INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 8, No. 5, p. 262-273, 2016

RESEARCH PAPER OPEN ACCESS

# Mitochondrial DNA analysis of five Pathan tribes from Pakistan

**Sadia Tabassum[1*], Habib Ahmad[2], Muhammad Ilyas[2], Ikram Muhammad[2], Amanda Owings[3], Hongjie Lee[3]**

*[*1]Department of Zoology, Hazara University Mansehra, Pakistan*

*[2]Department of genetics, Hazara University Mansehra, Pakistan*

*[3]Department of Anthropology, University of Illinois at Urbana Champaign, USA*

**Key words:** mtDNA HVI, Pathans of Charsada and Mardan, Population data.

http://dx.doi.org/10.12692/ijb/8.5.261-273 Article published on May 30, 2016

## Abstract

Substantial genetic diversity exists in Pakistani human population and understanding the evolution of this diversity is complicated due to several waves of migration from populations in the North and Northwest. Pathans are one of the largest ethnic groups of Pakistan inhabiting the vast geographical areas, specifically the northwestern part.

In the present study, we assessed the extent of genetic diversity using mtDNA sequence analysis of HVI (Hyper variable I) control region of the Pathan populations of Mardan and Charsada districts. A total of 165 buccal swabs were collected from five major populations of the two geographically adjacent districts, Charsada and Mardan. Mitochondrial control region HVI data was generated for all the samples. mtDNA haplogroups were assigned to each sample using a phylotree (www.phylotree.org). Principle Coordinate (PCoA) plot was generated by combining our data set with other published datasets from neighboring populations of central Asia, Middle East, Europe and South Asia. The most frequent mtDNA HVI macro haplogroups R (63.4%), M (26.8%) and N (8.6%) were observed among Pathan populations. Some novel mitochondrial haplogroups (mtDNA Hgs) of M and R sub-clades have also been detected that had not been observed in previous investigation of Pathans. The clustering pattern when compared on the basis of mitochondrial HVI data sets of different neighboring populations depicts less female mediated gene flow among populations. In addition, the mitochondrial genetic structure of some sub-populations may be influenced by Turkish invasion.

* **Corresponding Author:** Sadia Tabassum ✉ saadia.tabassum81@gmail.com

**Introduction**

Pakistan occupies an important geographic location on the map that has been an attractive voyaged rout followed by the drifters out of Africa. Pakistan is considered to be one of the first areas inhabited by the followers of costal rout (Walport *et al.*, 2000: Qamar *et al.*, 1999). Its population possess rich socio-linguistic and cultural history. The genetic pool of Pakistani population exhibits prominent genetic imprints of several waves of migration from Europe, East Asia, Central Asia and Middle East. The genetic differentiation and structuring of the population is contributed mainly by commonly practiced endogamy (Quintana *et al.*, 2004).

Almost 12 ethnic groups speaking more than twenty vernaculars are residing across Pakistan. Pathans comprising the third largest ethnic group among Punjabi, Sindhi, Saraiki, Mahajir, Baluchi, Hindkowans and Chitralis inhabit the vast area of Khyber Pakhtunkhwa (KPK) province in Northwest Pakistan (Caroe, 1957). Pakistani Pathans exhibit a heterogeneous origin with an admixture from West Eurasian, South Asian and East Asian source population (Rakha *et al.*, 2011). So far all the previously investigated Pathans were sampled generally with no sub-tribal discrimination and were considered as a single population. This study focused on the populations of Mardan and its neighboring area of Charsada district formerly called Pushkalavati that has been an important hub for the Alexander's army and had been an important probable rout for the west to eastward migration of different invaders into the India subcontinent. The major populations inhabiting this area are Mohmand (MM, MD), Muhammadzai (MZ), Yousafzai (YS) and Kakakhel Mian (KM). They might be influenced by the geographic and ethnographic background, foreign invasions and patrilocality. On the basis of various oral traditions about the origin of these populations, we expect them to be more homologous genetically being the descendants of Bani Israel except for Kakakhel Mian (KM) who claims their descent from Arabs (Caroe, 1984; Elahi, 1996).

The aim of this study was to dissect the genetic structure of these Pathan populations and to evaluate the extent of variation within and among the populations of the two geographically adjacent areas and the already investigated indigenous Pakistani population, using mtDNA HVI control region as a marker and to elucidate the genetic contribution of the neighboring populations.

**Materials and methods**

Buccal swabs were collected from 165 unrelated males of Muhammadzai (MZ) (n=28), Mohmand (MM) (n=31), Kakakhel Mian (KM) (n=35) populations from Charsada district and Yousafzai (YS) (n=36) and Mohmand (MD) (n=35) populations from Mardan district after obtaining informed consent from all the donors (Fig.1).

