INT. J. BIOL. BIOTECH., 15 (3): 453-458, 2018.

GENETIC DIVERSITY IN MITOCHONDRIAL DNA CONTROL REGION AMONG SUB-ETHNIC GROUPS OF PUNJABI POPULATION FROM PAKISTAN

Abida Iqbal Ansari, Obaid Y. Khan and Maqsood Ali Ansari*

Department of Genetics, University of Karachi, Karachi-75270, Pakistan

***Corresponding Author:** Email: maansari@uok.edu.pk; +923343135626

ABSTRACT

Mitochondrial DNA is a putative tool used for genetic identification of human and tracing migration and sketching their evolution. To investigate the genetic structure of Pakistani population with reference to mitochondrial DNA control region and to explore the uniparental contribution, 50 maternally unrelated individuals with Punjabi ethnic background were screened. Forty eight haplotypes were identified that were assigned to 12 major haplogroups. Haplogroup M (38%) and haplogroup U (24%) formed the predominant groups. Haplogroup data showed that Punjabi population is mainly structured through South Asian and West Eurasian matrilineal lineages. Estimation of average pairwise genetic distance within and between 6 different Punjabi groups (Rajput, Awan, Arain, Jatt, Mughal and Gujjar) and phylogenetic analysis devised Arain as most genetically diverse and evolutionary distant division, followed by Mughal among the studied sub-ethnicities. This study is worthy contribution to augment population specific forensic database and deepens insight to the matrilineal evolution.

Keywords: Haplotypes, Haplogroups, Matrilineal Lineages, Forensic database, Phylogeny

INTRODUCTION

The DNA of Human Mitochondria is circular, 16,569 bp long, having thirty-seven genes and a control region (CR) which is a significantly large non-coding portion of the mitochondrial genome (Anderson *et al*., 1981; Andrews *et al*., 1999). Human mitochondrial DNA (mtDNA) analysis has been achieved the worthy position in population studies for its unique features like absence of recombination which promotes the clonal transmission down the generations (Stearns and Hoekstra, 2005), matrilineal mode of inheritance (Sato and Sato, 2013) allows to trace the maternal ancestor even as back to 171KYA to "Mitochondrial Eve" (Slate and Gemmell, 2004) and higher rate of mutation in control region (Van Oven and Kayser, 2009). Assuming the occurrence of mutational events at constant rate scientists have established mtDNA molecular clock to estimate the course of evolution (Slate and Gemmell, 2004). Therefore, archaeologists and anthropologists are able to interpret history of human populations through their genetic makeup. Also, the stability against exonucleases and high copy number increase the reproducibility of mtDNA genome even from ancient or degraded samples (Hoong and Lek, 2005). Number of mtDNA per mitochondrion ranges between 2-10 (Satoh and Kuroiwa, 1991) while number of mitochondria per cell is as low as 83 ± 17 and as high as 667 ± 80, dependent on the cell type (Robin and Wong, 1988). These features aid forensic investigations where the condition of samples is usually compromised. mtDNA control region (CR) sizes ¬1.1kb, carries three hypervariable regions (HVR-I, HVR-II and HVR-III). Polymorphisms in these regions play significant role in distinguishing two maternally unrelated individuals (Melton and Nelson, 2001).

Pakistan is located in South Asia, characterizes a wide spectrum of different ethnic groups out of which Sindhi, Balochi, Punjabi and Pathan are the four major groups. Punjabi is the largest (www.worldpopulationreview.com) and tremendously diversified ethnic group of Pakistan, contains tribes; such as Jatt, Arain, Awan, Rajput, Gujjar, Mughal, Chinioti etc (Rose *et al.*, 1911) and speak a representative language, Punjabi (Brown and Ogilvie, 2010). The linguistic, social and traditional fences escalate the indigenous marriages thus the heterogeneity conspicuously increase between and decrease within various ethnic groups (Quintana-Murci *et al*., 2004).

The present study was conducted to explore the mtDNA perspective of Punjabi population from Pakistan.

Materials and Methods:

Sampling:

In ETDA-Vacutainer 2-3cc of whole blood sample was collected from 50 healthy and unrelated individuals with Punjabi ethnic background, irrespective of age and sex. Signed consents were obtained from each participant.

