Times

Sarah P Otto, *University of British Columbia, Vancouver, British Columbia, Canada*
Michael C Whitlock, *University of British Columbia, Vancouver, British Columbia, Canada*

The fixation probability of an allele is the probability that it will eventually be the ancestor of all the alleles within a population at that locus. The time to fixation is the number of generations that it takes for the allele to progress from its initial frequency to fixation.

Advanced article

Article contents

doi: 10.1038/npg.els.0005464

Introduction

Within a population, any allele must ultimately persist or be lost. If, sometime in the future, all members of the population are descendants of the allele, then that allele is said to have fixed within the population. If, sometime in the future, no members of the population are descendants of the allele, then that allele is said to have been lost. Using mathematical models, the probability of fixation and the time until fixation have been calculated under a range of assumptions. To some extent, these two quantities are related to most of the important processes in evolutionary biology. Evolution by natural selection relies on the greater fixation probability of beneficial alleles than deleterious alleles. The rate of adaptation depends on the fixation probabilities and times for alleles that increase fitness. The relative efficacy of selection and random genetic drift can be determined from a comparison of the fixation probabilities of selected and neutral alleles. The extinction risk of small populations depends on the fixation rates of beneficial and deleterious mutations. At a molecular level, the rate of substitution within deoxyribonucleic acid (DNA) sequences depends on the rate at which new mutations appear within a population and on their fixation probability. Thus, an understanding of fixation probabilities and times provides an important basis for a clear understanding of evolution itself. (***See*** Fitness and Selection; Genetic Drift.)

Probability of Fixation

Let us consider the fate (fixation or loss) of an allele (A) that arises within a population; A may represent a nucleotide mutation, an insertion or deletion, or any other genetic change. The probability of fixation (Ψ) of allele A depends on several factors, the most important of which is whether the allele has a positive, negative or neutral effect on fitness. If the average fitness of individuals who carry one copy of A is $(1 + s)$ times greater than the average fitness of individuals who carry no copies of A, then the fitness effect of allele A is measured by s, the selection coefficient. (***See*** Mutational Change in Evolution.)

When s is zero, the allele is said to be selectively neutral. For a neutral allele, the probability of fixation is equal to p, its initial allele frequency. Consider the fact that if one looks at some point a long time in the future, all members of the population must descend from one and only one allele currently present because no polymorphism can last forever in a finite population. As long as the A alleles always have the same average fitness as the non-A alleles, the chance that an A allele will be the lucky ancestor of all descendants far in the future is simply equal to the chance that an allele chosen at random from the initial population is the A allele, which is p. If, initially, there is one copy of the A allele in a population of census size N and ploidy level c ($c = 1$ for haploids, $c = 2$ for diploids, etc.), $p = 1/(cN)$ and hence the fixation probability for a single-copy neutral allele, Ψ_{Neutral}, is $1/(cN)$.

When allele A is beneficial ($s > 0$), it is more likely to persist over time because selection acts to increase its frequency. Haldane (1927) first derived the fixation probability for a single beneficial allele that appears within a population using a branching process model. Haldane assumed that the total number of alleles within a population (cN) is very large and constant (so that the average number of offspring per parent equals 1). Furthermore, he assumed that the number (j) of A-bearing offspring born to a parent carrying allele A would be Poisson distributed with mean $(1 + s)$. Consequently, for A eventually to be lost, which occurs with probability $1 - \Psi$, each of the j offspring copies of the A allele must also be lost eventually, which occurs with probability $(1 - \Psi)^j$ under the assumption that the fate of each offspring allele is independent of the others. This insight allowed Haldane to write

down an equation that the probability of fixation must obey:

$$1-\Psi=\sum_{j=0}^{\infty}\text{Poisson}[1+s](1-\Psi)^{j}=\sum_{j=0}^{\infty}e^{-(1+s)}\frac{(1+s)^{j}}{j!}(1-\Psi)^{j}=e^{-(1+s)\Psi} \quad (1)$$

If selection is weak, the solution to the above equation is approximately

$$\Psi\approx 2s \quad (2)$$

This remarkable result proves that selection is not omnipotent. Selection cannot ensure the fixation of every beneficial allele, because there is always the chance that a parent may fail to pass on the allele; for example, a new allele that confers a 1% fitness benefit ($s = 0.01$) is expected to fix only 1.97% of the time using eqn (1) (or approximately 2% of the time using eqn (2)).

