DERLEME REVIEW

Statistical Measures for Genetic Differentiation: Review

Genetik Farklılaşma İçin İstatistiksel Ölçüler

İsmet DOĞAN,[a]
Nurhan DOĞAN[a]

[a]Department of Biostatistics and Medical Informatics, Afyon Kocatepe University Faculty of Medicine, Afyonkarahisar

Geliş Tarihi/*Received:* 13.01.2016
Kabul Tarihi/*Accepted:* 31.03.2016

Yazışma Adresi/*Correspondence:*
Nurhan DOĞAN
Afyon Kocatepe University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Afyonkarahisar, TÜRKİYE/TURKEY
idogan@aku.edu.tr

ABSTRACT A primary goal of empirical population genetic studies is the identification, quantification and comparison of genetic differentiation among loci, individuals, populations, species and studies. Determining the genetic structure of natural populations forms an important part of population genetics and has many applications in evolutionary biology, conservation, forensics and plant and animal breeding. Population differentiation is a fundamental process of evolution, and many evolutionary studies, such as population genetics, phylogeography and conservation biology, all require the inference of population differentiation. Estimates of the fixation index have been used as measures of population differentiation for many decades. The method most frequently used to assess population structure is the calculation of F_{ST}, a summary statistic first introduced by Wright. Fixation measurements of the genetic differentiation among subpopulations are fundamental parameters in population genetics, with many valuable applications in molecular biology, evolutionary biology, conservation and forensic. A number of related indices of genetic differentiation have been subsequently derived in link with the natures of the diagnostic genetic markers such as G_{ST}, Φ_{ST}, R_{ST}. This paper is intended to be a review of the genetic differentiation indices that population geneticists frequently use. For this purpose seven fixation indices were investigated. In the literature, different indices are commonly used to quantify population differentiation, and none of them can be considered beter than others in all respects.

Key Words: Population structure; genetic differentiation; fixation indices; allele frequency

ÖZET Deneysel populasyon genetik çalışmalarının temel amacı, lokuslar, kişiler, popülasyonlar, türler ve çalışmalar arasındaki genetik farklılaşmayı karşılaştırmak, tanımlamak ve rakamsal olarak ifade etmektir. Doğal popülasyonların genetik yapısının belirlenmesi, popülasyon genetiğinin önemli bir parçasını oluşturmaktadır ve konu ile ilgili evrimsel biyoloji, adli tıp, genlerin korunması, bitki ve hayvan ıslahı alanlarında çok sayıda uygulama bulunmaktadır. Popülasyon farklılaşması evrimin temel sürecidir ve popülasyon genetiği, filocoğrafya ve koruma biyolojisi gibi alanlarda yapılan birçok evrimsel çalışma popülasyon farklılaşmasının belirlenmesini gerektirmektedir. Sabitleme indekslerine ait tahminler, popülasyon farklılaşmasının bir ölçüsü olarak uzun yıllardan beri kullanılmaktadır. Popülasyon yapısının belirlenmesi amacıyla en sık kullanılan yöntem, ilk kez Wright tarafından ileri sürülen F_{ST} istatistiğinin hesaplanmasıdır. Sabitleme indekslerine ait tahminler, popülasyon farklılaşmasının bir ölçüsü olarak uzun yıllardan beri kullanılmaktadır. Alt popülasyonlar arasındaki genetik farklılaşmaya ait sabitleme ölçümleri moleküler biyoloji, evrimsel biyoloji genlerin korunması ve adli tıp alanlarındaki bir çok değerli uygulamalar ile popülasyon genetiğinde temel parametrelerdir. Tanı koymada kullanılan genetik göstergelerin doğası ile bağlantılı olarak G_{ST}, Φ_{ST}, R_{ST}. vb. bir dizi genetik farklılaşma indeksi sonradan türetilmiştir. Bu çalışmada popülasyon genetikçilerinin sıklıkla kullandıkları genetik farklılaşma indekslerinin gözden geçirilmesi amaçlanmıştır. Bu amaçla yedi farklı sabitleme indeksi incelenmiştir. Literatürde, popülasyon farklılaşmasını ölçmek için genel olarak farklı indeksler kullanılmaktadır ve bunlardan hiç biri, bütün açılardan diğerlerinden daha iyi olarak kabul edilebilir değildir.