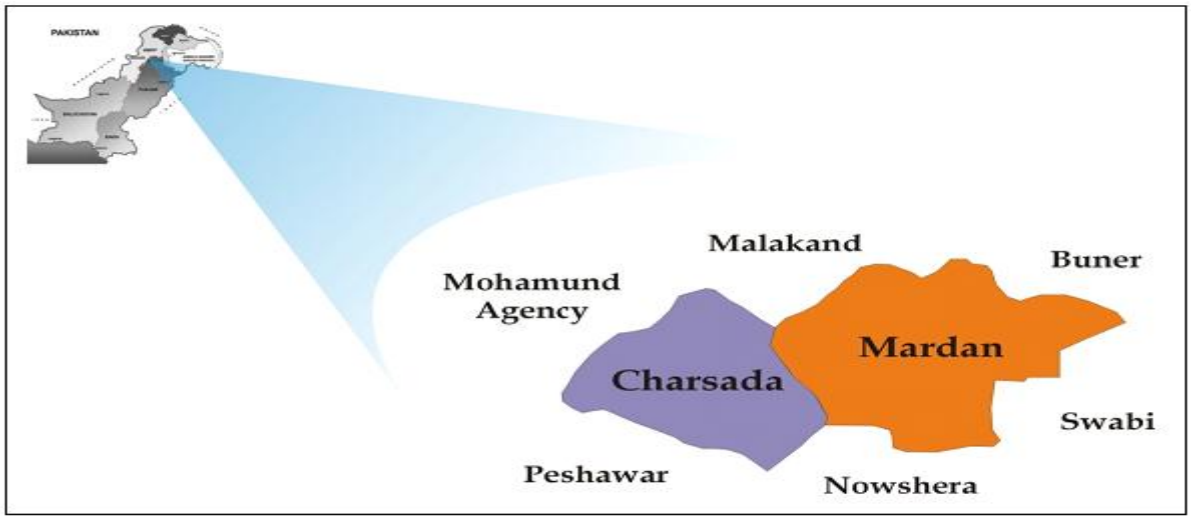


**Fig. 1.** Location map of Charsada and Mardan district (Khyberpakhtunkhwa) Pakistan.

*DNA Isolation*

DNA extraction was carried out using Phenol Chloroform protocol developed by Marisi and Sergio (Marisi and Sergio, 2007).

*Amplification and sequencing of mitochondrial DNA*

DNA of 165 samples from all five groups was amplified for mtDNA HVI sequence data and variable positions were determined from np 16024 – 16365. 10ng of Selected DNA sequences from all five populations were amplified with 5µM of 15971F and 484R primers, 10mM of dNTPs, 50mM $MgCl_2$ and 5U of Taq polymerase. 35 cycles of reaction were performed with denaturation at 94°C for 3 min, 30sec annealing at 94°C, 30 sec extension at 54°C and final extension at 72°C for 10 min.

PCR product was purified with Shrimp Alkaline Phosphatase and Exonuclease I following Malhi's lab protocol (Malhi *et al.*, 2010). Sequencing was carried out using Big Dye™ terminator cycle sequencing kit of Applied Biosystem on Genetic analyzer 3730.

Sequences were read using the program Sequencher 4.7 and all sequence data was aligned with revised Cambridge reference sequence (rCRS) (rCRS; Andrews *et al.*, 1999) using BioEdit program (Hall, 1999) and haplogrup was assigned to each sample using phylotree build 16 (www.phylotree.org) (Van *et al.*, 2009) and mitomaster (http:/www.mitomap.org) (Lott *et al.*, 2013).

To investigate the genetic diversity of five major Pathan tribes of Charsada and Mardan districts and to compare with the other Pathan populations in the neighboring area of Mohmand agency and Swat valley and populations from neighboring countries, their mtDNA HVI sequence data was exploited.

*Comparative populations*

For mtDNA analysis the comparative data from populations of India [Afridi, Tamil] (Metspalu *et al.*, 2004), Israel [Druze, Bedauin] (Hunley *et al.*, 2009), Greece [Greeks] (Irwin *et al.*, 2008), Russian Federation [Yakut] (Hunley *et al.*, 2009), Iran [Armenian, Azeris, Persian, Qashqis] (Derenko *et al.*, 2013), Iraq [Iraqi] (Abu *et al.*,2008), China [Uighur] (Hunley *et al.*, 2009), Afghanistan, Turkey [Turkish] (Comas *et al.*, 1996), indigenous populations from Pakistan [Pathan, Kalash, Brusho, Balouch, Brahui] (Hunley *et al.*, 2009) was used.

*Statistical analysis*

Genetic diversity estimates were calculated for all the five Pathan populations of the two areas separately as well as collectively using a formula $[n\ (1-\sum pi^2)/n-1]$, where $n$ is the number of samples and $pi$ is the frequency of the $i^{th}$ haplotype in the population (Nei. 1987). Genetic distances (Matrix of Slatkin linearized $F_{ST}$) were obtained by using ARLEQUIN Ver. 3.5.2.2 software for our data and for other global populations. Of all pairwise $F_{ST}$ values a distance matrix was constructed that was later used to construct a PCoA plot using GENALEX program (Peakall *et al.*, 2006). Median Joining (MJ) networks were constructed using NETWORK 4.6.1.3 (http://www.fluxus-engineering.com) by giving each locus a weight according to its estimated mutation rate (Keyser *et al.*, 2003). Mitomaster and phylotree were used to predict the mitochondrial haplogroups (http://www.mitomap.org) (Lott *et al.*, 2013).

