OALib

A Fundamental Statistical Tools Application for Livestock Diversity Studies from Microsatellite Data—A Mini Review

Nedup Dorji[1], Monchai Daugjinda[2]

[1]Department of Animal Science, College of Natural Resources, Royal University of Bhutan, Thimphu, Bhutan
[2]Department of Animal Science, Khon Kaen University, Khon Kaen, Thailand
Email: nedup@cnr.edu.bt

Abstract

The conservation of genetic variation is essential for meeting future unpredictable breeding requirements. Therefore, the foremost priority must be given to assessment of the population variation status. Among several markers, a DNA based microsatellite is a choice for the population studies, currently. We reviewed some important statistical tools for assessing the farm animal biodiversity. Firstly, the reliability of the results which is mainly based on the number of microsatellite loci genotyped and unrelated sample size per population was discussed. Secondly, we argued genetic parameters which are estimated while reporting the state of population variations. Thirdly, supplementing information to variations was also presented. Fourthly, we briefly discussed the types and constructing phylogenetic tree and the role of principal component analysis. Finally, we presented the important web-based statistical packages for genetic data analysis. This shall assist in devising appropriate conservation strategy for our farm animal genetic resources.

Keywords

1. Introduction

DNA markers can be easily detected using acrylamide gel technique and they have been a common tool for population studies. The DNA maker should have ideal features such as neutral, co-dominant multallelic, high

heterozygosity, evenly distributed throughout the genome and species specific. For last few decades, DNA microsatellite has been popularly used to assess farm animal biodiversity status [1]. Almost all the base maps for farm animals are well documented [2] adding greater value for the population variation studies. Microsatellites is also called as simple tandem repeats or simple sequence repeats because they are perfect or nearly perfect repeats and the tandem nucleotide repeat may be mono-, di-, tri-, penta- and hexa-nucleotides [3].

The varied farm animal breeds that we are familiar today are because of evolution and selection to suit wide range of agro-ecological conditions and human needs [4]. Genetic variations can be defined as diversity of alleles and genotypes present in a population [5]. The variation between individuals, populations and breeds represent the genetic diversity of animal species [6]. However, many breeds of farm animals have either gone extinct or threatened to extinction. For example, inbreeding depression has been documented for Danish sheep [7], Holestian cattle [8], pig [9], purebred Hampshire and Duroc swine [10], White Leghorn [11] and Spanish purebred horses [12]. The endangered population should be protected by conserving variations [9] for future unpredictable breeding requirements [1] [5] [13] [14]. Thus, genetic characterization of the breeds is important to apply in devising appropriate breeding and conservation strategies of farm animal genetic resources [1] [15]. The approach involves elucidating genetic variations, genetic differentiation and genetic distance within and among different populations. Genetic uniqueness should not be the only criterion for conservation reason. The population must be evaluated data based on demography, performance recording and breed description [16], economic value, adaptive characteristics and available of local raw materials [14].

Several studies have been carried out to report and address the status of biodiversity of farm animal species for decades irrelevant of the country's economy. The diversity studies are supported mainly by Food and Agriculture Organization (FAO). In 1995, the International Society for Animal Genetics (ISAG) based FAO was formed as the advisory group for animal genetic diversity studies. In 2004, [15] listed recommended marker for population variation study (available at http://www.fao.org/docrep/meeting/021/j1998e.pdf). More importantly, FAO launched a global internet-based database (domestic animal diversity information service, available at http://dad.fao.org/) to disseminate the available animal genetic resources and to form the working network of geneticists, conservationists and researchers. Therefore, we gathered some essential information associated with population diversity studies in farm animals focusing in developing countries.

