AcademicPres

Available online at www.notulaebiologicae.ro

Print ISSN 2067-3205; Electronic 2067-3264

Not Sci Biol 2 (4) 2010, 12-17

Notulae Scientia Biologicae

Comparative Study of Various Genetic Distance Measures between Populations for the ABO Gene

Supriyo CHAKRABORTY

North Carolina State University, Bioinformatics Research Centre, 318 Ricks Hall, 1 Lampe Drive, Raleigh, NC 27695, United States of America; supriyoch_2008@rediffmail.com

Abstract

Quantification of the genetic distance between populations is essential in many genetic research programs. Several formulae have been proposed for the estimation of the genetic distance between populations using gene frequency data. Nei's D has been the most widely used genetic distance measure in different research programs. But the selection of a suitable measure to estimate genetic distance between real-world human populations is a very difficult task. The present study was undertaken to estimate the genetic distance between Barak Valley Muslims (BVM) and other twenty four nations with the ABO blood group gene frequency data using seven different formulae, as well as to estimate the correlation coefficients between distance measures and to work out the regression equations. Seven genetic distance measures namely Nei's D, Nei's Nm, La, Nei's Da, Dc, Re and Nei's Ne were calculated between BVM and other 24 nations. Correlation coefficients of Nei's D with other measures were determined to find out which other genetic distance measures were similar to Nei's D. Linear regression equations of Nei's D with other distance measures were determined. Nei's D showed a highly significant (p=0.01) positive correlation with Cavalli-Sforza and Edwards chord distance Dc (0.90), Reynolds Re (0.90), Nei's Da (0.74) and Nei's Ne (0.63) but a negative correlation with Nei's Nm and La. Since Nei's D had very high positive correlation with Dc and Re distance measures, any one of these measures could be reliably used in genetic analysis instead of all the three measures for estimating genetic distance between populations.

Keywords: comparison, gene frequency, genetic distance measure

Introduction

Quantification of the genetic distance between populations is instrumental in many genetic research programs. A large number of formulae have been proposed for this purpose. However, the 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 (Libiger *et al*., 2009).

Nei's standard genetic distance has been the most widely used genetic distance measure between populations. Since several formulae have already been proposed for genetic distance measurement, it is essential to identify which genetic measures show a close similarity with Nei's D measure. The present study was undertaken to estimate the genetic distance between Barak Valley Muslims (BVM) and each of other 24 populations for ABO blood group gene frequency data using seven different genetic distance measures. These seven measures were Nei's D, Nei's Nm, Latter's La, Nei's Da, Cavalli-Sforza and Edwards Dc (RE), Reynolds Re and Nei's Ne. To identify the distance measure(s) that shows similarity with Nei's D, a correlation analysis was performed between the estimates of Nei's D and other distance measures. Regression equations of different distance measures on Nei's D were worked out to determine the value of a particular distance measure with a given value of Nei's D.

Materials and methods

In this study, ABO blood group distribution data of 25 populations excluding Barak Valley Muslims were obtained from the published literature and websites. The ABO blood group distribution data in Barak Valley Muslims were estimated by the author (Chakraborty, 2010). The frequencies of O, A and B alleles belonging to ABO blood group system for each population were estimated from ABO blood group phenotyping data using the formulae suggested by Hedrick (2005) as given below:

$$\mathrm{A} = 1 - \sqrt{\frac{N_{22} + N_{23} + N_{33}}{N}}$$

$$\mathrm{B} = 1 - \sqrt{\frac{N_{11} + N_{13} + N_{33}}{N}}$$

$$\mathrm{O} = \sqrt{\frac{N_{33}}{N}}$$

Where N= Total individuals
$N_{11}+N_{13}$= Individuals having "A" blood group
$N_{22}+N_{23}$= Individuals having "B" blood group
N_{33}= Individuals having "O' blood group

Received 19 October 2010; accepted 15 November 2010

Genetic distance measurement

The ABO gene frequency data (Tab. 1) were used to estimate the genetic distance between Barak Valley Muslims and each of the remaining 24 populations using seven distance measures as given below.

