

LETTER TO THE EDITOR

Open Access



Erroneous calculations of weight of DNA evidence may lead to miscarriage of justice in Pakistan

Nouman Rasool

Dear Editor,

The correct weightage of DNA evidence can be calculated by knowing the behavior of DNA marker in the relevant population. Allele frequencies, the prevalence of micro variants in the population, the pattern of inheritance, and the difference of allele frequencies among various subpopulations residing at a particular geographical area are the important factors for evaluating the weight of evidence. Inbreeding, population substructures and migration increase genetic relatedness between individuals (Steele and Balding 2014). In the presence of the abovementioned scenario, a genetically related individual already sharing more alleles compared to the unrelated individuals becomes genotypically more related and chances of inaccurate calculation of random match probability increase many folds. In Pakistan, there are two functional forensic laboratories, which are performing forensic DNA analysis in civil and criminal cases. One is the National Forensic Science Agency (NFSA), Islamabad, while the other is the Punjab Forensic Science Agency (PFSA), Lahore. It has passed nearly 9 years since their establishment, but no population database of allele frequencies of STR markers has been developed by either of the laboratories. NFSA is not currently doing any statistical calculations for forensic casework (Mateen et al. 2018). The forensic report generated by DNA department of NFSA contains allelic match between the profile obtained from evidence and profile(s) obtained from the victim and/or suspect without any statistical calculation of estimating the probability of a match. PFSA calculates “Random Match Probability” for forensic casework and relationship testing. In the absence of local population database of STR loci, PFSA is using Caucasian STR allele frequencies of US-FBI database, USA, to calculate match probabilities. Diverse subpopulations dwelling across the

country are genotypically and phenotypically different from each other. Allele frequencies of few populations residing in Pakistan are reported by researchers, but allele frequencies of all population substructures are not yet elucidated. Allele frequencies of Balochi, Bruhi, Burusho, Kalash, Makrani, Pathan, Sindhi, Saraiki, and Gujjar have already been reported and are available at (http://spsmart.cesga.es/search.php?dataSet=strs_local) (Jilani et al. 2016; Shan et al. 2016), while some other subpopulations, e.g., Kalash, Hindkowan, Shina, Kalyu, Balti, and Kashmiri, are not explored yet for determining allele frequencies in Pakistan. A strict caste/clan system prevails in the country where marriages occur only within the family or within the same caste. It seldom happens that one marries to a person who does not belong to one's caste. In such cases, few alleles become more prevalent in a particular subpopulation and few become rare. When match probabilities are calculated using Caucasian allele frequencies from the US-FBI database, chances of error increase in finding accurate statistical results. This could be worse when two related individuals such as full siblings or half-siblings or even cousins are involved in a crime or when a mixed DNA profile of three or more than three contributors is obtained from an evidence (Lohmueller and Rudin 2013; Toscanini et al. 2012). In estimating relatedness in the case of a missing person, legitimacy of a child, or property disputes, application of an unrelated dataset of allele frequencies would result in erroneous estimations.

No consolidated population DNA database containing allele frequencies of STR loci for Pakistani population is available for forensic casework and relationship testing. As such, the government has not provided funding to any institution for the establishment of a national population DNA database. There are various international databases available online for calculating allele frequencies for STR loci. STRbase (<http://strbase.org>) is developed by the European Network of Forensic Science Institutes (ENFSI) for calculating random match probabilities in European populations. These calculations are

Correspondence: noumanrasool@gmail.com

Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan

made using 16 STR loci (Gill et al. 2006; Parson and Roewer 2010). STRBase (<http://www.cstl.nist.gov/strbase>) developed by the National Institute of Standards and Technology (NIST) is a comprehensive database, which provides information frequencies of common and rare alleles. This also uses up to 16 loci and calculations are made using Omnipop software which is linked with this database (Ruitberg et al. 2001). ALFRED, the ALlele FREquency Database (<https://alfred.med.yale.edu/>), is a large allele frequency database with a sample size of $N > 660,000$ individuals. This fairly large database is used to calculate allele frequencies of populations residing in Europe and outside Europe (Rajeevan et al. 2011). Similarly, pop. STR (<http://spsmart.cesga.es/popstr.php>), PopAffiliator (http://cracs.fc.up.pt/_nf/popaffiliator2), and ALLSTR**R* (<http://allstr.de>) are online available databases of allele frequencies of STR loci for statistical calculations (Amigo et al. 2009; Bodner et al. 2016; Pereira et al. 2011). But these databases are not a true depiction of allelic frequencies of all subpopulations residing in Pakistan.

The allele frequencies calculated for European and American populations are different from those for population substructures living in Asia due to evolutionary and cultural history (Venables et al. 2016). The available STR databases developed in other jurisdictions have different genetic characteristics compared to the genetic makeup of Pakistani populations. Although, it is recommended that such calculations can be used by applying a correction to avoid variation in sub-population (Holsinger and Weir 2009), if these frequencies do not represent the true allele frequencies of other population, such corrections would be all in vain. Therefore, STR database containing all population substructures need to be established for calculating the correct weight of DNA evidence. This database should have an appropriate size and representation of each sub-population having $F_{ST} < 0.05$. The subpopulations having $F_{ST} < 0.05$ can be grouped into one population (Council 1996). This database can be used by all national forensic labs to estimate match probabilities in forensic casework and relationship testing.