2. Population Sample Size and Number of Loci Genotyping

The reliability of data derived depends on the sample size and the number of microsatellite loci genotyped (**Table 1**). FAO/ISAG advisory group on animal genetic variation suggested at least 30 individuals per population should be sampled for population studies [1]. Later FAO recommended sample size minimum of 25 unrelated animals per breed [1] [17] and later to 50 animals per breed [1] for estimating the allele frequency accurately. But, [18] suggested based on their similar findings as [19] that about 20 microsatellites with 24 unrelated samples per population is adequate for the study.

3. Parameters for Diversity Studies

3.1. Microsatellite Allele Diversity and Heterozygosity

The common parameters used to examine the genetic variations within a population are allele diversity (also called as allele richness and is expressed as mean number of alleles per locus, MNA), observed and expected frequency of heterozygosity [20]-[22]. The total number of alleles (TNA) from the populations within the same species forms a collection of gene, called gene pool. TNA are scored simply by counting the number of bands after electrophoresis. The bands are alleles which separated scored based on their molecular weight. When the study involves in identifying the populations and intended for conservation [22] [23] [32], the allele diversity examination is a fundamental form of measuring the genetic variations [17] [20] [22] [23]. The MNA may be estimated as:

$$\text{MNA} = \frac{1}{n}\sum_{i=1}^{n} k_i \; .$$

where n, total number of loci; k_i, total number of alleles detected in a locus.

Observed frequency of heterozygosity (H_O) and expected frequency of heterozygosity (H_E) also called as gene

Table 1. Preferences of markers, sample sizes, phylogenetic tree for previously chicken genetic diversity study.

Chicken		Markers		Tree	Reference
Populations	Sample sizes	Types	No. of loci genotyped		
52 populations including Red Jungle Fowl	50	Microsatellite	22	-	[6]
Nine Japanese chicken breeds	24	Microsatellite	20	NJ	[19]
Nine native Japanese long tailed chicken	35 - 48	Microsatellite	40	NJ	[18]
Thai native chicken	30	Microsatellite	20	NJ	[13]

NJ: neighbour-joining method; UPGMA: unweighted pair group method with arithmetic mean.

diversity [24] are important parameters in representing and extending the population variations [5] [20] because the MNA depends on sample size [23] and sampling method. H_O and H_E [24] is analyzed as:

$$H_O = \frac{\text{Direct count of heterozygote}}{\text{Total number of individuals surveyed}}$$

$$H_E = 1 - \sum_{i=1}^{k} p_i^2 ;$$

where p_i, frequency of allele i among k alleles in the analyzed population.

While H_E is estimated heterozygosity for the population based on the allele frequency of the sampled individuals given that the population is in Hardy Weinberg Equilibrium (HWE). Later is usually reported to describe the population variation [20]. When the heterozygosity value is high, it represents greater genetic variability and *vice versa*.

Allele frequency is the presence of relative allele at the locus in population which is expressed either in percentage or in proportion. Larger inclusion of the animal sample sizes, more the allele frequency accuracy can be drawn [23]. This estimated allele frequency of the loci is purposed for determining genetic variation and genetic distance. The frequency of let say A_1 allele is designated by p_i and can be estimated as:

$$p_i = \frac{\left(2N_{ii} + \sum_{i \neq j} N_{ij} \right)}{2N} ;$$

where N_{ii}, number of the A_{ii} homozygote; N_{ij}, number of the A_{ij} heterozygote and $2N$, total alleles in a population.

The relationship between H_O and H_E (for autosomes) is very important in determining the breeding system; when $H_O = H_E$, the population are more likely to be in random mating, $H_O < H_E$, the mode of mating system in a population is said to inbreeding and when $H_E < H_O$, the population may not be under inbreeding.

Nevertheless, heterozygosity as well as MNA should be considered for indicating the genetic variation because heterozygosity overestimates in presence of large allele numbers and/or small population [25].

