



## Research paper

## EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu &amp; Kashmir, Pakistan



Allah Rakha<sup>a,b,\*</sup>, Min-Sheng Peng<sup>c</sup>, Rui Bi<sup>a</sup>, Jiao-Jiao Song<sup>c</sup>, Zeenat Salahudin<sup>b</sup>, Atif Adan<sup>b</sup>, Muhammad Israr<sup>d</sup>, Yong-Gang Yao<sup>a</sup>

<sup>a</sup> Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China

<sup>b</sup> Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan

<sup>c</sup> State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

<sup>d</sup> Department of Forensic Studies, University of Swat, Swat, Pakistan

## ARTICLE INFO

## Article history:

Received 20 May 2016

Received in revised form 24 August 2016

Accepted 26 August 2016

Available online 28 August 2016

## Keywords:

Pakistan

Kashmiri

mtDNA

Control region

Forensic database

## ABSTRACT

The mitochondrial DNA (mtDNA) control region (nucleotide position 16024–576) sequences were generated through Sanger sequencing method for 317 self-identified Kashmiris from all districts of Azad Jammu & Kashmir Pakistan. The population sample set showed a total of 251 haplotypes, with a relatively high haplotype diversity (0.9977) and a low random match probability (0.54%). The containing matrilineal lineages belonging to three different phylogeographic origins of Western Eurasian (48.9%), South Asian (47.0%) and East Asian (4.1%). The present study was compared to previous data from Pakistan and other worldwide populations (Central Asia, Western Asia, and East & Southeast Asia). The dataset is made available through EMPOP under accession number EMP00679 and will serve as an mtDNA reference database in forensic casework in Pakistan.

© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Mitochondrial DNA analysis has become a very useful tool for human evolutionary studies and especially forensic casework in several circumstances when standard nuclear markers cannot be applied [1,2]. Forensic casework involving mtDNA depends on relevant but authentic databases for estimating the probability of random match. The EMPOP, at present, provides the best quality data representation from all over the world based on logical and phylogenetic measures admissible for forensic purposes [3].

Large-scale investigations of archaeological sites in Central Asia, Northern Pakistan and India revealed a typographical affinity between their cultures going as far back as major pre- and proto-historic periods. The legend relates to an early periodic movement of tribal people from Central Asia to the Kashmir Valley during the cold season when the valley was comparatively warm, which were later replaced by the influx of Aryans from the Punjab. There is historical evidence to the settlement of immigrants from Persia,

Greece, Turkistan and Tibet, China. With the advent of Islam there was an influx of a large number of Sufis and Sayyids in the 14th century [4]. Colonies of Mughals, Pathans, Punjabis, and Paharis settled within comparatively recent times throughout the Jammu and Kashmir. Kashmiris through out the Jammu and Kashmir (India) and Azad Jammu and Kashmir (Pakistan) speak the Kashmiri language. By origin it is a Dravidian Burushaski language, but it has become predominately Indo-Aryan in character. Reflecting the history of area, the Kashmiri vocabulary is mixed, containing Dardic, Sanskrit, Punjabi, and Persian elements [5]. Recent waves of immigrants to Azad Jammu and Kashmir have also introduced Punjabi and Pashto to the main languages [6].

The available mtDNA sequence data from Pakistan is scarce, fragmentary and limited to a few samples from main ethnic groups [7–9]. However, limited number of sample size inhibits the accurate characterization of any population for forensic and genetic purposes. Moreover, there is no representation of Kashmiris from Azad Jammu and Kashmir in any of reference population databases. The present study is intended to characterize the diversity of the matrilineal lineages of current inhabitants of Azad Jammu and Kashmir by analyzing the entire mitochondrial DNA control region. With this analysis, we seek to contribute new mtDNA haplotype data, taking into account that the development

\* Corresponding author at: Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China.

E-mail addresses: [dnaexpert@me.com](mailto:dnaexpert@me.com), [dnatypist@gmail.com](mailto:dnatypist@gmail.com) (A. Rakha).

and improvement of databases constitute a major goal for consolidating the use of mtDNA for forensic purposes. We further analyzed the haplogroup distribution in Kashmiris from Azad Jammu and Kashmir (Pakistan) to corroborate, from the perspective of female genetic lineages, the ancestry composition of this highly mixed population.

The present work constitutes of 317 entire mtDNA control region sequences from randomly selected Kashmiris sampled at eight districts of Jammu & Kashmir, Pakistan. The generated mtDNA population data is deposited to the EMPOP database under the accession number EMP00679 for the worldwide use, and particularly as a reference database for mtDNA applications in forensic and missing person casework in Pakistan.

## 2. Materials and methods

### 2.1. Samples

Blood samples were collected from 317 unrelated Kashmiri males and females residing in different parts of Azad Jammu & Kashmir, Pakistan (Fig. 1).

Only individuals with self-reported Kashmiri origin of at least two generations back on the maternal side were included. Written informed consent was obtained from all the volunteer donors. In order to have full representation, samples were collected from different towns and cities of Azad Jammu & Kashmir. Personal information was treated anonymously. In addition, sample collection was conducted in accordance with the Institutional Review Board of University of Health Sciences, Lahore.

### 2.2. DNA extraction, amplification and sequencing

Genomic DNA extraction was carried out with the Axygen® AxyPrep™ Blood Genomic DNA Miniprep Kit following the manufacturer's protocol (Axygen Biosciences; CA, USA). The entire control region from nt16024 to nt576 was amplified by using primers (Table S4 in Supplementary material) as reported earlier [10]. PCR reactions were performed in 30 mL of reaction mixture containing 3  $\mu$ L 10  $\times$  LA PCR Buffer II ( $Mg^{2+}$  Plus), 1.5 units of TaKaRa LA Taq (TaKaRa Bio Inc., Dalian, China), 400  $\mu$ M of each

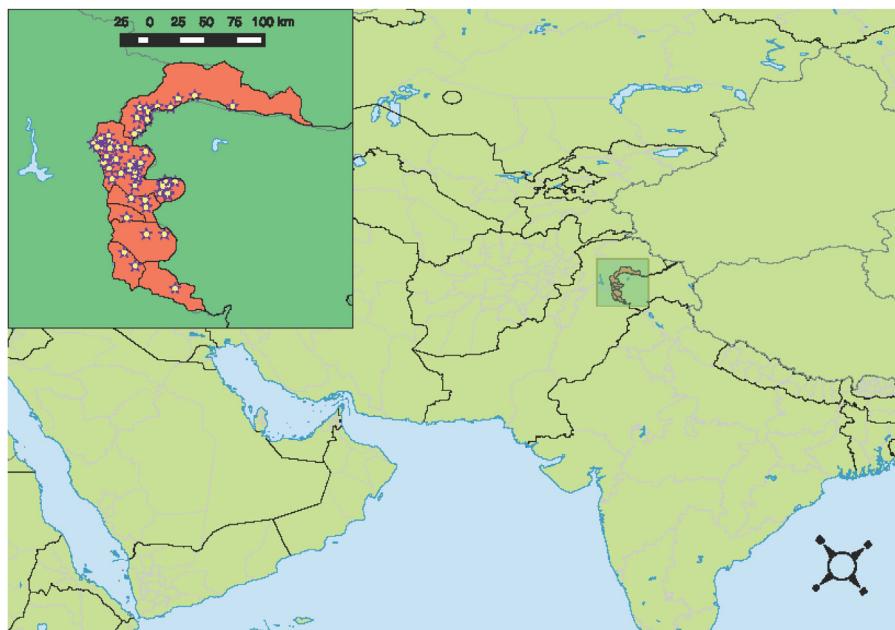
dNTP, 0.2  $\mu$ M of each primer, and 10 ng DNA. The amplification was run on the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) under following conditions: one denaturation cycle of 94 °C for 5 min; 30 amplification cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 2 min; and one full extension cycle of 72 °C for 10 min. The PCR products were purified with Exonuclease I (TaKaRa) and Shrimp Alkaline Phosphatase (TaKaRa) in 10X Exonuclease I Buffer (TaKaRa) following reaction compositions recommended by manufacturer, incubating at 37 °C for 45 min, followed by enzyme deactivation at 95 °C for 15 min. Sequencing by capillary electrophoresis using the BigDye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit was performed on Applied Biosystems 3730xl DNA Analyzer (Thermo Fisher Scientific) according to manufacturer's manual.

