Forensic Science International: Genetics Supplement Series 7 (2019) 87–89

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

GENETIC PORTRAIT OF THE PUNJABI POPULATION FROM PAKISTAN USING THE PRECISION ID ANCESTRY PANEL

Muhammad Adnan Shan[a,b,*], Mie Refn[a], Niels Morling[a], Claus Børsting[a], Vania Pereira[a]

[a] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[b] Centre for Applied Molecular Biology, University of the Punjab, Lahore, Pakistan

ARTICLE INFO

Keywords:
AIMs
Ancestry
Punjabi
PCA plot
Geno geographer

ABSTRACT

Prediction of geographical ancestry using genetic markers has a great potential in forensic genetics and may be used as an investigative lead in crime casework or missing person identification. Exploration of AIMs in Pakistan is interesting due to the distinct subpopulations with multidirectional ancestry from different groups. In the current study, 87 individuals from the Punjabi population from Pakistan were investigated using the Precision ID Ancestry Panel (Thermo Fisher Scientific) to assess whether it was possible to diff ;erentiate Punjabi individuals from other populations. With this panel, it is revealed that Punjabis are admixed and cannot be distinguished from other populations in South Central Asia and the Middle East.

1. Introduction

Genetic markers that present marked allele frequency differences among populations can be used as ancestry informative markers (AIMs). Prediction of geographical ancestry using AIMs has a great potential in forensic genetics and may be used as an investigative lead in crime casework or missing person identification [1–3]. Exploration of AIMs in Pakistan is interesting due to the distinct subpopulations with multidirectional ancestry from neighbouring states and native groups [4,5]. Punjab is the second largest and most populous of the four provinces of Pakistan. It has a population of more than 90 million corresponding to 46 % of the Pakistani population [6]. Punjab is located in the north-western part of the Indian plate at the Indus River system. The Punjabi population is the largest ethnic group in Pakistan. It consists of a heterogeneous population group with various tribes, clans, and communities. The native language of the province is Punjabi. Various ethnic groups settled in this region and formed the Indus Valley Civilization in the bronze age 3,300 to 1,300 BCE [4–5].

In this work, the Precision ID Ancestry Panel (Thermo Fisher Scientific) [7,8] was used to genotype Punjabi individuals and estimate their ancestry. The panel includes 165 autosomal AIMs for genoogeographic prediction. The marker set is a combination of 55 SNPs of the Kidd AISNP panel [9] and 123 AISNPs from the Seldin panel [10,11] with 13 SNPs overlapping [12]. The most likely population of origin was investigated with GenoGeographer [13,14], a tool that calculates the population likelihoods of a profile for each reference population included in the database (Sub-Saharan Africa, Somalia, North Africa, Europe, Middle East, South Central Asia, and Greenland).

1.1. Materials and methods

A total of 87 unrelated Punjabi individuals were typed for 165 ancestry informative markers using the Precision ID Ancestry Panel (Thermo Fisher Scientific). The Arlequin v.3.5.2.2 software [15] was used to estimate deviations from Hardy-Weinberg equilibrium (HWE). Data were compared to those of populations studied for the same markers (data kindly provided by the Kidd Lab, assembled from publicly available data). The SNP rs10954737 was excluded from the inter-population comparisons due to lack of data of the reference populations. Principal component analysis (PCA) was performed using an in-house developed Python script. The software GenoGeographer was used to calculate z-scores, likelihoods and likelihood ratios to infer the most likely population of Punjabis.

2. Results and Discussion

All studied 165 AIMs markers were in HWE after Bonferroni correction for multiple testing, except for one locus (rs310644; p-value_0.0001). Punjabis clustered with individuals from South Central Asia and the Middle East as visualised in the PCA plot (Fig. 1). The

* Corresponding author at: Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
E-mail address: muhammad.shan@sund.ku.dk (M.A. Shan).

https://doi.org/10.1016/j.fsigss.2019.09.034
Received 11 September 2019; Accepted 22 September 2019
Available online 17 October 2019
1875-1768/ © 2019 Elsevier B.V. All rights reserved.