Let the genetic distance for '*m*' loci with '*v*' alleles per locus be studied in populations 1 and 2 with n_1 and n_2 individuals having $\overline{n}$ as the average number of individuals. Let $\overline{P}_{lu1}$ and $\overline{P}_{lu2}$ be the frequencies of allele '*u*' at locus '*l*' in population 1 and 2, respectively and let P_{lu1} and P_{lu2} be the number of individuals that carry allele '*u*' at locus '*l*' in populations 1 and 2 respectively, then seven distance measures can be estimated as follows:

Nei's standard genetic distance (D) between two populations without bias correction according to Nei (1972) is estimated as:

$$\mathrm{D} = -\ln\left[\frac{\sum_l \sum_u \overline{P}_{lu1}\overline{P}_{lu2}}{\left(\sum_l \sum_u \overline{P}^2_{lu1} \sum_l \sum_u \overline{P}^2_{lu2}\right)^{1/2}}\right]$$

Nei's minimum distance (Nm) is given by the following equation:

$$\mathrm{Nm} = {}^1\!/_2 \sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu1}\sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu2} - \sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu1}.\overline{P}_{lu2}$$

Latter's distance (La) according to Latter (1972) is given by:

$$\mathrm{La} = \frac{{}^1\!/_2 \sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu1}\sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu2} - \sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu1.}\overline{P}_{lu2}}{1 - \sum_{l=1}^{m}\sum_{u=1}^{v}\overline{P}_{lu1.}\overline{P}_{lu2}}$$

Nei's Da distance according to Nei *et al.* (1983) is given by:

$$\mathrm{Da} = 1 - \frac{1}{m}\sum_{l=1}^{m}\sum_{u=1}^{v}\sqrt{\overline{P}_{lu1}\overline{P}_{lu2}}$$

Cavalli-Sforza and Edwards chord distance (Dc or CE) according to Cavalli-Sforza and Edwards (1967) is given by:

$$\mathrm{Dc} = {}^2\!/_{3.1416m}\sum_{l=1}^{m}\left[2\left\{1 - \sum_{u=1}^{v}\left(P_{lu1}.P_{lu2}\right)^{1/2}\right\}\right]^{1/2}$$

Reynolds genetic distance (Re) according to Reynolds *et al.* (1983) is given by:

$$\mathrm{Re} = \frac{\sum_l\left\{{}^1\!/_2\sum_u\left(\overline{P}_{lu1} - \overline{P}_{lu2}\right)^2 - \frac{1}{2\left(2\overline{n} - 1\right)}\left[2 - \sum_u \overline{P}^2_{lu1} + \overline{P}^2_{lu2}\right]\right\}}{\sum_l\left(1 - \sum_u \overline{P}_{lu1}\overline{P}_{lu2}\right)}$$

Tab. 1. Estimates of allele frequencies of ABO gene in Barak Valley Muslims (BVM) and other 24 populations/nations

Sl. No.	Population	Allele Frequency				Reference*
		O	A	B	Total	
1	Kenya	0.69	0.17	0.14	1.00	Anees and Mirza (2005)
2	Sudan	0.81	0.11	0.08	1.00	www.bloodbook
3	Saudi Arabia	0.58	0.21	0.21	1.00	- do -
4	India (Overall)	0.62	0.16	0.22	1.00	- do -
5	Sri Lanka	0.69	0.16	0.15	1.00	- do -
6	West Indonesia	0.69	0.10	0.21	1.00	Brequet *et al.* (1986)
7	Borneo (Malaysia)	0.62	0.22	0.16	1.00	Kamil *et al.* (2010)
8	South China	0.53	0.23	0.24	1.00	www.bloodbook
9	Australia	0.78	0.22	-	1.00	- do -
10	Bulgaria	0.57	0.31	0.12	1.00	- do -
11	Hungary	0.60	0.27	0.13	1.00	- do -
12	Austria	0.60	0.30	0.10	1.00	- do -
13	Pakistan	0.74	0.12	0.14	1.00	- do -
14	Central Asia (Uzbekistan)	0.56	0.25	0.19	1.00	Revavov *et al.* (1983)
15	Eastern Europe(Poland)	0.57	0.28	0.15	1.00	www.bloodbook
16	Siberia	0.57	0.16	0.27	1.00	- do -
17	Russia	0.57	0.25	0.18	1.00	- do -
18	Alaska	.62	0.29	0.09	1.00	- do -
19	USA (Whites)	0.67	0.25	0.08	1.00	- do -
20	Britain	0.69	0.26	0.05	1.00	- do -
21	Norway	0.62	0.32	0.06	1.00	- do -
22	Sweden	0.62	0.31	0.07	1.00	- do -
23	Iceland	0.74	0.19	0.07	1.00	- do -
24	Denmark	0.64	0.27	0.09	1.00	- do -
25	Barak Valley Muslims	0.63	0.18	0.19	1.00	Chakraborty (2010)