3.2. Hardy Weinberg Equilibrium

Testing Hardy Weinberg Equilibrium (HWE) and estimating fixation indices are parameters supplementing to investigate the population variation. The Hardy Weinberg law defines a direct relationship between allele and genotypic proportions in a population [26]. This law states that for all loci except for sex chromosomes in diploid organisms, HWE is attained from one to next generation with the underlying conditions that the population is not under any gene force (non-random mating, selection for genotypes, migration and mutations), infinite population size and equal fertility of parent genotypes [20]. For example, gene force may result to either decrease or increase in heterozygotes of the population [27].

Goodness of fit such as chi square ratio (χ^2) is commonly used to test if the population is in HWE for its simplicity which is based mainly on sample sizes [26] [27] and allele numbers [26] [28].

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \; ;$$

when O_{ij}, observed frequency; E_{ij}, expected frequency.

The degree of freedom for multiallele such as microsatellites is determined as:

$$df = \frac{k\left(k-1\right)}{2}$$

for k^{th} number of alleles detected.

Whenever calculated chi square is less or equal to the tabular chi square value, the null hypothesis is accepted, which means allele frequencies for a given locus are in HWE. If chi square estimate is greater than the tabular value, the locus is said to be deviating from HWE which could be explained by either or combination of gene forces.

3.3. Wright Fixation Indices

In 1951 [29] introduced Wright's fixation coefficient of an individual within a subpopulation (F_{IS}), fixation co-efficient of an individual within the total population (F_{IT}) and fixation coefficient of a subpopulation within the total population (F_{ST}). Wright's coefficients are used to analyze the degree of subpopulation division and breeding system when two gametes are drawn randomly from each subpopulation [29]. The coefficients are interpreted based on allele frequencies distribution among the populations [30].

F_{IS} (also called as inbreeding coefficient) indicates if individuals of the studied subpopulation are under non-random mating system. In other words, F_{IS} denotes the degree of HWE departure in subpopulations due to genetic inbreeding [17]. The F_{IS} value ranged from −1 to 1 for maximum outbreeding (excess heterozygote) and inbreeding (deficient heterozygote), respectively. It may be calculated by:

$$F_{IS} = \frac{\left(H_S - H_I\right)}{H_S}$$

where H_S, H_E of mean from subpopulations; H_I, H_O mean of subpopulation.

Fixation index (F_{ST}) is the most preferred index for examining the degree of genetic divergence (or differentiation) in populations based on allele frequency [17] [21] [30] [31] and to check the existence of gene flow [32]. Moreover, F_{ST} is mainly used for detecting the effect of genetic drift relative to other evolutionary forces [27]. Subpopulation division always results in heterozygotes deficiency therefore, F_{ST} ranges from 0 to 1 (always positive) [17]. When there is no subpopulation division, it is assumed to have identical allele frequency and fixation of different alleles and the F_{ST} value will be 0. Based on the value estimated, the degree of subpopulation may be hesitantly stated as: beginning (0 to <0.05), moderate (0.05 to <0.15), strong (0.15 ≤ 0.25), very strong (>0.25) or different species (1) [17]. F_{ST} is estimated by [33] [34] as;

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where, H_T, H_E overall population; H_S, H_E of mean from subpopulation.

F_{IT} is rarely used as to extent HWE deviation in the total population. Besides, the value may indicate maximum outbreeding and inbreeding for the value −1 to 1, respectively. It may be estimated as:

$$F_{IT} = \frac{\left(H_T - H_I\right)}{H_T}$$

where, H_T, H_E overall population; H_I, H_O of mean from subpopulation.

3.4. Genetic Relationships and Phylogenetic Tree

Genetic distance determines the gene differences between two populations [35] reflecting the degree of closeness of gene frequency. When the allele frequencies are similar the genetic distance approaches to zero and *vice*

versa [20]. Although, the estimation of genetic distance is based on allele frequencies, free availability of software has influenced the application of estimation [1]. There are various methods to measure the genetic distance for known allele frequencies providing interesting data [36]. [31] reviewed and suggested three genetic distances; the Nei's standard genetic distance [37], Nei *et al.* chord genetic distance [38] and Weir and Cockerham [39] for their popularity and valuable properties. Similarly, [1] reported that three genetic distances used commonly in FAO project for reporting status of biodiversity was Nei's standard genetic distance, Nei *et al.* chord distance and Reynolds genetic distance with 74%, 53% and 30%, respectively of the total survey.