Grouping of Sub-ethnicities:

The 50 randomly collected Punjabi individuals were grouped according to their sub-ethnic backgrounds and the groups covering 10% or greater proportion of the sampled population, included; Rajput, Awan, Arain, Jatt, Mughal and Gujjar; were selected for comparative analysis (Fig. **1**).

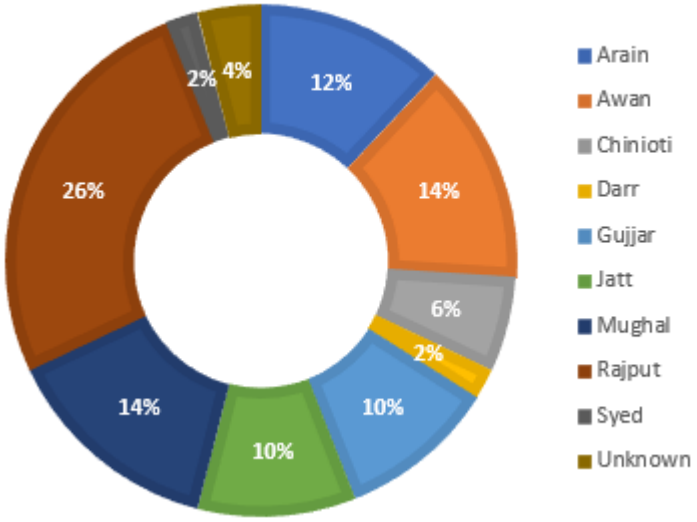

Fig.1. Distribution of sub-ethnicities in 50 Punjabi samples.

DNA Extraction and Quantification:

Genomic DNA was extracted using whole blood through commercially available MasterPure^{TM} Complete DNA and RNA Purification Kit. Extracted DNA was stored for analysis at -20^{0}C. Quantity and quality of DNA was assessed by UV-Spectrophotometry and Agarose Gel Electrophoresis.

Amplification of Control Region:

For amplification of mtDNA HVR-I (445bp) following primer set was used,

HV-I Forward: 5'-CTCCACCATTAGCACCCAAAG-3'

HV-I Reverse: 5'-CACCATCCTCCGTGAAATCA-3'

Co-amplification of HVR-II and III (591bp) was performed using the following primer pair,

HV-II+III Forward: 5'-GGTCTATCACCCTATTAACCAC-3'

HV-II+III Reverse: 5'-TATGTAGCTTACCTCCTCAA-3'

Primer sequences were obtained in accordance with Hoong and Lek 2005. The PCR was performed in MultiGene™ OptiMax Thermal Cycler (Labnet International, Inc., USA) with reaction volume of 25ul; contained 1X Taq Master Mix (Bioron, Life science), 0.1ug of each primer and 20ng of template DNA. The cycling conditions are given in Table 1. PCR products were purified with ExoSAP-IT® (Affymetrix, Inc., USB® products, USA) following manufacturer's guidelines. Purified products were sequenced using same primer sets from Macrogen Inc., Seoul Korea.

Table 1. PCR Cycling Conditions.

Initial Denaturation	94^{0}C for 2 min	
Denaturation	94^{0}C for 30sec	
Annealing	57^{0}C for 30sec	35cycles
Extension	72^{0}C for 1min	
Final Extension	72^{0}C for 10min	
Hold	4^{0}C	

Data Analysis:

DNA sequence alignment was made with revised Cambridge Reference Sequence (Andrews et al., 1999) using ClustalW. Major Haplogroups corresponding to PhyloTree v.17 (http://www.phylotree.org) were determined. MEGA 7 (Kumar *et al.*, 2016), a software for evolutionary analysis, was used to evaluate average pairwise genetic distances between and within groups of Punjabi individuals. Furthermore, Phylogenetic relationship was figured out by constructing neighbour-joining (NJ) and Minimum Evolutionary (ME) tree using MEGA 7 (Kumar *et al*., 2016).

