

## ORIGINAL ARTICLE

# ABO Blood Groups are Efficient in Delineating Relationships across Linguistic Subpopulations of Pakistan

Nuzhat Aisha Akram<sup>1</sup>, Nazia Jameel<sup>2</sup>, Muhammad Ikram Ali<sup>3</sup>, Iffat Raza<sup>4</sup>, Amina Raza<sup>5</sup>

<sup>1</sup>Department of Forensic Medicine and Toxicology, Karachi Institute of Medical Sciences, Karachi, <sup>2</sup>Department of Community Medicine, Baqai Medical University, Karachi, <sup>3</sup>Department of Forensic Medicine and Toxicology, Ziauddin University, <sup>4</sup>Department of Anatomy, Karachi Institute of Medical Sciences, Karachi, <sup>5</sup>Department of Biochemistry, Karachi Institute of Medical Sciences, Karachi, Pakistan.

## ABSTRACT

**Background:** Varying frequencies of ABO blood groups serve objectively to characterize populations just as the traditional characters are observed and measured by anthropologist. The study aimed to delineate genealogical relationships across the five major linguistic subpopulations of Pakistan using ABO blood groups phenotype and gene frequencies.

**Methods:** This was a prospective multiple cohort study design. The study was conducted during 2009-2014. In Karachi. Cohorts were formed because of mother tongue and termed as linguistic subpopulations. There were five major cohorts namely Balochi, Punjabi, Pashto, Sindhi and Urdu. Punjabi, Sindhi and Urdu belong to Indo Aryan class while Balochi and Pashto belong to Indo Iranian class of languages. Blood samples were collected from 2,327 unrelated individuals chosen through random sampling from different sampling units. ABO blood groups were determined using antisera. ABO gene frequencies were estimated through Bernstein methods. Cluster and Correspondence analyses were performed to explore genealogical relationships between the five linguistic subpopulations.

**Results:** Gene B was the most common across all the subpopulations except Balochis which showed relatively higher frequency of gene A. Dendrogram of cluster analysis and bi-plot of correspondence analysis showed closeness of three Indo Aryan linguistic subpopulations namely Punjabi, Sindhi and Urdu, while Indo Iranian linguistic subpopulations Balochi and Pashto stood out from Indo Aryan cluster.

**Conclusion:** ABO blood groups efficiently delineated genealogical relationship between the five linguistic subpopulations of Pakistan. Randomization, un-relatedness and a larger sample size enhanced the reliability of the results. Inclusion of more genetic markers may further elaborate the relationships between the subpopulations.

**Keywords:** ABO Phenotype; Gene Frequencies; Linguistic Subpopulations; Cluster and Correspondence Analyses; Genealogical Relationships.

**Corresponding Author:**

**Dr. Nuzhat Aisha Akram**

Department of Forensic Medicine and Toxicology,  
Karachi Institute of Medical Sciences,  
Karachi, Pakistan.

Email: smr.akram@gmail.com

doi.org/10.36283/PJMD9-1/007

## INTRODUCTION

The finding of blood group ABO by Landsteiner in the early part of twentieth century is one of the most important discoveries of medical science in both the clinical and scientific fields<sup>1-5</sup>. ABO polymorphism was investigated frequently to form anthropological criterion to reveal unknown and morphologically invisible relationships between different

populations<sup>3,6,7</sup>.

The world distribution patterns of ABO blood groups vary across populations<sup>8,9</sup>. A geographical gradient of blood groups was revealed in various research studies with a comparatively higher frequencies of blood group B in Asian populations than blood group A, AB or O. However, majority of them approached populations without considering their

ethnic and linguistic backgrounds hence ignored population stratification. Population strata may have ABO blood group frequencies that may differ significantly from overall picture of the whole population.

