This Week's Citation Classic®

A statistical method for estimating average heterozygosity and genetic distance from a small number of individuals was developed. It was shown that, as long as the number of genetic loci examined is large, the number of individuals to be used can be very small. [The SCI® indicates that this paper has been cited in over 490 publications.]

Heterozygosity and Genetic Distance

Masatoshi Nei
Center for Demographic and
Population Genetics
University of Texas Health Science Center
Houston, TX 77225

October 2, 1989

In the mid-1960s protein electrophoresis revolutionized the study of population genetics and evolution. This technique revealed an extensive amount of genetic polymorphism existing in natural populations and, at the same time, provided a powerful tool to examine the evolutionary relationships of closely related species.[1,2] In 1972 I developed a statistical method by which one can measure the genetic distance between populations in terms of the number of gene substitutions per locus.[3] This method immediately became popular among investigators. However, it was based on a large-sample theory of statistics, reflecting the then prevailing notion that a large sample size is required to estimate allele frequencies.

However, to determine allele frequencies for a large number of loci, electrophoresis was a time-consuming and quite expensive technique. So, experimentalists often used a very small number of individuals. In some extreme cases, only one or two individuals were used from each species. Previously, Arun K. Roychoudhury and I had shown that for estimating genetic distance a large number of loci should be examined even if the number of individuals per locus is small,[4] but we did not anticipate that experimentalists would use such a small number. Naturally, experimentalists using a very small number were criticized by fellow investigators and theoreticians, and a controversy arose over the number of individuals to be used.[5]

In 1977 I became interested in this problem and decided to solve it from the theoretical point of view. The results of my study showed that a very small sample size is indeed sufficient if the average heterozygosity (a measure of genetic variation within populations) is low and a large number of loci are examined. When the average heterozygosity is high, however, more individuals (say more than 10) are necessary. So, this paper put an end to a controversy that was going on in evolutionary biology. At the same time, I developed a statistical method for obtaining unbiased estimates of average heterozygosity and genetic distance and their standard errors from small samples.

I believe the reason for the frequent citation of this paper is that it provided a theoretical justification for the use of a small number of individuals at a time when electrophoretic study of evolutionary relationships of organisms was popular. The simplicity of the statistical method presented was also welcomed by experimentalists. Mathematically, it is a relatively simple extension of my paper with Roychoudhury,[4] but it is useful for practical purposes. In biometry, simplicity and utility usually go together.

When I published this paper, I thought that future researchers would cite it more often than my 1972 paper. This prediction proved to be wrong. Although this paper has been cited reasonably well, researchers have cited the 1972 paper more often. It seems that they want to cite the first original paper, even if they are actually using a method given in a later paper.

1. Lewontin R C. The genetic basis of evolutionary change. New York: Columbia University Press, 1974. 346 p.
(Cited 1,265 times.)
2. Nei M. Molecular population genetics and evolution. Amsterdam, The Netherlands: North-Holland, 1975. 288 p.
(Cited 880 times.)
3. —————. Genetic distance between populations. Amer. Naturalist 106:283-92, 1972. (Cited 1,345 times.) [See also:
Nei M. Citation Classic. (Barrett J T, comp.) Contemporary classics in plant, animal, and environmental sciences.
Philadelphia: ISI Press, 1986. p. 320.]
4. Nei M & Roychoudhury A K. Sampling variances of heterozygosity and genetic distance. Genetics 76:379-90, 1974.
(Cited 245 times.)
5. Gorman G C & Renzi J. Genetic distance and heterozygosity estimates in electrophoretic studies: effects of sample size.
Copeia 1979:242-9. (Cited 135 times.)