Nei's standard genetic distance was suitable for reconstructing phyologenetic tree while Nei *et al*. chord distance for quantification of branch length [1]. However, for very closely related populations, Reynolds distance may be suitable [40]. The most commonly used to measure the genetic relatedness is Nei's unbiased genetic distance which is likely to infer the genetic distance for the relative time divergence for neutral mutation model and it is calculated as [41]:

$$D = -\text{In}\left[\frac{\sum_m \sum_i p_1 m_i p_2 m_i}{\left(\sum_m \sum_i p_1 m_i^2\right)^{1/2} \left(\sum_m \sum_i p_2 m_1^2\right)^{1/2}}\right]$$

where; m, total loci; I, number of alleles at m^{th} locus; $p_1 m_i$, frequency of the ith allele at the mth locus in population one.

Dendrogram using information from morphology, fossils, biochemical and genotypic data might be reconstructed. Dendrogram is a common method of representing the relatedness of genes in studied populations.

There are several methods for reconstructing the phylogenetic tree namely: Unweighted Pair Group Method with Arithmetic mean (UPGMA), Neighbour Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML). FAO survey revealed about 86% of the study used NJ method and least for ML for animal diversity [1] revealing the popular of the NJ method in reconstruction of a phylogenetic tree. The first two methods are called distance matrix methods that measures the distance between all of the sequence pairs of genes, individuals, population, species and tax as under examination [28] therefore, multiple sequence alignment as a data input is necessary. NJ does not assume constant evolution unlike UPGMA, it is popularly used than UPGMA [28]. Moreover, NJ method was favoured by various earlier investigators to represent the general relationships among the chicken populations (**Table 1**).

The third and fourth methods are character state method providing information about attributes of genes, individuals, populations and species [28]. MP involves the identification of minimum number of evolutionary events to explain the differences among the groups of phylogenetic tree and it was used first for the morphological study [42]. MP is applicable when there is relatively small number of sequence and species. However, it does not account any assumption [42]. The common method of ML for the phylogenetic inference was first suggested by Cavalli-Sforza and Edwards in 1967 and its application is widely accepted recently because it is based on evolutionary model unlike other biological sequences [43] based on substitutions, single-base insertions and single base deletions but it demands high computation [28].

Unlike genetic distance by population, principal component analysis (PCA) is based on individual genetic distance. PCA is powerful statistical technique [44] used to visualize the correct stratification and NJ clustering [45] as several studies indicated that the phylogenetic tree may not be clear. Individual Dice's genetic distance is preferred over Jaccard coefficient because former puts more weight to shared bands than the later [17].

4. Web-Based Data Analysis Software

There are many online software programs available for the population genetic data analysis. The GENEPOP [46] [47] (available at http://genepop.curtin.edu.au/) is popularly used to estimate heterozygosity, HWE, linkage disequilibrium and population differentiation. Moreover, Arlequin (available at http://cmpg.unibe.ch/software/arlequin3/) and POPGENE (available at http://www.ualberta.ca/~fyeh/popgene.html) perform heterozogsity, HWE and Nei distance estimation [17].

Although, web-based software packages are available for the genetic distance analysis and construction of phylogenetic, the free available determines their usage especially in developing countries. For example, [1] re-

ported about 67% of the total survey used PHYLIP (available at http://evolution.genetics.washington.edu/phylip.html) followed by DISPAN (available at http://evolution.genetics.washington.edu/phylip/software.dist.html). Furthermore, STRUCTURE (available at http://pritchardlab.stanford.edu/publications.html) is a free statistical package becoming more common to determine population structure.

