Genetics and Molecular Research

HVSI polymorphism indicates multiple origins of mtDNA in the Hazarewal population of Northern Pakistan

N. Akbar[1,3], H. Ahmad[1], M.S. Nadeem[1], B.E. Hemphill[2], K. Muhammad[1], W. Ahmad[1] and M. Ilyas[3]

[1]Department of Genetics, Hazara University, Garden Campus, Mansehra, Pakistan
[2]Department of Anthropology, University of Alaska Fairbanks, Fairbanks, AK, USA
[3]Center for Human Genetics, Hazara University, Mansehra, Pakistan

Corresponding author: H. Ahmad
E-mail: drhahmad@gmail.com

Genet. Mol. Res. 15 (2): gmr.15027167
Received November 23, 2015
Accepted February 1, 2016
Published June 24, 2016
DOI http://dx.doi.org/10.4238/gmr.15027167

ABSTRACT. Mitochondrial DNA (mtDNA) is an important tool used to explore ethnogenetics and the evolutionary history of human populations. In this study, hypervariable segment I (HVSI) from mtDNA was analyzed to establish the genetic lineage of the Hazarewal populations residing in the Mansehra and Abbottabad districts of Northern Pakistan. HVSI was extracted from genetic specimens obtained from 225 unrelated male and female individuals belonging to seven distinct Pakistani ethnic groups (31 Abbassis, 44 Awans, 38 Gujars, 16 Jadoons, 23 Karlals, 33 Syeds, and 40 Tanolis). Eighty-three haplogroups, 39 of which were unique, were identified; haplogroup H was predominantly represented (in 40% of the people), followed by haplogroups M (21.78%), R (16.89%), N (15.56%), L (3.11%), and HV (2.67%). The results revealed a sex-biased genetic contribution from putative West Eurasian, South Asian, and Sub-Saharan populations to

the genetic lineage of Hazarewal ancestry, with the effect of Eurasians being predominant. The HVSI nucleotide sequences exhibited some characteristic deletion mutations between 16,022 and 16,193 bp, which is characteristic of specific ethnic groups. HVSI sequence homology showed that Hazarewal populations fall into three major clusters: Syeds and Awans sorted out into cluster I; Tanolis, Gujars, and Karlals segregated in cluster II; and Abbassis and Jadoons in cluster III. Here, we have reported the firsthand genetic information and evolutionary sketch of the selected populations residing alongside the historical Silk Route, which provides a baseline for collating the origin, route of migration, and phylogenetics of the population.

Key words: Pakistan; Hazara Division; Ethnicity; mtDNA; Phylogenetics

INTRODUCTION

The heritage of Pakistan, which can be traced back to the Indus valley civilization in 2500 BC, has been very well preserved (Mezzera and Aftab, 2009; Ali et al., 2010). Pakistan has a population exceeding 170 million (Jaffery and Sadaqat, 2006), which has been divided into 31 administrative divisions. The Hazara Division (Figure 1), in the northeastern region of Pakistan, is inhabited by people of various ethnic groups, including the Awan, Dhund, Gujar, Jadoon, Kashmiri, Kharral, Kohistani, Pathan, Syed, Swati, Tanoli, and Turk (Rose, 1907; Watson, 1908; Truman, 1909; William, 1910; Rose, 1911). The local ethnic groups of this Division speak several languages, including Hindko, Pushto, Gojri, and Kohistani (Qadeer, 2006). The origin of the inhabitants of the Hazara region is polyphyletic in nature, as evidenced by the differences in their culture and languages (Awan, 2009), and therefore needs scientific exploration. The conformance of the historical review with the results of archeological analyses is one method with which this can be accomplished. Another, more reliable method would be to analyze the phylogenetic data extracted from sequenced DNA. For example, human DNA analysis has been successfully used in the past to explore the historical movements of populations (Cann et al., 1987; Cavalli-Sforza et al., 1994).