Blood groups have been linked to certain diseases including cancer like threatening illnesses<sup>10-14</sup>. These groups were also used to reconstruct population genealogical structure<sup>15-17</sup>. Recently ABO blood group studies investigated evolutionary history of populations<sup>18-21</sup>. It is worth noticing that the gene frequencies are found to be similar in populations who are known to be related. If the gene frequencies for the two components of a mixed population are known, the gene frequencies in the mixture tell us the proportions of the two original population's present<sup>21,22</sup>.

The study of the blood groups of human populations in general is becoming a planned and integrated investigation<sup>22</sup>. People of the Indian subcontinent have a long history of population heterogeneity with respect to its complex cultural, social, and even genetic structure. These effects are magnified with repeated invasion and migration of the populations from other areas. Systemic forces e.g., mutation, migration, selection, admixture and rate of population increase also affect population structure and equilibrium. Population genetic studies have also shown some substructures in Pakistani population for genetic variants, including genetic diseases, blood groups and DNA polymorphism<sup>23-26</sup>. We may hope that population studies will one day include examinations of many more markers and in collaboration with other human studies; we will be able to study the population structure in detail. This study was aimed to delineate genealogical relationships across the five major linguistic subpopulations of Pakistan using ABO blood groups phenotype and gene frequencies.

**METHODS**

Blood samples were collected from 2,966 unrelated

individuals living in Karachi. Individuals from each linguistic subpopulation were sampled randomly from different sampling units for ABO blood grouping using anti sera A, B and D. Sampling units were chosen through systematic random sampling from each of the five districts (south, central, west, and east) of the city. Each of the samples was provided with an informed consent and a questionnaire regarding his or her ethnicity and mother tongue up to second generation. Ethical approval for the present study was given by the Board of Advanced Studies and Research (BASR), University of Karachi.

Each linguistic subpopulation was based on individuals' mother tongue. Either sample size was calculated using a frequency estimate of 0.5 for of the Gene A, B or O which gives a maximum sample size needed<sup>27</sup>. The calculation showed a sample size of 96 with a precision of 0.1 and 384 with a precision of 0.05. Therefore, only those linguistic subpopulations were included in the study, which showed a sample size of more than 100. Only five major linguistic subpopulations who fulfilled this criteria were included making a total sample size of 2,327. These subpopulations were Balochi, Punjabi, Pashto, Sindhi and Urdu. A comprehensive sampling strategy, randomization and a large sample size (n=2, 327) yielded the proportions of linguistic subpopulations that were roughly the same as their proportion in the total population of Karachi. The data obtained from different sampling units was compiled and analyzed using Pearson chi square test and exploratory statistical tests namely Cluster and Correspondence analyses to delineate the relationships between different linguistic subpopulations of Karachi.

