

This Week's Citation Classic

Nei M. Genetic distance between populations. *Amer. Naturalist* 106:283-92, 1972. [Div. Biological and Medical Sciences, Brown Univ., Providence, RI]

A mathematical method for estimating the number of codon (amino acid) substitutions per locus (protein) between two populations is developed by using information on isozyme polymorphism. It is proposed that the genetic distance should be measured in terms of this number. [The *SCI*[®] indicates that this paper has been cited over 210 times since 1972.]

Masatoshi Nei
Center for Demographic and
Population Genetics
University of Texas
PO Box 20334
Houston, TX 77025

August 21, 1979

"In 1968 when I was in Japan, Hubby and Throckmorton published a paper on the proportion of different isozymes between two sibling or non-sibling species of *Drosophila*.¹ Looking at the summary of the paper I wondered whether this sort of data would permit any estimate of the number of amino acid differences per protein between species, but did not have time to read it because of other urgent work. In 1969 I moved to Brown University, Providence, Rhode Island, and just before Christmas, Marty Tracey, then a graduate student, discussed Hubby and Throckmorton's paper in a seminar. After his presentation I raised the question I conceived a year before. Everyone in the seminar said that it would be impossible to estimate the number of amino acid differences, since a pair of proteins which are electrophoretically different may involve more than one amino acid difference. However, I was confident, since I knew that the number of amino acid substitutions in evolution roughly follows the Poisson process. Indeed, by the late evening on the same day I had completed most of the

theoretical work, and next morning I showed it to one of my colleagues. However, it took several months to complete a paper on this subject, and the paper was published in 1971.²

"This first theory had some defects; in particular it neglected the genetic polymorphism within populations which was quite common. Around September, 1970, while I was working on the mathematical theory of gene identity within and between populations, I came to realize that the problem of polymorphism can be solved if we normalize the between-population gene identity by the within-population gene identity. I then defined the genetic distance as the negative logarithm of this normalized identity of genes. The genetic distance defined in this way measures the average number of gene substitutions per locus, and is therefore suited for studying the genetic differentiation of populations. Immediately after the paper on this improved method was published, many authors started to use it.

"I believe the reasons for the frequent citation of the paper are: (1) The simplicity and generality of the method presented, (2) the clear biological meaning of the unit of the distance proposed, and (3) the growing number of studies on genetic differentiation of populations by means of electrophoresis. Most of the distance measures proposed by that time were based on the geometric distance between two populations represented in a multidimensional space, and none of them were proportional to the number of gene substitutions. Later it became clear that the magnitude of my genetic distance generally increases as the rank of the taxa being compared becomes higher. Some authors have proposed that the definition of species should be made by considering this distance when other criteria do not give conclusive results."

1. Hubby J L & Throckmorton L H. Protein differences in *Drosophila*. IV. A study of sibling species. *Amer. Naturalist* 102:193-205, 1968.
2. Nei M. Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. *Amer. Naturalist* 105:385-98, 1971.