### 2.3. Haplogroup assignment

Forward and reverse sequences were aligned and compared using SeqManNGen® version 12.0 (DNASTAR, Madison, WI) with the revised Cambridge Reference Sequences (rCRS) [11]. Quality of sequences was examined manually, and two analysts independently annotated deviations from the reference sequence. The recommended nomenclature for mtDNA typing was used for alignment of variants [12]. The haplogroup assignments were carried out using Mitotool ([www.mitotool.org](http://www.mitotool.org)) [13], Haplogrep ([www.haplogrep.uibk.ac.at](http://www.haplogrep.uibk.ac.at)) [14], and EMMA ([www.empop.online](http://www.empop.online)) [15] based on PhyloTree builds 16 and 17 [16] with referring to the additional guidelines [1]. Haplogroup assignments were re-evaluated by manual inference and conservative MRCA status was assigned to each sequence to improve the predictions. All 317 mtDNA haplotypes were confirmed and validated by the EMPOP curators, being now available from EMPOP browser with the accession number (EMP00679) [3]. The sequences are also available on GenBank via accession numbers KX084069–KX084385.

### 2.4. Data analysis

For all calculations, insertions at nt16193, nt309, nt315, and nt573 were ignored unless otherwise mentioned. The number and



**Fig. 1.** Geographic location of Azad Jammu & Kashmir and inset map depicting the birthplaces of samples used for this study.

**Table 1**

Forensic parameters and diversity indices of mtDNA control region.

	Kashmiri	Kashmiri (+1C considered at position 571)	Pathan [19]	Makrani [20]
No. of Samples	317	317	230	99
No. of Haplotypes	251	252	192	71
No. of Polymorphic sites including indels	230	231	215	142
Discrimination Capacity	0.7918	0.7949	0.8348	0.7172
Random Match Probability	0.0054	0.0053	0.0066	0.0195
Haplotype Diversity (Hd)	0.9977	0.9978	0.9978	0.9905
Average Number of Pairwise differences	10.447	10.447	11.144	12.11
Nucleotide Diversity (Pi)	0.00932	0.00932	0.00993	0.0108

Sequences in Pakistani Ethnic Groups.

types of haplotypes were calculated by direct counting method. Discrimination capacity was estimated by dividing the number of different haplotypes by total number of individuals, and random match probability was calculated as the sum of squared haplotypes' frequencies.

Based on the haplogroup statuses, haplotypes were divided into three broad ancestry groups: West Eurasian (H, HV, I, J, K, N, R, R0, T, U1a, U5a, U7, U8a, W, X), South Asian (M, U2, U4), and East Asian (A, B, C, D, F, G) [16]. Molecular diversity indices, pairwise differences between and within populations, haplotype frequencies, analysis of molecular variance (AMOVA) and pairwise  $F_{ST}$  values were calculated using Arlequin 3.5.1.2 (Computational and Molecular Population Genetics Lab, Institute of Ecology, and Evolution, University of Berne, Bern, Switzerland) [17].  $F_{ST}$  calculations were made under the Kimura 2-parameter model [18]. In addition, the Kashmiri data was compared the previously published data of other ethnic groups from Pakistan (Pathan [19], Makrani [20]) and worldwide populations (Uzbek, Kazakh, Turkmen, Kyrgyz, Afghan, Russian, Tajik [21], Chinese [22–24], Thai [25], Vietnamese [26], Laos [27], Kuwaiti [28], Iraqi [29], Egyptian [30], Dubai [31]) data for which entire control region (16024–576) data could be obtained from the literature. In order to have uniformity and non-arbitrary comparative analysis from published data, consistent treatment of selecting base on the basis of phylogenetic weightage was applied for the sequence heteroplasmic positions. Pairwise  $F_{ST}$  matrix was imported into R statistical package (<http://www.R-project.org/>) for plotting heatmap (ComplexHeatmap) and principal components analysis (FactoMineR). This study followed the requirements of this journal [32,33] and the ISFG recommendations for forensic population data publication.

### 3. Results and discussion

To establish reference database for mtDNA applications in forensic and missing person cases, here we present a high quality mtDNA control region sequences from 317 Kashmiri individuals sampled from all administrative parts of Jammu & Kashmir, Pakistan. The mtDNA haplotypes of 317 individuals are given in Table S1 in Supplementary material and also available from the EMPOP (EMP00679). Data shows a high genetic diversity (0.9977) and a low random match probability (0.0054) among Kashmiris. Point Heteroplasmies were identified in 7 individuals at 6 different positions (Table S5 in Supplementary material). Length heteroplasmies were common in HV1, HVII, and HVIII homopolymeric cytosine stretches.

#### 3.1. Haplotypes and haplogroups information

A total of 251 distinct haplotypes characterized by 230 variable sites were observed (Table 1). There were 209 haplotypes observed once, whereas the most common among all was shared by 8

individuals (2.5%). While considering the phylogenetic weight of additions at nucleotide position 573 with reference to rCRS (normalized number with uniform single addition wherever observed), a total of 252 haplotypes were identified (Table 1).

All the samples (317) were assigned to 156 different haplogroups/sub-haplogroups based on PhyloTree Build 17 [16], albeit at different levels of certainty, as only the control region information was considered for haplogrouping. The most frequent haplogroup was U7a (6.6%), followed by M30 + 16234 (3.8%) and M65a + @16311 (3.5%) (Table 2).

At a broader status of haplogroups, the most frequent was M (34.7%), followed by U (24%) and H (11%) (Table 2). Present data set from the Kashmiri ethnic group had noticeable contributions from three phylogeographical origins: a total of 48.9% of samples were categorized as belonging to West Eurasian ancestry haplogroups, followed by South Asian (47.0%), and East Asian (4.1%) (Fig. 2).

#### 3.2. Random match probability, pairwise comparisons, and population structure

At a broader phylogenetic resolution, the mtDNA control region haplogroup composition of Kashmiris from Azad Jammu & Kashmir was similar to that of Pathans from Khyber Pakhtunkhwa [19], but different from Makrani from southern parts of Pakistan [20] (Fig. 2). The comparison of pairwise  $F_{ST}$  distances showed no statistically significant difference in mtDNA distribution between Pathans of Khyber Pakhtunkhwa and Kashmiri sampled from Azad Jammu & Kashmir. On the other hand, Makranis were found to have significantly different distribution of mtDNA sequences. Analysis of Molecular Variance (AMOVA) revealed 97.57% of variance is attributable to differences within populations and 2.43% represents differences among populations (Table S2a in Supplementary material). There were 14 haplotypes found to be shared between Kashmiri and Pathan, and only 4 haplotypes between Kashmiri and Makrani (Table S3 in Supplementary material). Kashmir as a whole is famous for its highly variable geographical landscape and human demographic history. Pathans are the main inhabitants of adjacent province and surrounding northern parts of Pakistan. One explanation for such a low difference could be relatively recent high migrations (of females). Future studies based on extensive sampling and more genetic markers will shed more lights into matter.

