

Ewens W J. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3:87-112, 1972.

[Department of Zoology, University of Texas, Austin, TX]

To test the theory that observed genetic variation in natural populations arises randomly rather than through selection, it is necessary to find what patterns of variation would arise purely randomly. These patterns have several unexpected and counterintuitive features. [The *SC*¹ indicates that this paper has been cited in over 195 publications.]

What Does Genetic Randomness Look Like?

Warren J. Ewens
Department of Biology
University of Pennsylvania
Philadelphia, PA 19104-6018

January 18, 1989

What are the observable consequences of randomness? This question has been asked in all disciplines where random processes affect significantly, or even entirely, those phenomena being studied. It arose in evolutionary genetics in the late 1960s when it was claimed that most of the genetic variation in natural populations, which was then being observed for the first time in detail, does not have a selective basis but arises only as a result of random frequency changes of selectively equivalent gene types (alleles). This claim, called the neutral theory, was controversial when it was first put forward, partly because at the time it was quite unknown what pattern of gene frequencies would be expected if the theory was true.

One frequently stated view was that if there are no selective differences between different alleles, then all those alleles observed in a sample should appear with about equal frequency. Under this view, substantially unequal frequencies would imply selective differences between allelic types. However, this view had no theoretical support and, further, did not lead to any explicit testing procedure for the neutral theory.

In early 1971 I was able, with the help of my colleagues Charles Langley, James Crow, and Sam Karlin, to find the complete probability distribution of the number of allelic types in a sample of genes, together with their frequencies. This distribution depends on the size N of the population from which the sample was taken as well as the mutation rate u to new alleles but does so only through the product Nu , which is of course unknown and quite different from one population to another. This distribution has a number of surprising features. For example, it shows that under the neutral theory the least likely configuration to be observed in a sample of genes is one in which all allelic types arising in the sample occur with equal frequency, implying that the view described above is, in a sense, as incorrect as possible. This occurs because the different allelic types observed in a sample taken at one given time arose by mutation at different times in the past evolution of the population and thus tend, at one given time, to have different frequencies. Historical features such as this are central in biological evolution but not in other areas of science, where they do not arise and thus where "randomness" has quite different consequences. The distribution also shows that all the information concerning Nu in a sample of genes resides in the number of allelic types in the sample, not their frequencies, invalidating the then-common practice of estimating Nu using only the frequencies. This also implies that, given the number of alleles observed in the sample, the probability distribution of the frequencies does not depend on the unknown quantity Nu and can be written down explicitly. These frequencies can thus be used to test the neutral theory. Many tests using this fact have been carried out,¹ some supporting the theory, others not. Thus, although the patterns of gene frequencies expected under the neutral theory are now known, the controversy concerning the theory itself continues.

It is interesting that the same distribution that arises for the frequencies of neutral gene frequencies applied also for random permutations,² random functions,³ ecological theory, the distribution of the prime factors of integers,⁴ urn models,^{5,6} and exchangeability theory.³

1. Watterson G A. Heterosis or neutrality? *Genetics* 85:789-811, 1977. (Cited 60 times.)
2. Kingman K F C. Uses of exchangeability. *Ann. Probab.* 6:183-97, 1978. (Cited 25 times.)
3. Aldous D J. *Exchangeability and related topics*. Berlin, FRG: Springer-Verlag, 1985. 198 p. (Cited 15 times.)
4. Vershik A M. The asymptotic distribution of factorizations of natural numbers into prime divisors. *Sov. Math. Dokl.* 34:57-61, 1986.
5. Donnelly P & Tavaré S. The ages of alleles and a coalescent. *Advan. Appl. Probab.* 18:1-19, 1986.
6. Hoppe F. The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* 25:123-59, 1987.