RESULTS AND DISCUSSION

Sequence Diversity and Haplogroup Assignment:

Sequence analysis of HVRI (16024-16519) and HVRII+III (64-576) of 50 individuals revealed that HVRI was more polymorphic, possessed 73 (20.3%) variable sites while HVRII+III showed 60 (11.5%). A few sites showed high level of polymorphism, like position 195, with C in 9 sequences (18%) and A in 1 (2%) and position 529 exhibited a T and A, each with frequency of 2%. Also 195G was reported in Sindhi population while variation at position 529 was not evident in Sindhi (Bhatti *et al*., 2017**b**), Saraiki (Hayat *et al*., 2015), Hazara (Rakha *et al*., 2017), Kashmiri (Rakha *et al*., 2016) and previously studied Punjabi population (Bhatti *et al.*, 2017**a**). None of characteristic deletions GTA at 16061-16063 and CCC at 16191-16193 were shown in present study as were reported in Gujjars and Awans from Mansehra, respectively (Akbar *et al*., 2016). The data produced 48 haplotypes which were congregated to 12 major-haplogroups according to nomenclature in PhyloTree build 17 (http://www.phylotree.org). Frequency of each major-haplogroup is presented in Fig. **2**. Haplogroups with South Asian lineage were found in highest proportion i.e. M (38%); R (10%) followed by West Eurasian U (24%); J, HV (6% each); I (4%); H, W, T (2% each). Haplogroup M is derived from sub-Saharan African Haplogroup L3 (Quintana-Murci *et al*., 1999). The diverse distribution of haplogroup M is traced to South Asia (Quintana-Murci et al., 2004). It has been illustrated that Haplogroup M of mtDNA significantly indicates migration of ancient man to Indian subcontinent following Southern path (Maji *et al.*, 2009). The recent studies showed higher frequency of haplogroup M in various sub-populations of Pakistan; 42% in Sindhi (Bhatti *et al*., 2017**b**), 42% in Punjabi from previous study (Bhatti *et al*., 2017**a**), 24.7% in Kashmiri (Rakha *et al*., 2016), whereas in Saraiki U (Hayat *et al*., 2015), in Makrani L (Siddiqi *et al*., 2014) and H in Hazara (Rakha *et al.*, 2017) and Hazarewal (Akbar *et al.*, 2016) were frequently observed haplogroups. The unisex genetic contribution to multiethnic groups structured mainly through South Asian and West Eurasian mtDNA lineages.

Estimation of Genetic Distances:

Mean pairwise genetic distance between sub-groups of Punjabi population was performed using Kimura 2-parameter (Kimura, 1980), applied with 1000 bootstrap replications to compute standard errors. Rajput and Jatt were noticed much closer to each other (d=0.0105 ± 0.002) while Arain and Gujjar were relatively distant (d=0.0152 ± 0.002). Genetic diversity among Arain and Gujjar were computed as 0.978 and 0.947 respectively thus Arain has shown more number of haplotypes than Gujjars (Bhatti *et al*., 2017**a**). Overall, on the basis of nucleotide substitutions, evolutionary distance between each group found was not more than 1.52%. Mean genetic distance estimates within groups indicate evolutionary closeness (0.9%) in Jatt whereas Arain shows comparatively greater distance (1.6%) (Table 2).

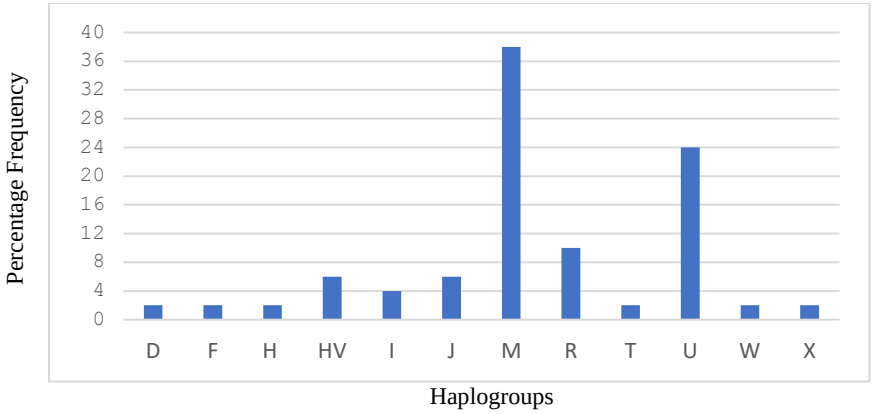

Fig.2. Frequency (%) distribution of major haplogroups in 50 Punjabi Individuals as per PhyloTree build**.**