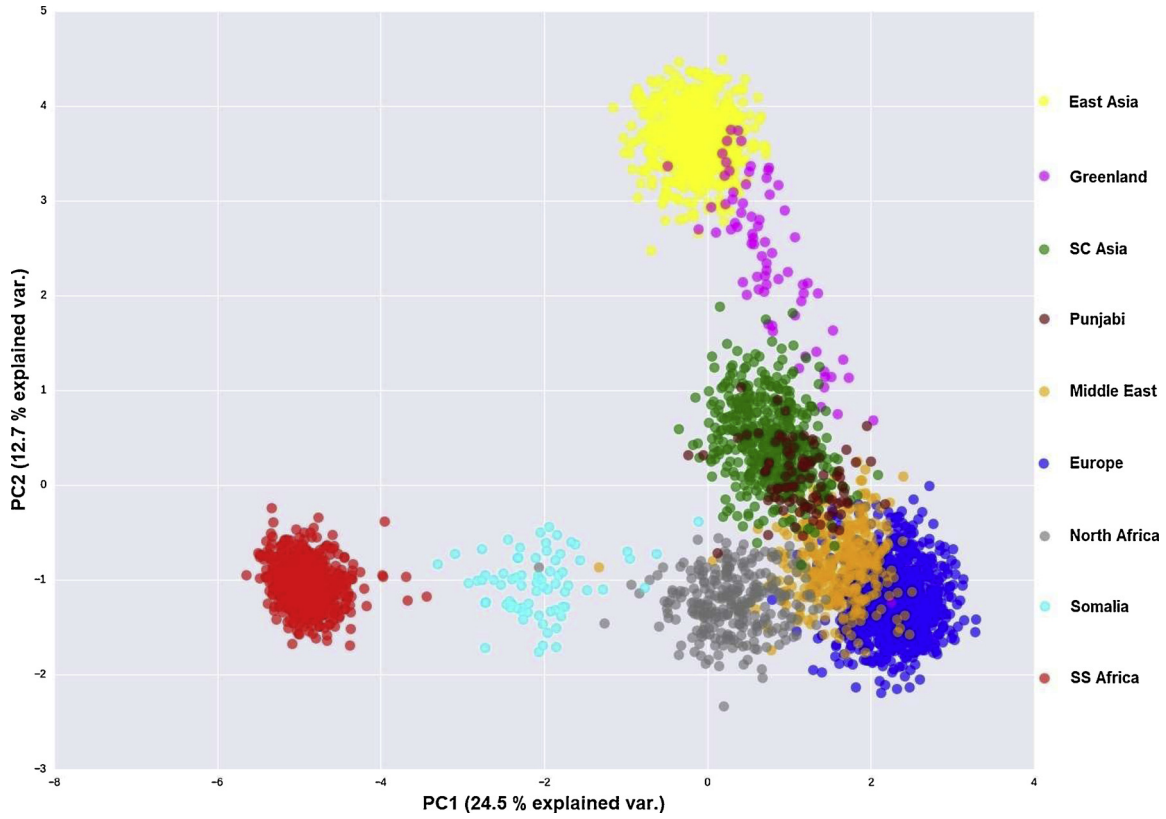

Fig. 1. PCA plot of the Punjabi individuals (shown in brown) and selected reference populations. Each coloured symbol represents an individual according to the population membership. Abbreviations used: SC Asia: South-Central Asia, SS Africa: Sub-Saharan Africa.

Table 1
Likelihood ratios for 16 randomly selected individuals typed with the Precision ID Ancestry Panel.

Individual	Most likely Metapopulation *	SC Asia/ME	SC Asia/N Africa	SC Asia/E Asia
01-PUN	SC Asia	7.67E+07	2.34E+14	5.16E+25
06-PUN	SC Asia	5.28E+01	1.75E+06	2.33E+36
08-PUN	SC Asia	1.13E+06	1.09E+08	2.96E+30
09-PUN	SC Asia	1.55E+02	1.76E+06	1.10E+40
12-PUN	SC Asia	1.69E+05	3.51E+12	9.48E+29
15-PUN	SC Asia	3.85E+05	1.08E+08	1.84E+32
21-PUN	SC Asia	5.51E+04	4.27E+08	6.47E+30
24-PUN	SC Asia	1.86E+04	1.78E+05	1.61E+36
30-PUN	SC Asia	5.24E+07	5.32E+13	7.00E+30
37-PUN	SC Asia	2.73E+03	8.87E+07	9.48E+29
46-PUN	SC Asia	1.01E+06	2.08E+12	1.42E+25
64-PUN	SC Asia	4.46E+07	3.83E+12	5.51E+26
72-PUN	SC Asia	6.08E+03	1.46E+09	1.99E+35
80-PUN	SC Asia	1.39E+07	1.39E+12	7.80E+19
91-PUN	SC Asia	5.32E+01	2.02E+09	7.10E+31
99-PUN	SC Asia	2.30E+04	7.59E+06	3.27E+33

