Points of View

Syst. Zool., 33(2):238–240, 1984

Misuse and Modification of Nei's Genetic Distance

DAVID M. HILLIS

Museum of Natural History and Department of Systematics and Ecology, The University of Kansas, Lawrence, Kansas 66045

The use and misuse of various measures of genetic "distance" in electrophoretic studies of systematics have recently been the subject of considerable attention (Mickevich and Johnson, 1976; Throckmorton, 1978; Baverstock et al., 1979; Avise et al., 1980; Farris, 1981; Mickevich and Mitter, 1981; Patton et al., 1981; Sites et al., 1981; Thorpe, 1982). Despite the problems associated with the use of genetic distances in the construction of phylogenetic trees, the convenience and general utility of such measures have resulted in their continued and widespread application. The most commonly-used such distance is Nei's *D* (Hedrick, 1983). This measure purportedly assesses the "accumulated number of gene substitutions per locus" *if* "the rate of gene substitution per locus is the same for all loci" (Nei, 1972:283). However, the assumption of equal rates of gene substitution at all loci is rarely, if ever, considered by investigators before employing Nei's *D*. Because this assumption is probably never met in an interspecific electrophoretic survey that involves more than one locus (Sarich, 1977; Thorpe, 1982), it is important to examine the effects of not meeting this assumption.

Nei (1972) defined the normalized identity between two randomly mating diploid populations as

$$I_j = \Sigma x_i y_i / \sqrt{\Sigma x^2_i \Sigma y^2_i},$$

where x_i and y_i are the frequencies of the ith allele at the jth locus in populations X and Y, respectively. He then defined the normalized identity of genes between X and Y with respect to all loci as

$$I = J_{xy} / \sqrt{J_x J_y},$$

where J_{xy}, J_x, and J_y are the arithmetic means of $\Sigma x_i y_i$, Σx^2_i, and Σy^2_i over all loci, respectively. Nei's distance is then defined as

$$D = -\log_e I.$$

Nei (1972:284) further stated that "theoretically it is possible to compute the arithmetic mean of I_j rather than the above quantity [I], but the genetic interpretation of the arithmetic mean is not simple." Nei (1972) suggested that, "when the rate of gene substitution varies with locus and all I_j's are large, a more appropriate measure of genetic distance is given by" employing the geometric means (rather than the arithmetic means) of $\Sigma x_i y_i$, Σx^2_i, and Σy^2_i. This distance measure was named D' by Nei (1972).

Consider a case involving the electrophoretic examination of two species, X and Y (Table 1). It is found that, at one locus, X and Y share the same alleles (either at identical or varying frequencies), whereas at a second locus X and Y share none of the same alleles. This situation is encountered in interspecific electrophoretic studies, because virtually all such analyses include examination of enzymes that differ in substitution rate by an order of magnitude or more (Sarich, 1977; Thorpe, 1982). Table 1 shows three possible variations on this situation. In the first case, X_1

TABLE 1. Three hypothetical cases of allelic frequencies of two taxa at two gene loci.

Locus and allele	Case 1		Case 2		Case 3	
	X_1	Y_1	X_2	Y_2	X_3	Y_3
1 a	1.0	1.0	1.0	1.0	0.5	0.5
b	—	—	—	—	0.5	0.5
2 a	1.0	—	0.5	—	1.0	—
b	—	—	0.5	—	—	—
c	—	1.0	—	0.5	—	1.0
d	—	—	—	0.5	—	—

TABLE 2. Genetic distances for the three cases in Table 1. All of these distances are between pairs of taxa that have identical allelic frequencies at one locus and no alleles in common at a second locus.

Measure	Case 1	Case 2	Case 3
D	0.69	0.41	1.10
D'	∞	∞	∞
D_R	0.50	0.35	0.50
D^*	0.69	0.69	0.69

and Y_1 are fixed for the same allele at locus 1 and are fixed for different alleles at locus 2. In the second case, X_2 and Y_2 are also fixed for the same allele at locus 1; at locus 2, both X_2 and Y_2 are polymorphic but share none of the same alleles. In the third case, X_3 and Y_3 share a polymorphism at locus 1 and are fixed for different alleles at locus 2.

In all three of the comparisons in Table 1, species X shares with Y the same alleles at locus 1 and none of the same alleles at locus 2. In other words, I_j for the first locus is 1 and for the second locus is 0 in all three cases. Given these situations, there is no basis on which to consider X more recently diverged from Y in case 1, 2, or 3. Yet Nei's D ranges from 0.41 for case 2 to 1.10 for case 3 (Table 2). This range encompasses approximately the average difference between genera within a family on the one hand and between sibling species on the other (Ayala, 1975; Avise, 1976; Thorpe, 1982). The reason for this discrepancy is the unmet assumption of equal rates of change for loci, as required by Nei (1972). That this assumption is not met is obvious in that one locus has remained unchanged while a second has become diagnostic. Nei's D' (1972) is also useless in this situation, because I_j for locus 2 is smaller than the 0.7 suggested by Nei as a lower limit for this measure.

The problem outlined above is not limited to Nei's genetic distance measure. The other index of genetic distance that is commonly used in systematic applications, Rogers' (1972) D_R, also varies depending on the amount of polymorphism present in the two species (Table 2). As noted by Wright (1978), Rogers' D_R decreases between two species with no alleles in common as polymorphism increases. Because of this characteristic, it is possible for two species with no alleles in common to have a *smaller* D_R than two species which share half of their alleles.