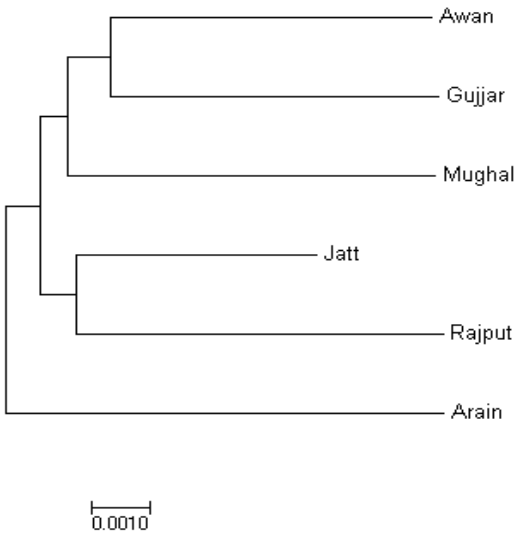

Fig.3. Neighbour-joining tree, branches are drawn to scale. The optimal tree with the sum of branch length 0.038 is shown. The evolutionary distances were computed in the units of number of base substitution per site. All positions with gaps and missing data were eliminated.

Table 2. Mean Pairwise Genetic distances between sub-ethnic groups are shown in lower half and the diagonal represents within group distances (SE=0.002 for each value).

	Mughal	Jatt	Awan	Rajput	Arain	Gujjar
Mughal	***0.014***					
Jatt	0.0115	***0.009***				
Awan	0.0129	0.0114	***0.012***			
Rajput	0.0139	0.0105	0.0136	***0.013***		
Arain	0.0149	0.0130	0.0148	0.0150	***0.016***	
Gujjar	0.0125	0.0117	0.0112	0.0137	0.0152	***0.011***

Phylogenetic Analysis:

Phylogenetic analysis portrayed the evolutionary pattern for the six Punjabi sub-ethnicities. Neighbor-joining tree construction (Saitou and Nei, 1987) displayed that Awan are closely related to Gujjar, and Rajput have a close affinity with Jatt, whereas Arain located differentially at the basal branch exhibited higher diversity among rest of the groups (Fig. **3**). Minimum evolution tree was also constructed as per (Rzhetsky and Nei, 1991) presented the same topology as by neighbor-joining (NJ).

GenBank Accession Numbers:

The mtDNA sequence data can be accessed from GenBank-NCBI through following accession numbers: MG491220 to 48, MG407182 to 201 and MG386497

Conclusion

mtDNA Single Nucleotide Polymorphism (SNP) data from 50 unrelated Punjabi individuals revealed high degree of genetic diversity. The six Punjabi sub-ethnic groups presented in this study found to be close and admixture of each other. We did not witness any tribe specific mutation in the control region. This is the preliminary study conducted with the collected data, further analysis is intended to be done with more detailed parameters, taking the Punjabi samples as a single group.

Acknowledgement

We highly acknowledge the individuals who donated their blood samples to the study. Also, we would like to thank Dr. Khitab Gul and Mr. Noman for their valuable efforts in sample collection.

REFERENCES

Akbar, N., H. Ahmad, M. S. Nadeem, B. E. Hemphill, K. Muhammad, W. Ahmad and M. Ilyas (2016). HVSI polymorphism indicates multiple origins of mtDNA in the Hazarewal population of Northern Pakistan. *Genet Mol Res.*, 15(2): doi: 10.4238/gmr.15027167

Anderson, S., A.T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, ...and I.G. Young (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290: 457. doi: 10.1038/290457a0

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull and N. Howell (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.,* 23(2): 147. doi: 10.1038/13779

Bhatti, S., S. Abbas, M. Aslamkhan, M. Attimonelli, M. S. Trinidad, H. H. Aydin,, . . .and G.R. Gonzalez (2017a). Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA Part A*, 1-13. doi: 10.1080/24701394.2017.1350951

Bhatti, S., M. Aslamkhan, M. Attimonelli, S. Abbas and H.H. Aydin (2017b). Mitochondrial DNA variation in the Sindh population of Pakistan. *Australian Journal of Forensic Sciences,* 49(2): 201-216. do Hi: 10.1080/00450618.2016.1144788

Brown, K. and S. Ogilvie (2010). *Concise Encyclopedia of Languages of the World* (ISBN 978-0-08-087775-4 ed., pp. 522–523): Elsevier.