The pairwise  $F_{ST}$  and respective  $p$ -values among three Pakistani ethnic groups Pathan [19,20] including Kashmiri samples generated here, and fifteen regional populations of Central Asia, Western Asia and East & Southeast Asia [21–31] are given in Table S6 in Supplementary material. Heatmap (based on pairwise  $F_{ST}$  values) and relevant clustering analysis showed regional pattern of distribution of mtDNA control region sequences (Fig. 3).

From the PCA plot (based on pairwise  $F_{ST}$  values), it is evident that the populations geographically closer to northern parts of Pakistan clustered with Kashmiri and Pathans. Whereas, the

**Table 2**

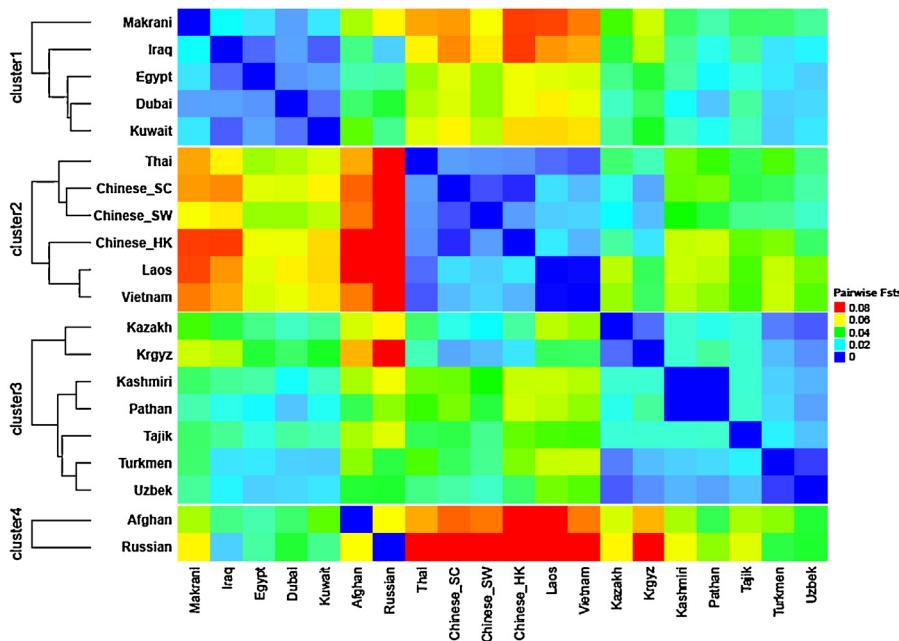
Haplogroup frequencies of 317 Kashmiris from AJ &amp; K, Pakistan.

Broad Haplogroup	Number	Percent	Haplogroup	Number	Percentage
A	2	0.6	A17	1	0.3
			A19	1	0.3
B	1	0.3	B4c1a1c	1	0.3
C	4	1.3	C4a2c	1	0.3
			C4a3b	3	0.9
D	2	0.6	D4b2b	1	0.3
			D6a1	1	0.3
F	3	0.9	F1c1a	1	0.3
			F1d	2	0.6
G	1	0.3	G3a1'2	1	0.3
H	34	10.7	H*	13	4.1
			H1	2	0.6
			H14a	1	0.3
			H15	1	0.3
			H15a1b	1	0.3
			H17a1	2	0.6
			H17c	2	0.6
			H1ap1	1	0.3
			H1e3	1	0.3
			H1f + 16093	1	0.3
			H3v	1	0.3
			H41a	1	0.3
			H5	1	0.3
			H6	1	0.3
			H66a	1	0.3
			H6b	3	0.9
			H7a2	1	0.3
HV	6	1.9	HV14	3	0.9
			HV2	1	0.3
			HV6	2	0.6
I	3	0.9	I1	1	0.3
			I1f	1	0.3
			I6a	1	0.3
J	11	3.5	J1b1a1	1	0.3
			J1b1b	8	2.5
			J1d	2	0.6
K	1	0.3	K2a5	1	0.3
M	110	34.7	M*	1	0.3
			M18c	1	0.3
			M22a	2	0.6
			M2a1a	3	0.9
			M3	5	1.6
			M30	3	0.9
			M30 + 16234	12	3.8
			M30b	6	1.9
			M30c	1	0.3
			M30c1	6	1.9
			M30d	2	0.6
			M30e	1	0.3
			M30g	1	0.3
			M31a2	2	0.6
			M33a2a	1	0.3
			M35b + 16304	1	0.3
			M37e	2	0.6
			M37e2	1	0.3
			M3a	1	0.3
			M3a1 + 204	6	1.9
			M3c + 152	1	0.3
			M3c1b	1	0.3
			M3c2	3	0.9
			M3d	1	0.3
			M4	10	3.2
			M49d	1	0.3
			M4b	2	0.6
			M5	4	1.3
			M5a1	1	0.3
			M5a1a	2	0.6
			M5a2a	2	0.6
			M5a2a1a	2	0.6
			M5b2	1	0.3
			M5d	1	0.3
			M6	2	0.6
			M65	1	0.3
			M65a	1	0.3
			M65a + @16311	11	3.5
			M65b	4	1.3

**Table 2 (Continued)**

Broad Haplogroup	Number	Percent	Haplogroup	Number	Percentage
N	6	1.9	M6a1b	1	0.3
			N1a1a1a	2	0.6
			N1a1a1a1a	1	0.3
			N1a1b	1	0.3
			N1a2	2	0.6
R	19	6	R	4	1.3
			R30a1b	1	0.3
			R30b1	7	2.2
			R30b2	3	0.9
			R6	3	0.9
			R6b	1	0.3
R0	7	2.2	R0	5	1.6
			R0a + 60.1T	2	0.6
T	14	4.4	T1a	1	0.3
			T1a1'3	1	0.3
			T2	5	1.6
			T2b	1	0.3
			T2b34	1	0.3
			T2c1	3	0.9
			T2e	1	0.3
			T2e2	1	0.3
			U1a	2	0.6
			U1a1a1	1	0.3
U	77	24.3	U2	1	0.3
			U2 + 152	2	0.6
			U2a	3	0.9
			U2a1a	1	0.3
			U2b	5	1.6
			U2b1	1	0.3
			U2b2	5	1.6
			U2c'd	7	2.2
			U2e1	2	0.6
			U2e1h	1	0.3
			U2e2	1	0.3
			U4	1	0.3
			U4a2c1	1	0.3
			U4b1a1a1	2	0.6
			U5a1	1	0.3
			U5a1b + 16362	1	0.3
			U7	10	3.2
			U7a	21	6.6
W	14	4.4	U7a3	1	0.3
			U7a3	1	0.3
			U7a3b	4	1.3
			U7a4	1	0.3
			U8b1a1	1	0.3
			W	1	0.3
			W + 194	4	1.3
			W3a1 + 199	1	0.3
			W4	1	0.3
			W5a	1	0.3
X	2	0.6	W5a1a	1	0.3
			W6	4	1.3
			W6b1	1	0.3
X2	2	0.6	X2	1	0.3
			X2d	1	0.3

**Fig. 2.** Percentage composition of phylogeographical lineages in Pakistani populations.



**Fig. 3.** Heatmap matrix of pairwise  $F_{ST}$  values for 20 populations including Kashmiri.

Makranis being in south of Pakistan are found to be clustered with other West Asian populations (Fig. S1 in Supplementary material).

The Kashmiri dataset was compared to other 15 populations by computing AMOVA (Table S2b in Supplementary material). The majority of observed variance (96.38%) was attributable to differences within populations, and only 3.62% represented differences among populations.