Kimura (1957, 1962) generalized this result to include deleterious alleles ($s < 0$), arbitrary population size and arbitrary initial allele frequency. By using a diffusion analysis, Kimura found that the probability that an A allele would eventually fix is

$$\Psi\approx\frac{1-\exp[-2cN_{e}sp]}{1-\exp[-2cN_{e}s]} \quad (3)$$

In this equation, N_e is the 'variance effective population size', which predicts the degree of genetic drift. Equation (3) is written assuming that allele A has the same effect on fitness (s) regardless of its genetic background. For diploids, this implies that allele A is additive (codominant), but dominance may also be incorporated (see Crow and Kimura, 1970, p. 427). (***See*** Diffusion Theory; Effective Population Size.)

Equation (3) is extremely powerful. Although the diffusion approximation technically holds only for weak selection and a large population size, in practice it is approximately correct even for very small populations and large selection coefficients (**Table 1**). Equation (3) also describes the probability that a deleterious allele will drift to fixation despite selection against it; for example, a new allele that causes a 1% selective disadvantage will fix with probability 0.00037 within a diploid population of size 100 but will essentially never fix ($\Psi\approx 8\times 10^{-19}$) in a diploid population of size 1000 (putting $c=2$, $p=1/(2N)$, $N_e=N$ and $s=-0.01$ into eqn (3)). By taking the limit of eqn (3) when $s=0$, we can also regain the result that $\Psi_{\text{Neutral}}=p$. When N_e is large and s is small but positive, the fixation probability (3) approaches

$$\Psi\approx 2s(N_{e}/N) \quad (4)$$

Thus, factors that increase the extent of random genetic drift (i.e. decrease N_e relative to N) will decrease the chance that a beneficial allele will become established within a population. Such factors include a skewed sex ratio and a high variance in reproductive success.

The above results assume a population that is constant in size, homogeneous over space, randomly mating and unaffected by selection at other loci, but the models have been extended in a number of important ways, which make the results more applicable to natural populations. Fixation probability can be significantly affected by population growth and decline (Ewens, 1967; Otto and Whitlock, 1997), spatial population structure (Barton, 1993; Whitlock, 2003), inbreeding (Caballero and Hill, 1992) and

Table 1 Probability of fixation (Ψ) for an allele initially present in one copy in a diploid population ($c=2$; $N_e=N$)

Population size (N)	Selection (s)	Ψ_{Observed}	$\Psi_{\text{Branching process}}$	$\Psi_{\text{Diffusion}}$
10	0.001	0.0510	0.0020	0.0510
	0.01	0.0600	0.0197	0.0601
	0.1	0.1787	0.1761	0.1847
	1	0.7902	0.7968	0.8647
100	0.001	0.0061	0.0020	0.0061
	0.01	0.0201	0.0197	0.0202
	0.1	0.1760	0.1761	0.1813
	1	0.7958 (0.0004)	0.7968	0.8647
1000	0.001	0.0021 (0.00005)	0.0020	0.0020
	0.01	0.0193 (0.0001)	0.0197	0.0198
	0.1	0.1760 (0.0004)	0.1761	0.1813
	1	0.7975 (0.0004)	0.7968	0.8647

The 'observed' probability of fixation was obtained from an exact matrix calculation, when possible, or by simulation with 10^6 replicates (followed by standard errors in parentheses). Ψ based on a branching process (eqn (1)) is accurate for strong selection and/or large populations ($Ns > \approx 1$). Ψ based on diffusion analyses (eqn (3)) is accurate except with very strong selection ($s > \approx 0.1$).

selection at linked loci (Hill and Robertson, 1966; Barton, 1995).