The properties of mitochondrial DNAs (mtDNAs), including their lack of recombination, fast evolutionary rate, maternal inheritance, higher copy number per cell, and higher population-specific polymorphisms, makes them distinct from nuclear DNAs and valuable genetic markers in population genetics and molecular anthropology (Bandelt et al., 2001; Forster et al., 2002; Pakendorf and Stoneking, 2005; Fang et al., 2015). Specifically, mtDNA could be a reliable tool for sketching the early history, origin, and pathways of migration from the maternal aspect. Therefore, the primary purpose of this study was to obtain mtDNA data from a geographical region that lies alongside the historical Silk Road (the only route linking the Middle East, Iran, and Pakistan to China and the pathway of several invaders from Afghanistan and central Asia). This investigation was believed to generate more data to help trace the migration and expansion of modern humans in this area. In this study, we analyzed the mtDNA of people belonging to the various ethnic groups in central Hazara comprising the Abbottabad and Mansehra districts, area of 6558 km^2, and a population of 3 million (Ghulam, 2003; Ali, 2006).

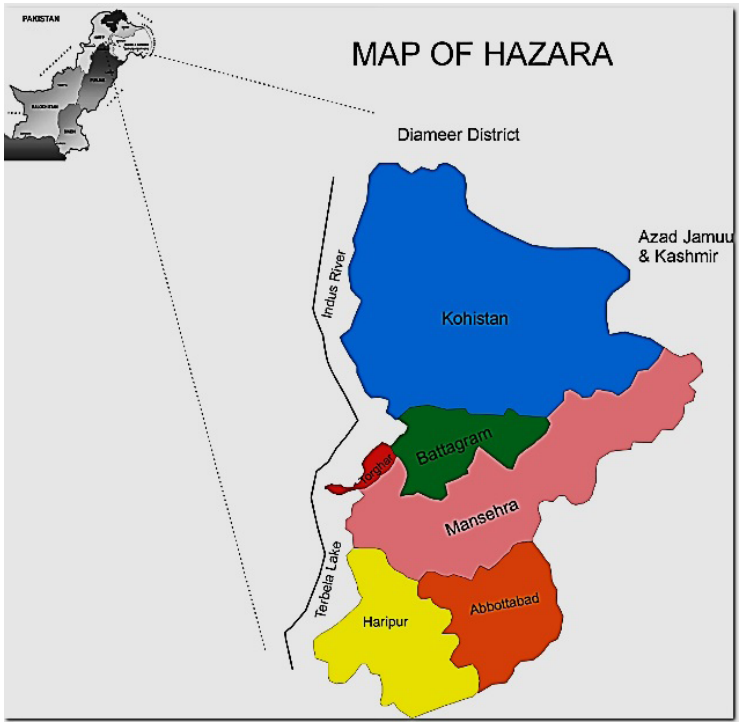

Figure 1. Geographic map of Hazara Division in the Khyber Pakhtunkhwa Province of Pakistan, showing the main geographic locations of sampling (Mansehra and Abbottabad).

MATERIAL AND METHODS

Sampling and DNA extraction

This phylogeographic study included major tribal populations residing in the northwestern regions of Pakistan, encompassing the Hazara division of the Khyber Pakhtunkhwa Province located to the west of Azad Jammu and Kashmir, north of Punjab, east of Tarbela dam, and south of Gilgit Baltistan (Figure 1). Saliva samples were obtained from 225 unrelated healthy individuals belonging to seven different ethnic tribes, including the Awans, Syeds, Abbassis, Tanolis, Karlals, Jadoons, and Gujars (Table 1). Signed informed consent forms were obtained from all individuals. DNA was isolated using a phenol:chloroform protocol modified from that used in a previous study (Ralser et al., 2006).

Nucleotide sequence analysis of mitochondrial hypervariable segment I (*HVSI*)

The extracted DNA was used as a template for PCR amplification of the 451-bp fragment at the 15,974-16,425-bp region of mtDNA, containing HVSI. The 25-µL PCR mixture was comprised of 1X Taq buffer, 2.5 mM dNTPs, 2.0 mM $MgCl_2$, 10 pM each of the forward and reverse primers, 2.5 U *Taq* DNA polymerase (Enzynomics, Korea, Cat. P050B), and 30-40 ng template DNA. The following primers were used for PCR

amplification: forward, 5'-CTCCACCATTAGCACCCAAAGCTAAG-3' and reverse, 5'-GATATTGATTTCACGGAGGATGGTGGTC-3'. The reaction was performed in an Applied BioSystems-2720 thermocycler (Applied Biosystems, Foster City, CA); the reaction conditions were set as follows: initial denaturation at 94°C for 4 min; 35 cycles of denaturation at 94°C for 40 s, annealing at 56°C for 1 min, and extension at 72°C for 1 min, and a final extension at 72°C for 5 min. The nucleotide sequences of purified PCR fragments obtained from individuals of different tribal populations were commercially analyzed by Macrogen (Geumcheon, South Korea).