5. Conclusion

Based on the past researches, one may generalize that the use of statistical tools for population variation studies is dependent on its convenience and the availability of statistical software. The sampling size and genotyping method is also another important criterion while requiring for the characterization of farm animals. There are many web-based statistical packages for genetic data analysis but, the free available decides their usage. However, it is believed that Single Nucleotide Polymorphisms (SNP) to be future suitable marker for the diversity studies.

References

[1] Baumung, R., Simianer, H. and Hoffmann, I. (2004) Genetic Diversity Studies in Farm Animals—A Survey. *Journal of Animal Breeding and Genetics*, **121**, 361-373. http://dx.doi.org/10.1111/j.1439-0388.2004.00479.x

[2] Soller, M., Weigend, S., Romanov, M.N., Dekkers, J.C.M. and Lamont, S.J. (2006) Strategies to Assess Structural Variation in the Chicken Genome and Its Association with Biodiversity and Biological Performances. *Poultry Science*, **85**, 2061-2078. http://dx.doi.org/10.1093/ps/85.12.2061

[3] Ellagra, H. (2004) Microsatellites: Simple Sequences with Complex Evolution. *Nature Review Genetics*, **5**, 435-445. http://dx.doi.org/10.1038/nrg1348

[4] Muchadeyi, F.G., Wollny, C.B.A., Eding, H., Weigend, S., Makuza, S.M. and Simianer, H. (2007) Variation in Village Chicken Production Systems among Agro-Ecological Zones of Zimbabwe. *Tropical Animal Health and Production*, **39**, 413-461. http://dx.doi.org/10.1007/s11250-007-9050-0

[5] Toro, M.A. and Caballero, A. (2005) Characterization and Conservation of Genetic Diversity in Subdivided Populations. *Philosophical Transactions of the Royal Society B*: *Biological Sciences*, **360**, 1367-1378. http://dx.doi.org/10.1098/rstb.2005.1680

[6] Hillel, J., Martien, A.M., Groenen, A.M., Boichard, M.T., Abraham, B., Korol, A.B., David, V., Kirzhner, V.M., Burke, T., Barre-Dirief, A., Crooijmans, R.P.M.A., Elo, K., Feldman, M.W., Paul, J., Freidlin, P.J., Mäki-Tanila, A., Oortwijn, W., Thomson, P., Vigna, A., Wimmers, K. and Weigend, S. (2003) Biodiversity of 52 Chicken Populations Assessed by Microsatellite Typing of DNA Pools. *Genetic Selection Evolution*, **35**, 533-557. http://dx.doi.org/10.1186/1297-9686-35-6-533

[7] Norberg, E. and Sørensen, A.C. (2007) Inbreeding Trend and Inbreeding Depression in the Danish Populations of Texel, Shropshire, and Oxford Down. *Journal of Animal Science*, **85**, 299-304. http://dx.doi.org/10.2527/jas.2006-257

[8] Miglior, F., Burnside, E.B. and Hohenboken, W.D. (1994) Heterogeneity among Families of Holestien Cattle in Inbreeding Depression for Production Traits. *Proceedings of the 5th WCGALP*, **18**, 479-482.

[9] Rathje, T.A. (2000) Strategies to Manage Inbreeding Accumulation in Swine Breeding Company Nucleus Herds: Some Case Studies. *Journal of Animal Science*, **79**, 1-8.

[10] Culbertson, M.S., Mabry, J.W., Misztal, I. and Bertrand, J.K. (1997) Effect of Inbreeding and Outbreeding in Purebred and Hampshire and Duroc Swine. *The Professional Animal Scientist*, **13**, 194-197.