Mantel tests were performed to measure the degree of correlation between mtDNA based Genetic distance among population and geographical distance.

**Results and discussion**

A total of 93 haplotypes were identified in 165 samples (table. 1). 38.7% of the haplotypes were found to be shared among all the populations and 57 haplotypes (61.3%) were singletons. Genetic diversity in all the samples amplified for HVI control region was 0.993 which is similar to the already investigated Pathans from the nearby area of FATA (Rakha *et al.*, 2011). It remained highest among all the populations when compared with the already published datasets of indigenous ethnic groups (table. 2) and Match probability was 0.987. Haplotype (gene) diversity for all the five populations ranges from 0.97814 for Kakakhel Mian (KM) to 0.994 for Muhammadzai (MZ).

**Table 1.** List of SNPs defining mtDNA HVI Haplogroups in five Pathan populations of Charsada and Mardan district.

| Macro Haplogroup | Haplogroup | Frequency | Origin | No.of variants | SNPs |
|---|---|---|---|---|---|
| R | U2e | 4 | S.Asian | 7 | A16051G, G16129C, A16182C, A16183C, T16189C, A16194C, C16197G |
| R | H2(T152C T16311C) | 5 | SW.Asian/Middle Eastern | 1 | T16311C |
| M | M3 | 7 | S.Asian | 4 | T16126C, C16223T, T16311C, T16519C |
| R | J1b1a1 | 5 | Middle eastern | 7 | C16069T, T16126C, G16145A, T16172C, C16222T, C16261T, G16474T |
| R | U5a2 | 1 | S.Asian | 5 | C16192T, C16256T, C16270T, C16278T, G16526A |
| R | U2c | 2 | S.Asian | 2 | A16051G, C16179T |
| R | I1 | 1 | SW.Asian/Middle Eastern | 8 | G16129A, A16180G, C16223T, A16254G, A16293C, T16311C, G16391A, T16519C |
| R | R6(G16129A) | 3 | SW.Asian/Middle Eastern | 4 | G16129A, G16213A, T16362C, T16519C |
| R | J1c | 1 | Middle eastern | 6 | C16069T, T16126C, C16168T, C16266T, C16278T, T16311C |
| R | U6c | 3 | S.Asian | 5 | G16129A, C16169T, A16183C, T16189C, A16194C |
| M | M5a1 | 1 | S.Asian | 5 | G16129A, C16223T, C16234T, C16291T, T16519C |
| R | H1 | 1 | SW.Asian/Middle Eastern | 2 | A16051G, C16239T |
| R | J1b5b | 2 | Middle eastern | 7 | C16069T, T16126C, G16145A, C16261T, T16263C, C16290T, T16519C |
| R | U2e1f | 1 | S.Asian | 11 | A16051G, G16129C, A16182C, A16183C, T16189C, A16194T, C16221T, T16311C, T16325C, T16362C, C16447T |
| M | M33 | 1 | S.Asian | 4 | C16223T, T16324C, T16357C, C16527T |
| M | M6a1b | 2 | S.Asian | 5 | C16188T, C16223T, T16231C, T16362C, T16519C |
| R | R8a1a3 | 2 | SW.Asian/Middle Eastern | 4 | C16259T, C16292T, A16497G, T16519C |
| M | M4 | 3 | S.Asian | 7 | T16086C, C16111T, G16145A, C16223T, C16261T, T16311C, T16519C |
| R | H6 | 1 | SW.Asian/Middle Eastern | 2 | T16362C, A16482G |
| N | N7 | 2 | C.Asian | 3 | G16129A, C16223T, T16519C |
| R | U2e1h | 2 | S.Asian | 6 | A16051G, G16129C, A16183C, C16193CC, T16362C, T16519C |
| M | C4a | 1 | C.Asian | 5 | T16093C, G16129A, T16298C, C16327T, T16519C |
| R | R2 | 5 | S.Asian | 2 | C16071T, T16519C |
| R | H10((T16093C)) | 3 | SW.Asian/Middle Eastern | 4 | T16093C, C16278T, G16319A, T16519C |
| R | H2a2a1d | 2 | SW.Asian/Middle Eastern | 3 | T16172C, A16182AC, A16183C |
| N | N1a1a1a | 1 | Middle eastern | 7 | C16147A, T16172C, C16223T, C16248T, C16320T, C16355T, T16519C |
| R | U2a | 1 | S.Asian | 3 | A16051G, A16206C, T16311C |
| R | H2a3 | 1 | SW.Asian/Middle Eastern | 5 | T16093C, C16223T, C16234T, G16274A, T16519C |
| N | W4 | 2 | IndoEuropian | 5 | T16086C, C16223T, C16286T, C16292T, T16519C |
| N | W | 2 | IndoEuropian | 5 | C16192T, C16223T, A16284G, C16292T, T16519C |
| R | H2a2a1g | 6 | SW.Asian/Middle Eastern | 4 | T16172C, A16182AC, A16183C, T16189C |
| M | G2a1d2 | 1 | E.Asian | 4 | C16223T, A16227G, C16278T, T16362C |
| R | U4a3 | 1 | S.Asian | 3 | T16356C, T16362C, T16519C |
| R | U7 | 4 | S.Asian | 5 | A16309G, A16318T, A16343G, T16362C, T16519C |
| R | U5a1b(T16362C) | 2 | S.Asian | 6 | C16192T, C16256T, C16270T, T16362C, A16399G, T16519C |
| M | M5 | 2 | S.Asian | 3 | G16129A, C16223T, T16519C |
| M | M65a(C16311T) | 2 | S.Asian | 5 | T16126C, C16223T, A16289G, A16399G, T16519C |
| M | M5a2a1a | 1 | S.Asian | 4 | G16129A, C16223T, A16265C, T16519C |
| R | H94 | 1 | SW.Asian/Middle Eastern | 2 | C16339T, C16355T |