Table 1. Summary of the sample size and place of collection.

No.	Sampling site	Sampling size	Ethnic group
1	Adola and Dalola/Abbottabad	31	Abbassis
2	Mansehra and Dhodial/Mansehra	44	Awans
3	Attar Sheesha/Mansehra	38	Gujars
4	Havalian and Nawan shehr/Abbottabad	16	Jadoons
5	Barmi Gali and Nathia gali/Abbottabad	23	Karlarrs
6	Mansehra, Dhodial, and Shinkiari/Mansehra	33	Syeds
7	Lassan Nawab/Mansehra	40	Tanolis

HVSI-based genetic haplogroup polymorphism

The basic nucleotide data were aligned and compared to the revised Cambridge reference sequence (CRS) (Anderson et al., 1981; Andrews et al., 1999). The genetic haplotypes of *HVSI* sequences were analyzed using the online database MTHAP (http://dna.jameslick.com/mthap/). All haplotypes were compared to previously reported haplotypes from different ethnic groups of Pakistan, as well as the rest of the world. Sequences were aligned using BioEdit (Hall, 1999). The percentage frequency of different genetic haplogroups was determined individually in all populations. All sequences were finally subjected to a quasi-median (QM) network analysis using the NETWORK software, available on the EMPOP website (www.empop.org) (Parson et al., 2004; Parson and Dür, 2007).

Tribe-specific nucleotide polymorphism

The *HVSI* nucleotide sequences were aligned to the modified CRS (Anderson et al., 1981; Andrews et al., 1999) and characteristic deleted nucleotides were identified. Sequences obtained from all tribes were separately compared to the standard sequence, and the sequence homology of *HVSI* sequences between different populations was determined.

RESULTS

HVSI sequence-based frequency of haplotypes

Whole genomic DNA was extracted from the saliva of 225 unrelated individuals, amplified by PCR, and purified on an agarose gel. Analysis of the *HVSI* sequence in the mtDNA by PCR revealed the p83 different haplotypes. The haplogroup frequencies observed in these samples are summarized in Table 2. Thirty-nine (39/83; 47%) haplogroup were

observed once, 15 (15/83; 18%) were observed twice, and 12 (12/83; 14.5%) were observed thrice. The haplogroup H was predominantly observed in these individuals (89; 39.7%), followed by haplogroup M (49/89; 22%). The latter haplogroup is predominantly observed in South Asians (Figure 2). The haplogroup H was observed in people belonging to the Abbassis (29.63%), Awan (22.73%), Gujar (58.97%), Tanoli (35%), Syeds (48.48%), Jadoon (56.25%), as well as Karlals (25%) tribes. The sister clade T, on the other hand, was quite frequent in Abbassis (11.11%), and the haplogroups U7 and N1 were frequently observed in Awans (9.09%), Tanoli (5%), and Syeds (9.09%), and Karlals (20.83%), respectively. Moreover, the prevalent haplogroup M (in Pakistan) was observed in Abbassis (22.22%), Awans (15.91%), Gujars (7.69%), Tanolis (22.50%), Syeds (21.21%), Jadoons (6.25%), and Karlals (16.67%).

Table 2. Frequency of haplogroups in the seven tribes residing in the Hazara division of Khyber Pakhtunkhwa Province, and comparison with the haplogroups previously reported in Pakistani populations.

