Genetic Distance Measures: Review

Article · April 2016

DOI: 10.5336/biostatic.2015-49517

CITATIONS
0

READS
708

2 authors:

Ismet Dogan
Afyon Kocatepe University
54 PUBLICATIONS **336** CITATIONS
SEE PROFILE

Nurhan Dogan
Afyon Kocatepe University
37 PUBLICATIONS **249** CITATIONS
SEE PROFILE

DERLEME REVIEW

Genetic Distance Measures: Review

Genetik Uzaklık Ölçüleri

İsmet DOĞAN,[a]
Nurhan DOĞAN[a]

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

Geliş Tarihi/*Received:* 30.12.2015
Kabul Tarihi/*Accepted:* 02.02.2016

Yazışma Adresi/*Correspondence:*
Nurhan DOĞAN
Afyon Kocatepe University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Afyonkarahisar, TÜRKİYE/TURKEY
nurhandogan@hotmail.com

ABSTRACT The major unsolved problem of descriptive population genetics is an adequate specification of the genetic difference between two closely related species as compared to the genetic difference between two populations of the same species. Traditional population genetic analyses deal with the distribution of allele frequencies between and within populations. From these frequencies several measures of population structure can be estimated, one of the most widely used being the genetic distance measures. Genetic distance is the degree of gene difference (genomic difference) between species or populations that is measured by some numerical method. Genetic distance measures have already been established as one of the major tools for analyzing data on gene differentiation between populations. Many genetic distances have been developed, of which a few remain in regular use. Each of these genetic distances has unique evolutionary and statistical properties, and evolutionary relationships inferred from each genetic distance can be quite different. Quantification of the genetic distance between populations is instrumental in many genetic research initiatives, and a large number of formulas for this purpose have been proposed. However, selection of an appropriate measure for assessing genetic distance between real-world human populations that diverged as a result of mechanisms that are not fully known can be a challenging task. In this study twenty six distance measures were investigated. For macroevolutionary comparisons, Nei's measures are probably the best. In microevolutionary studies, when sample sizes are approximately equal and the differences in gene frequency are great, Edward's E^2 is preferable. If sample sizes are quite variable and gene frequencies do not differ greatly, Sanghvi's G^2 would be most appropriate.

Key Words: Genetic distance; allel frequency; gene frequency

ÖZET Tanımlayıcı popülasyon genetiğinde henüz çözümlenememiş en önemli sorunlardan biri de aynı türe ait iki popülasyon arasındaki veya birbiri ile yakın ilişkili iki tür arasındaki genetik farklılığın yeteri kadar belirlenememesidir. Popülasyon genetiğinde kullanılan geleneksel analizler, çoğunlukla popülasyonlar içi veya popülasyonlar arası alel frekansların dağılımı ile ilgilenmektedir. Alel frekanslar kullanılarak popülasyon yapısı ile ilgili çeşitli ölçülere ait değerler tahmin edilebilmektedir. Değeri tahmin edilmek istenen en yaygın ölçülerden birisi de genetik uzaklık ölçüleridir. Genetik uzaklık, türler veya popülasyonlar arasındaki gen farklılığının (genomik farklılık) bazı sayısal yöntemlerle ölçülen derecesidir. Genetik uzaklık ölçüleri popülasyonlar arasındaki gen farklılığı ile ilgili verilerin analizi için geliştirilmiş en önemli araçlardan biridir. Düzenli olarak kullanılan az sayıda genetik uzaklık ölçüsü olmasına rağmen, geliştirilmiş çok sayıda uzaklık ölçüsü bulunmaktadır. Literatürde yer alan genetik uzaklık ölçülerinin her biri kendine ait istatistiksel ve gelişimsel özelliklere sahiptir. Gelişimsel özellikler bakımından genetik uzaklıklar, birbirlerinden oldukça farklılık göstermektedir. Genetik araştırmaların çoğunda popülasyonlar arasındaki genetik uzaklık bir gösterge olarak hesaplanmaktadır. Genetik uzaklığın hesaplanmasında kullanılan çok sayıda formül önerilmiştir. Ancak, tam olarak bilinmeyen mekanizmalardan dolayı ayrışan gerçek dünyadaki insan popülasyonları arasındaki genetik uzaklığın değerlendirilmesi için uygun ölçünün seçilmesi oldukça zor bir görevdir. Bu çalışmada yirmi altı farklı uzaklık ölçüsü incelenmiştir. Makro evrimsel karşılaştırmalar için Nei tarafından önerilen ölçülerin, örneklem genişliklerinin birbirine yakın ve gen frekanslarındaki farklılıkların büyük olduğu mikro evrimsel çalışmalarda Edward tarafından önerilen E^2 ölçüsünün, örneklem genişliklerinin değişken ve gen frekanslarındaki farklılıkların büyük olmadığı mikro evrimsel çalışmalarda ise Sanghvi tarafından önerilen G^2 ölçüsünün kullanılması önerilmektedir.