| | | | | | |
|---|---|---|---|---|---|
| N | N | 6 | C.Asian | 3 | C16223T, C16292T, T16519C |
| R | H5 | 4 | SW.Asian/Middle Eastern | 3 | C16266T, T16304C, T16519C |
| M | D4q | 1 | E.Asian | 6 | C16223T, C16256T, T16311C, G16319A, T16362C, T16519C |
| M | M52a | 1 | S.Asian | 5 | C16223T, A16275G, G16303A, G16390A, T16519C |
| M | M71 | 1 | S.Asian | 2 | C16223T, T16271C |
| R | T1a | 1 | W.Eurasian | 6 | T16126C, A16163G, C16186T, T16189C, C16294T, T16519C |
| R | P6 | 1 | SE.Asian | 6 | C16221T, T16311C, T16325C, T16362C, C16447T, T16519C |
| M | M2a1 | 1 | S.Asian | 7 | T16075C, C16223T, C16270T, G16274A, G16319A, T16352C, T16519C |
| R | H15a | 2 | SW.Asian/Middle Eastern | 1 | C16184T |
| R | T1a1'3 | 2 | W.Eurasian | 6 | T16126C, A16163G, C16186T, T16189d, C16294T, T16519C |
| M | M49 | 2 | S.Asian | 3 | C16223T, C16234T, T16519C |
| R | U2e3 | 1 | S.Asian | 4 | C16168T, C16234T, C16260T, T16362C |
| R | H13a1d | 1 | SW.Asian/Middle Eastern | 1 | C16234T |
| N | Y2 | 1 | C.Asian | 5 | T16126C, C16223T, C16266T, T16311C, T16519C |
| R | H2a2b | 1 | SW.Asian/Middle Eastern | 2 | C16291T, T16311C |
| R | H1e | 2 | SW.Asian/Middle Eastern | 3 | G16129A, A16182AC, A16183C |
| R | R0a | 2 | S.Asian | 3 | T16126C, T16362C, T16519C |
| R | H4a | 1 | SW.Asian/Middle Eastern | 1 | C16287T |
| M | M18b | 1 | S.Asian | 5 | A16160G, C16223T, A16318T, T16325C, T16519C |
| R | P4a | 1 | SE.Asian | 3 | A16037G, C16111T, G16319A |
| R | H14b1 | 1 | SW.Asian/Middle Eastern | 2 | T16126C, T16519C |
| M | C4a1(G16129A) | 4 | C.Asian | 5 | G16129A, C16223T, T16298C, C16327T, T16519C |
| M | M2b | 1 | S.Asian | 1 | C16169CC |
| R | H1bt | 1 | SW.Asian/Middle Eastern | 6 | C16292T, C16355T, T16406d, A16497G, T16519C, C16527T |
| R | F1c1a | 1 | SE.Asian | 3 | C16111T, G16129A, T16304C |
| N | W6 | 2 | IndoEuropian | 6 | C16192T, C16223T, C16292T, T16325C, C16465T, T16519C |
| R | H13a1a1d | 1 | SW.Asian/Middle Eastern | 6 | T16086C, G16145A, C16173T, C16261T, T16311C, T16519C |
| R | I6a | 1 | SW.Asian/Middle Eastern | 8 | G16129A, C16223T, A16293C, T16311C, T16362C, G16391A, C16447T, T16519C |
| R | H3p | 1 | SW.Asian/Middle Eastern | 3 | C16069G, C16222T, A16269G |
| R | J1b3 | 1 | Middle eastern | 6 | C16069T, T16126C, G16145A, C16222T, A16235G, C16261T |
| M | M39a | 1 | S.Asian | 1 | C16353T |
| R | T1 | 1 | W.Eurasian | 7 | T16093C, T16126C, A16163G, C16186T, T16189C, C16294T, T16519C |
| R | H14a | 1 | SW.Asian/Middle Eastern | 3 | C16256T, C16270T, T16352C |
| R | T2e | 1 | W.Eurasian | 5 | T16126C, G16153A, C16294T, C16296T, T16519C |
| R | U2b2 | 1 | S.Asian | 2 | A16051G, C16239T |
| M | M2a1a | 2 | S.Asian | 6 | C16223T, C16270T, G16319A, T16352C, C16449T, C16451T |
| R | H2a2a1c | 4 | SW.Asian/Middle Eastern | 8 | A16051G, T16086C, C16259A, C16267T, C16291T, A16300G, A16326G, C16353T |
| R | JT | 1 | Middle eastern | 3 | T16086C, T16126C, T16519C |
| R | HV2 | 1 | W.Eurasian | 2 | T16217C, C16446T |
| R | U1a2 | 1 | S.Asian | 4 | G16129A, T16189C, C16192T, A16202C |
| R | J1b4a | 1 | Middle eastern | 6 | C16069T, T16126C, G16145A, C16218T, C16261T, C16287T |
| R | H17a1 | 1 | SW.Asian/Middle | 4 | G16129A, C16223T, C16291T, T16519C |