Population	N	H*	HV	R	J	T1	T2	U4	U7	N1	X	M	L	Other	Study
Baluchis	39	20.50	10.30	-	7.70	-	-	2.60	2.60	5.20	-	33.30	2.60	7.70	Ottoni et al. (2011)
Hazarewal	23	13.00	4.30	-	-	-	-	8.70	4.30	-	-	30.30	-	26.10	Ottoni et al. (2011)
Pakistanis	100	12.00	4.00	2.00	1.00	-	1.00	-	5.00	3.00	1.00	49.00	1.00	8.00	Ottoni et al. (2011)
Pathans	44	4.50	2.30	6.80	6.80	4.50	4.50	-	-	4.60	-	29.50	-	9.10	Ottoni et al. (2011)
Sindhis	23	8.70	4.30	-	-	-	-	-	8.70	8.70	-	30.40	-	-	Ottoni et al. (2011)
Hazara, Abbassis	27	29.63	-	11.11	-	-	11.11	-	3.70	7.41	-	22.22	3.70	11.11	Present study
Hazara, Awans	44	22.73	2.27	6.82	2.27	4.55	4.55	2.27	9.09	13.64	2.27	15.91	2.27	11.36	Present study
Hazara, Gujars	39	58.97	-	5.13	2.56	-	2.56	-	2.56	7.69	2.56	7.69	-	10.26	Present study
Hazara, Tanolis	40	35.00	5.00	2.50	2.50	-	2.50	-	5.00	7.50	5.00	22.50	7.50	5.00	Present study
Hazara, Syeds	33	48.48	6.06	-	3.03	-	-	-	9.09	3.03	-	21.21	3.03	6.06	Present study
Hazara, Jadoons	16	56.25	-	6.25	-	-	-	-	-	-	-	6.25	-	31.25	Present study
Hazara Karlarrs	24	25.00	-	-	-	-	-	-	-	20.83	8.33	16.67	4.17	25.00	Present study

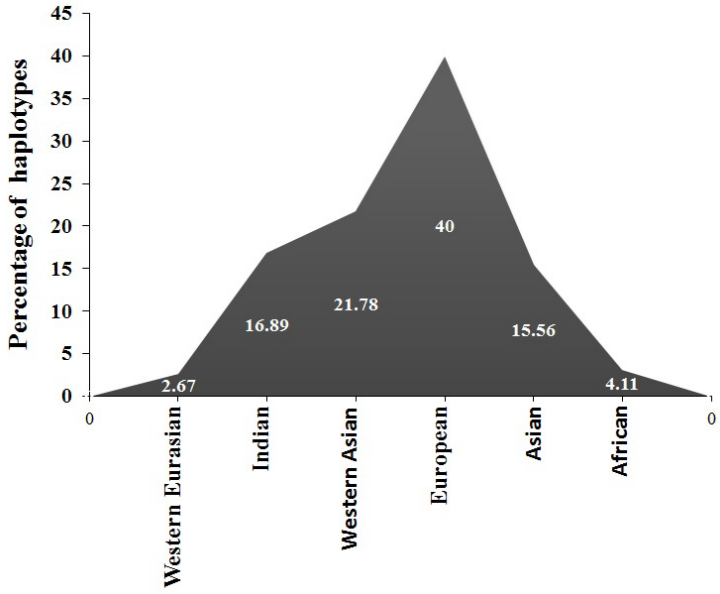

Figure 2. Percentage influence of different ethnic populations on mitochondrial DNA of tribes in the Hazara division. Macrohaplogroups observed in the seven tribes indicate the presence of haplotypes H (European; 40%), M (Western Asian; 21.78%), R (Indian; 16.89%), N (Asian; 15.56%), L (African; 3.11%), and HV (Western Eurasian; 2.67%).

Tribe-specific nucleotides

We compared the HVSI consensus sequences from each tribe against the CRS and observed a 42 to 93% homology; this was then used to construct a relatedness tree (Figure 3). The nucleotide sequences of individuals from all tribes were compared to the normal human mtDNA sequence, as well as the nucleotide sequences of individuals from other populations. We observed some deletions that were conserved among the populations. For example, thymine (T) at position 16,022 was missing in the Syed tribe from Mansehra district (**S1 Figure**), while the Gujar population showed a GTA deletion at 16,061-16,063. On the contrary, the Tanolis, Awans, and Abbassis showed a 13-nucleotide deletion (GAAGCAGATTTG) from 16,035 to 16,047, CCC deletion at 16,191-16,193, and a deletion in the fragment ATTTGGG between 16,043 and 16,049, respectively (**S2**, **S3**, **S4**, and **S5 Figures**). These deletions can be applied as characteristic features of the populations under investigation.