Anahtar Kelimeler: Genetik uzaklık; alel frekans, gen frekansı

doi: 10.5336/biostatic.2015-49517

Turkiye Klinikleri J Biostat 2016;8(1):87-93

A continuing problem in population biology is to estimate and explain genetic differences among the populations constituting a species. Such differences are generally measured as genetic distances.[1] Genetic distances have frequently been used in assessments of species status of closely related taxa and in a diversity of evolutionary studies. The objective measurement of the similarity between populations has concerned geneticists, taxonomists, anthropologists, and plant and animal breeders for a long time. Initially comparisons between populations were made on the basis of morphological measurements. More recently immunological and electrophoretic data, gene frequencies, and amino acid and DNA sequences have been used for such purposes.[2] Many questions of evolutionary interest require that genetic differences between populations be expressed as a single statistic, often called "genetic distance". Genetic distances are used, for example, to evaluate the degree of genetic differentiation achieved during the speciation process or at other stages of evolutionary divergence. Genetic distances also are used in the construction of phenograms or cladograms and have indeed provided valuable information for the reconstruction of phylogenetic history on the basis of extant species.[3] Studies of phylogenetic relationships among very closely related species are often hampered by a lack of variation. The estimation of relationships among such closely related taxa, or the estimation of relationships within a species, would be easier if faster evolving characters were used.[4] Traditional population genetic analyses deal with the distribution of allele frequencies between and within populations. From these frequencies several measures of population structure can be estimated, one of the most widely used being the genetic distance measures. Genetic distance is the degree of gene difference (genomic difference) between species or populations that is measured by some numerical method. Genetic distance measures have already been established as one of the major tools for analyzing data on gene differentiation between populations.[5]

The genetic composition of a population can often be described in terms of the frequencies, or relative abundances, in which alternative alleles are found. Usually, it is more convenient to analyze the data in terms of relative frequency than in terms of the observed numbers. For genotypes, the genotype frequency in a population is the proportion of organisms that have the particular genoype. For each allele, the allele frequency is the proportion of all alleles that are of the specified type. The concepts will be illustrated by using the human MN blood groups because this genetic systems is exceptionally simple. There are three possible phenotypes M, MN and N corresponding to the combination of M and N antigens that can be present on the surface of red blood cells. These antigens are unrelated to ABO and other red-cell antigens. The M, MN and N phenotypes correspond to three genotypes of one gene: MM, MN and NN, respectively. In a study of a British population, a sample of 1000 people yielded 298 M, 489 MN and 213 N phenotypes. From the one to one correspondence between genotype and phenotype in this system, the genotypes can be directly inferred to be 298 MM, 489 MN and 213 NN. M and N alleles break down in the following way:

$$298\ MM\ persons = 596\ M\ alleles,$$

$$489\ MN\ persons = 489\ M\ alleles + 489\ N\ alleles,$$

$$213\ NN\ persons = 426\ N\ alleles,$$

$$Totals = 1085\ M\ alleles + 915\ N\ alleles$$

In the MN example, the genotype frequencies are obtained by dividing the observed numbers by the total sample size, in this case 1000. Therefore, the genotype frequencies are 0.298 MM, 0.489 MN and 0.213 NN. Similarly, the allele frequencies are obtained by dividing the observed number of each allele by

the total number of alleles, in this case 2000, so;

$$Allele\ Frequency\ of\ M = 1085/2000 = 0.5425$$
$$Allele\ Frequency\ of\ N = 915/2000 = 0.4575$$

Allele and genotype frequencies must always be between 0 and 1. Allele frequencies are often more useful than genotype frequencies because genes, not genotypes, form the bridge between generations. [6] A large number of formulas for quantifying genetic distance that consider allele frequency differences between two populations have been proposed. These formulas are derived from and are based on a variety of assumptions, which may assist researchers in their choice of a measure as well as in the interpretation of results from analyses that use a specific measure.[7]

Gene frequencies and hence genetic variability tend to remain unchanged in such a population, generation after generation, because of the persistence of genes and the symmetry of the Mendelian mechanism.[8] Analysis of genotypic data from neutral loci is an important method for describing the patterns of genetic variation within species and inferring the evolutionary processes that give rise to those patterns. Genotypic data are notoriously multivariate: the frequency of each allele at each locus is usually different in each population. Genetic distances are metrics that summarize these differences in an overall measure of differentiation for a pair of populations.[9]

The general genetic distance between two taxa is a distance between the sets of DNA-related data chosen to represent them. Many genetic distances have been developed to summarize allele frequency differences between populations. This paper is intended to be a review of the genetic distance literature that population geneticists frequently use. For this purpose twenty six distance measures were investigated.

A number of proposed measures of genetic similarity and distance measures can be estimated as follows: a *population* is represented by a double-indexed vector $x = (x_{ij})$ with $\sum_{j=1}^{n} m_j$ components, where x_{ij} is the frequency of the *i*-th *allele* (the label for a state of a gene) at the *j*-th gene locus (the position of a gene on a chromosome), m_j is the number of alleles at the *j*-th locus, and r is the number of considered loci. Since x_{ij} is the frequency, we have $x_{ij} \geq 0$ and $\sum_{i=1}^{m_j} x_{ij} = 1$. Denote by Σ summation over all i and j.

NEI STANDART GENETIC DISTANCE[7,10]

$$N_s = -ln\left[\frac{\sum_j \sum_i x_{ij} y_{ij}}{(\sum_j \sum_i x_{ij}^2 \sum_j \sum_i y_{ij}^2)^{1/2}}\right]$$

NEI STANDART GENETIC DISTANCE WITH BIAS CORRECTION[7,11]

$$N_b = -ln\left[\frac{(2\bar{n}-1)\sum_j \sum_i x_{ij} y_{ij}}{\left[\sum_j (2\bar{n}\sum_i x_{ij}^2 - 1)\sum_j (2\bar{n}\sum_i y_{ij}^2 - 1)\right]^{1/2}}\right]$$

$$\bar{n} = \sum_{i=1}^{2} n_i - \frac{\sum_i n_i^2}{\sum_i n_i}$$

$\bar{n}$ is the average number of individuals.

NEI MINIMUM DISTANCE[12]

$$N_m = \frac{1}{2r}\sum (x_{ij} - y_{ij})^2$$

NEI GEOMETRIC DISTANCE[7,10]

$$N_g = 1 - \frac{1}{r}\sum_{l=1}^{r}\sum_{u=1}^{v}(p_{lu1}p_{lu2})^{1/2}$$

p_{lu1} and p_{lu2} are the number of individuals that carry allele u at locus l in populations 1 and 2 respectively.