[11] Sewalem, A., Johansson, K., Wilhelmson, M. and Lillpers, K. (1999) Inbreeding and Inbreeding Depression on Reproduction and Production Traits of White Leghorn Lines Selected for Egg Production Traits. *British Poultry Science*, **40**, 203-208. http://dx.doi.org/10.1080/00071669987601

[12] Gómez, M.D., Valera, M., Molina, A., Gutiérrez, J.P. and Goyache, F. (2009) Assessment of Inbreeding Depression for Body Measurements in Spanish Purebred (Andalusian) Horses. *Livestock Science*, **122**, 149-155. http://dx.doi.org/10.1016/j.livsci.2008.08.007

[13] Dorji, N., Daujinda, M. and Phasuk, Y. (2011) Genetic Characterization of Thai Indigenous Chickens Compared with Commercial Lines. *Tropical Animal Health and Production*, **43**, 779-785. http://dx.doi.org/10.1007/s11250-010-9763-3

[14] Nidup, K. and Moran, C. (2011) Genetic Diversity of Domestic Pigs as Revealed by Microsatellites: A Mini Review. *Genomics and Quantitative Genetics*, **2**, 5-18.

[15] Food and Agriculture Organization/ISAG (2004) Measurement of Domestic Animal Diversity—A Review of Recent

Diversity Studies. http://www.fao.org/docrep/meeting/021/j1998e.pdf

[16] Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E. and Weigend, S. (2010) Genetic Diversity in Farm Animals—A Review. *Animal Genetics*, **41**, 6-31. http://dx.doi.org/10.1111/j.1365-2052.2010.02038.x

[17] Nassiry, M.R., Javanmard, A. and Tohidi, R. (2009) Application of Statistical Procedures for Analysis of Genetic Diversity in Domestic Animal Populations. *American Journal of Animal & Veterinary Sciences*, **4**, 136-141. http://dx.doi.org/10.3844/ajavsp.2009.136.141

[18] Tadano, R., Sekino, M., Nishibori, M. and Tsudzuki, M. (2007) Microsatellite Marker Analysis for the Genetic Relationships among Japanese Long-Tailed Chicken Breeds. *Poultry Science*, **86**, 460-469. http://dx.doi.org/10.1093/ps/86.3.460

[19] Osman, S.A., Sekino, M., Nishibori, M., Kawamoto, Y., Kinoshita, K., Yamamoto, Y. and Tsudzuki, M. (2004) Genetic Variability and Relationships of Japanese Native Chickens Assessed by Means of Microsatellite Profiling Approach-Focusing on the Oh-Shamo (Japanese Large Game) and Its Related Breeds. *Journal of Poultry Science*, **41**, 94-109. http://dx.doi.org/10.2141/jpsa.41.94

[20] Frankham, R., Ballou, J.D. and Briscoe, D.A. (2004) A Primer of Conservation Genetics. Cambridge University Press, London. http://dx.doi.org/10.1017/CBO9780511817359

[21] Hanotte, O. and Jianlin, H. (2005) Genetic Characterization of Livestock Populations and Its Use in Conservation Decision-Making. *The Role of Biotechnology*, Turin, 5-7 March 2005, 131-136.

[22] Toro, M.A., Fernandez, J. and Caballero, A. (2009) Molecular Characterization of Breeds and Its Use in Conservation. *Livestock Science*, **120**, 174-195. http://dx.doi.org/10.1016/j.livsci.2008.07.003

[23] Kalinowski, S.T. (2004) Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conservation Genetic*, **5**, 539-543. http://dx.doi.org/10.1023/B:COGE.0000041021.91777.1a

[24] Nei, M. (1978) Estimation of Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, **89**, 583-590.

[25] Allendorf, F.W. (1986) Genetic Drift and the Loss of Alleles versus Heterozygosity. *Zoo Biology*, **5**, 181-190. http://dx.doi.org/10.1002/zoo.1430050212

[26] Lauretto, M.S., Nakano, F., Faria Jr., S.R., Pereira, C.A.B. and Stern, J.M. (2009) A Straightforward Multiallelic Significance Test for the Hardy-Weinberg Equilibrium Law. *Genetics and Molecular Biology*, **32**, 619-625. http://dx.doi.org/10.1590/S1415-47572009000300028

[27] Rousset, F. and Raymond, M. (1997) Statistical Analyses of Population Genetic Data: New Tools, Old Concept. *Trends in Ecology & Evolution*, **12**, 313-317. http://dx.doi.org/10.1016/S0169-5347(97)01104-X

[28] Weir, B.S. (1996) Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Inc., Sunderland.