Hayat, S., T. Akhtar, M. H. Siddiqi, A. Rakha, N. Haider, M. Tayyab, , . . and F. M. Khan (2015). Mitochondrial DNA control region sequences study in Saraiki population from Pakistan. *Leg Med (Tokyo)*, 17(2): 140-144. doi: 10.1016/j.legalmed.2014.10.010

Hoong, L. L. and C. Lek (2005). Genetic polymorphisms in mitochondrial DNA hypervariable regions I, II and III of the Malaysian population. *Asia Pacific Journal of Molecular Biology and Biotechnology*, 13(2): 79-85.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol, 1*6: 111-120.

Kumar, S., G. Stecher and K. Tamura (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.,* 33(7): 1870-1874. doi: 10.1093/molbev/msw054

Maji, S., S. Krithika and T.S. Vasulu (2009). Phylogeographic distribution of mitochondrial DNA macrohaplogroup M in India. *J Genet*, 88(1): 127-139.

Melton, T. and K. Nelson (2001). Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States. *Croat Med J.*, 42(3): 298-303.

Quintana-Murci, L., R. Chaix, R.S. Wells, M. D. Behar, H. Sayar, R. Scozzari, , . . .and K. McElreavey (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet.,* 74(5): 827-845. doi: 10.1086/383236

Quintana-Murci, L., O. Semino, H.J., Bandelt, G. Passarino, K. McElreavey and A.S. Santachiara-Benerecetti (1999). Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet.*, 23(4): 437-441. doi: 10.1038/70550

Rakha, A., Fatima, M.-S. Peng, A. Adan, R. Bi, M. Yasmin and Y.-G. Yao (2017). mtDNA sequence diversity of Hazara ethnic group from Pakistan. *Forensic Science International: Genetics,* 30: e1-e5. doi: 10.1016/j.fsigen.2017.07.004

Rakha, A., M.-S.Peng, R. Bi, J.-J. Song, Z. Salahudin, A. Adan, . . . and Y.-G. Yao (2016). EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu & Kashmir, Pakistan. *Forensic Science International: Genetics*, 25: 125-131. doi: 10.1016/j.fsigen.2016.08.009

Robin, E. D. and R. Wong (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol.,* 136(3): 507-513. doi: 10.1002/jcp.1041360316

Rose, H. A. I., Denzil, Maclagan and Edward Douglas (1911). *A glossary of the tribes and castes of the Punjab and North-West frontier province* (Vol. II).

Rzhetsky, A. and M. Nei (1991). *A Simple Method for Estimating and Testing Minimum-Evolution Trees* (Vol. 9).

Saitou, N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.,* 4(4): 406-425. doi: 10.1093/oxfordjournals.molbev.a040454

Sato, M. and K. Sato (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(8): 1979-1984. doi: https://doi.org/10.1016/j.bbamcr.2013.03.010

Satoh, M. and T. Kuroiwa (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res.,* 196(1): 137-140.

Siddiqi, M. H., T. Akhtar, A. Rakha, G. Abbas, A. Ali, N. Haider, . . . and F.M. Khan (2014). Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. *Legal Medicine,* 17(2): 134-139. doi: 10.1016/j.legalmed.2014.09.007

Slate, J. and N.J. Gemmell (2004). Eve 'n' Steve: recombination of human mitochondrial DNA. *Trends in Ecology & Evolution,* 19(11): 561-563. doi: https://doi.org/10.1016/j.tree.2004.09.001

Stearns, S. C. and R.F. Hoekstra (2005). *Evolution: An Introduction* (second ed.). Oxford: Oxford University Press.

Van Oven, M. and M. Kayser (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.,* 30(2): E386-394. doi: 10.1002/humu.20921

www.worldpopulationreview.com

www.phylotree.org

(*Accepted for publication June 2018*)