LATTER'S DISTANCE[7,13]

$$La = \frac{\frac{1}{2}\sum_j \sum_i x_{ij} \sum_j \sum_i y_{ij} - \sum_j \sum_i x_{ij} y_{ij}}{1 - \sum_j \sum_i x_{ij} y_{ij}}$$

CAVALLI-SFORZA-EDWARDS CHORD DISTANCE[7,14]

$$CE = \frac{2}{\pi r}\sum_{l=1}^{r}\left\{2\left[1 - \sum_{u=1}^{m_j}(p_{lu1}p_{lu2})^{1/2}\right]\right\}^{1/2}$$

p_{lu1} and p_{lu2} are the number of individuals that carry allele u at locus l in populations 1 and 2 respectively.

EDWARDS DISTANCE [15]

$$E^2 = \frac{8\left[1 - \sum_{i=1}^{k}\sqrt{X_i Y_i}\right]}{\left[1 + \sum_{i=1}^{k}\sqrt{(X_i/k)}\right]\left[1 + \sum_{i=1}^{k}\sqrt{(Y_i/k)}\right]}$$

X_i and Y_i are the gene frequencies of the ith allele at a k-allelic locus in each of two populations.

SANGHVI DISTANCE [16]

$$G^2 = 100\frac{\sum_{j=1}^{r}\sum_{k=1}^{s_j+1}\left[\frac{\left(p_{1jk} - p_{jk}\right)^2}{p_{jk}} + \frac{\left(p_{2jk} - p_{jk}\right)^2}{p_{jk}}\right]}{\sum_{j=1}^{r} s_j}$$

$$p_{jk} = \frac{\left(p_{1jk} + p_{2jk}\right)}{2}$$

p_{1jk} and p_{2jk} are the proportions in the kth class for the jth character in populations P_1 and P_2 respectively and $s_j + 1$ is the number of classes for the jth character.

BALAKRISHNAN-SANGHVI DISTANCES[17]

The proportions p_{ijk} are the maximum likelihood estimates of the proportions of the kth class, $k = 1,2,\dots,s_j + 1$, of the jth character S_j, $j = 1,2,\dots,r$, in the ith population P_i, $i = 1,2,\dots,q$. The dispersion matrix of the estimates p_{ijk} is given by A_{ij}, where

$$A_{ijk} = p_{ijk}\left(1 - p_{ijk}\right)/n_{ij}, \quad k = l,$$
$$= -p_{ijk}p_{ijl}/n_{ij} \qquad k \neq l,$$
$$k, l = 1,2,\dots,s_j + 1,$$

and $n_{ij} = sample\ size$ for S_j from P_i. Since $\sum_{k=1}^{s_j+1} p_{ijk} = 1$, the rows and columns of A_{ij} add up to zero. The common dispersion matrix of the variables X_{jk} over q populations is estimated by C_j, where

$$C_{jkl} = \sum_{i=1}^{q} n_{ij}^2 A_{ijkl} / \sum_{i=1}^{q} n_{ij}$$

The distance between two populations P_m and P_n,

$$B_{mn}^2 = \sum_{j=1}^{r}\sum_{k=1}^{s_j}\sum_{l=1}^{s_j} C_j^{kl} d_{jk} d_{jl}$$

$$d_{jk} = p_{mjk} - p_{njk}$$
$$C_j^{kl} = C_j^{-1}$$

Another measure for the distance,

$$G_c^2 = \sum_{j=1}^{r}\sum_{k=1}^{s_j}\sum_{l=1}^{s_j} C_j^{kl} d_{jk} d_{jl}$$

$$C_{jkl} = p_{jk}\left(1 - p_{jk}\right), \quad k = l,$$
$$= -p_{jk}p_{jl} \qquad k \neq l,$$
$$k, l = 1,2,\dots,s_j + 1$$

