

Population Differentiation: Measures of

Giorgio Bertorelle, *University of Ferrara, Ferrara, Italy*

Advanced

Note to author: Please supply a definition: a two-sentence definition of the title subject.

0456.1 In most species, individuals tend to find a partner and reproduce within a limited group of conspecifics. That behavior may be for different reasons (geographic distance, physical or ecological barriers, differences in language or religion, etc.), but the genetic consequence is always the same: the allelic frequencies drift almost independently within homogeneous groups or subpopulations, and new mutations do not freely spread beyond the limits of the group where they originated. This process results in a measurable amount of genetic divergence between groups, i.e. population differentiation.

0456.2 Different indices are commonly used to quantify population differentiation, and none of them can be considered better than others in all respects. Without doubt, however, the most celebrated and well known index is Wright's F_{st} . Initially defined assuming a simple biallelic locus and no mutation, F_{st} measures the correlation between pairs of homologous genes within subpopulations relative to pairs of genes taken at random from the total population. An easier but still rigorous definition says that F_{st} is the fraction of heterozygosity expected in the total population that is not observed within the subpopulations. In practice, values of F_{st} close to 0 indicate that allele frequencies are very similar in different subpopulations, and values close to 1 suggest that genetic drift has had the time to randomly fix different alleles in different groups.

0456.3 Several estimators of F_{st} with slightly different properties and applicability (and different names) have been proposed to accommodate different situations and different types of markers. Some of them (G_{st} and θ , for example) consider only the allele frequencies and are preferable when subpopulations diverge under a pure drift process. Others (N_{st} , R_{st} , Φ_{st}) incorporate some measure of the evolutionary distance between alleles (usually estimated from nucleotide or, for microsatellites, length differences), and are therefore more appropriate when population differentiation is accompanied by mutational events. Recent discussions on this topic can be found in Nei and Kumar (2000) and Excoffier (2001).

Whereas F_{st} or similar indices summarize with a single number the overall (average) level of divergence among several subpopulations, specific relationships between pairs of groups can be quantified using genetic distances.

0456.5 Earlier distances (Rogers' distance, for example) were mainly derived for their statistical properties. Based on the geometric representation of the allelic frequencies of subpopulations in a multidimensional space, these distances are not directly related to the process of population differentiation. More recent distances tend, on the other hand, to be based on genetic models of population divergence. Among them, the allele-sharing distance and Nei's standard genetic distance D are preferable when the evolutionary relationship between alleles cannot be estimated, whereas the number of net substitutions D_A (and successive improvement suitable for unequal group sizes) and the squared difference in the mean number of repeats $(\delta\mu)^2$ are more suitable for deoxyribonucleic acid (DNA) sequences and microsatellites respectively. Pairwise F_{st} , with all its analogs, is commonly used as a genetic distance in several different circumstances. A clear description of genetic distances can be found in Beaumont *et al.* (1998). (See A0440.)

0456.6 The final goal of a good measure of differentiation is not the simple identification of the relative level of genetic divergence between two or more groups, but the estimation of some evolutionarily relevant parameters. In particular, assuming a nonequilibrium model of isolated subpopulations, some measures of differentiation can be used to estimate the divergence time, i.e. the time in the past when the subpopulations arose from a single ancestral population (assuming that the mutation rate or population sizes, or both, are known). Pairs of divergence times, or directly genetic distances, are also commonly used to reconstruct phylogenetic trees. If an equilibrium model of subpopulations exchanging migrants is more appropriate, and genetic differentiation is therefore not increasing with time, estimates of gene flow can be obtained from measures of population divergence. In the last few

years, maximum likelihood methods have successfully competed with population differentiation measures in this effort to reconstruct demographic and historical processes.

References

- Beaumont MA, Ibrahim KA, Boursot P and Bruford MW (1998) Measuring genetic distance. In: Karp A, Isaac PG and Ingram DS (eds) *Molecular Tools for Screening Biodiversity*, pp. 315–327. London: Chapman and Hall.
- Excoffier L (2001) Analysis of population subdivision. In: Balding DJ, Bishop M and Cannings C (eds) *Handbook of Statistical Genetics*, pp. 271–302. Chichester: John Wiley.
- Nei M and Kumar S (2000) *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Keywords

population structure, F_{st} , genetic distance, divergence time, migration