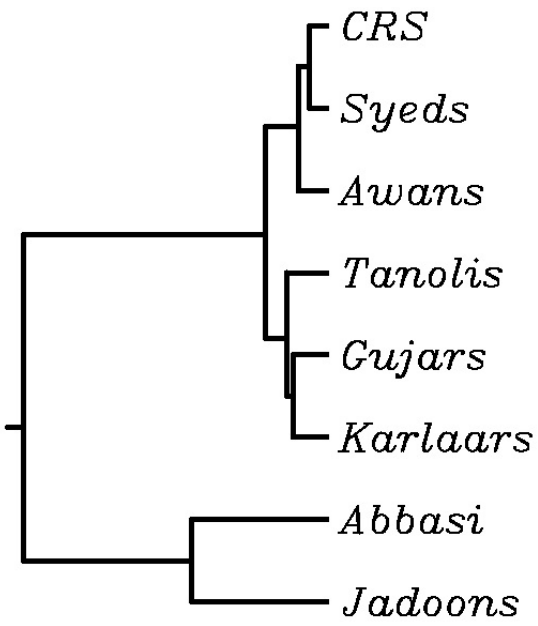

Figure 3. Affinities among the mitochondrial DNA of local populations. The relatedness of different tribes based on the HVSI sequence, based on the Cambridge reference sequence.

Sequence data quality assessment

The QM network allows us to examine the quality of mtDNA data by graphically representing the genetic structure of the lineages in a data set (Figure 4). Data sets from all seven populations were combined and scrutinized by a QM network analysis, in order to identify unusual polymorphisms. Networks visualize mtDNA data in a way that can help trace the evolutionary patterns of mtDNA phylogeny. The dataset was filtered using EMPOPspeedyWE (www.empop.org), which resulted in a star-shaped HV-I network without alarming reticulations.

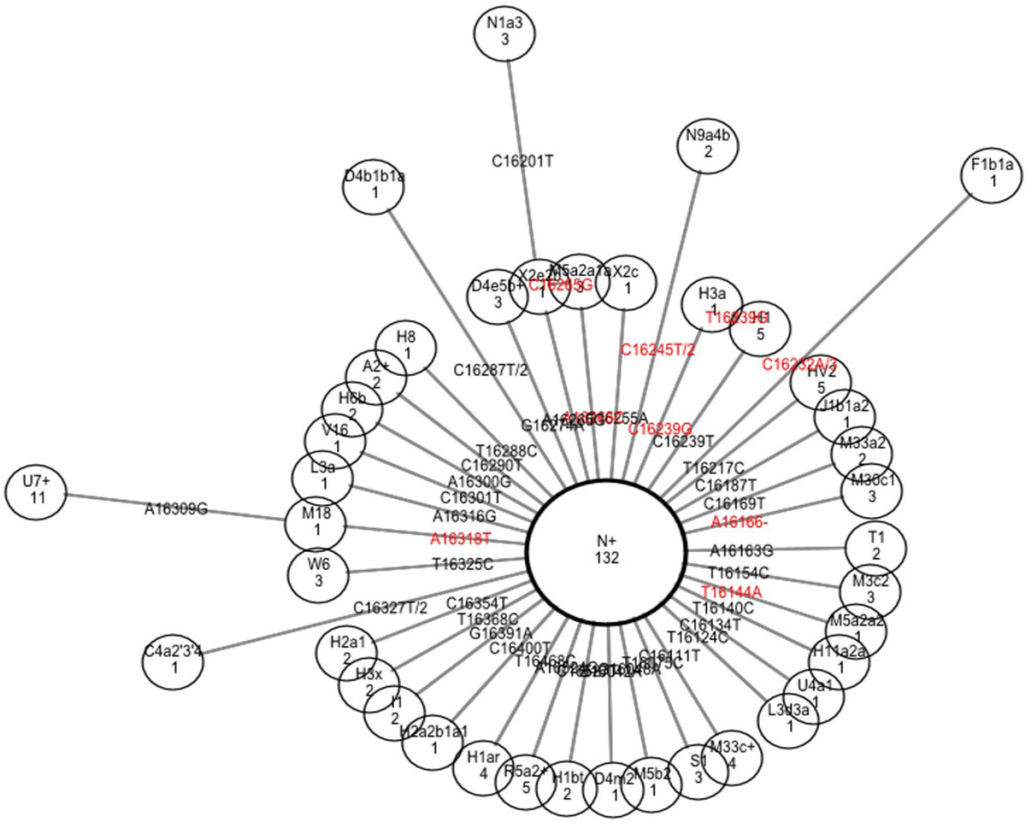

Figure 4. Phylogenetic clustering of the mitochondrial DNA haplotypes based on *HVSI* sequence observed in Hazarewal populations. Haplotypes are shown with circles, with the number of carriers indicated.