| | | | | | |
|---|---|---|---|---|---|
| | | | Eastern | | |
| M | C4a2'3'4 | 1 | C.Asian | 6 | C16223T, T16297C, T16298C, C16327T, T16357C, T16519C |
| R | U2a1a | 1 | S.Asian | 6 | A16051G, T16154C, A16206C, A16230G, T16311C, T16519C |
| R | T2b2b | 1 | W.Eurasian | 5 | C16111A, T16126C, C16294T, C16296T, T16519C |
| M | M3c2 | 1 | S.Asian | 4 | T16126C, T16154C, C16223T, T16519C |
| M | C4a4b | 1 | C.Asian | 8 | T16086C, G16129A, C16150T, C16223T, T16298C, C16327T, T16357C, T16519C |
| R | H1k | 1 | SW.Asian/Middle Eastern | 4 | A16051G, T16189C, C16290T, C16292T |
| R | J1b | 2 | Middle eastern | 5 | C16069T, T16126C, G16145A, C16261T, T16519C |
| R | R | 1 | S.Asian | 3 | C16292T, A16497G, T16519C |
| N | N1b1 | 1 | C.Asian | 7 | G16145A, C16176G, C16223T, C16256T, A16309G, G16390A, T16519C |
| L3 | L3 | 1 | E.African | 5 | T16093C, G16129A, C16223T, A16305T, T16519C |
| M | M38d | 1 | S.Asian | 5 | G16129A, C16223T, C16266T, T16311C, T16519C |
| R | R6b | 2 | E.Asian | 7 | C16179T, A16227G, C16245T, C16266T, G16274A, C16278T, T16362C |

**Table 2.** Comparison of mtDNA Genetic diversity among various ethnic groups of Pakistan.Path(Pathan), Mkr(Makrani), Pth(Pathan), Bal(Baluch), Brh(Brahui), Haz(Hazara), Bsh(Burusho), Ksh(Kalash), Par(Parsi), Snd(Sindhi), Sar(Saraiki).

| Parameters | Path (this study) | Mkr (Siddiqi *et al.* 2015) | Pth (Rakha *et al.*2011) | Bal (Quintana *et al.* 2004) | Brh (Quintana *et al.* 2004) | Haz (Quintana *et al.* 2004) | Bsh (Quintana *et al.* 2004) | Ksh (Quintana *et al.* 2004) | Par (Quintana *et al.* 2004) | Snd (Quintana *et al.* 2004) | Sar (Hayat *et al.* 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of Sample | 165 | 100 | 230 | 39 | 38 | 23 | 44 | 44 | 44 | 23 | 85 |
| No. of haplotypes | 93 | 70 | 157 | 26 | 22 | 21 | 32 | 12 | 22 | 21 | 63 |
| No of unique haplotypes | 57 | 54 | 128 | 18 | 15 | 19 | 25 | 5 | 12 | 19 | 58 |
| Genetic Diversity | 0.993 | 0.968 | 0.993 | 0.974 | 0.952 | 0.992 | 0.980 | 0.851 | 0.950 | 0.992 | 0.957 |

Nucleotide diversity within the five populations ranges from 0.010874 +/- 0.005893 for Mohmand from Charsada (MM) to 0.021940 +/- 0.012445 for Kakakhel Mian (KM) (table.3). Other statistical parameters for the populations of the two areas are given in table 4.

To characterize the maternal genetic variation among the populations, haplogroup frequencies were calculated. Among the 93 haplotypes, 63.4% of the samples belong to the haplogroup R while haplogroup M, N and L were found with the frequency of 26.8%, 8.6% and 1.1% respectively (Fig.2).

The mtDNA Macro haplogroups found in these populations are R, M, N and L. The first three haplogroups are thought to be originated around 60000-75000 years ago in South Asia (Kivisild *et al.*, 2003). Therefore, the presence of these haplogroups suggests a South Asian maternal origin of these populations. In the present study, the mitochondrial haplogroups of South Asian origin have highest proportion of 37.6%, comprising M3 (7.5%), R2(5.3%), U7, U2e (4.3%), M4, U6c (3.2%), U2c, M65a, M6a1b, U5a1b, M49, M5, M2a1a, M6a1b, R0a, U2e1h (2.2%) and U5a2, U2e3, U4a3, U2e1f, U2b2, U2a1a, U2a, M52a, M5a1, M33, M38d, M39a, M2b, M18b, M71(1.1%).