#### 4. Conclusion

The primary goal of this study was to develop forensic quality mtDNA control region reference database from Kashmiris broadly sampled from the entire region of Azad Jammu & Kashmir, Pakistan. Following the best practices for mtDNA population databases, 317 complete mtDNA control region haplotypes are made available to all users through the EMPOP (EMP00679). Additionally, phylogeographical findings based on haplogroup composition confirmed a high genetic heterogeneity of the Kashmiri population as a supposed consequence of the impact of different waves of migrations to this region. This mtDNA heterogeneity at the ethnicity and provincial levels demands extended and more reliable databases to improve forensic and genetic analyses in the Pakistani populations. Our study provides reasonably good start towards the expansion of Pakistani mtDNA reference database rendering more accurate estimates of random match probability in routine forensic mtDNA analysis.

#### Conflicts of interest

The authors declare no conflict of interest.

#### Acknowledgements

We would like to thank all volunteers for providing their samples used in this study, Salah-ud-din, Ghulam Yasin, Ayad Akhtar, Javaid Akhtar Chaudhry for their valuable support for the sampling, Li-Li Kong, Lu-Xiu Yang, Yu Li, Saadia Noreen, Fatima Gull for technical assistance, Hansi Weissensteiner (HAPLOGREP,

Medical University Innsbruck, Austria) and EMPOP (University of Innsbruck, Austria) for haplogroup assignments. This project was supported by the President's International Fellowship Initiative (PIFI) Program of the Chinese Academy of Sciences and the National Natural Science Foundation of China (30925021). M.-S.P. thanks the Youth Innovation Promotion Association, Chinese Academy of Sciences, for its support.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2016.08.009>.

#### References

- [1] H.-J. Bandelt, M. van Oven, A. Salas, Haplogrouping mitochondrial DNA sequences in legal medicine/forensic genetics, *Int. J. Legal Med.* 126 (2012) 901–916, doi:<http://dx.doi.org/10.1007/s00414-012-0762-y>.
- [2] T. Kivisild, Maternal ancestry and population history from whole mitochondrial genomes, *Investig. Genet.* 6 (2015) 3, doi:<http://dx.doi.org/10.1186/s13323-015-0022-2>.
- [3] W. Parson, A. Dür, EMPOP—a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92, doi:<http://dx.doi.org/10.1016/j.fsigen.2007.01.018>.
- [4] P.N.K. Bamzai, Culture and Political History of Kashmir, M.D. Publications Pvt. Ltd., 1994.
- [5] P.I. Ahmad, A Hand Book on Azad Jammu & Kashmir, (2003).
- [6] Languages: Kashmiri-Language, <http://www.britannica.com/topic/Kashmiri-language> (accessed 07.05.16).
- [7] R. Cordaux, N. Saha, G.R. Bentley, R. Aunger, S.M. Sirajuddin, M. Stoneking, Mitochondrial DNA analysis reveals diverse histories of tribal populations from India, *Eur. J. Hum. Genet.* 11 (2003) 253–264, doi:<http://dx.doi.org/10.1038/sj.ejhg.5200949>.
- [8] L. Quintana-Murci, R. Chaix, R.S. Wells, D.M. Behar, H. Sayar, R. Scozzari, et al., Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor, *Am. J. Hum. Genet.* 74 (2004) 827–845, doi:<http://dx.doi.org/10.1086/383236>.
- [9] M. Metspalu, T. Kivisild, E. Metspalu, J. Parik, G. Hudjashov, K. Kaldma, et al., 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 (2004) 26, doi:<http://dx.doi.org/10.1186/1471-2156-5-26>.
- [10] H.-W. Wang, X. Jia, Y. Ji, Q.-P. Kong, Q. Zhang, Y.-G. Yao, et al., Strikingly different penetrance of LHON in two Chinese families with primary mutation G11778A is independent of mtDNA haplogroup background and secondary mutation

- G13708A, *Mutat. Res.* 643 (2008) 48–53, doi:<http://dx.doi.org/10.1016/j.mrfmmm.2008.06.004>.
- [11] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowler, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147, doi:<http://dx.doi.org/10.1038/13779>.
- [12] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, et al., DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci. Int. Genet.* 13 (2014) 134–142, doi:<http://dx.doi.org/10.1016/j.fsigen.2014.07.010>.
- [13] L. Fan, Y.-G. Yao, MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations, *Mitochondrion* 11 (2011) 351–356, doi:<http://dx.doi.org/10.1016/j.mito.2010.09.013>.
- [14] H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, et al., HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing, *Nucleic Acids Res.* (2016) gkw233, doi:<http://dx.doi.org/10.1093/nar/gkw233>.
- [15] A.W. Röck, A. Dür, M. van Oven, W. Parson, Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA), *Forensic Sci. Int. Genet.* 7 (2013) 601–609, doi:<http://dx.doi.org/10.1016/j.fsigen.2013.07.005>.
- [16] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E386–94, doi:<http://dx.doi.org/10.1002/humu.20921>.
- [17] 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, doi:<http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x>.
- [18] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* 16 (1980) 111–120.
- [19] A. Rakha, K.-J. Shin, J.A. Yoon, N.Y. Kim, M.H. Siddique, I.S. Yang, et al., Forensic and genetic characterization of mtDNA from Pathans of Pakistan, *Int. J. Legal Med.* 125 (2011) 841–848, doi:<http://dx.doi.org/10.1007/s00414-010-0540-7>.
- [20] M.H. Siddiqi, T. Akhtar, A. Rakha, G. Abbas, A. Ali, N. Haider, et al., Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data, *Leg. Med. (Tokyo)* 17 (2015) 134–139, doi:<http://dx.doi.org/10.1016/j.legalmed.2014.09.007>.
- [21] J.A. Irwin, A. Ikramov, J. Saunier, M. Bodner, S. Amory, A. Röck, et al., The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns, *Int. J. Legal Med.* 124 (2010) 195–204, doi:<http://dx.doi.org/10.1007/s00414-009-0406-z>.
- [22] D. Wang, L.-Y. Su, A.-M. Zhang, Y.-Y. Li, X.-A. Li, L.-L. Chen, et al., Mitochondrial DNA copy number, but not haplogroup, confers a genetic susceptibility to leprosy in Han Chinese from Southwest China, *PLoS One* 7 (2012) e38848, doi:<http://dx.doi.org/10.1371/journal.pone.0038848>.
- [23] W. Zhang, J. Tang, A.-M. Zhang, M.-S. Peng, H.-B. Xie, L. Tan, et al., A matrilineal genetic legacy from the last glacial maximum confers susceptibility to schizophrenia in Han Chinese, *J. Genet. Genomics* 41 (2014) 397–407, doi:<http://dx.doi.org/10.1016/j.jgg.2014.05.004>.
- [24] J.A. Irwin, J.L. Saunier, P. Beh, K.M. Strouss, C.D. Painter, T.J. Parsons, Mitochondrial DNA control region variation in a population sample from Hong Kong, China, *Forensic Sci. Int. Genet.* 3 (2009) e119–25, doi:<http://dx.doi.org/10.1016/j.fsigen.2008.10.008>.
- [25] B. Zimmermann, M. Bodner, S. Amory, L. Fendt, A. Röck, D. Horst, et al., Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai), *Int. J. Legal Med.* 123 (2009) 495–501, doi:<http://dx.doi.org/10.1007/s00414-009-0373-4>.
- [26] J.A. Irwin, J.L. Saunier, K.M. Strouss, T.M. Diegoli, K.A. Sturk, J.E. O'Callaghan, et al., Mitochondrial control region sequences from a Vietnamese population sample, *Int. J. Legal Med.* 122 (2008) 257–259, doi:<http://dx.doi.org/10.1007/s00414-007-0205-3>.
- [27] M. Bodner, B. Zimmermann, A. Röck, A. Kloss-Brandstätter, D. Horst, B. Horst, et al., Southeast Asian diversity: first insights into the complex mtDNA structure of Laos, *BMC Evol. Biol.* 11 (2011) 49, doi:<http://dx.doi.org/10.1186/1471-2148-11-49>.
- [28] M. Scheible, M. Alenizi, K. Sturk-Andreaggi, M.D. Coble, S. Ismael, J.A. Irwin, Mitochondrial DNA control region variation in a Kuwaiti population sample, *Forensic Sci. Int. Genet.* 5 (2011) e112–3, doi:<http://dx.doi.org/10.1016/j.fsigen.2011.04.001>.
- [29] N. Al-Zahery, J. Saunier, K. Ellingson, W. Parson, T.J. Parsons, J.A. Irwin, Characterization of mitochondrial DNA control region lineages in Iraq, *Int. J. Legal Med.* 127 (2013) 373–375, doi:<http://dx.doi.org/10.1007/s00414-012-0757-8>.
- [30] J.L. Saunier, J.A. Irwin, K.M. Strouss, H. Ragab, K.A. Sturk, T.J. Parsons, Mitochondrial control region sequences from an Egyptian population sample, *Forensic Sci. Int. Genet.* 3 (2009) e97–103, doi:<http://dx.doi.org/10.1016/j.fsigen.2008.09.004>.
- [31] F. Alshamali, A. Brandstätter, B. Zimmermann, W. Parson, Mitochondrial DNA control region variation in dubai, United Arab Emirates, *Forensic Sci. Int. Genet.* 2 (2008) e9–10, doi:<http://dx.doi.org/10.1016/j.fsigen.2007.08.005>.
- [32] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* 7 (2013) 217–220, doi:<http://dx.doi.org/10.1016/j.fsigen.2013.01.001>.
- [33] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., Update of the guidelines for the publication of genetic population data, *Forensic Sci. Int. Genet.* 10 (2014) A1–A2, doi:<http://dx.doi.org/10.1016/j.fsigen.2014.01.004>.

