A Program for Calculating Nei's Genetic Distances and Their Jackknifed Confidence Intervals

H. A. Lessios

Nei's (1972) Standard Genetic Distance is the most widely used index for quantifying electrophoretically determined genetic differences between populations. Mueller and Ayala's (1982) method for calculating confidence intervals by jackknifing has made it possible to determine whether two or more genetic distances are significantly different from each other. However, the jackknifing procedure is computationally tedious and can realistically only be calculated by computer. Sattler and Hilburn (1985) published a program for calculating Hillis's (1984) modification of Nei's genetic distance and for comparing two groups for significance of their differences, but their program does not calculate Nei's Standard Genetic Distance and will not compare average distances between two pairs of groups. Nei et al. (1985) developed an alternative method for calculating confidence intervals of indices of genetic distance, but it is not based on jackknifing and is applicable only to the construction of phylogenetic trees using the unweighted pair-group method.

The DBASE III+® computer program described here accepts as input a table of the frequency of each allele at every locus in each population, calculates Nei's (1972) Standard Genetic Distance D and Hillis's (1984) modified genetic distance D^* between populations, and outputs them in matrix form. Then it jackknives these values by dropping one locus at a time and calculating the pseudovalue that results from this elimination. Means and variances of the pseudovalues are calculated, as well as covariances between all possible pairwise distances. If the user wants to determine the significance of differences between two groups, the program uses the calculated parameters to compute and compare average within-group distances with average between-group distances. In addition, any two average distances between two groups (including groups consisting of one population each) can be compared. The statistical methods outlined by Mueller and Ayala (1982) are used throughout. At the user's option, loci for which data are missing can be eliminated from all comparisons between populations, or they can be included in pairwise comparisons of populations in which they are present. All calculated statistics, including pseudovalues, mean distances and their variance, covariances between all distances, U values, and their confidence intervals, are saved to disk in DBASE III+ files. Reports can be generated to the printer of all these files, which can also be converted to ASCII form for processing by other programs.

Given enough disk space, the program can deal with an effectively infinite number of loci and alleles (the loci × alleles combination can reach the 1 billion maximum records allowed by DBASE III+). However, because covariances between all combinations of genetic distances are saved in a matrix form, and because DBASE III+ has a limit of 128 fields (columns), the program cannot handle more than 16 populations at a time. The program was written for the MS-DOS operating system, and will also run on any program compatible with DBASE III+ (such as FOXBASE®). A listing of the program on disk or printed copy is available from the author. The program, and two ancillary ones for creating the frequency table and converting DBASE files to text files, along with their documentation, occupy 88 KB of disk space and contain a total of 2,588 lines of code. Examples occupy another 50 KB of disk space. They can be obtained from the author by sending a 5¼ in 1.2 MB or 360 KB formatted disk, or a 3½ in 1.44 MB or 720 KB formatted disk.

References

Hillis DM, 1984. Misuse and modification of Nei's genetic distance. Syst Zool 33:238–240.

Mueller LD and Ayala FJ, 1982. Estimation and interpretation of genetic distance in empirical studies. Genet Res Camb 40:127–137.

Nei M, 1972. Genetic distance between populations. Amer Nat 106:283–292.

Nei M, Stephens JC, and Saitou S, 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol Biol Evol 2:66–85.

Sattler PW and Hilburn LR, 1985. A program for calculating genetic distance, and its use in determining significant differences in genetic similarity between two groups of populations. J Hered 76:400.