DISCUSSION

The geographical area constituting Pakistan has been a site of extensive and repetitive invasions, and incorporates several ancient trade routes. Several previous studies have analyzed the mtDNA of various Indian populations, thereby providing some insight into the genetic structure of the populations in India (Macaulay et al., 1999; Basu et al., 2003; Kivisild et al., 2003; Roy et al., 2003). However, there has been no systematic investigation into the genetic composition and affinities of northwestern Pakistani populations. In this study, we analyzed the polymorphism of HVSI haplotypes in seven tribes of the Hazara division, a part of northern Pakistan (Figure 1). mtDNA was extracted from 225 unrelated individuals. Eighty-three haplogroups (39 unique) were identified, among which haplogroups H (40%), M (21.78%), R (16.89%), N (15.56%), L (3.11%), and HV (2.67%) were predominant. The findings of mtDNA haplogroup analysis suggest a genetic sex-biased contribution from putative west Eurasian, south Asian, and Sub-Saharan populations in the genetic lineage of Hazarewal ancestry, with the most predominant effect being attributed to Eurasians. The haplogroups H, M, R, N, L, and HV are specifically associated with European, western Asian, Indian, Asian, Eurasian, and African populations, respectively. Therefore, the mtDNA was

believed to have been influenced by multiple populations from across the world. Previous studies have indicated that all mtDNA lineages in India and Pakistan are partially linked to the Western Eurasian mtDNA phylogeny, while including abundant basal branches that are absent in Europe (Kivisild et al., 1999). This population may have entered Pakistan from the west via Iran or the north via central Asia (Quintana-Murci et al., 2004). The gene flow is more limited in the opposite direction and not very far-reaching. For example, 4 haplogroups, one each of R5 and N5, and two of M, were identified from among 42 mtDNA lineages from central Iran; however, only one distinct lineage from the macrohaplogroup M from among 95 identified in Northern and western Iran potentially belong to the South Asian haplogroups (Quintana-Murci et al., 2004). Only one R5 lineage has been reported in Iraqi samples (Al-Zahery et al., 2003). Such lineages were virtually absent further north-west, in Caucasus and Turkey. In the Central Asian data set of Comas et al. (2004), only six of 232 lineages, 2 U2a, 1 U2c, 2 R5, and possibly 1 M4, belonged to South Asian haplogroups, whereas the Western Eurasian mtDNA lineage could be assembled into multiple haplogroups (HV, N1I, N2W, R1, R2, JT, UK, and X) (Torroni et al., 2006; Macaulay et al., 1999; Tambets et al., 2004). Lineage analysis of mtDNA phylogeny shows that all mtDNA haplogroups outside of Africa are descendants of either haplogroup M or its sibling haplogroup N (Metspalu et al., 2004). The geographical distributions of M and N, excluding African migrations and their subsequent habitation throughout the rest of the world, is often considered to be a single major prehistoric migration of humans out of Africa, which shows that the M and N haplogroups were part of this habitation process (Macaulay et al., 2005). Moreover, haplogroup M has not penetrated west of the Indus Valley, although it is present at high frequencies in south Pakistani and Indian populations. The polynucleotide sequences obtained in this study were compared to the human mtDNA sequence available in GenBank. Person-to-person variations in the HVSI were recorded, and their importance in the forensic analyses and population genetics were recorded. The results revealed that the genetic pattern of the seven tribes of Mansehra and Abbottabad districts is similar to that observed in West Eurasia in a previous report (Richards et al., 2000; Achilli et al., 2004; Metspalu et al., 2004; Pereira et al., 2005; Rajkumar et al., 2005). Moreover, we discovered that the ancestral node of the phylogenetic tree of all mtDNA types was typically in Central Asia; Middle Eastern and European phylogenies were also found in South Asia at relatively high frequencies. The datasets from all seven populations, combined and scrutinized by QM network for the identification of unusual polymorphisms, showed that the mtDNA data was consistent with anthropological and linguistic theories that suggested that the migration of early Indo-European-speaking farmers took place from West Asia into Europe and India (Renfrew, 1992). Furthermore, variations in the frequency distribution of sub clades or macrohaplogroup were attributed to the accumulation of particular mutations in the tribes because of geographic or social isolation of the people; alternately, it was attributed to the migration/colonization distance of the people residing in this area.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This paper forms a section of the PhD study of the principal author. The authors are