**Table 3.** Genetic diversity indices of five Pathan populations.

| populations | No. of samples (N) | No. of Haplotypes | Unique haplotypes | Haplotype (Gene) diversity | Nucleotide Diversity |
|---|---|---|---|---|---|
| MZ | 28 | 26 | 24 | 0.994 | 0.012736 +/- 0.007952 |
| KM | 35 | 26 | 19 | 0.97814 | 0.021940 +/- 0.012445 |
| MM | 31 | 26 | 21 | 0.989 | 0.010874 +/- 0.005893 |
| YS | 36 | 32 | 29 | 0.993 | 0.015125 +/- 0.009150 |
| MD | 35 | 27 | 20 | 0.98951 | 0.014993 +/- 0.009093 |

**Table 4.** Statistic parameters of five populations of Charsada and Mardan. MZ (Muhammadzai), KM (Kakakhel Mian), MM (Mohmand from Charsada), YS (Yousafzai), MD (Mohmand from Mardan), S.d (Standard deviation).

| Population | MZ(28) | KM(35) | MM(31) | YS(36) | MD(35) | Mean | S.d |
|---|---|---|---|---|---|---|---|
| No. of Transitions | 22 | 18 | 35 | 15 | 20 | 22.0 | 7.714 |
| No. of Transversions | 4 | 6 | 3 | 2 | 4 | 3.8 | 1.483 |
| No. of Substitutions | 26 | 24 | 38 | 17 | 24 | 25.8 | 7.629 |
| No. of Indels | 0 | 2 | 4 | 1 | 2 | 1.8 | 1.483 |

The second major haplogroups identified were of Southwest Asian or Middle Eastern origin (36.5%) including H2a2a1g (6.4%), J1b1a1, H2 (5.2%), H2a2a1c, H5 (4.3), H10, R6 (3.2), H15a, H1e, H2a2a1d, R8a1a3, J1b, J1b5b (2.2%). The third most prevalent mitochondrial haplogroups are of Central Asian origin (8.6%) followed by West Eurasian (6.5%), east Asian and Indo-Europeans (3.2%) and the least frequent haplogroup was found L (1.1%) of East African origin.

**Table 5.** AMOVA Analysis (mtDNA HVI Haplotypes) of five populations from two districts (groups).

| Source of Variation | Degree of freedom | Sum of Squares | Variance Components | Percentage of Variation |
|---|---|---|---|---|
| Among Groups | 1 | 1.270 | -0.02156 Va | -1.57 |
| Among population within Groups | 3 | 9.600 | 0.04977 Vb | 3.63 |
| Within populations | 182 | 244.445 | 1.34311 Vc | 97.94 |
| Total | 186 | 255.316 | 1.37131 | |

The high frequency of macro haplogroup M had been recorded in Asia, particularly in India, Bangladesh, Nepal and Tibet reaching up to 60 to 80% (Rajkumar *et al.*, 2005). Mitochondrial haplogroup M3 being the sub-clad of HgM, has been found in South Asia, with highest frequency in West India and Pakistan *(Metspalu et al., 2004). It has been reported in Pakistani Pathan population with frequency of 7.8% (Rakha et al., 2011).*

**Table 6.** Result of Mantel Test.

| mtDNA and Geography | | Coefficient | P value |
|---|---|---|---|
| Correlation Coefficient | mtDNA / Geography | -0.07 | 0.46 |

The frequency of these mtDNA macro haplogroups found in the populations we genotyped in present study is in close agreement with the one already reported in Pakistani Pathans with a little variation in frequency of sub-clades. The Haplogroup F1c1a, G2a1d2, H10, H13a1, H14b1, H15a, H17, H3P, H4a, I6a, J1c, JT, L3, N1a1a1a, N7, P4a, P6, U2e, U6c, W4 and Y2 are the novel haplogroups for Pathan population, have not been found in the previously studied Pathans by Rakha *et al.*, 2011.

Mitochondrial haplogroups HV9, M65, H1k, U2a1, M71, M52 and U5a1b were confined to the populations of Mardan only while haplogroups U7, M18 and P6 were restricted to Charsada populations only and absent in Mardan. The high frequencies of local lineages in these populations confirm their descent as a consequence of Paleolithic population expansion (McElreavey *et al.*, 2005).

To obtain insight into the genetic structure of the studied populations Median joining Networks were constructed for all the major haplogroups based on the frequency of haplotypes in these five populations using program NETWORK 4.6.1.3. Little sharing of haplotypes was observed among populations in

haplogroup R and a star haplotype with total frequency of 47.8% in MM, 17.4% in MD, 13.04% in MZ and KM and 8.7% in YS was found in this haplogroup sequences (Fig.3a). While very little sharing of haplotypes was observed in Hg M and no sharing was recorded in Hg N (Fig.3b, c).