Table S1. Mitochondrial DNA control region (16024-576) haplotypes and estimated haplogroups for Kashmiri ethnic group from Azad Jammu & Kashmir, Pakistan

Sample ID	MRCA	Polymorphisms (delimited with tabs)																												
KPR_067	U1a	16182C	16183C	16189C	16249C	16362C	73G	263G	285T	309.1C	309.2C	315.1C																		
KPR_068	H1ap1	16129A	16189C	16311C	16519C	152C	263G	309.1C	309.2C	315.1C																				
KPR_069	M4	16086C	16145A	16189C	16223T	16261T	16311C	16519C	73G	152C	263G	315.1C	489C																	
KPR_070	H6b	16300G	16362C	16482G	239C	263G	309.1C	315.1C																						
KPR_071	U2c'd	16051G	16129A	16179T	16234T	16247G	73G	152C	263G	309.1C	315.1C	480C																		
KPR_072	H*	16519C	146C	263G	315.1C	460.1T																								
KPR_073	M35b+16304	16223T	16304C	16362C	16519C	73G	199C	263G	309.1C	315.1C	489C																			
KPR_074	R	16519C	73G	263G	309.1C	315.1C																								
KPR_075	M5	16129A	16223T	16271C	16519C	73G	263G	309.1C	315.1C	489C																				
KPR_076	U7a3	16136C	16207G	16309G	16318T	16519C	73G	151T	152C	263G	309.1C	315.1C	523d	524d																
KPR_077	M3c1b	16179T	16183C	16189C	16193.1C	16223T	16294T	16519C	73G	152C	263G	309.1C	309.2C	315.1C	482C	489C	523d	524d												
KPR_078	M4	16086C	16145A	16148T	16223T	16261T	16311C	16519C	16527T	73G	263G	315.1C	489C																	
KPR_079	M65a+@16311	16223T	16289G	16519C	73G	263G	315.1C	489C	511T																					
KPR_080	H1f+16093	16093C	16183C	16189C	16519C	182T	263G	309.1C	309.2C	315.1C																				
KPR_081	M30+16234	16172C	16223T	16234T	16519C	73G	195A	263G	309.1C	315.1C	489C	523d	524d																	
KPR_082	U2e2	16051G	16092C	16129C	16153A	16182C	16183C	16189C	16261T	16362C	16519C	73G	151T	152C	217C	263G	309d	315.1C	508G											
KPR_083	U7a	16309G	16318C	16519C	73G	151T	152C	263G	315.1C	523d	524d																			
KPR_084	T2c1	16126C	16286T	16292T	16294T	16296T	16519C	73G	152C	263G	315.1C																			
KPR_085	U7a	16309G	16318C	16519C	73G	151T	152C	263G	309.1C	315.1C	523d	524d																		
KPR_086	M30c1	16166d	16223T	16519C	73G	146C	195A	204C	263G	315.1C	489C	523d	524d																	
KPR_087	U2a	16051G	16129A	16206C	16291T	16311C	44.1C	73G	146C	263G	309.1C	315.1C																		
KPR_088	U7	16092C	16183C	16189C	16309G	16318C	16519C	73G	151T	152C	263G	315.1C	523d	524d	573.1C															
KPR_089	H15a1b	16124C	16184T	44.1C	55C	57C	146C	263G	309.1C	309.2C	315.1C																			
KPR_090	J1b1b	16069T	16126C	16145A	16189C	16261T	16519C	73G	263G	271T	295T	315.1C	462T	489C	523d	524d														
KPR_091	U7	16297C	16309G	16318T	16519C	73G	152C	263G	309.1C	315.1C	523d	524d																		
KPR_092	W3a1+199	16169T	16223T	16292T	16311C	16519C	73G	189G	194T	195C	199C	204C	207A	263G	309.1C	315.1C														
KPR_093	H*	16311C	16362C	152C	263G	315.1C																								
KPR_094	F1c1a	16111T	16129A	16304C	16519C	73G	234G	249d	263G	315.1C	523d	524d																		
KPR_095	U7a	16085T	16309G	16318C	16519C	73G	151T	152C	263G	315.1C	523d	524d																		
KPR_096	H17a1	16129A	16291T	193G	198T	263G	309.1C	315.1C																						
KPR_097	M5b2	16048A	16129A	16223T	16519C	73G	263G	309.1C	309.2C	315.1C	356.1C	489C	523d	524d																
KPR_098	U2+152	16051G	16093C	16206T	16362C	44.1C	73G	152C	263G	309.1C	315.1C																			
KPR_099	U7a	16274A	16309G	16318T	16343G	16390A	16519C	73G	151T	152C	263G	315.1C	523d	524d																
KPR_100	R30b1	16172C	16182C	16183C	16189C	16234T	16261T	16298C	16299G	73G	152C	257G	263G	299d	315.1C	373G	480C													
KPR_101	C4a3b	16223T	16298C	16311C	16327T	16357C	16519C	73G	249d	263G	310C	489C																		
KPR_102	N1a1b	16223T	16311C	16391A	16519C	73G	143A	199C	204C	250C	263G	309.1C	309.2C	315.1C	573.1C	5														