Anahtar Kelimeler: Genetik varyasyon; genetik farklılaşma; sabitleşme indeksleri; allel frekansı

doi: 10.5336/biostatic.2016-50192

Turkiye Klinikleri J Biostat 2016;8(2):180-6

The knowledge of how genetic variation is partitioned among populations may have important implications not only in evolutionary biology and ecology, but also in conservation biology. Hence, reliable estimates of population differentiation are crucial to understand the connectivity among populations and represent important tools to develop conservation strategies.[1] One important goal of population genetic studies is to estimate the amount of genetic differentiation among populations in order to draw conclusions on the demographic history. A common measure for the degree of genetic differentiation is the fixation index F_{ST}, first defined by Wright (1951). It has become a fundamental parameter in population genetics, with numerous valuable applications in molecular ecology, evolutionary biology and conservation biology. Since Wright, many more differentiation statistics conceptually similar to F_{ST} have been proposed to deal with highly polymorphic markers such as microsatellites (e.g. G_{ST} and R_{ST}) and DNA sequences (e.g. Φ_{ST} and R_{ST}).[2]

A number of workers have used or proposed measures to compare different populations using gene frequency data for a number of loci. To test the amount of difference between two populations one can use a F statistic which gives the correlation between random gametes within the two populations, relative to that of the gametes in the two populations combined. For a locus with two alleles, F_{ST} is equal to the ratio of the actual variance of the gene frequencies between the two populations.[3]

In population genetics it is common to analyze population structure to test hypotheses concerning gene flow and isolation within species. The most frequently used method consists of estimating F_{ST}, a measure of population differentiation first developed by Wright. Since the original work of Wright, several authors have proposed methods to estimate F_{ST}, leading to a number of F_{ST} analogues such as G_{ST}, R_{ST}, Φ_{ST}, $G_{ST}^{'}$, D and φ_{ST}. One thing that these F_{ST} analogues have in common is that their values are dependent on the amount of within-population genetic variation; high levels of genetic variation therefore generally lead to lower F_{ST} estimates than low levels of variation. This dependency makes it difficult to compare genetic markers with different mutation rates or species with different effective population sizes.[4]

Despite the development of alternative approaches such as methods assigning individuals to populations, differentiation estimators remain the most commonly used tools to describe population structuring. The main reason behind this popularity stems from their direct link to the biologically relevant number of effective migrants.[1]

This paper is intended to be a review of the genetic differentiation measures that population geneticists frequently use. For this purpose seven fixation indices were investigated.

WRIGHT'S F_{ST}

Wright's (1951) fixation indice, F_{ST}, is the parameter most widely used to describe population structure. Wright defined the fixation indice as a correlation between uniting gametes. His treatment is restricted to neutral diallelic loci; it is somewhat artificial (because numerical values are assigned to gametes) and not entirely clear.[5] Two of the most commonly used definitions for F_{ST} at a given locus are based on the variance of allele frequencies between populations, and on the probability of identity by descent. While $\bar{p}$ is the average frequency of an allele, σ_S^2 denotes the weighted variance in the frequency. These weights are determined by the size of the subpopulations (S). And σ_T^2 stands for the variance of the allelic state in the total population (T). F_{ST} is defined as[6]

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}\,(1-\bar{p}\,)}$$

Wright (1951) showed that the amount of genetic differentiation among populations has a predictable relationship to the rates of important evolutionary processes (migration, mutation and drift). F_{ST} is a convenient measure of genetic differentiation, and as a result F_{ST} and related statistics are among the most widely used descriptive

statistics in population and evolutionary genetics. But F_{ST} is more than a descriptive statistic and a measure of genetic differentiation. F_{ST} is directly related to the variance in allele frequency among populations and, conversely, to the degree of resemblance among individuals within populations. Wright's F_{ST} ranges from 0.0 to 1.0. If F_{ST} is small, it means that the allele frequencies within each population are similar; if it is large, it means that the allele frequencies are different. Estimates of F_{ST} are also important in association mapping of human disease genes and in forensic science.[7]