Population abbreviations: SC Asia: South-Central Asia, ME: Middle East, N Africa: North Africa, E. Asia: East Asia.
Columns with LRs were coloured according to size: Bright red contains: Low LR, dark red contains: High LR.
*Based on the databases available in GenoGeographer [5;6].

GenoGeographer tool was used to infer ancestry and calculate the weight of the evidence (examples in Table 1). Likelihood ratios = P (DNA | H1)/P(DNA | H2) were calculated for 74 out of the 87 tested individuals (z-score ≤1.64). For 13 out of the 87 Punjabis (14.9%), no appropriate reference population was found in the database (z score > 1.64). Thus, ancestry inference could not be done. Of the 74 individuals with z-score ≤1.64, the most likely population of origin was South Central Asia (n = 71) or Middle East (n = 3).

3. Conclusions

The present study represents an attempt to evaluate the genetic composition of the Punjabi population. PCA indicated that the Punjabi population is an admixed population with genetic ancestry components

similar to those of other South Central Asian populations and the Middle East. The same was observed with Geno Geographer: The LRs in Table 1 revealed that Punjabis are more closely related to South Central Asian and Middle Eastern populations than to East Asian populations. To be able to distinguish Punjabi from these populations will most likely require a larger set of SNPs and/or a second tier panel specifically designed for these populations.

References

[1] I. Halder, M. Shriver, M. Thomas, et al., A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications, Hum. Mutat 29 (5) (2008) 648–658.

[2] R. Pereira, C. Phillips, N. Pinto, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, PLoS One 7 (1) (2012) e29684.

[3] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet 18 (2015) 49–65.

[4] M.D. Petraglia, B. Allchin, The Evolution and History of Human Populations in South Asia: Inter-disciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics, Springer Science & Business Media 6 (2007) ISBN 978-1-4020-5562-1.

[5] R.P. Wright, The Ancient Indus: Urbanism, Economy, and Society, Cambridge University Press, 2009, pp. 44–51 ISBN 978-0-521-57652-9.

[6] 2017 Census Archived 15 October 2017 at the Wayback Machine.

[7] V. Pereira, H.S. Mogensen, C. Børsting, et al., Evaluation of the Precision ID Ancestry Panel for crime casework: a SNP typing assay developed for typing of 165 ancestral informative markers, Forensic Sci. Int. Genet 28 (2017) 138–145.

[8] G.E. Themudo, H.S. Mogensen, C. Børsting, et al., Frequencies of HID-ion ampliseq ancestry panel markers among Greenlanders, Forensic Sci. Int. Genet 24 (2016) 60–64.

[9] K.K. Kidd, W.C. Speed, A.J. Pakstis, et al., Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet 10 (2014) 23–32.

[10] R. Kosoy, R. Nassir, C. Tian, et al., Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, Hum. Mutat 30 (1) (2009) 69–78 (2009).

[11] R. Nassir, R. Kosoy, C. Tian, et al., An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels, BMC Genet 10 (1) (2009) 39.

[12] Thermo Fisher Scientific, Ion AmpliSeq™ library preparation for human identification applications, Thermo Fisher Scientific Inc., Carlsbad, 2015.

[13] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, et al., Weight of the evidence of genetic investigations of ancestry informative markers, Theor. Popul. Biol 120 (2018) 1-10.

[14] H.S. Mogensen, T. Tvedebrink, C. Børsting, et al., Ancestry prediction efficiency of the software GenoGeographer using a z-score method and the ancestry informative markers in the Precision ID Ancestry Panel, Forensic Sci. Int. Genet (2019) in press.

[15] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol Ecol Resour 10 (2010) 564–567.