Sample ID	MRCA	Polymorphisms (delimited with tabs)																									
KPR_201	U2a1a	16051G	16154C	16206C	16230G	16311C	73G	263G	309.1C	309.2C	315.1C																
KPR_202	D6a1	16189C	16223T	16274A	16362C	73G	263G	309.1C	315.1C	489C																	
KPR_203	H1e3	16129A	16189C	16519C	93G	263G	309.1C	309.2C	315.1C																		
KPR_204	M4	16086C	16145A	16148T	16223T	16261T	16311C	16519C	73G	204C	207A	263G	315.1C	489C													
KPR_205	M65a+@16311	16166G	16185T	16223T	16289G	16362C	16519C	73G	195C	263G	315.1C	489C	511T														
KPR_206	T2	16126C	16172C	16294T	16296T	16325C	16519C	73G	195C	263G	309.1C	309.2C	315.1C														
KPR_207	H15	16092C	16519C	55C	57C	152C	263G	309.1C	309.2C	315.1C	523d	524d															
KPR_208	J1b1b	16069T	16126C	16145A	16189C	16261T	16519C	73G	263G	271T	295T	315.1C	462T	489C	523d	524d											
KPR_209	H*	16519C	152C	263G	309.1C	315.1C																					
KPR_210	M65a+@16311	16223T	16234T	16289G	16519C	73G	263G	309.1C	315.1C	489C	511T																
KPR_211	U2b2	16051G	16239T	16352C	16353T	73G	146C	234G	263G	315.1C																	
KPR_212	U1a1a1	16129A	16169T	16182C	16183C	16189C	16249C	16274A	16380T	16519C	73G	146C	152C	263G	285T	309.1C	315.1C	385G	523d	524d	573.1C	573.2C					
KPR_213	U2c'd	16051G	16169T	16234T	16278T	16304C	16519C	73G	152C	217C	263G	315.1C															
KPR_214	M65	16223T	16311C	16519C	73G	263G	309.1C	315.1C	489C	511T																	
KPR_215	M30g	16223T	16304C	16519C	73G	195A	204C	263G	309T	315.1C	489C	523d	524d														
KPR_216	T1a	16126C	16163G	16172C	16186T	16189C	16294T	16519C	73G	152C	195C	263G	315.1C														
KPR_217	M4	16136C	16145A	16223T	16261T	16304C	16311C	16519C	73G	263G	315.1C	489C															
KPR_218	M31a2	16126C	16145A	16223T	73G	146C	195C	263G	315.1C	489C																	
KPR_219	M30+16234	16223T	16234T	16256T	16519C	73G	195A	263G	309.1C	315.1C	489C	523d	524d														
KPR_220	D4b2b	16037G	16223T	16304C	16362C	16519C	73G	194T	263G	309.1C	315.1C	489C	523d	524d													
KPR_221	U2c'd	16051G	16126C	16178C	16179T	16183C	16189C	16193.1C	16234T	16247G	73G	146C	152C	263G	315.1C	573d											
KPR_222	T2c1	16111A	16126C	16292T	16294T	16296T	16519C	73G	263G	309.1C	315.1C																
KPR_223	R30b1	16172C	16182C	16183C	16189C	16234T	16261T	16298C	16299G	73G	152C	257G	263G	299d	309d	315.1C	373G	480C									
KPR_224	M30	16189C	16190T	16193.1C	16223T	16260T	16293G	16519C	73G	195A	263G	315.1C	489C	523d	524d												
KPR_225	M37e2	16111T	16182C	16183C	16189C	16223T	16224C	16295T	16519C	73G	182T	263G	309.1C	315.1C	489C												
KPR_226	U7	16318T	16519C	73G	152C	263G	309.1C	315.1C	523d	524d																	
KPR_227	R6b	16169T	16179T	16227G	16245T	16266T	16278T	16362C	16519C	73G	94A	195C	246C	263G	309.1C	309.2C	315.1C	523d	524d								
KPR_228	M3a1+204	16126C	16223T	16311C	16519C	73G	204C	217C	263G	315.1C	482C	489C															
KPR_229	U7a	16189C	16266T	16309G	16318T	16519C	73G	151T	152C	263G	315.1C	523d	524d														
KPR_230	M37e	16086C	16111T	16184T	16189C	16223T	16274A	16295T	16519C	73G	263G	309.1C	315.1C	489C													
KPR_231	M22a	16126C	16223T	16290T	16519C	73G	263G	309.1C	309.2C	315.1C	489C																
KPR_232	H7a2	16176T	16184T	16357C	16497G	16519C	146C	263G	309.1C	315.1C																	
KPR_233	R0	16519C	263G	315.1C																							
KPR_234	T2	16126C	16294T	16296T	16519C	73G	200G	263G	315.1C																		
KPR_235	M22a	16126C	16223T	16290T	16519C	73G	263G	309.1C	315.1C	489C																	
KPR_236	M33a2a	16169T	16172C	16223T	16519C	73G	146C	263G	309.1C	315.1C	462T	489C															
KPR_237	M5a2a	16129A	16209C	16223T	16519C	73G	151T	263G	315.1C	489C																	
KPR_238	H6b	16300G	16362C	16482G	239C	263G	309.1C	315.1C																			
KPR_239	M65a+@16311	16185T	16223T	16289G	16362C	16519C	73G	263G	315.																		

Sample ID	MRCA	Polymorphisms (delimited with tabs)														
KPR_268	M3a	16126C	16223T	16311C	16519C	73G	150T	204C	263G	315.1C	482C	489C				
KPR_269	M30c	16183C	16189C	16223T	16519C	73G	146C	195A	200G	263G	309.1C	315.1C	489C	523d	524d	
KPR_270	M30d	16179d	16223T	16302G	16519C	73G	195A	263G	315.1C	489C	523d	524d				
KPR_271	M4b	16086C	16145A	16192T	16223T	16261T	16311C	16519C	73G	263G	315.1C	489C				
KPR_272	R30b1	16183C	16189C	16298C	16299G	73G	152C	195C	263G	299d	309.1C	309.2T	315.1C	373G		
KPR_273	N1a2	16192T	16223T	16274A	16301T	16309G	16311C	16356C	73G	199C	204C	263G	315.1C	523d	524d	
KPR_274	U2+152	16051G	16319A	16519C	73G	152C	263G	309.1C	315.1C	489C						
KPR_275	H*	151T	152C	263G	315.1C											
KPR_276	J1b1b	16069T	16126C	16145A	16183C	16189C	16261T	16519C	73G	263G	271T	295T	315.1C	462T	489C	523d
KPR_277	M3a1+204	16223T	16311C	16519C	73G	204C	217C	263G	309.1C	315.1C	482C	489C				
KPR_278	M3a1+204	16126C	16223T	16519C	73G	146C	204C	263G	309.1C	315.1C	482C	489C				
KPR_279	U7a	16085T	16309G	16318C	16519C	73G	151T	152C	263G	315.1C	523d	524d				
KPR_280	U7a	16085T	16309G	16318C	16519C	73G	151T	152C	263G	315.1C	523d	524d				
KPR_281	M3	16126C	16223T	16519C	73G	263G	315.1C	482C	489C							
KPR_282	M5a1	16129A	16223T	16291T	16519C	73G	263G	309.1C	315.1C	489C						
KPR_283	U7a	16309G	16318C	16519C	73G	151T	152C	263G	309.1C	309.2C	315.1C	523d	524d			
KPR_284	M3	16126C	16185T	16223T	16519C	73G	195C	263G	315.1C	482C	489C					
KPR_285	M5a2a1a	16129A	16223T	16265C	16344T	16519C	73G	263G	315.1C	374G	489C					
KPR_286	M5a1a	16129A	16223T	16291T	16519C	73G	152C	263G	315.1C	334C	489C					
KPR_287	R30b1	16183C	16189C	16298C	16299G	73G	152C	195C	263G	299d	309.1C	315.1C	373G			
KPR_288	U2	16051G	16092C	16192T	16325C	73G	195C	263G	309.1C	315.1C						
KPR_289	U7a	16309G	16318T	16519C	73G	151T	152C	263G	315.1C	523d	524d					
KPR_290	U7a	16093C	16274A	16309G	16318T	16343G	16390A	16519C	73G	151T	152C	263G	315.1C	523d	524d	
KPR_291	M3c2	16126C	16154C	16223T	16224C	16519C	73G	195C	263G	309.1C	315.1C	482C	489C	523d	524d	
KPR_292	H6	16519C	239C	263G	315.1C											
KPR_293	U8b1a1	16129A	16183C	16189C	16193.1C	16519C	73G	185A	195C	263G	309.1C	309.2C	315.1C			
KPR_294	M30+16234	16223T	16234T	16519C	73G	195A	263G	309.1C	315.1C	489C	523d	524d				
KPR_295	W+194	16223T	16240G	16292T	16519C	73G	189G	194T	195C	204C	207A	263G	309.1C	315.1C		
KPR_296	H3v	16519C	93G	152C	263G	315.1C	408A	513A								
KPR_297	U4	16356C	16519C	73G	195C	263G	309.1C	315.1C	499A							
KPR_298	R	16519C	73G	146C	263G	309.1C	315.1C									
KPR_299	M3c2	16126C	16154C	16223T	16519C	73G	263G	315.1C	482C	489C	523d	524d				
KPR_300	M30d	16179d	16223T	16302G	16519C	73G	195A	263G	309.1C	309.2C	315.1C	489C	523d	524d		
KPR_301	R0	16519C	263G	315.1C												
KPR_302	H*	16519C	263G	315.1C	513A											
KPR_303	T2e	16126C	16153A	16189C	16294T	16296T	16519C	73G	150T	263G	309.1C	315.1C				
KPR_304	U7a3b	16140C	16207G	16309G	16318T	16362C	16519C	73G	151T	152C	195C	263G	315.1C	523d	524d	
KPR_305	U1a	16182C	16183C	16189C	16249C	16257T	16519C	73G	204C	263G	285T	309.1C	315.1C			
KPR_306	N1a2	16189C	16223T	16286T	16301T	16356C	73G	199C	204C	263G	315.1C	523d	524d			
KPR_307	U7a	16140C	16207G	16309G	16318T	16362C	16519C	73G	151T	152C	195C	263G	315.1C	523d	524d	
KPR_308	M30b	16223T	16278T	16519C	73G	152C	195A	263G	315.1C	489C	523d	524d				
KPR_309	U7a	16309G	16318T	16519C	73G	151T	152C	263G	315.1C	523d	524d					
KPR_310	R	16497G	16519C	73G	152C	263G	315.1C	480C								
KPR_311	U7a	16309G	16318C	16519C	73G	151T	152C	263G	309.1C	315.1C	523d	524d				
KPR_312	M6a1b	16178C	16188T	16223T	16231C	16362C	16519C	73G	146C	185A	263G	315.1C	461T	489C		
KPR_313	M30+16234	16223T	16234T	16362C	16519C	73G	195A	263G	315.1C	489C	523d	524d				
KPR_314	M30+16234	16223T	16234T	16362C	16519C	73G	195A	263G	315.1C	489C	523d	524d				
KPR_315	M65a+@16311	16223T	16276C	16289G	16519C	73G	185A	263G	309.1C	315.1C	489C	511T				
KPR_316	H14a	16256T	16352C	263G	309.1C	315.1C										
KPR_317	M31a2	16126C	16145A	16223T	73G	146C	195C	263G	315.1C	489C						