F_{ST} is a measure of population divergence. It measures variation between populations vs. within populations. One can calculate a global measure, assuming that all populations are equally diverged from an ancestral population, or one can calculate F_{ST} for specific populations or for pairs of populations while utilizing data from all populations. F_{ST} may be calculated for single genetic markers. For multiallelic markers, such as microsatellites, this is useful, but single-nucleotide polymorphisms (SNPs) contain much less information when taken one at a time, and thus it is advantageous to calculate averages over windows of markers or even over the whole genome. The advantage of windowed F_{ST} is that it can be used to find regions of the genome that show different patterns of divergence, indicative of selective forces at work during human history. Another measure of human evolutionary history is haplotype diversity.[8]

NEI'S G_{ST}

G_{ST} can quantify differentiation fairly well when heterozygosity is low whatever the causes (e.g. low mutation rate, low initial heterozygosity of the ancestral population or short split time); however, when heterozygosity is high (whatever the causes, e.g. high mutation rate, high initial heterozygosity) and gene flow is moderate to strong G_{ST} often fails to measure differentiation.[9] Consider a diploid population which is subdivided into s subpopulations, and assume that there are r alleles $(A_1, A_2, \dots, A_r)$ segregating in the population. Let p_{ik} be the frequency of allele A_k in the ith subpopulation, and P_{ikl} be the frequency of genotype A_kA_l in this subpopulation. Nei (1977) has defined fixation indice G_{ST} in the following way.[10]

$$G_{ST} = 1 - H_S/H_T$$

where,

$$H_S = 1 - \sum_{k=1}^{r} \overline{p_k^2} \qquad H_T = 1 - \sum_{k=1}^{r} \bar{p}_k^2 \qquad (1)$$

here,

$$\overline{p_k^2} = \sum_{i=1}^{s} w_i p_{ik}^2 \, \bar{p}_k = \sum_{i=1}^{s} w_i p_{ik}$$

in which w_i is the relative size of the ith subpopulation with $\sum_{i=1}^{s} w_i = 1$. In most instances w_i is not known, but $w_i = 1/s$ may be assumed, because population size is quite transitory and geneticists are interested in gene frequency differences disregarding the effect of population size. H_S and H_T represent the expected heterozygosities under Hardy-Weinberg equilibrium or gene diversities within subpopulations and in the total population, respectively.[11]

SLATKIN'S R_{ST}

Microsatellite loci are often highly polymorphic and relatively easy to survey and hence offer the hope of greater understanding of population structure. The question is how to make the best use of allele frequencies at microsatellite loci. Slatkin (1995), introduce a statistic (R_{ST}), analogous to Wright's F_{ST} that can be used to estimate effective migration rates or times since population divergence.[12]

$$R_{ST} = \frac{\bar{S} - S_W}{\bar{S}}$$

$$\bar{S} = \frac{2n-1}{2nd_S - 1} S_W + \frac{2n(d_S - 1)}{2nd_S - 1} S_B$$

$$S_W = \frac{1}{d_S} \sum_{j=1}^{d_S} \frac{2}{2n(2n-1)} \sum_{i<i'} (a_{ij} - a_{i'j})^2$$

$$S_B = \frac{2}{(2n)^2 d_S (d_S - 1)} \sum_{j<j'} \sum_{i<i'} (a_{ij} - a_{i'j'})^2$$

a_{ij}: allele size of the ith copy $(i = 1, 2, \dots, 2n)$ in the jth population $(j = 1, 2, \dots, d_S)$

WEIR AND COCKERHAM'S Φ_{ST}

Weir and Cockerham (1984) based their study of population structure on the analysis of the variance and covariances of indicator variables for allelic state, and they related their parameters to fixation indices and measures of identity by descent.[13] Although Weir and Cockerham's analysis is more lucid and general than Wright's, it is disturbing that negative variance components may occur if mates are less closely related than the average within subpopulations.[5] If $\tilde{p}_i$ is the frequency of allele A in the sample of size n_i from population $i\,(i = 1, 2, \dots, r)$ and $\tilde{h}_i$ is the observed proportion of individuals heterozygous for allele A, then

$$\Phi_{ST} = \frac{a}{a+b+c}$$

$a = \frac{\bar{n}}{n_c}\left\{s^2 - \frac{1}{\bar{n}-1}\left[\bar{p}(1-\bar{p}) - \frac{r-1}{r}s^2 - \frac{1}{4}\bar{h}\right]\right\}$; variance for between populations,