*Detailed reference in text

Nei's geometric distance (Ne) based on genotype frequency data (not gene frequency) is given by:

$$\mathrm{Ne} = 1 - \tfrac{1}{m}\sum_{l=1}^{m}\sum_{u=1}^{v}\left(P_{lu1}P_{lu2}\right)^{1/2}$$

Correlation and regression analysis

Correlation coefficient between any two distance measures was calculated as per Harris *et al.* (2007). Correlation coefficient was tested by the 't' test for significance at p=0.01 and 0.05. Linear regression equation of a distance measure (as dependent variable) on Nei's D as independent variable was estimated by the method of least squares as per Harris *et al.* (2007).

Results and discussion

The Barak Valley Area, named after the mighty river Barak flowing through the area, is located in southern part of Assam state in North East India. The valley has been inhabited by one of the major endogamous religious groups, the Muslims, for several centuries. Barak Valley has a total population of about 3.21 million including Hindus, Muslims and Christians with a land area of 6,992 square kilometers. This region is characterized by undulating topography with wide plain area, low lying water logged tracts and hillocks. The climate of the Barak valley is subtropical, warm and humid with average annual rainfall of 318cm and 146 rainy days. Nearly 80% of the total population depends on agriculture for livelihood.

Gene frequency

The frequencies of O, A and B alleles of ABO gene of different nations/populations were estimated from the ABO blood group distribution data of each population (Tab. 1). In general, the frequency of O allele was the highest in all the populations. B allele was not reported in Australians.

Genetic distance between populations

The estimates of various genetic distance measures (expressed in percent) between Barak Valley Muslims (BVM) and each of the twenty-four populations were calculated on the basis of ABO gene frequency data (Tab. 2).

Nei's D estimate was the lowest (0.0015) between BVM and India (in general) indicating the lowest genetic distance but highest genetic identity between these two populations for ABO gene. On the other hand, the highest Nei's D value (0.0395) was found between BVM and Australia suggesting greatest genetic distance but lowest genetic identity between these two populations for ABO gene out of 24 combinations. Nei's geometric distance (Ne), except all other genetic distance measures, was calculated on the basis of genotypic data estimated from ABO gene frequency.

Tab. 2. Estimates of seven different genetic distance measures between BVM and other nations

Sl. No.	Combination	Nei's D	Nei's Nm	La	Nei's Da	Cavalli-Sforza Dc	Reynolds Re	Nei's Ne
1	BVM-Kenya	0.0045	0.0081	0.0159	0.0027	0.0464	-0.0038	-0.0549
2	BVM-Sudan	0.0268	-0.0453	-0.0996	0.0216	0.1325	0.0051	-0.0169
3	BVM-Saudi Arab	0.0032	0.0 569	0.1021	0.0013	0.0330	-0.0042	-0.0762
4	BVM-India	0.0015	0.0 388	0.0720	0.0009	0.0265	-0.0047	-0.0691
5	BVM-Sri Lanka	0.0038	0.0080	0.0157	0.0022	0.0418	-0.0039	-0.0552
6	BVM-West Indonesia	0.0083	0.0074	0.0146	0.0068	0.0741	-0.0029	-0.0505
7	BVM-Borneo	0.0027	0.0 394	0.0730	0.0017	0.0368	-0.0045	-0.0689
8	BVM-South China	0.0138	0.0 791	0.1366	0.0051	0.0646	-0.0021	-0.0808
9	BVM-Australia	0.0395	-0.0310	-0.0661	0.1000	0.2847	0.0073	0.0239
10	BVM-Bulgaria	0.0276	0.0623	0.1108	0.0135	0.1047	-0.00002	-0.0662
11	BVM-Hungary	0.0136	0.0487	0.0888	0.0076	0.0783	-0.0025	-0.0669
12	BVM-Austria	0.0253	0.0490	0.0893	0.0150	0.1101	-0.0004	-0.0599
13	BVM-Pakistan	0.0113	-0.0144	-0.0297	0.0071	0.0761	-0.0013	-0.0412
14	BVM-Central Asia	0.0093	0.0661	0.1168	0.0039	0.0562	-0.0031	-0.0773
15	BVM-Eastern Europe (Poland)	0.0161	0.0620	0.1103	0.0074	0.0776	-0.0020	-0.0722
16	BVM-Siberia	0.0106	0.0608	0.1084	0.0045	0.0607	-0.0029	-0.0740
17	BVM-Russia	0.0084	0.0617	0.1098	0.0037	0.0547	-0.0033	-0.0758
18	BVM-Alaska (Eskimo)	0.0235	0.0401	0.0742	0.0158	0.1131	-0.0006	-0.0559
19	BVM-USA Whites	0.0175	0.0177	0.0342	0.0149	0.1098	-0.0013	-0.0483
20	BVM-Britain	0.0261	0.0090	0.0177	0.0269	0.1476	0.0009	-0.0349
21	BVM-Norway	0.0379	0.0404	0.0748	0.0282	0.1513	0.0022	-0.0447
22	BVM-Sweden	0.0328	0.0403	0.0746	0.0235	0.1379	0.0012	-0.0489
23	BVM-Iceland	0.0185	-0.0137	-0.0282	0.0170	0.1172	0.0004	-0.0339
24	BVM-Denmark	0.0187	0.0311	0.0586	0.0138	0.1058	-0.0014	-0.0543