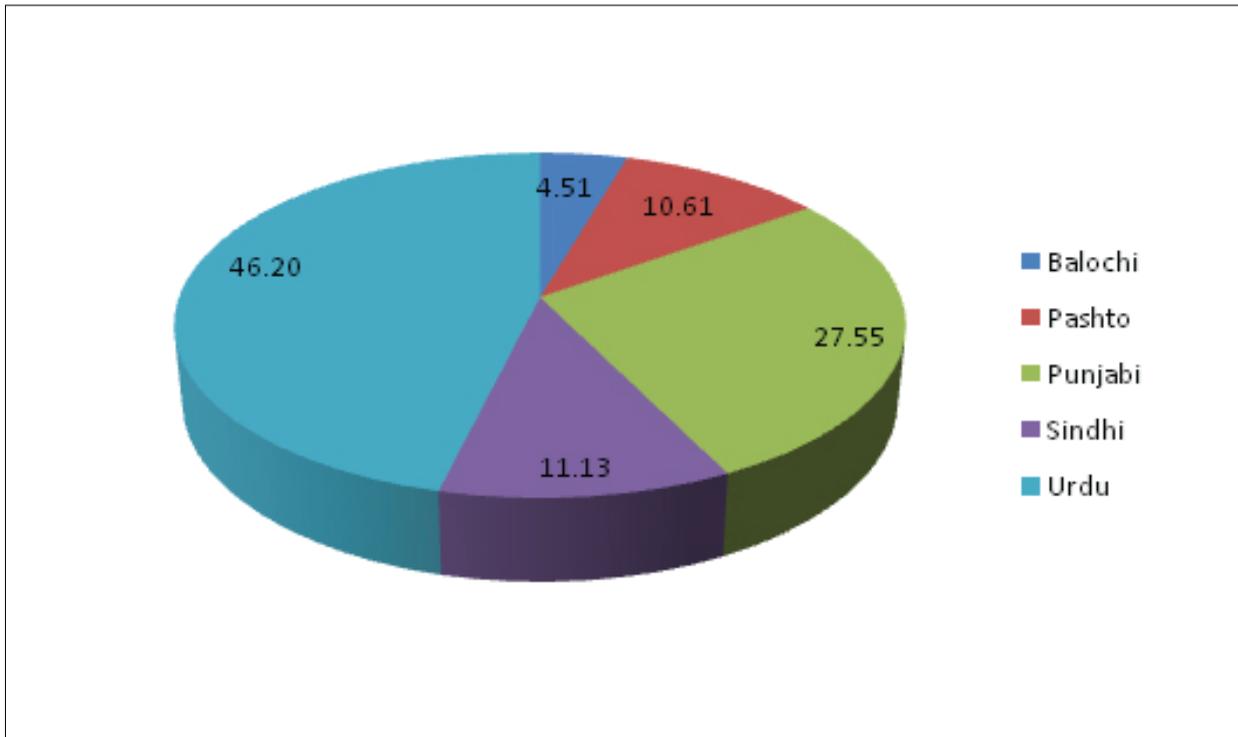
**RESULTS**

There were five linguistic subpopulations included in the study. Sample size of each subpopulation is shown in Table 1. As the individuals were sampled randomly, and total sample size was large, their proportion more or less reflected the actual proportion in the city census population (Figure 1).

**Table 1: Demographic information of mother tongue based linguistic subpopulations of Karachi.**

S. No.	Linguistic subpopulation	Language Family	Language Class	Sample Size	Percentage population in Karachi*
1	Balochi	Indo European	Indo Iranian	105	4.34
2	Pashto	Indo European	Indo Iranian	247	11.42
3	Punjabi	Indo European	Indo Aryan	641	13.94
4	Sindhi	Indo European	Indo Aryan	259	7.22
5	Urdu	Indo European	Indo Aryan	1075	48.52

\* According to 1998 census.



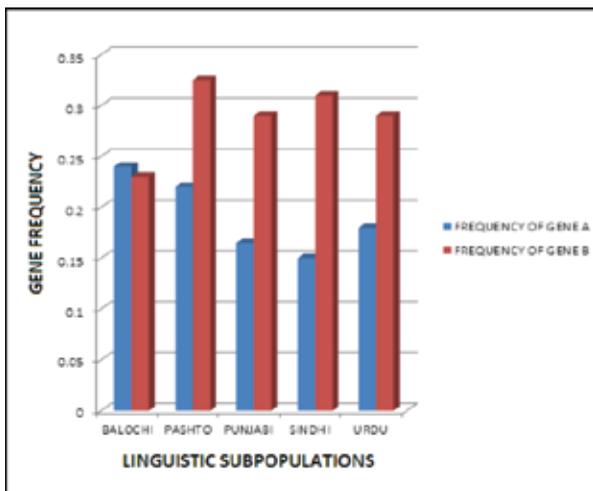
**Figure 1: Proportion (%) of each linguistic subpopulation in the present study.**

Phenotype and gene frequencies of ABO blood group for each linguistic subpopulation were calculated from the data using SPSS version 16

(Table 2). Gene frequencies were estimated by the simple Bernstein method<sup>28</sup> (Figure 2).