$b = \frac{\bar{n}}{\bar{n}-1}\left[\bar{p}(1-\bar{p}) - \frac{r-1}{r}s^2 - \frac{2\bar{n}-1}{4\bar{n}}\bar{h}\right]$; variance for between individuals within populations,

$c = \frac{1}{2}\bar{h}$; variance for between gametes within individuals,

$\bar{n} = \sum_i n_i/r$, the average sample size,

$n_c = \frac{(r\bar{n} - \sum_i n_i^2/r\bar{n})}{r-1} = \bar{n}(1 - c^2/r)$; with c^2 the squared coefficient of variation of sample sizes

$\bar{p} = \sum_i \frac{n_i\tilde{p}_i}{r\bar{n}}$; the average sample frequency of allele A,

$s^2 = \sum_i \frac{n_i(\tilde{p}_i - \bar{p})^2}{(r-1)\bar{n}}$; the sample variance of allele A frequencies over populations,

$\bar{h} = \sum_i \frac{n_i\tilde{h}_i}{r\bar{n}}$; the average heterozygote frequency for allele A.

HEDRICK'S $G_{ST}^{'}$

Hedrick pointed out that G_{ST} does not vary between zero and one, but rather between zero and $G_{ST(max)}$, a maximum value that depends on H_S and the number of demes that were sampled.[14] Hedrick (2005) used the original Nei's definition of G_{ST} and that its maximum value ($G_{ST(max)}$) is a function of the expected heterozygosity, H_S, and the number of sampled populations k

$$G_{ST(max)} = \frac{(k-1)(1-H_S)}{k-1+H_S}$$

Hedrick then defined the standardized G_{ST}, which he called $G_{ST}^{'}$ as[15]

$$G_{ST}^{'} = \frac{G_{ST}}{G_{ST(max)}} = \frac{G_{ST}(k-1+H_S)}{(k-1)(1-H_S)}$$

JOST'S D

Jost (2008) developed a new framework for estimating genetic differentiation.[16] Instead of using heterozygosity, Jost based his statistic D on the effective number of alleles. Jost's D does not depend on the effective population size. This can be regarded as an advantage when absolute levels of population differentiation are compared among species with very different sizes, but also as a disadvantage since the different divergence dynamics in small and large populations is not considered. Jost's D for a locus can be written as[17]

$$D = \left(\frac{k}{k-1}\right)\left(\frac{H_T - H_S}{1-H_S}\right)$$

k is the number of subpopulations,H_S and H_T were given in equation (1).

EXCOFFIER, SMOUSE AND QUATTRO'S φ_{ST}

The AMOVA (Analysis of MOlecular VAriance - as designated by Excoffier et al. (1992)) framework draws from a rich literature on genetic differentiation. The focus of the original seminal work on AMOVA was to derive a framework for partitioning total variance in allele frequencies (across multiple loci) within and among different strata (within populations, among populations, within subpopulations, and among subpopulations) by defining genetic distances between haplotypic data.[18] AMOVA is a method for studying molecular variation within a species. This technique treats genetic distances as deviations from a group mean position, and uses the squared deviations as variances. The resulting test statistic φ_{ST} is analogous to Wright's F_{ST} AMOVA is a particular approach that partitions genetic variation among individuals within populations and among populations. It can

also be used to partition variation at higher levels of structure in nested analyses (e.g., by geographic region or species). The AMOVA approach generates squared Euclidean distances, (δ_{jk}^2), between pairs of individuals and then partitions this variation at different levels of structure – within and among populations, among regions, among species, etc. – depending on the hierarchy that is available for testing.[19] The analysis of molecular variance was initially introduced as an extension of the analysis of gene frequencies for molecular haplotypes in an essentially haploid system. The typical input for AMOVA consisted of a matrix of pairwise Euclidean distances, between all multisite haplotypes and files containing the frequency of those haplotypes within each population.[20] To understand the working principle of AMOVA, a linear