grateful to all donors of genetic material for their valuable cooperation, as well as the Higher Education Commission of Pakistan for financially supporting the study under the project, entitled "Ethnogenetic Elaboration of KP through Dental Morphology and DNA analysis".

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, et al. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* 75: 910-918. http://dx.doi.org/10.1086/425590

Ali A, Ahmed H and Idrees M (2010). Molecular epidemiology of Hepatitis C virus genotypes in Khyber Pakhtoonkhaw of Pakistan. *Virol. J.* 7: 203. http://dx.doi.org/10.1186/1743-422X-7-203

Ali I (2006). Mapping and documentation of the cultural assets of Kaghan Valley, Mansehra. UNESCO, Islamabad.

Al-Zahery N, Semino O, Benuzzi G, Magri C, et al. (2003). Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* 28: 458-472. http://dx.doi.org/10.1016/S1055-7903(03)00039-3

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465. http://dx.doi.org/10.1038/290457a0

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, et al. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23: 147. http://dx.doi.org/10.1038/13779

Awan MSK (2009). History of Awan. Al- Faisal Nashran, Lahore.

Bandelt HJ, Lahermo P, Richards M and Macaulay V (2001). Detecting errors in mtDNA data by phylogenetic analysis. *Int. J. Legal Med.* 115: 64-69. http://dx.doi.org/10.1007/s004140100228

Basu A, Mukherjee N, Roy S, Sengupta S, et al. (2003). Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* 13: 2277-2290. http://dx.doi.org/10.1101/gr.1413403

Cann RL, Stoneking M and Wilson AC (1987). Mitochondrial DNA and human evolution. *Nature* 325: 31-36. http://dx.doi.org/10.1038/325031a0

Cavalli-Sforza LL, Menozzi P and Piazza A (1994). The history and geography of human genes. Princeton University Press, Princeton.

Comas D, Plaza S, Wells RS, Yuldaseva N, et al. (2004). Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur. J. Hum. Genet.* 12: 495-504. http://dx.doi.org/10.1038/sj.ejhg.5201160

Fang Y, Yang HY, Shi YH, Cui JH, et al. (2015). Mitochondrial DNA haplogroups and somatic mutations are associated with lung cancer in patients from Southwest China. *Genet. Mol. Res.* 14: 5031-5043. http://dx.doi.org/10.4238/2015.May.12.6

Forster P, Calì F, Röhl A, Metspalu E, et al. (2002). Continental and subcontinental distributions of mtDNA control region types. *Int. J. Legal Med.* 116: 99-108. http://dx.doi.org/10.1007/s00414-001-0261-z

Ghulam M (2003). Mansehra- an Introduction. In: Science-Religion Dialogue (Mansehra, Paksitan: Hazara Society for Science-Religion Dialogue). Spring 50-161.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Jaffery NB and Sadaqat M (2006). NFC Awards: Commentary and Agenda. *Pak. Econ. Soc. Rev.* 44: 209-234.

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, et al. (1999). Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* 9: 1331-1334. http://dx.doi.org/10.1016/S0960-9822(00)80057-3

Kivisild T, Rootsi S, Metspalu M, Mastana S, et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72: 313-332. http://dx.doi.org/10.1086/346068

Macaulay V, Richards M, Hickey E, Vega E, et al. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64: 232-249. http://dx.doi.org/10.1086/302204

Macaulay V, Hill C, Achilli A, Rengo C, et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034-1036. http://dx.doi.org/10.1126/science.1109792

Metspalu M, Kivisild T, Metspalu E, Parik J, et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* 5: 26. http://dx.doi.org/10.1186/1471-2156-5-26

Mezzera M and Aftab S (2009). Democratisation and transitional justice cluster country case study: Pakistan. Pakistan state-society analysis. 16.