Table S2.

a. AMOVA for Pakistani ethnic groups (Kashmiri, Pathan, Makrani)

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	2	70.132	0.14807	2.43
Within populations	643	3820.524	5.94172	97.57
Total	645	3890.656	6.08979	

b. AMOVA results for Kashmiri and 15 regional populations

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	15	869.77	0.2269	3.62
Within populations	3672	22178.289	6.03984	96.38
Total	3687	23048.059	6.26674	

**Table S3. Relative Frequencies of haplotypes Shared by Kashmiris with Makrani and Pathan**

Hap_ID	Kashmiri (317)	Makrani (99)	Pathan (230)	MRCA	Haplotype														
Hap_1	0.0158	0.0202	0.013	R0	16519C	263G	315.1C												
Hap_2	0.00315	0.0101	0.0087	M3a1+204	16126C	16223T	16311C	16519C	73G	204C	217C	263G	315.1C	482C	489C				
Hap_3	0.00315	0.0303	0	U2b1	16051G	16172C	73G	263G	309.1C	315.1C									
Hap_4	0.00315	0.0202	0.00435	M4	16145A	16176T	16223T	16261T	16311C	16519C	73G	263G	309.1C	315.1C	489C				
Hap_5	0.00315	0	0.00435	W6b1	16189C	16223T	16292T	16325C	16355T	16519C	73G	143A	189G	194T	195C	204C	207A	263G	309.1C
Hap_6	0.00631	0	0.00435	W6	16192T	16223T	16266T	16292T	16325C	16519C	73G	189G	194T	195C	204C	207A	263G	309.1C	315.1C
Hap_7	0.00315	0	0.00435	U2c'd	16051G	16093C	16234T	73G	146C	152C	263G	315.1C	524.1A	524.2C					
Hap_8	0.00315	0	0.00435	T2b	16126C	16294T	16296T	16304C	16519C	73G	263G	309.1C	309.2C	315.1C					
Hap_9	0.00315	0	0.0087	M5a1	16129A	16223T	16291T	16519C	73G	263G	309.1C	315.1C	489C						
Hap_10	0.00315	0	0.0174	R	16519C	73G	263G	309.1C	315.1C										
Hap_11	0.00315	0	0.013	H*	16311C	93G	152C	263G	309.1C	315.1C									
Hap_12	0.00315	0	0.00435	M5	16129A	16223T	16271C	16519C	73G	263G	309.1C	315.1C	489C						
Hap_13	0.00315	0	0.00435	M65a+@16311	16185T	16223T	16289G	16362C	16519C	73G	263G	315.1C	489C	511T					
Hap_14	0.00631	0	0.00435	U2b2	16051G	16209C	16239T	16352C	16353T	73G	146C	152C	234G	263G	309.1C	315.1C			
Hap_15	0.00315	0	0.00435	HV14	16311C	263G	309.1C	315.1C	480C										
Hap_16	0.00315	0	0.00435	H*	16311C	152C	263G	309.1C	315.1C										
Hap_17	0.00315	0	0.00435	H*	16311C	16362C	152C	263G	315.1C										
Hap_18	0.00315	0	0.00435	M4	16124C	16145A	16176T	16223T	16261T	16311C	16519C	73G	146C	263G	309.1C	315.1C	489C		
Hap_19	0.00631	0	0.00435	M5a2a	16129A	16209C	16223T	16519C	73G	151T	263G	315.1C	489C						

**Table S4: Primers for mtDNA control region sequence analysis**

Primers	Sequence (5' to 3')	Usage
L15594	CGCCTACACAATTCTCCGATC	PCR and Sequencing
H901	ACTTGGGTTAACGTGTGACC	PCR and Sequencing
L15996	CTCCACCATTAGCACCCAAAGC	Sequencing
L16209	CCCCATGCTTACAAGCAAGT	Sequencing
L29	GGTCTATCACCCATTAAACCAC	Sequencing
L333	GCTTCTGGCCACAGCACT	Sequencing
H16498	CCTGAAGTAGGAACCAAGATG	Sequencing
H408	CTGTTAAAAGTGCATACCGCCA	Sequencing