model can be defined in the form:

$$p_{jig} = p + a_g + b_{ig} + c_{jig}$$

where p_{jig} indexes the jth individual ($j = 1, 2, \dots, N_{ig}$) in the ith population ($i = 1, 2, \dots, I_g$) in the gth group ($g = 1, 2, \dots, G$) and p is the unknown expectation of p_{jig} averaged over the whole study. The effects are a for group, b for populations and c for individuals within populations. The effects have the associated variance components σ_a^2, σ_b^2 and σ_c^2 respectively.[21] The corresponding sums of squares are,

$$SSD(WP) = \sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{\sum_{j=1}^{N_{ig}}\sum_{k=1}^{N_{ig}}\delta_{jk}^2}{2N_{ig}}$$

$$SSD(APWG) = \sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{I_g}\sum_{j=1}^{N_{ig}}\sum_{i'=1}^{I_g}\sum_{k=1}^{N_{i'g}}\delta_{jk}^2}{\sum_{i=1}^{I_g}2N_{ig}} - \sum_{i=1}^{I_g}\frac{\sum_{J=1}^{N_{ig}}\sum_{k=1}^{N_{ig}}\delta_{jk}^2}{2N_{ig}}\right)$$

$$SSD(AG) = \left(\frac{\sum_{j=1}^{N_{ig}}\sum_{k=1}^{N_{ig}}\delta_{jk}^2}{2N_{ig}} - \sum_{g=1}^{G}\frac{\sum_{i=1}^{I_g}\sum_{j=1}^{N_{ig}}\sum_{i'=1}^{I_g}\sum_{k=1}^{N_{i'g}}\delta_{jk}^2}{\sum_{i=1}^{I_g}2N_{ig}}\right)$$

The mean squared deviations (MSD) are then obtained by dividing such sum of squared deviations (SSD) by the appropriate degrees of freedom as reported in Table 1.

The n coefficients in Table 1 represent the average sample sizes of particular hierarchical levels, allowing for unequal sample sizes,

$$n = \frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig} - \sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{I_g}N_{ig}^2}{\sum_{i=1}^{I_g}N_{ig}}\right)}{\sum_{g=1}^{G}I_g}$$

$$n' = \frac{\sum_{g=1}^{G}\left(\frac{\sum_{j=1}^{I_g}N_{jg}^2}{\sum_{i=1}^{I_g}N_{ig}}\right) - \frac{\sum_{g=1}^{G}\sum_{j=1}^{I_g}N_{jg}^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}}}{G-1}$$

$$n'' = \frac{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig} - \frac{\sum_{g=1}^{G}\left(\sum_{j=1}^{I_g}N_{jg}\right)^2}{\sum_{g=1}^{G}\sum_{i=1}^{I_g}N_{ig}}}{G-1}$$

The variance components (σ^2's) of each hierarchical level are extracted by equating the mean squares (MSD's) to their expectations. Using variance components, φ_{ST} value is obtained from the following formula,

$$\varphi_{ST} = \frac{\sigma_a^2+\sigma_b^2}{\sigma^2} = \frac{\sigma_a^2+\sigma_b^2}{\sigma_a^2+\sigma_b^2+\sigma_c^2}$$

RESULTS

$F_{ST}$is among the most widely used measures for genetic differentiation and plays a central role in ecological and evolutionary genetic studies. It is commonly thought that large sample sizes are required in order to precisely infer F_{ST} and that

TABLE 1: General design for hierarchical analysis of molecular variance (AMOVA).[21]

Source of Variation	Degrees of Freedom	Mean Squared Deviation	Expected Mean Squared Deviation
Among Groups/ Regions	$G-1$	$MSD(AG)$	$\sigma_c^2 + n'\sigma_b^2 + n''\sigma_a^2$
Among Populations within Groups/Regions	$\sum_{g=1}^{G}I_g - G$	$MSD(APWG)$	$\sigma_c^2 + n\sigma_b^2$
Among Individuals within Populations	$N - \sum_{g=1}^{G}I_g$	$MSD(WP)$	σ_c^2
Total	$N-1$		