**Table 2: Sample size and ABO blood group counts and percentages (in parentheses) for each of the five linguistic subpopulations included in the study.**

S. No.	Linguistic subpopulation	Blood group A n (%)	Blood group B n (%)	Blood group AB n (%)	Blood group O n (%)	Total
1	Balochi	35 (33.33)	33 (31.43)	10 (9.52)	27 (25.71)	105
2	Urdu	244 (22.7)	428 (39.81)	116 (10.79)	287 (26.7)	1075
3	Pashto	69 (27.94)	109 (44.13)	27 (10.93)	42 (17)	247
4	Punjabi	125 (19.5)	257 (40.09)	66 (10.3)	193 (30.11)	641
5	Sindhi	54 (20.85)	118 (45.56)	17 (6.56)	70 (27.03)	259
	Agglomerated population	527 (22.65)	945 (40.61)	236 (10.14)	619 (26.6)	2327



**Figure 2: Frequencies of Gene A and Gene B across all the linguistic subpopulations of the present study.**

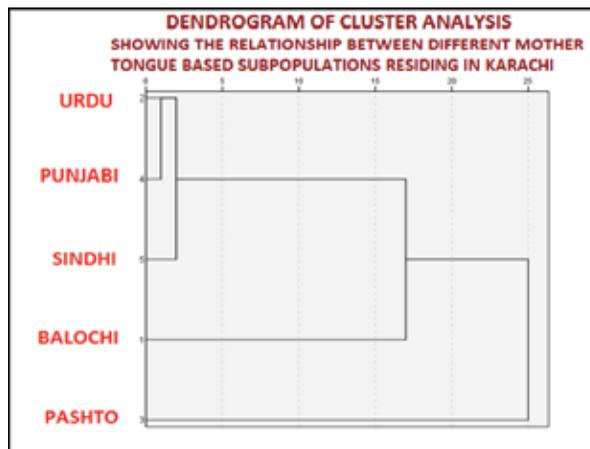
Null hypothesis of independence across five linguistic subpopulations was tested using Chi square test of independence. The test was performed using SPSS version 16.0. A significant p value (< 0.05) showed that the two categorical variables are associated (Table 3).

**Table 3: Chi square test of independence shows significant association (p value <0.05) for the distribution of blood groups ABO across the five linguistic subpopulations.**

S.No.	Tests	Value	Degree of Freedom	p-value
1	Pearson Chi square	28.56	12	0.005
2	Likelihood Ratio	29.77	12	0.003

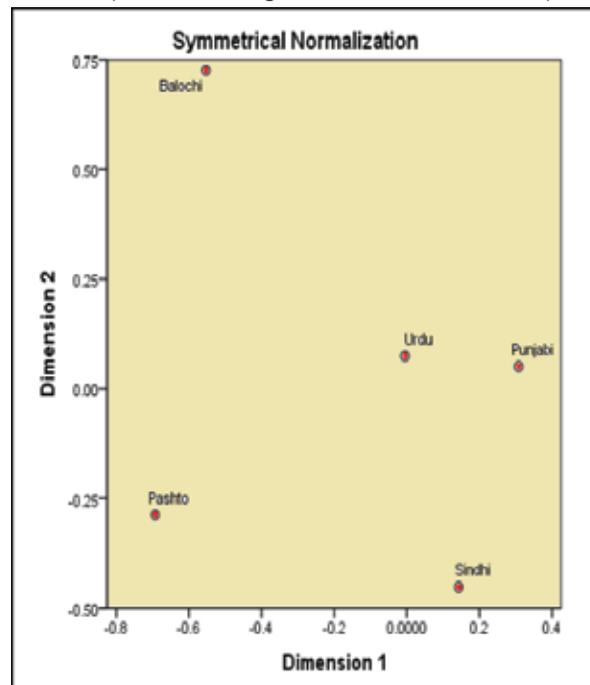
Exploratory analyses were performed to delineate relationship across linguistic subpopulations of Karachi by constructing a population tree (dendrogram) of them. For this purpose, data was explored through two multivariate statistical approaches, namely Cluster and Correspondence analyses.