$$p_{jk} = \sum_{i=1}^{q} n_{ij}p_{ijk} \Big/ \sum_{i=1}^{q} n_{ij}$$

STEINBERG DISTANCE[17]

$$D_k^2 = \sum_{j=1}^{n}\sum_{k=1}^{s_j}\sum_{l=1}^{s_j} A_j^{kl} d_{jk} d_{jl}$$

$$d_{jk} = p_{1jk} - p_{2jk}$$
$$A_j^{kl} = \left(A_{jkl}\right)^{-1}$$

$$A_{jkl} = p_{jk}\left(1 - p_{jk}\right), \quad k = l,$$
$$= -p_{jk}p_{jl} \qquad k \neq l,$$
$$k, l = 1,2,\dots,s_j$$

$$p_{jk} = \frac{1}{2}\left(p_{1jk} + p_{2jk}\right)$$

CAVALLI-SFORZA-BODMER DISTANCE[18]

$$f = \frac{4\sum_j\left[1 - \sum_i\left(x_{ij}y_{ij}\right)^{1/2}\right]}{\sum_j\left(m_j - 1\right)}$$

$$Distance = -ln(1 - f)$$

ROGERS' DISTANCE[19]

$$D_R = \frac{1}{n}\sum_j\left[\frac{1}{2}\sum_i\left(x_{ij} - y_{ij}\right)^2\right]^{1/2}$$

REYNOLDS-WEIR-COCKERHAM DISTANCE[7,20]

$$Re = \frac{\sum_j\left\{\frac{1}{2}\sum_i\left(x_{ij} - y_{ij}\right)^2 - \frac{1}{2(2\bar{n} - 1)}\left[2 - \sum_i x_{ij}^2 + \sum_i y_{ij}^2\right]\right\}}{\sum_j\left(1 - \sum_i x_{ij}y_{ij}\right)}$$

GOLDSTEIN-LINARES-CAVALLI-SFORZA-FELDMAN DISTANCE[4]

$$(\delta\mu)^2 = \frac{\sum_{j=1}^{r}\left(\mu_{X_j} - \mu_{Y_j}\right)^2}{r}$$

$$\mu_{X_j} = \sum_i i\, x_{ij}$$

$$\mu_{Y_j} = \sum_i i\, y_{ij}$$

FELDMAN-BERGMAN-POLLOCK-GOLDSTEIN DISTANCE[21]

$$D_1 = log\left(1 - \frac{\sum_i (\delta\mu)_i^2}{M}\right)$$

M is the average value of the distance at maximal divergence.

SHRIVER-JIN-BOERWINKLE-DEKA-FERRELL-CHAKRABORTY DISTANCE[22]

$$D_{SW} = \frac{1}{r}\sum_{k=1}^{r}\sum_{l\le i,j\le m_k} |i-j|\left(2x_{ik}y_{jk} - x_{ik}x_{jk} - y_{ik}y_{jk}\right)$$

HEDRICK SIMILARITY COEFFICIENT [23]

$$I_{x.y} = \frac{\sum_{j=1}^{n_j} p_{j.x}p_{j.y}}{\frac{1}{2}\left(\sum_{j=1}^{n} p_{j.x}^2 + \sum_{j=1}^{n} p_{j.y}^2\right)}$$

n_j *is the number of genotypes at the locus.*

$p_{j.x}$ and $p_{j.y}$ are the frequencies of the jth genotype in populations X and Y respectively.