Time to Fixation

Even when an allele ultimately rises to fixation within a population, it may take many generations, especially when selection is weak or absent. For a neutral allele ($s = 0$) initially present in one copy, the average time to fixation is approximately $2cN_e$ generations, conditional on the allele fixing, a result first derived by Kimura and Ohta (1969). Note that this result is approximately the same as that obtained from coalescence theory for the number of generations, looking back in time, until all members of a current population trace back to a single allele (i.e. the time until the first common ancestor). Either looking forwards or backwards in time, it takes approximately $2cN_e$ generations for a neutral allele to spread through drift from a single copy to fixation. (***See*** Coalescence Theory.)

Selection causes beneficial alleles to rise more rapidly to fixation. For an additive allele initially present in one copy, the average time to fixation satisfies

$$\bar{t} \approx \int_{1/(cN_e)}^{1} \frac{\left(e^{2cN_e sx} - 1\right)\left(e^{2cN_e s(1-x)} - 1\right)}{sx(1-x)\left(e^{2cN_e s} - 1\right)} dx \qquad (5)$$

(Ewens, 1979, p. 151, where results for arbitrary dominance may also be found). For very weak selection ($s < 1/(cN_e)$), eqn (5) approaches $2cN_e$ generations, the fixation time for a neutral allele. Conversely, for very strong selection ($s \gg 0$), eqn (5) approaches

$$\bar{t} \approx \frac{2\ln(cN_e - 1)}{s} \qquad (6)$$

Equation (6) also equals the time for a beneficial allele to rise from frequency $1/(cN_e)$ to $1-1/(cN_e)$ in a model that ignores drift (see Crow and Kimura, 1970, eqn (5.3.13)).

Thus, when selection is weak relative to the force of drift, the fixation time for a beneficial allele is of the order of the population size, whereas when selection is very strong the fixation time is much shorter and is inversely proportional to the strength of selection (**Figure 1**). What is much more surprising is that the time to fixation, conditional on fixation, is the same for deleterious and beneficial alleles. Although deleterious alleles are much less likely to fix, when they do rise to fixation, the time that it takes is also given by eqn (5) (Maruyama and Kimura, 1974). This is because drift must, by chance, occur more rapidly if it is to offset stronger selection against a deleterious allele.

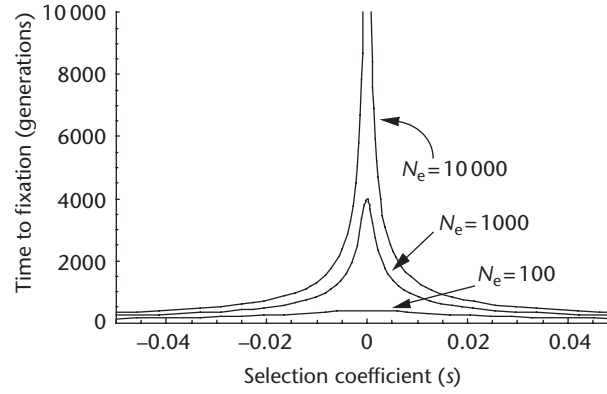

Figure 1 Mean time to fixation for additive alleles in a diploid population (from eqn 5). The average time to fixation, conditioned on the fact that the allele does fix, is a decreasing function of $|s|$. As the strength of selection increases, both beneficial and deleterious alleles, if they fix, will fix faster. Alleles will also fix faster, on average, in populations of smaller effective size. Note that the mean time to fixation for neutral alleles ($s = 0$) is $4N_e$ (for $N_e = 10\,000$, this time is 40 000 generations).