AG: Among Groups; $APWG$: Among Populations Within Groups; WP: Within Populations; I_g: The number of individuals in gth group.

small sample sizes lead to overestimation of genetic differentiation.[22] Wright's F_{ST}, truly measures the differentiation due to demographic factors only (migration and subpopulation size), and can be interpreted as such and compared across studies. However, it is difficult to calculate from marker data when mutations are important.[2]

The most widely applied statistic is G_{ST}, proposed by Nei for measuring differentiation from multiallelic markers. The development and wide application of microsatellites have made G_{ST} ever more popular, but also its weakness more prominent. The high mutation rate and thus high polymorphism of microsatellites lead to a high within subpopulation heterozygosity, and thus a low G_{ST} because it is upper bounded by the average within subpopulation homozygosity. This is not a problem as long as the differentiation at the focal microsatellite loci is concerned; G_{ST} provides an unbiased measurement of the actual level of differentiation in allele frequency at these particular loci due to all evolutionary forces, including migration, drift, mutation and selection. Nei's G_{ST} measures the differentiation at a locus due to all evolutionary forces, including genetic drift, migration, selection and mutation. As a result, G_{ST} should be interpreted in terms of demographic factors only when mutation and selection are unimportant.[2]

Weir and Cockerham's method is based on the following assumptions. 1) Conceptually, there are infinitely many populations which are derived from the same ancestral population at the same time, and s populations are sampled from this ensemble population. All fixation indices are defined in terms of the ensemble population, and the fixation indices are estimated from genotype frequency data from the s populations sampled. 2) All populations are statistically independent, and no mutation, no migration, and no selection are assumed. However, certain types of migration and selection can be incorporated as long as the statistical independence is maintened. 3) The same population size is maintened for all populations and for all generations. Under this assumptions Weir and Cockerham's statistical method seems to be correct. However, the problem lies with the validity of the assumptions. Obviously, most natural populations do not satisfy any of these assumptions.[23]

Jost argued that the additive partitioning that is used for G_{ST}, where the total diversity is the sum of the within-population and among-population diversity, is inadequate to describe the among-population diversity. The second problem recognized by Jost is that the expected heterozygosity is an unsuitable metric for describing the diversity, leading to unintuitive results. A disadvantage of this diversity index (D) is that it depends on the sample size, so rarefaction to a standard sample size is needed before estimates can be compared.[24] Jost's D is not a proper measure of genetic differentiation. It is highly dependent on the initial gene diversity of the marker loci, is highly sensitive to how alleles and loci are defined and how data are analysed, does not always increase monotonically with divergence time and with drift, is highly dependent on the unknown parameter of the number of subpopulations. Rather, the extent of differentiation depends on the magnitude of difference in allele frequency, which is measured by F_{ST} and G_{ST} but not D.[2] When many alleles are found within populations, but few of them are shared among populations, F_{ST} and related indices greatly underestimate the level of population differentiation. D and $G_{ST}^{'}$ metrics have been proposed to overcome this problem.[25]

Hedrick's standardized genetic differentiation measure, based on G_{ST}, that is independent of the amount of genetic variation and therefore suitable for comparisons between studies that employ different genetic markers. However, Hedrick's standardized G_{ST}, $G_{ST}^{'}$, is based on Nei's original definition of G_{ST}, which has no bias correction for sample size or number of populations sampled. Therefore, estimates of Hedrick's $G_{ST}^{'}$ may also have a bias when calculated for small sample sizes or a small number of populations sizes.[4]

The AMOVA is a powerful statistical method for the description of factors influencing the structure of populations. AMOVA differs from analysis of variance in that it can accommodate different evolutionary assuptions without modifying the basic structure of the analysis, and in that hypotheses are tested using permutational methods so that normal distribution assumption is not required.[26] AMOVA is currently the most commonly reported analysis of genetic differentiation in the literature, because it incorporates more genetic information into the analysis and is viewed as superior to allele/haplotype based methods.[27]