Complete linkage (furthest neighbor) and squared Euclidean distance were used for the analysis. Agglomeration schedule and the dendrogram for the clusters so formed showed the similarities between the linguistic subpopulations of Karachi (Figure 3). Dendrogram showed a distinct cluster of Punjabi, Sindhi and Urdu subpopulations, while Balochi and Pashto were a bit apart from the former three subpopulations.



**Figure 3: Dendrogram of Cluster Analysis showing the relationships across linguistic subpopulations of Karachi.**

Correspondence analysis was used to describe the relationships between subpopulations in a low dimensional space. Three dimensions of solutions were chosen. These dimensions measured the variation between the subpopulations like the principle components of Principal Component Analysis (PCA). Dimensions were extracted like the principle components in the PCA. Symmetrical normalization was used to construct biplot between the subpopulations (Figure 4). Biplot of the Correspondence analysis showed the same relationship across the linguistic subpopulations as shown by the Dendrogram of the Cluster analysis.



**Figure 4: Biplot of Correspondence Analysis showing the relationships along the axes of Dimension 1 and Dimension 2 across the five linguistic subpopulations of Karachi.**

## DISCUSSION

Ethnic, linguistic, and cultural diversity is the sine qua non of the inhabitant populations of South Asia, making it the focus of interest for a wide variety of discipline including anthropology, linguistic, genetics and evolutionary studies. Pakistani population is known for its heterogeneity with respect to ethnicity, bradri and caste system, linguistic origin, cultural heritage, complex social structure and even the genetic variants. This population is said to be a sub-structured population<sup>29</sup>. The consanguinity rate was found to be as high as 25.9%<sup>30</sup>. Population genetic studies have also shown some substructure in our population for genetic variants, including genetic diseases, blood groups and DNA polymorphism. Therefore, it is mandatory to study the population in context of its substructure. In the present study mother tongue was taken as the basis of variation in the population. Chi square test of independence shows that the linguistic subpopulations are heterogeneous in nature for the distribution of ABO blood group frequencies. As ABO blood groups are expressions of one's DNA, therefore it can be concluded that these linguistic subpopulations vary from each other in their genetic makeup.

Three methods are used to estimate ABO gene frequencies from their phenotypic frequencies<sup>31</sup>. Efficiencies of all the three methods are almost the same but it depends on the number of alleles under consideration. Here the number of alleles ( $m$ ) is three i.e. A, B and O. For  $m=3$  simple Bernstein method is fully efficient<sup>32</sup>. Therefore, gene frequency estimation from simple Bernstein method was used for the analyses. Blood group B was found to be the most frequent blood group. This finding is in agreement with the previous studies, which claim the higher prevalence of blood group A in the west and that of blood group B in the east<sup>32, 33</sup>.

However, there is a subpopulation, which did not show the higher prevalence of blood group B. The observation is mainly confined to a single subpopulation i.e. Balochi. Blood group A is found to be more frequent in this subpopulation. Although the difference is smaller but this is, significant keeping in view the ratio of A to B is almost 1: 2 in other subpopulations. One can conclude that the subpopulations having a high prevalence of 'A' may have an evolutionary history of admixture with other world populations who have high frequency of blood group A. However, this finding needs to be explored extensively with a larger sample size and in conjunction with other human studies.

Pashto subpopulation, which like Balochi subpopulation belongs to Indo Iranian linguistic class, did not show the same frequency distribution of ABO blood group as did the Balochi subpopulation. This difference might be due to early migration of Pashto linguistic subpopulation to Karachi and their admixture with other Indo Aryan linguistic subpopulations residing in Karachi. However, the ratio of blood

group A to B is 2:3 in Pashto subpopulations as compared to Indo Aryan subpopulations that showed the ratio of A to B as 1:2. Moreover, Balochi subpopulation is thought to be a relatively isolated subpopulation, which remains confined to its clan due to strict cultural values and social norms. Dendrograms of cluster analysis and biplot of Correspondence analysis showed the clustering of the three subpopulations Punjabi, Sindhi, and Urdu. All these subpopulations belong to Indo Aryan class of Indo European group of languages. Remaining two subpopulations Pashto and Balochi belong to Indo Iranian class of Indo European group. This finding showed that ABO polymorphism could be used to trace out linguistic affiliations.