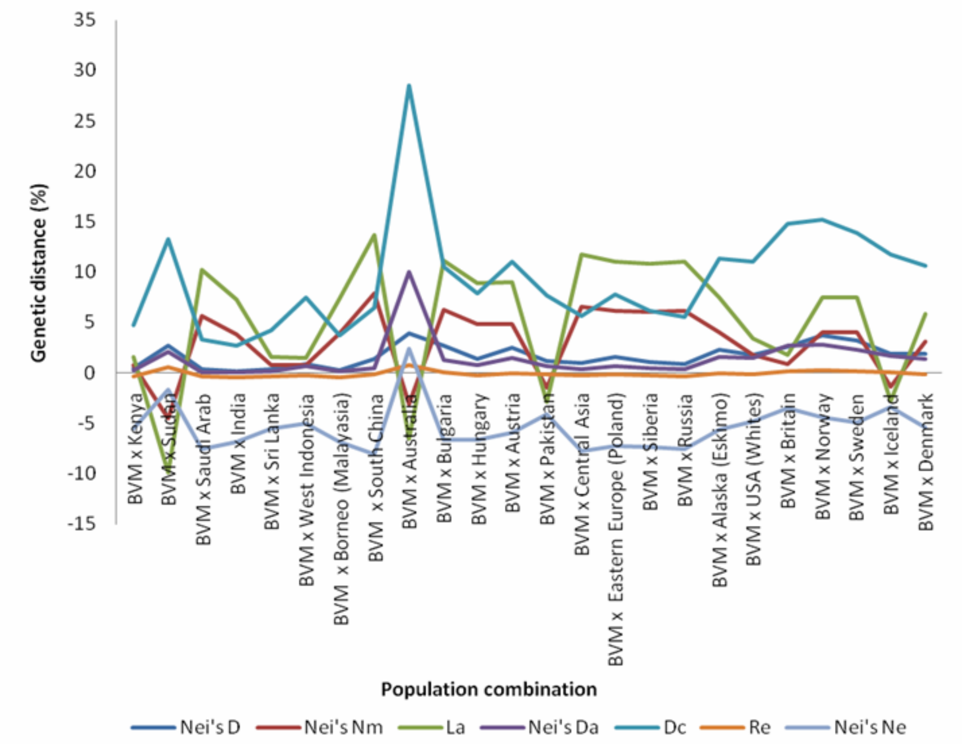

Fig. 1. Comparison of different genetic distance measures between BVM and other nations

Nei's minimum genetic distance (Nm) ranged from the lowest value 0.0074 between BVM and West Indonesia to the highest value 0.0791 between BVM and South China irrespective of sign. Similarly, Latter's distance (La) ranged from 0.0146 between BVM and West Indonesia to 0.1366 between BVM and South China.

Nei's Da estimate ranged from 0.0009 between BVM and India to 0.1000 between BVM and Australia. Cavalli-Sforza and Edwards chord distance (Dc) showed the range from 0.0265 between BVM and India to 0.2847 between BVM and Australia. Reynolds genetic distance (Re) ranged from the lowest estimate 0.00002 between BVM and Bulgaria to the highest value 0.0073 between BVM and Australia. Nei's Ne estimate ranged from the lowest value 0.0169 between BVM and Sudan to the highest value 0.0808 between BVM and South China.