Ottoni C, Ricaut FX, Vanderheyden N, Brucato N, et al. (2011). Mitochondrial analysis of a Byzantine population reveals

the differential impact of multiple historical events in South Anatolia. *Eur. J. Hum. Genet.* 19: 571-576.
Pakendorf B and Stoneking M (2005). Mitochondrial DNA and human evolution. *Annu. Rev. Genomics Hum. Genet.* 6: 165-183. http://dx.doi.org/10.1146/annurev.genom.6.080604.162249
Parson W and Dür A (2007). EMPOP--a forensic mtDNA database. *Forensic Sci. Int. Genet.* 1: 88-92. http://dx.doi.org/10.1016/j.fsigen.2007.01.018
Parson W, Brandstätter A, Pircher M, Steinlechner M, et al. (2004). EMPOP-the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. *Int. Cong.* 1261: 106-108.
Pereira L, Richards M, Goios A, Alonso A, et al. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* 15: 19-24. http://dx.doi.org/10.1101/gr.3182305
Qadeer MA (2006). Pakistan: Social and Cultural Transformations in a Muslim Nation. Taylor and Francis. 40.
Quintana-Murci L, Chaix R, Wells RS, Behar DM, et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am. J. Hum. Genet.* 74: 827-845. http://dx.doi.org/10.1086/383236
Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, et al. (2005). Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. *BMC Evol. Biol.* 5: 26. http://dx.doi.org/10.1186/1471-2148-5-26
Ralser M, Querfurth R, Warnatz HJ, Lehrach H, et al. (2006). An efficient and economic enhancer mix for PCR. *Biochem. Biophys. Res. Commun.* 347: 747-751. http://dx.doi.org/10.1016/j.bbrc.2006.06.151
Renfrew C (1992). Archaeology, genetics, and linguistic diversity. *Man (Lond)* 27: 445-478. http://dx.doi.org/10.2307/2803924
Richards M, Macaulay V, Hickey E, Vega E, et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67: 1251-1276. http://dx.doi.org/10.1016/S0002-9297(07)62954-1
Rose HA (1907). Hazara Gazettre. Sarhad Academy of Urdu. Abbottabad (reprint: 1983).
Rose HA (1911). Glossary of the tribes and castes of Punjab and NWFP. Civil and Military Gazette Press, Lahore.
Roy S, Thakur Mahadik C and Majumder PP (2003). Mitochondrial DNA variation in ranked caste groups of Maharashtra (India) and its implication on genetic relationships and origins. *Ann. Hum. Biol.* 30: 443-454. http://dx.doi.org/10.1080/0301446031000111410
Tambets K, Rootsi S, Kivisild T, Help H, et al. (2004). The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* 74: 661-682. http://dx.doi.org/10.1086/383203
Torroni A, Achilli A, Macaulay V, Richards M, et al. (2006). Harvesting the fruit of the human mtDNA tree. *Trends Genet.* 22: 339-345. http://dx.doi.org/10.1016/j.tig.2006.04.001
Truman M (1909). The interrelation of the dialects of the fourteen-edicts of Asoka. 1: General introduction and the dialect of the Shuhbazgarhi and Mansehra Redactions. *J. Am. Orient. Soc.* 301: 77-93.
Watson HD (1908). Gazetteer of the Hazara District. In: Civil Service, Settlement Officer. London: Chatto and Windus.
William I (1910). Gazetteer of the Hazara District. In: Review of H. D. Watson Journal of the Royal Asiatic Society New Series, 183-184.

Supplementary material

S1 Figure. Deletion mutation at position 16022 (T) observed in the ethnic Syed population of Mansehra District.

S2 Figure. Deletion of a long DNA fragment (16035-16047; GGAAGCAGATTTG) in the ethnic Tanoli population of Mansehra District.

S3 Figure. Three nucleotide deletion mutations (GTA) at position 16061-16063 bp found in the ethnic Gujar population of Mansehra District.

S4 Figure. A seven-nucleotide deletion (ATTTGGG) at position 16043-16049 seen in the ethnic Abbassi population of Abbottabad.

S5 Figure. A three-cytosine deletion (CCC) at position 16191-16193 observed in the ethnic Awan population residing in Mansehra.