## CONCLUSION

ABO blood group served as efficient genetic markers to unravel genealogical relationships between subpopulations of Pakistan. Similarities between subpopulations were shown either as distinct cluster in Dendrogram or closeness in the biplot of correspondence analysis. Similarities of subpopulation may reflect either a common ancestry or recent admixture. Keeping in view the linguistic origin and evolutionary history of Pakistani subpopulations both were the plausible explanations. However, the inference will be strong and reliable when the data include as many markers along with ABO blood groups and as large sample size as possible. In conjunction with the studies of other genetic markers, such as STRs, VNTRs, Alu, and SNPs etc., we will be able to draw a clearer picture of the population.

## ACKNOWLEDGEMENTS

We are thankful to the administration of hospitals, clinics, blood banks and other sampling units from where blood samples were collected after taking permission and sampled for ABO blood grouping for the present study.

## CONFLICT OF INTEREST

There is no conflict of interest among the authors.

## ETHICS APPROVAL

Ethical approval for the present study was given by the Board of Advanced Studies and Research (BASR), University of Karachi (Ref. No. BASR/Sc/-Gen/823).

## PATIENT CONSENT

Verbal and written informed consent was obtained from all the samples included in the present study.

## AUTHORS' CONTRIBUTIONS

NAA conceived the idea, designed the study, planned for data collection and analyzed the result as well as the final manuscript. NJ helped in data

collection, performed the bench work and studied the design for review. MIA helped in the data collection and literature review. IR helped in data entry, compilation, and analyzed the results. Further, AR helped in manuscript writing and bibliography.