Implications

The probability of fixation has many interesting implications in evolutionary biology. At a molecular level, the rate of nucleotide substitution will equal the rate of appearance of new alleles within a population times their probability of fixation. For a neutral site with a per generation mutation rate of μ per site (or μcN per population), the substitution rate will equal $(\mu cN) \times 1/(cN)$ or simply μ. This calculation predicts that the rate of substitution of neutral alleles should depend only on the mutation rate and not on population size, a fact that has been used to explain why the rate of substitution is often similar in different evolutionary lineages (the so-called 'molecular clock' hypothesis; Kimura, 1983). Higher (lower) substitution rates would be observed if the site were under positive (negative) selection, as can be calculated using eqn (3). (***See*** Molecular Clocks; Mutation Rate; Nucleotide Substitution: Rate.)

If the rate of appearance of beneficial alleles by mutation limits the rate of evolution, then the probability that these new alleles fix is critical to the evolutionary advance of a species. It is not yet known to what extent waiting for new mutations limits the rate of evolution, but potentially this is a large problem, especially for structural mutations such as insertions, deletions, and rearrangements. Furthermore, because alleles are more likely to fix if they have a large advantage, the alleles that do in fact fix in a population are much more likely to have large effect than would be expected by the mutational distribution.

This may explain in part why the genetic basis of many selected traits seems dominated by alleles of large effect (Orr and Coyne, 1992).

The fact that deleterious alleles can fix via genetic drift allows the possibility that the mean fitness of a population can decrease over evolutionary time. In small populations, the risk of decreasing fitness can be significant and may cause a 'mutational meltdown', where extinction results from the accumulation of deleterious mutations (Lynch *et al*., 1995). Furthermore, the rate of fixation of beneficial mutations in small populations may not be high enough for them to adapt to and persist in a changing environment. Thus, because populations with a small effective size fix more deleterious and fewer beneficial alleles over time than a larger population, they are doubly endangered by the effects of genetic drift on fixation rates.

See also

Coalescence Theory
Diffusion Theory

References

Barton NH (1993) The probability of fixation of a favoured allele in a subdivided population. *Genetical Research* **62**: 149–158.

Barton NH (1995) Linkage and the limits to natural selection. *Genetics* **140**: 821–841.

Caballero A and Hill WG (1992) Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* **131**: 493–507.

Crow JF and Kimura M (1970) *An Introduction to Population Genetics Theory*. New York, NY: Harper & Row.

Ewens WJ (1967) The probability of survival of a new mutant in a fluctuating environment. *Heredity* **22**: 438–443.

Ewens WJ (1979) *Mathematical Population Genetics*. Berlin: Springer-Verlag.

Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**: 838–844.

Hill WG and Robertson A (1966) The effect of linkage on the limits to artificial selection. *Genetical Research* **8**: 269–294.

Kimura M (1957) Some problems of stochastic processes in genetics. *Annals of Mathematical Statistics* **28**: 882–901.

Kimura M (1962) On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.

Kimura M and Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763–771.

Lynch M, Conery J and Bürger R (1995) Mutational meltdowns in sexual populations. *Evolution* **49**: 1067–1080.

Maruyama T and Kimura M (1974) A note on the speed of gene-frequency changes in reverse directions in a finite population. *Evolution* **28**: 161–163.

Orr HA and Coyne JA (1992) The genetics of adaptation: a reassessment. *American Naturalist* **140**: 725–742.

Otto SP and Whitlock MC (1997) The probability of fixation in populations of changing size. *Genetics* **146**: 723–733.

Whitlock MC (2003) Fixation probability and time in subdivided populations. *Genetics* (in press).

Further Reading

Fisher RA (1922) On the dominance ratio. *Proceedings of the Royal Society of Edinburgh* **42**: 321–341.

Fisher RA (1930) The distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh* **50**: 204–219.

Gerrish PJ and Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* **103**: 127–144.

Li WH (1997) *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Lynch M, Conery J and Bürger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist* **146**: 489–518.

Nurminsky DI (2001) Genes in sweeping competition. *Cellular and Molecular Life Sciences* **58**: 125–134.

Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* **52**: 935–949.

Poon A and Otto SP (2000) Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* **54**: 1467–1479.

Whitlock MC (2000) Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution* **54**: 1855–1861.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159.