Moreover, Thermo Fisher Scientific and Promega have marketed STR kits having more than 21 STR loci. Allele frequencies for Asians are available with GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific, South San Francisco, CA, USA) and PowerPlex® 21 System (Promega, Madison, WI, USA). There are two issues in these allele frequencies, i.e., first, the sample size is not large enough, and, second, frequencies of few alleles in many loci are not given. In Pakistan, forensic laboratories are considering to use these kits for forensic case work, paternity, and kinship analysis. The old databases do not have allele frequencies for newly introduced STR loci, so there is a need for calculating allele frequencies for new STR loci added in these systems.

It is already reported that while establishing paternity, chances of error is about 3% even when proper allele frequencies are applied for calculating the paternity index (Green and Mortera 2017; Karlsson et al. 2007). Forensic DNA analysis requires flawless statistical calculation for random match probability, likelihood ratio, and relationship testing excluding close relatives among donors. If allele frequencies do not truly represent a population sub-structure, the chance of erroneous inclusion increases to a great degree.

Lawyers and judiciary have little or no knowledge about the importance of calculating the probabilities and determining the weightage of forensic DNA evidence in a case. What they understand is the statement in a report that is “Match” or “Not Match”, and they are unable to understand the statements related to the probability of matching any other individuals having the same profile in the relevant population. For a better understanding of forensic DNA analysis reports, lawyers and judges must be educated on why and how match probabilities are used in these reports. In Pakistan, only few institutions are offering degree programs in Forensic Biology or Forensic Genetics, in which such forensic calculations are part of the curriculum. Few institutions/universities, i.e., CAMB, University of Punjab, Lahore; DoC, GC College University, Lahore; IBBT, University of Animal and Veterinary Science, Lahore; DLS, University of Management and Technology, Lahore; and University of Health Sciences, Lahore, are offering degree programs in Forensic Sciences. All degree programs are approved and accredited by the Higher Education Commission, Islamabad. All institutions, except GC College University, offer programs of Forensic Biology. GC College University offers a graduate program in Forensic Chemistry. It also offers a 1-year MS program for lawyers, judges, and personnel from law enforcing agencies and general public. Without understanding the exact weightage of DNA evidence, it is difficult to conduct a fair trial which may result in a miscarriage of justice. In the future, if these cases would be reanalyzed for statistical calculations using true allele frequencies, there are chances of finding a fair number of wrongful convictions across the country, based on forensic DNA reports. There is a great need to develop a national population DNA database for all subpopulations residing within the country.

Acknowledgements

Not applicable.

Authors' contributions

NR has analyzed the literature and reported the whole study. The author has read and approved the final manuscript.

Funding

No funding received for this study.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares that he/she has no competing interests.

Received: 20 February 2019 Accepted: 14 June 2019

Published online: 24 June 2019

References

- Amigo J, Phillips, C, Salas, T, Formoso LF, Carracedo Á, Lareu M (2009) pop. STR—an online population frequency browser for established and new forensic STRs. *Forensic Sci Int Genet Supplement Series* 2:361–362
- Bodner M et al. (2016) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Sci Int Genet* 24:97–102
- Gill P, Fereday L, Morling N, Schneider PM (2006) New multiplexes for Europe—amendments and clarification of strategic development. *Forensic Science International* 163(1-2):155–157
- Green PJ, Mortera J (2017) Paternity testing and other inference about relationships from DNA mixtures. *Forensic Sci Int Genet* 28:128–137
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics* 10:639–650
- Jilani A, Nadeem A, Tahir M, Rasool N (2016) Genetic analysis of the Saraiki population living in Pakistan. *J Can Soc Forensic Sci* 49:152–160
- Karlsson AO, Holmlund G, Egeland T, Mostad P (2007) DNA-testing for immigration cases: the risk of erroneous conclusions. *Forensic Sci Int* 172: 144–149
- Lohmueller KE, Rudin N (2013) Calculating the weight of evidence in low-template forensic DNA casework. *J Forensic Sci* 58:S243–S249
- Mateen RM, Tariq A, Rasool N (2018) Forensic science in Pakistan; present and future. *Egypt J Forensic Sci* 8:45
- National Research Council (1996). The evaluation of forensic DNA evidence. National Academies Press, Danvers.
- Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. *Int J Legal Med* 124:505–509
- Pereira L et al. (2011) PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. *Int J Legal Med* 125:629–636
- Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK (2011) ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res* 40:D1010–D1015
- Ruitberg CM, Reeder DJ, Butler JM (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res* 29:320–322
- Shan MA, Hussain M, Shafique M, Shahzad M, Perveen R, Idrees M (2016) Genetic distribution of 15 autosomal STR markers in the Punjabi population of Pakistan. *Int J Legal Med* 130:1487–1488
- Steele CD, Balding DJ (2014) Choice of population database for forensic DNA profile analysis. *Sci Justice* 54:487–493
- Toscanini U, Garcia-Magariños M, Berardi G, Egeland T, Raimondi E, Salas A (2012) Evaluating methods to correct for population stratification when estimating paternity indexes. *PLoS One* 7:e49832
- Venables SJ et al. (2016) Allele frequency data for 15 autosomal STR loci in eight Indonesian subpopulations. *Forensic Sci Int Genet* 20:45–52

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)