## REFERENCES

- Landsteiner K. Über agglutinationserscheinungen normalen menschlichen blutes. *Wien Klin Wschr.* 1901;14:1132-4.
- Mourant AE. The use of blood groups in the study of populations. *Ann N Y Acad Sci.* 1970; 169 (1): 194-198.
- Bodmer W. Genetic characterization of human populations: from ABO to a genetic map of the British people. *Genet.* 2015; 199(2): 267-279.
- Vala D, Nayyar AS, Pooja VK, Kartheeki B, Patel N, Vala D, Tanmay P. Determination of ABO blood grouping from dentine and pulp by absorption-elution technique. *Int J Orolac Biol.* 2017;1:70-80
- Kim JY, Park EY, Lee HJ. The study of ABO genotyping for forensic application. *Forensic Sci Int Genet.* 2017; 6:e570-e572.
- Apecu RO, Mulogo EM, Bagenda F, Byamungu A. ABO and rhesus blood group distribution among blood donors in rural south Uganda: a retrospective study. *BMC Res Notes.* 2016; 9: 513.
- Majumdar PP, Roy J. Distribution of ABO blood groups on the Indian subcontinent: A cluster analytic approach. *Curr Anthropol.* 1982; 23(5): 539-566.
- Canizalez-Román A, Campos-Romero A, Castro-Sánchez JA, López-Martínez MA, Andrade- Muñoz FJ, Cruz-Zamudio CK. Blood groups distribution and gene diversity of the ABO and Rh (D) loci in the Mexican population. *Biomed Res Int.* 2018;1925619.
- Kumar S, Modak PK, Ali SH, Barpanda SK, Gusain VS, Roy R. A retrospective study: ABO and Rh phenotype blood group distribution among blood donors in H.N.B. Base Hospital, Srinagar, Uttarakhand, India. *J Family Med Prim Care.* 2018;7:34-8
- Vasan SK, Rostgaard K, Majeed A, Ullum H, Titlestad KE, Pedersen OB, et al., ABO Blood group and risk of thromboembolic and arterial disease: a study of 1.5 million blood donors. *Circulation.* 2016; 133(15): 1449-1457.
- RummelSK and Ellsworth RE. The role of the histoblood ABO group in cancer. *Future Sci OA.* 2016; 2(2): 107.
- Fry AE, Griffiths MJ, Auburn S, Diakite M, Forton JT, Green A. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. *Human Mol Genet.* 2008; 17(4):567-76.
- Zhang C, Li Y, Wang L, Sun S, Liu G, Leng J, et al. Blood group AB is protective factor for gestational diabetes mellitus: A prospective population-based study in Tianjin, China. *Diabetes Metab Res Rev.* 2015; 31(6) 627-37.
- MaoY, Yang W, Qi Q, Yu F, Wang T, Zhang H, et al. Blood groups A and AB are associated with increased gastric cancer risk: evidence from a large genetic study and systematic review. *BMC Cancer.* 2019; 19:164.
- Bangham J. Blood groups and human groups: Collecting and calibrating genetic data after World War Two. *Stud Hist Philos Biol Biomed Sci.* 2014;47:74-86.
- Suárez-Díaz E. Indigenous populations in Mexico: Medical anthropology in the work of Ruben Lisker in the 1960s. *Stud Hist Philos Biol Biomed Sci.* 2014;47:108-17.
- Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science.* 2014; 344(6189): 1280-1285.
- Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A, et al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci USA.* 2009; 106(21): 8611-8616.
- Romero-Hidalgo S, Ochoa-Leyva A, Garcíaarubio A, Acuña-Alonso V, Antúnez-Argüelles E, Balcazar-Quintero M, et al. Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nat Commun.* 2017; 8(1).
- Adhikari K, Chacón-Duque JC, Mendoza-Revilla J, Fuentes-Guajardo M, and Ruiz-Linares A. The genetic diversity of the Americas. *Annu Rev Genomics Hum Genet.* 2017; 18(1): 277-296.
- Mourant AE. Blood Groups. *Symposia of the society for the study of human biology.* 1961;IV.
- Mourant AE. The use in anthropology of blood groups and other genetical characters. *J African His.* 1962; III, 2: 291-296.
- Firasat S., Khaliq S, Mohyuddin A, Papaioannou M, Smith CT, Underhill PA and Ayub Q. Y- chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan. *Eur J Hum Genet.* 2007; 15: 121-126.
- Mohyuddin A, Williams F, Mansoor A, Mehdi SQ and Middleton D. Distribution of HLA-A alleles in eight ethnic groups of Pakistan. *Tissue Antigen.* 2002; 61:286-291.
- Mian A, Farooq A. Distribution of ABO and RH blood group alleles in different populations of southern Punjab, Pakistan. *Anthropologischer Anzeiger.* 1999;33-9.
- Akram N, Farooqi S. A model based 'varimax' sampling strategy for a heterogeneous population. *Ann Hum Biol.* 2014; 41(2):118-26.
- Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. *Arch Orolac Sci.* 2006; 1: 9-14.
- Nam JM, Gart JJ. Bernstein and gene counting methods in generalized ABO like systems. *Ann Hum Genet.* London 1976; 39: 361- 373.
- Overall ADJ. The influence of the wahlund effect on the consanguinity hypothesis: Consequences for recessive disease incidence in a society structured Pakistani population. *Hum Heredity.* 2009; 67 9 9(2): 140-4.
- Kumar D. Genetic disorder of the Indian subcontinent. 2004. Springer.
- DeGroot MH. Efficiency of gene frequency estimates for the ABO system. (1955). Committee on Statistics, University of Chicago.
- Allan T.M. Hirsfeld and the ABO blood groups. *Brit J Prev Soc Med.* 1963; 17:166-171.
- Buettner-Janusch J. Natural selection in man: The ABO (H) blood group system. *Am Anthropol.* 1959; 61 (3): 437-451.