



Genetic peopling of Pakistan: Influence of consanguinity on population structure and forensic evaluation of traces

Ijaz Anwar*, Franco Taroni

School of Criminal Justice, University of Lausanne, 1015 Dorigny, Lausanne, Switzerland.

ARTICLE INFO

Keywords:
Consanguinity
Population genetics
DNA evidence evaluation

ABSTRACT

Pakistan is one of the most consanguineous country in the world [1] where the cousin marriages account for more than half of the total unions. To investigate the genetic evidence of consanguinity in Pakistani populations, 1020 samples were collected from the volunteers belonging to four different populations. The magnitude of the effect of population structure and consanguinity on the calculation of likelihood ratios were also explored.

1. Introduction

Pakistan is located at important cross-roads of human history. It has nurtured many civilizations and have seen several mass migrations. Historical records have shown that humans from Africa traversed to India, China and up to Australia passing through current day Pakistan [2]. Therefore, it keeps a unique mixture of ancient and modern civilizations. Pakistan is the sixth largest country in terms of population where people speak around 74 different languages [3]. The language barrier and some other factors force people to marry within their ethnicity. In Pakistan, 56% of the marriages are between first and second cousins only [1,4], which makes it interesting to study its genetic composition and population structure.

There has been some genetical studies on Pakistani populations, but with some constraints such as the limited number of samples and of loci analyzed and the extension of populations considered. An earlier study reported the genetic distance between three co-resident Pakistani populations to be 13% and suggested to use this value as a correction factor to calculate the match probabilities [5]. But soon it was revealed by another study that the previous results were in contradiction with published guidelines. During this period, the first study was quoted in many court cases where the defense suggested the use of F_{ST} equal to 0.13 as a correction factor. In one of the case, the paternity index reduced substantially from 1300 to just 350 [6].

So, there was a need to thoroughly study the genetic composition and structure of Pakistani populations, which ensures the robust sampling of an extended number of Pakistani populations and also covers most of the population. A more reliable and representative database of allele frequencies could be established. That can be used in forensic evaluation of biological evidence. The genetic and forensic parameters

of interest such as heterozygosity, coefficient of co-ancestry (F_{ST}) and coefficient of inbreeding (F_{IS}) of Pakistani populations can be used to explain the evolutionary development of these populations. Those parameters could be efficiently used in the evidential assessment of DNA results in forensic context.

Furthermore, both the ENFSI guidelines [7] and NRC recommendations [8] suggest the use of the likelihood ratio and the use of parameters to assess the probative value of forensic evidence, respectively. In case of a correspondence between the value of the genetic profiles obtained from the evidence material and the reference samples, should be supported by some underlying data describing the relevant population genetically.

2. Materials and methods

In order to study the structure of Pakistani populations and the impact of migrations and consanguinity, 1020 healthy individuals belonging to Punjabi, Saraiki, Pakhtun and Sindhi populations based on 15 autosomal STR loci (AmpF/STR Identifier® kit) were analyzed after getting the written consent form to participate in this study. Genetic Analyses were performed by using the FORSTAT [9] program and hierfstat [10] package of R language.

3. Results

Allele frequencies were calculated and population databases were established for each sub-population as well as for the whole country. In order to check the applicability of these STR markers to be used as a potential human identification tool, heterozygosity was calculated on each locus for each sub-population. Results have shown that mean

* Corresponding author:

E-mail addresses: ijaz.anwar@unil.ch (I. Anwar), franco.taroni@unil.ch (F. Taroni).

<https://doi.org/10.1016/j.fsigss.2019.09.089>

Received 17 September 2019; Accepted 26 September 2019

Available online 26 September 2019

1875-1768/ © 2019 Elsevier B.V. All rights reserved.

Observed Heterozygosity (H_o) ranged from 0.7267 in Sindhi population to 0.7805 in Pakhtun population, while the mean Gene Diversity (H_s) ranged from 0.7978 in Sindhi to 0.8038 in Saraiki population. Population specific coefficient of co-ancestry (F_{ST}) was observed to be -0.003 in Pakhtun population, 0.002 in Punjabi as well as in Saraiki and 0.005 in Sindhi population. This F_{ST} value was in contrast with one published earlier [5]. The genetic distance based on F_{ST} was lowest (0.0007) between Pakhtun and Sindhi, while it was highest (0.0024) between Punjabi and Sindhi populations.

With a view to explore the extent of consanguinity in Pakistani populations and to search for the genetic proofs, coefficient of inbreeding (F_{IS}) was calculated. Mean values for the coefficient of inbreeding ranged from 0.029 to 0.091 which is quite high for human populations.