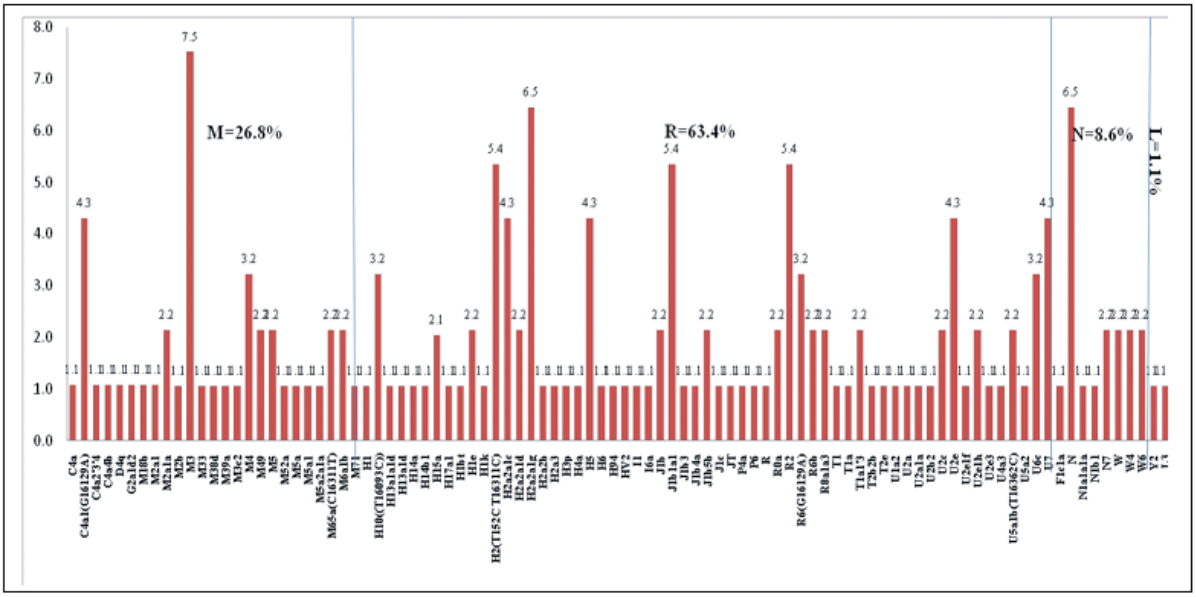


**Fig. 2.** mtDNA HVI Haplogroup frequencies among five Pathan populations of Charsada and Mardan (Pakistan).

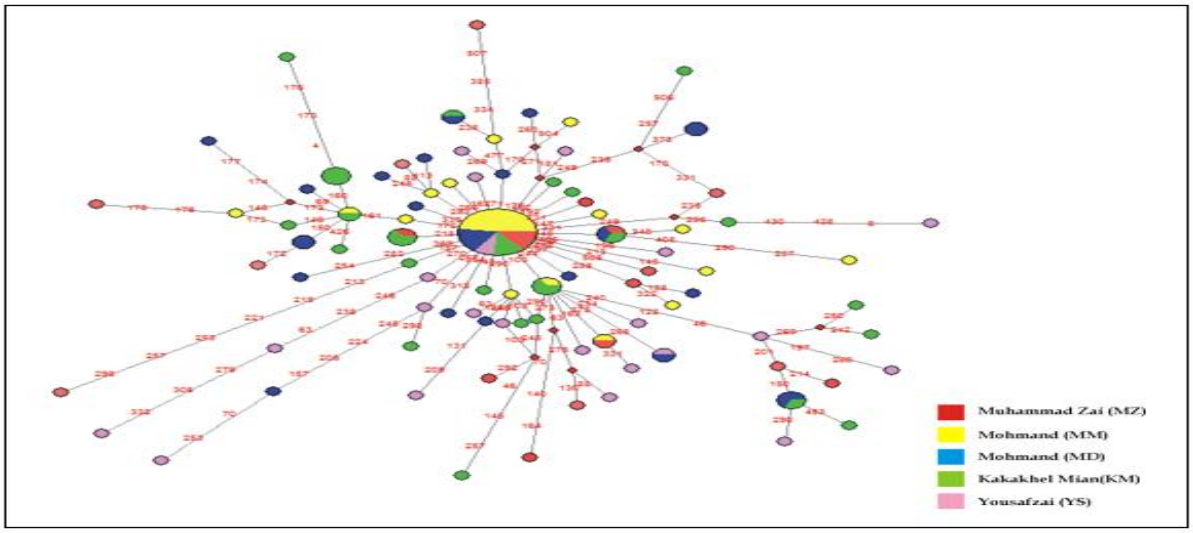


**Fig. 3a.** Median-Joining network of mtDNA haplogroup R in five populations of Charsada and Mardan. Areas of circles are proportional to the haplotype frequencies.

The proportion of genetic variation distributed within and between the populations of two districts was assessed by analysis of molecular variance (AMOVA). Percentage of variation among groups was -1.57% and variation among populations within the group was 3.63% while percentage of variation within populations was found to be 97.94% (table.4).

$F_{ST}$ values were calculated for all pair of populations in Arlequin V.3.5.2.2 (Excoffier *et al.*, 2010). No significant genetic distance ($F_{ST}$) was observed between population pairs except for the Kakakhel Mian and Mohmand. The PCoA plot provided information regarding the degree of association between different indigenous and global populations using comparative sequence data from Greece, Russia, Iraq, Iran, Israel, Xinjiang china, Afghanistan, Turkey, South India, North Indian Afridi, Balouchi,

Brahui, Brusho, Kalash and Pakistani Pathan population. Clustering pattern revealed that Mohmand (MD) from Mardan was clustered together with Pathans from Pakistan and Yousafzai (YS) shows affinity with Afridi Pathans from India. As most of the previously studied Pathans were sampled from Mohmand agency of FATA and were Mohmand so they exhibit homogeneity with Mohmand (MD) from Charsada, while Muhammadzai and Kakakhel Mian from Charsada are clustered together with Brahui population from Pakistan.