PREVOSTI-OCANA-ALONSO DISTANCE[24]

$$D_P = \frac{\sum\left|x_{ij} - y_{ij}\right|}{2r}$$

NEI-TAJIMA-TATENO DISTANCE[25]

$$D_A = \frac{1}{r}\sum_{j=1}^{r}\left(1 - \sum_{i=1}^{m_j}\sqrt{x_{ij}y_{ij}}\right)$$

BHATTACHARYYA DISTANCE[26]

$$\theta^2 = \left\{arccos\left[\sum_{i,j}\sqrt{x_{ij}y_{ij}}\right]\right\}^2$$

TOMIUK-LOESCHCKE DISTANCE[27]

$$D_{TL} = -\ln\left(I_{TL}\right)$$

$$I_{TL} = \frac{\sum_{i,j} x_{ij}y_{ij}}{r}\frac{\sum_{i,j} y_{ij}^x}{r}$$

where x_{ij}^y and y_{ij}^x are the frequencies of the ith allele at the jth locus in the populations X and Y, respectively, that are also present in their sister population. Distance can also be calculated using the following formula,[28]

$$D_{TL} = -ln\frac{1}{r}\sqrt{\sum x_{ij}\sum y_{ij}}$$

CAVALLI-SFORZA ARC DISTANCE [28]

$$D = \frac{2}{\pi}arccos\left(\sum\sqrt{x_{ij}y_{ij}}\right)$$

THORPE (DPS) DISTANCE [28]

$$Dps = -ln\frac{\sum min\left(x_{ij}, y_{ij}\right)}{\sum_{j=1}^{n} m_j}$$

AVERAGE SQUARE DISTANCE[28]

$$ASD = \frac{1}{r}\sum_{k=1}^{r}\left[\sum_{1\le i\le j\le m_j}(i-j)^2 x_{ik}y_{jk}\right]$$

SHARED ALLELE DISTANCE[28]

$$D_{SA} = \frac{\sum_{j=1}^{r} m_j}{2r}$$

FUZZY SET DISTANCE[28]

$$D_{Fuzzy} = \frac{\sum 1_{x_{ij}\neq y_{ij}}}{\sum_{j=1}^{r} m_j}$$

CONCLUSION

Several analogous measures have been developed to describe differentiation between populations. These measures have different mathematical foundations and represent distinct but related concepts. The relative merits of each measure have not been resolved, but, in practice, estimates of these statistics are generally similar.[9]

Using genetic distances as yardsticks for species limits is not only a problematic issue regarding cut-off values. A more basic question concerns the correction of distances to account for multiple substitutions at certain sites. In phylogenetic analyses, model selection is considered important as the substitution model always influences branch lengths and may consequently affect the tree topology. The improper use of uncorrected as well as under-corrected distances will lead to underestimation of the actual differences between long separated taxa. Another basic issue concerns the comparability of genetic distances. It is well known that different loci have different mean rates of evolution, and according-

ly levels of divergence between taxa depend on the loci being compared. However, this is often neglected, as studies based on different loci are frequently indiscriminately compared. Also different parts of the same gene are known to have different evolutionary rates this may confound comparisons based on different gene fragments, or overlapping fragments of unequal lengths even when they are from the same locus.[29]

No study, however, has conclusively shown that all genetic distance measures lead to the same conclusion about, for example, the ordering of populations based on their genetic similarity. In any event, most measures of genetic distance are highly correlated under most conditions. A large number of studies have reported high correlations among different distance measures used on the same set of data.[30]

With regard to the "best" genetic distance measure, it should be apparent that selection of an appropriate measure depends upon the characteristics of the data and upon one's theoretical perspective. For macroevolutionary comparisons, Nei's measures are probably the best, since they give the best estimates of divergence time and are related to a well-defined biological process-codon substitution. In microevolutionary studies, when sample sizes are approximately equal and the differences in gene frequency are great, Edward's E^2 is preferable. If sample sizes are quite variable and gene frequencies do not differ greatly, Sanghvi's G^2 would be most appropriate.[30]