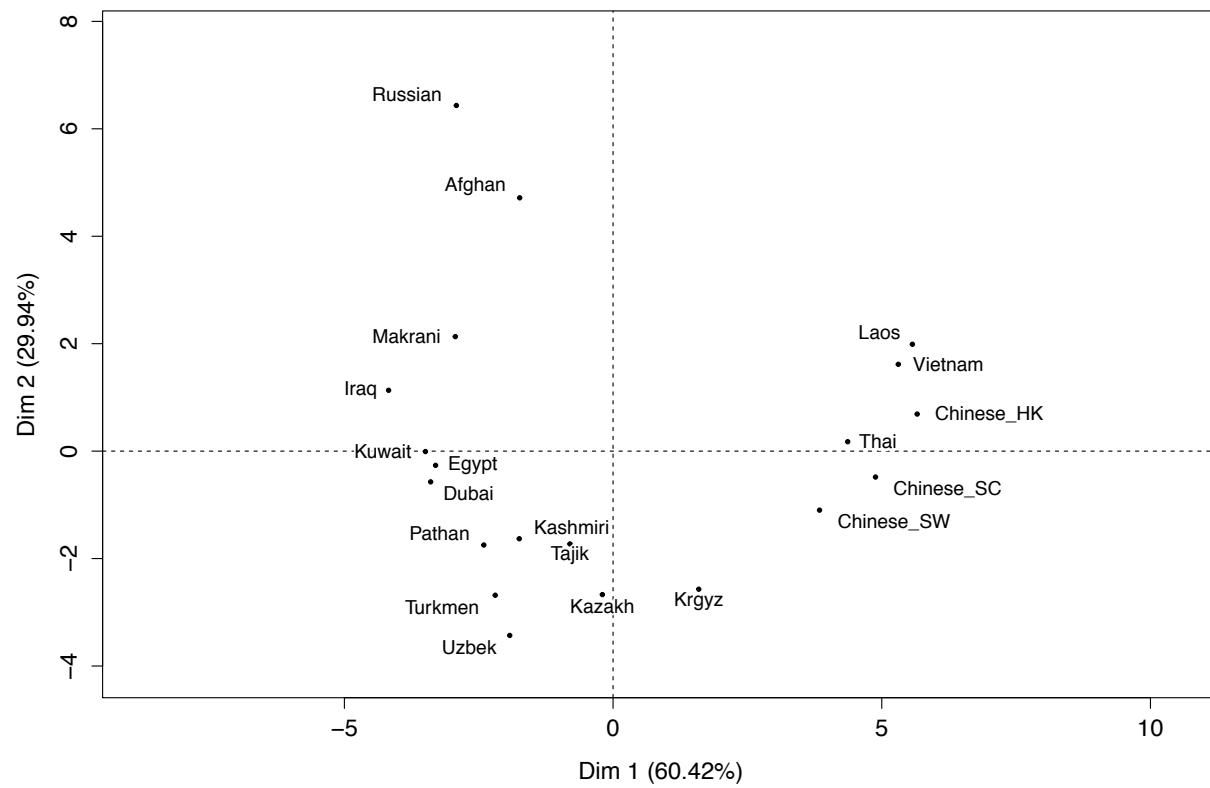
**Table S6. Pairwise differences calculated using Arlequin 3.5.1.2**

		Table S6a. Population pairwise Fst values																			
		Afghan	Chinese_HK	Dubai	Egypt	Kashmiri	Chinese_ZW	Iraq	Kazakh	Kuwait	Kyrgyz	Laos	Chinese_WD	Makrani	Pathan	Russian	Tajik	Thai	Turkmen	Uzbek	Vietnam
Afghan	0																				
Chinese_HK	0.08047	0																			
Dubai	0.03445	0.05766	0																		
Egypt	0.02809	0.05794	0.01036	0																	
Kashmiri	0.04942	0.05265	0.02055	0.02903	0																
Chinese_ZW	0.07575	0.00138	0.05541	0.05611	0.04369	0															
Iraq	0.03217	0.07841	0.01146	0.00631	0.03058	0.07232	0														
Kazakh	0.05452	0.02958	0.0261	0.03172	0.02465	0.02224	0.03779	0													
Kuwait	0.04283	0.06402	0.00722	0.01172	0.02611	0.06107	0.00531	0.03024	0.0066	0.03923	0										
Krygz	0.06815	0.01782	0.03501	0.03825	0.02449	0.01203	0.05095	0.02029	0.01796	0.05349	0.05789	0									
Laos	0.08092	0.01827	0.06156	0.05613	0.05302	0.01726	0.07124	0.05119	0.06443	0.0359	0										
Chinese_WD	0.07414	0.01103	0.0475	0.04758	0.04012	0.04046	0.06207	0.02018	0.05168	0.01416	0.01547	0									
Makrani	0.04931	0.07821	0.0114	0.01768	0.03481	0.07073	0.02029	0.04146	0.01796	0.05349	0.07789	0.05939	0								
Pathan	0.03648	0.05359	0.01468	0.01967	-0.00009	0.04476	0.02194	0.02189	0.02122	0.02996	0.05121	0.03795	0.0281	0							
Russian	0.05958	0.11475	0.0385	0.02938	0.05831	0.09953	0.01572	0.06093	0.0314	0.08075	0.10176	0.0914	0.06091	0.04663	0						
Tajik	0.04945	0.04323	0.02964	0.02314	0.02446	0.03722	0.03054	0.02401	0.02679	0.02434	0.04176	0.03129	0.03465	0.02498	0.05578	0					
Thai	0.0691	0.00977	0.05061	0.04822	0.04425	0.01108	0.06115	0.03404	0.05514	0.02732	0.00645	0.01024	0.06962	0.04101	0.0883	0.03644	0				
Turkmen	0.04628	0.04535	0.01579	0.01805	0.01568	0.03573	0.01771	0.00814	0.01545	0.014	0.05261	0.03083	0.03431	0.0166	0.03759	0.01889	0.04193	0			
Uzbek	0.03842	0.03492	0.01668	0.01574	0.01319	0.02886	0.01914	0.00506	0.01789	0.00976	0.04456	0.02547	0.0302	0.01144	0.03874	0.01461	0.0322	0.00272	0		
Vietnam	0.07372	0.01334	0.05718	0.05473	0.05064	0.01389	0.06921	0.04752	0.06293	0.03518	0.00014	0.01604	0.07364	0.04687	0.10288	0.04142	0.00472	0.05309	0.04254	0	
		Table S6b. Fst p-values																			
		Afghan	Chinese_HK	Dubai	Egypt	Kashmiri	Chinese_ZW	Iraq	Kazakh	Kuwait	Kyrgyz	Laos	Chinese_WD	Makrani	Pathan	Russian	Tajik	Thai	Turkmen	Uzbek	Vietnam
Afghan	*																				
Chinese_HK	0.00000±0.0000	*																			
Dubai	0.00000±0.0000	0.00000±0.0000	*																		
Egypt	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*																	
Kashmiri	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*																
Chinese_ZW	0.00000±0.0000	0.09570±0.0081	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*															
Iraq	0.00000±0.0000	0.00000±0.0010	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*														
Kazakh	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*													
Kuwait	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*												
Krygz	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*											
Laos	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*										
Chinese_WD	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*									
Makrani	0.00000±0.0000	0.00000±0.0000	0.00000±0.0010	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*							
Pathan	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.04687±0.0198	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*					
Russian	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	*				
Tajik	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000	0.00000±0.0000																	

Table S5. Point heteroplasmy observed

SampleID	Heteroplasmic Position	Haplotypes (based on dominant sequence variants)
KPR026	16311Y	16519C 263G 315.1C
KPR029	263R	16051G 16239T 73G 146C 234G 263G 315.1C 575T
KPR054	16181R	16147A 16172C 16189C 16223T 16248T 16320T 16355T 16519C 73G 152C 199C 204C 263G 315.1C 573.1C
KPR064	16111M	16223T 16234T 16243C 16244A 16290T 16519C 73G 186T 249d 263G 315.1C 489C
KPR122	16169Y	16129A 16169T 16178C 16223T 16270T 16319A 16352C 16519C 73G 143A 195C 204C 263G 309.1C 309.2C 315.1C 447G 489C
KPR156	152Y	16129A 16213A 16362C 16519C 73G 152C 195C 263G 309.1C 315.1C
KPR308	152Y	16223T 16278T 16519C 73G 152C 195A 263G 315.1C 489C 523d 524d

**Individuals factor map (PCA)**



**Fig S1.** Principal component analysis plot for all the compared populations based on  $F_{ST}$  values calculated by Kimura 2-parameter