G_{ST}, R_{ST}, Φ_{ST}, $G_{ST}^{'}$, D and φ_{ST} are the primary metrics utilized for empirically estimating and testing the magnitude of genetic divergence among populations. There is currently active discussion in the literature about which of these metrics are most appropriate for empirical surveys of genetic differentiation. All of these measures are highly sensitive to the diversity of alleles shared between populations. Overall, there is no single metric that best captures population genetic differentiation.[27]

REFERENCES

1. Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers. Mol Ecol 2002;11(2):155-65.
2. Wang J. On the measurements of genetic differentiation among populations. Genet Res (Camb) 2012;94(5):275-89.
3. Hedrick PW. A new approach to measuring genetic similarity. Evolution 1971;25(2):276-80.
4. Meirmans PG. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 2006; 60(11):2399-402.
5. Nagylaki T. Fixation indices in subdivided populations. Genetics 1998;148(3):1325-32.
6. Wright S. The genetical structure of populations. Ann Eugen 1951;15(1):323-54.
7. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet 2009;10(9):639-50.
8. Browning SR, Weir BS. Population structure with localized haplotype clusters. Genetics 2010;185(4):1337-44.
9. Liang M, Ya-Jie J, De-Xing Z. Statistical measures of genetic differentiation of populations: rationales, history and current states. Current Zoology 2015;61(5):886-97.
10. Nei M. F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 1977:41(2):225-33.
11. Nei M, Chesser RK. Estimation of fixation indices and gene diversities. Ann Hum Genet 1983;47(Pt 3):253-9.
12. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics 1995;139(1):457-62.
13. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984;38(6):1358-70.
14. Verity R, Nichols RA. What is genetic differentiation, and how should we measure it--GST, D, neither or both? Mol Ecol 2014;23(17):4216-25.
15. Hedrick, PW. A standardized genetic differentiation measure. Evolution 2005; 59(8):1633-8.
16. Jost L. GST and its relatives do not measure differentiation. Mol Ecol 2008:17(18); 4015-26.
17. Ryman N, Leimar O. GST is still a useful measure of genetic differentiation- a comment on Jost's D. Mol Ecol 2009; 18(10):2084-7.
18. Excoffier L. Analysis of population subdivision, analysis of molecular data: the AMOVA framework. In: Balding DJ, Bishop M, Canning C, eds. Handbook of Statistical Genetics, 3rd ed. Chichester: John Wiley & Sons; 2007. p.1001-4.
19. Kindt R, Muchugi A, Hansen OK, Kipruto H, Poole J, Dawson I, et al. Chapt. 8. Analysis of molecular variance (AMOVA). In: Dawson I, Hansen OK, eds. Molecular Markers for Tropical Trees: Statistical Analysis of Dominant Data. ICRAF Technical Manual No. 13. Nairobi: World Agroforestry Centre, Frederiksberg: Forest and Landscape Denmark; 2009. p.63-6.
20. Michalakis Y, Excoffier L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 1996;142(3):1061-4.
21. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 1992; 131(2):479-91.
22. Willing EM, Dreyer C, von Oosterhout C. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. PLoS One 2012;7(8): e42649.
23. Nei M. Definition and estimation of fixation indices. Evolution 1986;40(3):643-5.
24. Meirmans PG, Hedrick PW. Assessing population structure: F(ST) and related measures. Mol Ecol Resour 2011;11(1):5-18.
25. Mona S, Bertorelle G. Population differentiation: measures. eLS. Chichester: John Wiley & Sons Ltd; 2013. p.1-4.
26. Mengoni A, Bazzicalupo M. The statistical treatment of data and the **A**nalysis of **MO**lecular **VA**riance (AMOVA) in molecular microbial ecology. Ann Microbiol 2002;52(2):95-101.
27. Bird CE, Karl SA, Smouse PE, Toonen RJ. Detecting and measuring genetic differentiation. In: Held C, Koenemann S, Schubart CD, eds. Phylogeography and Population Genetics in Crustacea. 1st ed. Florida: CRC Press; 2011. p.31-56.