REFERENCES

1. Sokal RR. Genetic, geographic, and linguistic distances in Europe. Proc Natl Acad Sci U S A 1988;85(5):1722-6.
2. Goodman MM. Genetic distance: measuring dissimilarity among populations. Yearbook of Physical Anthropology 1974;17:1-38.
3. Mueller LD, Ayala FJ. Estimation and interpretation of genetic distance in empirical studies. Genet Res 1982;40(2):127-37.
4. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. An evaluation of genetic distances for use with microsatellite loci. Genetics 1995;139(1):463-71.
5. Chakraborty R, Tateno Y. Correlations between some measures of genetic distance. Evolution 1976;30(4):851-3.
6. Hartl DL, Jones EW. Allele frequencies and genotype frequencies. Genetics: Principles and Analysis. 4th ed. Sudbury, Massachusetts: Jones and Bartlett Publishers Inc.; 1998. p.628-9.
7. Libiger O, Nievergelt CM, Schork NJ. Comparison of genetic distance measures using human SNP genotype data. Hum Biol 2009;81(4):389-406.
8. Wright S. The genetical structure of populations. Ann Eugen 1951;15(4):323-54.
9. Kalinowski ST. Evolutionary and statistical properties of three genetic distances. Mol Ecol 2002;11(8):1263-73.
10. Nei M. Genetic distance between populations. Am Nat 1972;106(949):283-92.
11. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89(3):583-90.
12. Nei M. Genetic distance and molecular phylogeny. In: Ryman N, Utter F, eds. Population Genetics and Fishery Management. 1st ed. Seattle, WA: University of Washington Press; 1987. p.193-223.
13. Latter BD. Selection in finite populations with multiple alleles. 3. Genetic divergence with centripetal selection and mutation. Genetics 1972;70(3):475-90.
14. Cavalli-Sforza LL, Edwards WF. Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 1967;19(3 Pt 1):233-57.
15. Edwards WF. Distances between populations on the basis of gene frequencies. Biometrics 1971;27(4):873-81.
16. Sanghvi LD. Comparison of genetical and morphological methods for a study of biological differences. Am J Phys Anthropol 1953;11(3):385-404.
17. Balakrishnan V, Sanghvi LD. Distance between populations on the basis of attribute data. Biometrics 1968;24(4):859-65.
18. Cavalli-Sforza LL, Bodmer WF. Population structure. The Genetics of Human Populations. 1st ed. San Francisco: WH Freeman and Company; 1971. p.387-490.
19. Rogers JS. Measures of genetic similarity and genetic distance. In: Wheeler MR, ed. Studies in Genetics VII. 1st ed. Austin, Texas: University of Texas Publication; 1972. p.145-53.
20. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 1983;105(3):767-79.
21. Feldman MW, Bergman A, Pollock DD, Goldstein DB. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics 1997;145(1):207-16.
22. Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R. A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 1995;12(5):914-20.
23. Hedrick PW. A new approach to measuring genetic similarity. Evoluation 1971;25(2):276-80.
24. Prevosti A, Ocaña J, Alonso G. Distances between populations of Drosophila subobscura, based on chromosome arrangement frequencies. Theor Appl Genet 1975;45(6):231-41.
25. Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequ-ency data. J Mol Evol 1983; 19(2):153-70.

26. Bhattacharyya A. On a measure of divergence between two multinomial populations. Sankhyā: The Indian Journal of Statistics 1946;7(4):401-6.

27. Tomiuk J, Guldbrandtsen B, Loeschcke V. Population differentiation through mutation and drift--a comparison of genetic identity measures. Genetica 1998;102/103(1-6):545-58.

28. Deza MM, Deza E. Distances in natural sciences, distances in biology, genetic distances. Encyclopedia of Distances. 3rd ed. Heidelberg: Springer; 2014. p.435-8.

29. Fregin S, Haase M, Olsson U, Alström P. Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae. Mol Phylogenet Evol 2012;62(1):319-28.

30. Jorde LB. Analytical theory, the genetic structure of subdivided human populations, methods of genetic analysis, genetic distance measures. In: Mielke JH, Crawford MH, eds. Current Developments in Anthropological Genetics: Theory and Methods. 1st ed. New York: Plenum Press; 1980. p.137-44.