[29] Wright, S. (1951) The Genetical Structure of Populations. *Annals of Eugenics*, **15**, 323-354. http://dx.doi.org/10.1111/j.1469-1809.1949.tb02451.x

[30] Holsinger, K.E. and Weir, B.S. (2009) Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting F_{ST}. *Nature Reviews Genetics*, **10**, 639-650. http://dx.doi.org/10.1038/nrg2611

[31] Kalinowski, S.T. (2002) Evolutionary and Statistical Properties of Three Genetic Distances. *Molecular Ecology*, **11**, 1263-1273. http://dx.doi.org/10.1046/j.1365-294X.2002.01520.x

[32] Pearse, D.E. and Crandall, K.A. (2004) Beyond F_{ST}: Analysis of Population Genetic Data for Conservation. *Conservation Genetics*, **5**, 585-602. http://dx.doi.org/10.1007/s10592-003-1863-4

[33] Balloux, F. and Lugon-Moulin, L. (2002) The Estimation of Population Differentiation with Microsatellite Markers. *Molecular Ecology*, **11**, 155-165. http://dx.doi.org/10.1046/j.0962-1083.2001.01436.x

[34] Hedrick, P.W. (2005) A Standardized Genetic Differentiation Measure. *Evolution*, **59**, 1633-1638. http://dx.doi.org/10.1111/j.0014-3820.2005.tb01814.x

[35] Nei, M. and Roychoudhury, A.K. (1974) Sampling Variances of Heterozygosity and Genetic Distance. *Genetics*, **76**, 379-390.

[36] Ruane, J. (1999) A Critical Review of the Value of Genetic Distance Studies in Conservation of Animal Genetic Resources. *Animal Breeding and Genetics*, **116**, 317-323.

[37] Nei, M. (1972) Genetic Distance between Populations. *American Naturalist*, **106**, 283-292. http://dx.doi.org/10.1086/282771

[38] Nei, M., Tajima, F. and Tateno, Y. (1983) Accuracy of Estimated Phylogenetic Trees from Molecular Data. *Journal of Molecular Evolution*, **19**, 153-170. http://dx.doi.org/10.1007/BF02300753

[39] Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358-1370. http://dx.doi.org/10.2307/2408641

[40] Laval, G., San Cristobal, M. and Chevalet, C. (2001) Measuring Genetic Distances between Breeds: Use of Some Distances in Various Short Term Evolution Models. *Genetics Selection Evolution*, **34**, 481-507. http://dx.doi.org/10.1186/1297-9686-34-4-481

[41] Felsenstein, J. (1981) Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution*, **17**, 368-376. http://dx.doi.org/10.1007/BF01734359

[42] Nei, M. and Kumar, S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Inc., New York.

[43] Thorne, J.L., Kishino, H. and Felsenstein, J. (1991) An Evolutionary Model for Maximum Likelihood Alignment of DNA Sequences. *Journal of Molecular Evolution*, **33**, 114-124. http://dx.doi.org/10.1007/BF02193625

[44] Suhr, D.D. (2005) Principal Component Analysis vs. Exploratory Factor Analysis. *SUGI* 30 *Proceedings*, 203-230.

[45] Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nature Genetics*, **38**, 904-909. http://dx.doi.org/10.1038/ng1847

[46] Raymond, M. and Rousset, F. (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*, **86**, 248-249.

[47] Rousset, F. (2008) GENEPOP'007: A Complete Reimplementation of the GENEPOP Software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106. http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x