Several studies were carried out on genetic distance measurements across different populations. Genetic distance and gene diversity studies by Roy *et al.* (1990) among 10 endogamous groups in Chattisgarh, India using the gene frequency data of three genetic loci revealed that the gene differentiation among these population groups is only about 2 per cent (0.02).

Genetic differentiation studies in Indian populations by Papiha *et al.* (1982) revealed that genetic differentiation in India populations was low (0.26-1.70%). In Assam, genetic variation studies by Das (1979) among three caste populations namely Brahmin, Kalita and Kaibarta on the basis of the ABO blood groups and other anthropometric characters revealed that the Kaibarta stand apart from the Brahmin and the Kalita, who are similar to each other. Genetic study by Danker-Hopfe *et al.* (1988) among 13 Assamese populations including two Muslim groups for the distribution of anthropometric, anthroposcopic and dermatoglyphic traits revealed that the Muslims in Assam were distinguished between Marias (who seemed to be more closely related to Mongoloid populations) and Sheikhs (whose phenotypic appearance was more like that of Hindu caste groups).

Genetic distance studies by Roychoudhury *et al.* (1982) between Jews and Non-Jews using gene frequency data of nine blood groups and protein loci revealed that the Yemenite Jews have a high degree of genetic affinity to the Israeli Arabs and the Iranian Jews to the Iranians. Genetic distance studies by Triantaphyllidis *et al.* (1983) between the inhabitants of nine Mediterranean countries and the three major human races using the gene frequency data of several genetic markers suggested that the Algerians were closer to Negroids while the other Mediterraneans were closer to Caucasoids.

Genetic and taxonomic distance studies by Sokal (1988) among 3466 samples of human populations in Europe based on 97 allele frequencies and 10 cranial variables demonstrated that speakers of different language families in Europe differ genetically and that this difference remains even after geographic differentiation.

Correlation analysis

The estimates of correlation coefficients between any two distance measures (Tab. 3) revealed that Nei's

Tab. 3. Correlation coefficients of Nei's D with other genetic distance measures

Distance measure	Nei's Nm	La	Nei's Da	Dc	Re	Nei's Ne
Nei's D	-0.24	-0.25	0.74**	0.90**	0.90**	0.63**

** Significant at p=0.01

D showed highly significant (p=0.01) positive correlation with Cavalli-Sforza and Edwards chord distance Dc (0.90), Reynolds Re (0.90), Nei's Da (0.74) and Nei's Ne (0.63). This indicated great similarity between these four distance measures and any one of these measures could be used instead of all the four measures in genetic analysis. But due to very high magnitude of the positive correlation of Nei's D with Cavalli-Sforza and Edwards chord distance Dc and Reynolds Re, the use of any one out of these three measures would be more effective in genetic analysis. Nei's D showed non-significant negative correlation with Nei's minimum distance Nm and Latter's distance La.

Regression analysis

Nei's D is the most widely used genetic distance measure in research programs. Assuming Nei's D as a dependent variable and anyone of the remaining distance measures (Da, Dc, Re or Ne) as independent variable, the linear regression equations of the latter on Nei's D were estimated (Tab. 4). Since Nei's minimum distance Nm and Latter's distance La did not show significant correlation with Nei's D, hence Nm and La were not used as dependent variables in determining the linear regression equation with Nei's D.

Tab. 4. Linear regression equations of different distance measures on Nei's D

Dependent variable (y)	Independent variable (x)	Regression equation y = A+Bx
Nei's Da	Nei's D	Da = -0.80 + 1.34D
Cavalli-Sforza Dc	Nei's D	Dc = 1.91 + 4.44D
Reynolds Re	Nei's D	Re = -0.51 + 0.24D
Nei's Ne	Nei's D	Ne = -7.60 + 1.30D

These regression equations could be used to estimate the magnitude of the particular genetic distance measure with a given value of Nei's D between two populations. But the accuracy of the particular genetic estimates calculated from a given estimate of Nei's D using the above linear regression equations would decrease with the decreasing value of correlation coefficients. In the regression equation y = A+Bx, the B estimate represents the regression coefficient (slope) for linear regression and the regression constant A represents the magnitude of the y-intercept *i.e.* the distance from the origin to the point where the straight line intersects the y-axis.