Because of the above problems, two alternatives seem practical with regard to electrophoretic studies in which Nei's (1972) assumption of equal rates of change for all loci is not met. The first is to ignore Nei's D entirely. The second is to modify Nei's D so that it is not adversely affected by varying rates of change at different loci. This can be accomplished by redefining genetic identity as

$$I^* = \frac{\Sigma I_j}{L},$$

where L is the number of loci, and genetic distance as

$$D^* = -\log_e I^*.$$

D^* is not distorted by shared or unshared polymorphisms as is D when Nei's assumption is not met, and D^* is not distorted by small I_j's as is D'. In the examples in Table 1, although Nei's D ranges from 0.41 to 1.10, Rogers' D_R ranges from 0.35 to 0.50, and Nei's D' is distorted to infinity, D^* remains at 0.69 (the same as D without any polymorphisms) for all three cases. Without any additional information, the time since divergence must be assumed to be equivalent for all of these cases. Whereas D, D', and D_R all fail in this regard, D^* meets this qualification. Because Nei's assumption of equal rates of change for all loci is rarely (if ever) met in systematic evaluations of electropho-

retic data, D^* is a more appropriate measure for use in systematic applications.

Although the modification of genetic distance presented herein does solve one of the problems of applying Nei's D to systematic studies, some of the other problems raised against genetic distance measures remain. For instance, D^* does not account for multiple-step allelic changes, does not differentiate between derived versus ancestral similarities, and is not a metric distance. However, for those who wish to use a summary statistic, D^* appears to eliminate some of the problems of Nei's D, without violating its basic formulation.

ACKNOWLEDGMENTS

A number of people have either read the manuscript or have discussed problems of genetic distance with me; to these I owe a debt of thanks: John C. Avise, Peter R. Baverstock, Anthony A. Echelle, William E. Duellman, Darrel R. Frost, John S. Frost, Philip W. Hedrick, Argye I. Hillis, Richard L. Mayden, John C. Patton, James S. Rogers, H. Bradley Shaffer, Jack W. Sites, Jr., E. O. Wiley, and David A. Wright. Support of this work has been provided by the Center for Biomedical Research at The University of Kansas, by National Science Foundation grant DEB81-08266, and by a National Science Foundation fellowship.

REFERENCES

AVISE, J. C. 1976. Genetic differentiation during speciation. Pages 106–122 *in* Molecular evolution (F. J. Ayala, ed.). Sinauer, Sunderland, Massachusetts.

AVISE, J. C., J. C. PATTON, AND C. F. AQUADRO. 1980. Evolutionary genetics of birds. I. Relationships among North American thrushes and allies. Auk, 97:135–147.

AYALA, F. J. 1975. Genetic differentiation during the speciation process. Pages 1–78 *in* Evolutionary biology. Volume 8 (T. Dobzhansky, M. K. Hecht, and W. C. Steer, eds.). Plenum, New York.

BAVERSTOCK, P. R., S. R. COLE, B. J. RICHARDSON, AND C. H. S. WATTS. 1979. Electrophoresis and cladistics. Syst. Zool., 28:214–219.

FARRIS, J. S. 1981. Distance data in phylogenetic analysis. Pages 3–23 *in* Advances in cladistics. Proceedings of the First Meeting of the Willi Hennig Society (V. A. Funk and D. R. Brooks, eds.). New York Botanical Garden, Bronx.

HEDRICK, P. W. 1983. Genetics of populations. Science Books International, Boston.

MICKEVICH, M. F., AND M. S. JOHNSON. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. Syst. Zool., 25:260–270.

MICKEVICH, M. F., AND C. MITTER. 1981. Treating polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data. Pages 45–58 *in* Advances in cladistics. Proceedings of the First Meeting of the Willi Hennig Society (V. A. Funk and D. R. Brooks, eds.). New York Botanical Garden, Bronx.

NEI, M. 1972. Genetic distance between populations. Am. Nat., 106:283–292.

PATTON, J. C., R. J. BAKER, AND J. C. AVISE. 1981. Phenetic and cladistic analyses of biochemical evolution in peromyscine rodents. Pages 288–308 *in* Mammalian population genetics (M. H. Smith and J. Joule, eds.). Univ. Georgia Press, Athens.

ROGERS, J. S. 1972. Measures of genetic similarity and genetic distance. Studies in Genetics VI, Univ. Texas Publ. No. 7203:145–153.

SARICH, V. M. 1977. Rates, sample sizes, and the neutrality hypothesis for electrophoresis in evolutionary studies. Nature, 265:24–28.

SITES, J. W., JR., I. F. GREENBAUM, AND J. W. BICKHAM. 1981. Biochemical systematics of neotropical turtles of the genus *Rhinoclemmys* (Emydidae: Batagurinae). Herpetologica, 37:256–264.

THORPE, J. P. 1982. The molecular clock hypothesis: Biochemical evolution, genetic differentiation and systematics. Annu. Rev. Ecol. Syst., 13:139–168.

THROCKMORTON, L. H. 1978. Molecular phylogenetics. Pages 221–239 *in* Biosystematics in agriculture (R. H. Foote, L. Knutson, and P. L. Lentz, eds.). John Wiley, New York.

WRIGHT, S. 1978. Evolution and the genetics of populations. Volume 4. Variability within and among natural populations. Univ. Chicago Press, Chicago.

Received 20 September 1983; accepted 25 January 1984.