These genetic parameters can influence the assessment of a forensic case. To quantify the impact of coefficient of co-ancestry (F_{ST}) and the coefficient of inbreeding (F_{IS}) on the calculation of likelihood ratio of a match, we used the formulas suggested by Balding and Nichols [11] and Ayres & Overall [12]. It can be easily shown that the value of evidence is strongly reduced by integrating F_{ST} and F_{IS} values in the likelihood ratio calculation. This emphasizes the crucial need of genetical parameters such as F_{ST} and F_{IS} to assess DNA profiles in populations where high degree of consanguinity is present along with population sub structure.

4. Discussion

Pakistani populations have been living together since ages. There has been introduction of new alleles into these populations due to migrations, catastrophes, wars and search for the better socioeconomic opportunities. Yet these populations remained admixed with each other and maintained their unique genetic and geographical structure. High prevalence of first and second cousin marriages may have contributed to the formation of this admixture by keeping their gene pool intact in social boundaries. Ease of communication, unique cultural heritages and religious priorities have also allowed these population to marry within families. These inter-family marriages have produced greater similarities between the individuals of these populations. These similarities between individuals should be quantified and taken into account in every evidence evaluation involving DNA profiles.

5. Conclusion

Pakistan is a country with an important demographic history. Pakistani populations are highly consanguineous and they have

retained their gene pool by keeping the tradition of endogamous marriages intact, which is quite evident from high F_{IS} values. This unique behavior advocates for the use of F_{IS} values along with F_{ST} when calculating the conditional match probabilities. This will avoid overstatement about the value of the evidence.

Declaration of Competing Interest

Authors agree that there was no conflict of interest.

Acknowledgements

This research was supported by the Swiss Excellence Government Scholarship for PhD studies of foreign students in Switzerland. The authors would like to thank Shahid Hussain, Attaur Rehman, Manzoor Hussain, Jacques Linden, Lorenzo Gaborini and Tacha Hicks for their technical help in carrying out this project.

References

- [1] H. Hamamy, Consanguineous marriages, *J. Community Genet.* 3 (2011) 185–192, <https://doi.org/10.1007/s12687-011-0072-y>.
- [2] C.J. Bae, K. Douka, M.D. Petraglia, On the origin of modern humans: asian perspectives, *Science* 358 (2017), <https://doi.org/10.1126/science.aai9067> eai9067.
- [3] Pakistan, (Date Accessed 16 September 2019). <https://www.ethnologue.com/country/PK>.
- [4] Pakistan demographic and Health Survey, (Date Accessed 16 September 2019). <https://dhsprogram.com/pubs/pdf/FR290/FR290.pdf>.
- [5] L.A. Zhivotovskiy, S. Ahmed, W. Wang, A.H. Bittles, The forensic DNA implications of genetic differentiation between endogamous communities, *Forensic Sci. Int.* 119 (2001) 269–272, [https://doi.org/10.1016/s0379-0738\(00\)00442-4](https://doi.org/10.1016/s0379-0738(00)00442-4).
- [6] J.M. Curran, J. Buckleton, The appropriate use of subpopulation corrections for differences in endogamous communities, *Forensic Sci. Int.* 168 (2007) 106–111, <https://doi.org/10.1016/j.forsciint.2006.06.073>.
- [7] ENFSI, (Date Accessed 16 September 2019). http://enfsi.eu/wp-content/uploads/2016/09/ml_guideline.pdf.
- [8] NRC II, (Date Accessed 16 September 2019). https://www.ncbi.nlm.nih.gov/books/NBK232610/pdf/Bookshelf_NBK232610.pdf.
- [9] P.G. Ristow, M.E. D'Amato, Forensic statistics analysis toolbox (FORSTAT): a streamlined workflow for forensic statistics, *Forensic Sci. Int. Genet. Suppl. Ser.* 6 (2017) e52–e54, <https://doi.org/10.1016/j.fsigss.2017.09.006>.
- [10] J. Goudet, Hierfstat a package for r to compute and test hierarchical F-statistics, *Mol. Ecol. Notes* 5 (2005) 184–186, <https://doi.org/10.1111/j.1471-8286.2004.00828.x>.
- [11] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification relatedness, database selection and single bands, *Forensic Sci. Int.* 64 (1994) 125–140, [https://doi.org/10.1016/0379-0738\(94\)90222-4](https://doi.org/10.1016/0379-0738(94)90222-4).
- [12] K.L. Ayres, A.D.J. Overall, Allowing for within-subpopulation inbreeding in forensic match probabilities, *Forensic Sci. Int.* 103 (1999) 207–216, [https://doi.org/10.1016/s0379-0738\(99\)00087-0](https://doi.org/10.1016/s0379-0738(99)00087-0).