Conclusions

The present study revealed that the Barak Valley Muslims had the highest genetic distance from Australians for the ABO gene but the lowest from the Indians. Nei's D genetic distance measure showed a highly significant, positive correlation with other distance measures namely Cavalli-Sforza and Edwards chord distance Dc and Reynolds Re measures indicating great similarity between these three distance measures. But Nei's D measure showed a negative correlation with Nei's minimum distance Nm and Latter's distance La.

Acknowledgements

The first author is thankful to the Department of Biotechnology, Ministry of Science and Technology, Govt. of India, New Delhi for providing the DBT Overseas Associateship (No. BT/05/NE/2009) to undertake the present research work and also grateful to Prof. Z.B. Zeng, Director, BRC, NCSU, Raleigh, USA for his kind suggestions.

References

Anees, M. and M. S. Mirza (2005). Distribution of ABO and Rh blood group alleles in Gujrat region of Punjab, Pakistan. Proc. Pak. Acad. Sci. 42(4):233-238.

Brequet, G., R. Ney, H. Gerber and M. F. Garner (1986). Treponemal serology and blood groups on Bali Island, Indonesia. Genitourin Med. 62:298-301.

Chakraborty, S. (2010). Genetic analysis on frequency of alleles for Rh and ABO blood group systems in the Barak Valley populations of Assam. Notulae Scientia Biologicae 2(2):31-34.

Cavalli-Sforza, L. L. and A. W. F. Edwards (1967). Phylogenetic analysis: models and estimation procedure. Am. J. Hum. Genet. 19:233-257.

Danker-Hopfe, H., B. M. Das, H. Walter, P. B. Das and R. Das (1988). Anthropological studies in Assam, India - Differentiation processes among Assamese populations. Anthropol. Anz. 46(2):159-184.

Das, B. M. (1979). Physical variation in three Assamese castes. Anthropol. Anz. 37(3):204-210.

Harris, M., G. Taylor and J. Taylor (2007). Maths and Stats for the Life and Medical Sciences. Viva Books Pvt Ltd, New Delhi.

Hedrick, P. W. (2005). Genetics of populations (3rd Ed), Jones and Bartlett Publishers, Sudbury.

Kamil, M., Al-Jamal Han and N. M. Yusoff (2010). Association of ABO blood groups with diabetes mellitus. Lib. J. Med. 5 at http://www.journals.sfu.ca/coactions/index.php/ljm/article/view article/4847/5365.

Latter, B. D. (1972). Selection in infinite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. Genet. 70:475-490.

Libiger, O., C. M. Nievergelt and N. J. Schork (2009). Comparison of genetic distance measures using human SNP data. Hum. Biol. 81(4):389-406.

Nei, M. (1972). Genetic distance between populations. Am. Nat. 106:283-292.

Nei, M. (1973). The theory and estimation of genetic distance. In: N. E Morton (Eds.). Genetic structure in populations, University Press of Hawaii, Honolulu.

Nei, M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genet. 3:583-590.

Nei, M., F. Tajima and Y. Yateno (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19:150-173.

Papiha, S. S., B. N. Mukherjee, S. M. S. Chahal, K. C. Malhotra and D. F. Roberts (1982). Genetic heterogeneity and population structure in north-west India. Ann. Hum. Biol. 9(3):235-251.

Reynolds, J., B. S. Weir and C. C. Cockerham (1983). Estimation of the coancestry coefficient: Basis for a short term genetic distance. Genet. 105:767-779.

Roy, M., U. Datta, M. Mitra and C. S. Singhrol (1990). Genetic distance and gene diversity among ten endogamous groups in Chattisgarh, Central India. Int. J. Anthropol. 5(2):109-115.

Roychoudhury, A. K. (1982). Genetic distance between Jews and Non-Jews of four regions. Hum. Hered. 32(4):259-263.

Revavov, A. A., A. Asanov, I. N. Lunga and S. M. Bakhramov (1983). Frequencies of ABO system blood groups and haptoglobins in Uzbekistan-The problems of sampling studies. Genetika 19(7):1193-1197.

Sokal, R. R. (1988). Genetic, geographic and linguistic distances in Europe. Proc. Natl. Acad. Sci., USA 85:1722-1726.

Triantaphyllidis, C. D., A. Kouvatsi and L. Kaplanoglou (1983). The genetic distances between the inhabitants of nine Mediterranean countries and the three major human races. Hum. Hered. 33(2):137-139.

www.bloodbook.com/world-abo.html.