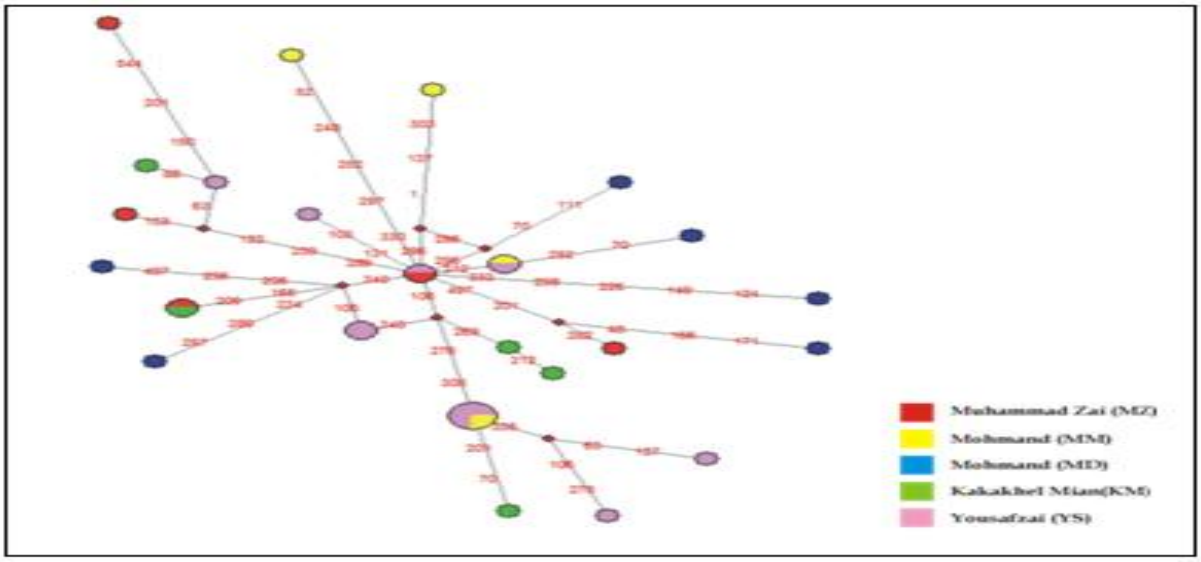


**Fig. 3b**. Median-Joining network of mtDNA haplogroup M in five populations of Charsada and Mardan. Areas of circles are proportional to the haplotype frequencies.

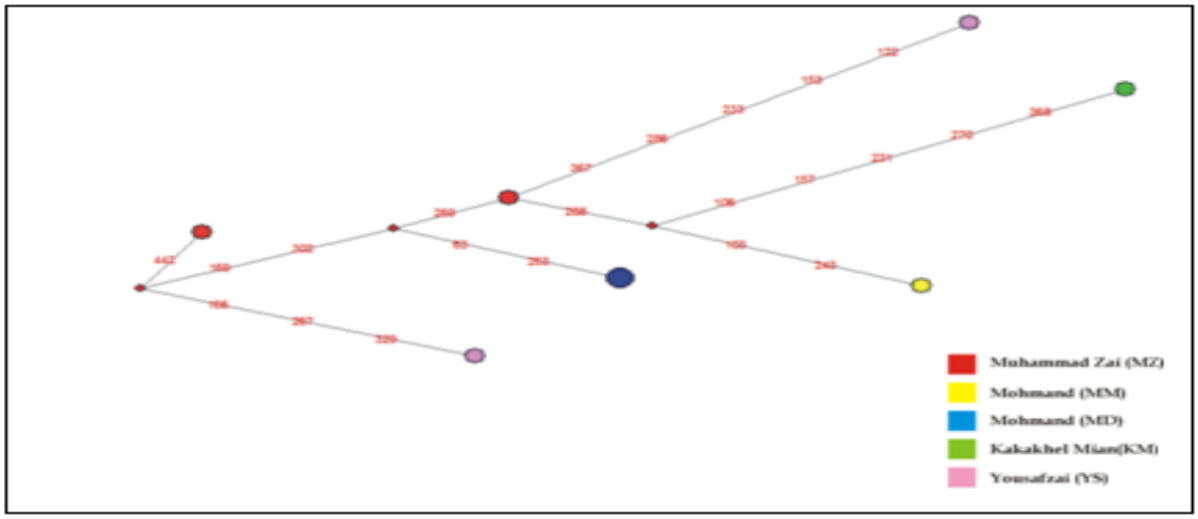


**Fig. 3c.** Median-Joining network of mtDNA haplogroup N in five populations of Charsada and Mardan. Areas of circles are proportional to the haplotype frequencies.

On the other hand Mohmand (MM) from Charsada clustered with Turkish population and this homogeneity depicts their admixed maternal lineage from Turkey and reflects the influence of Turkish invasion of this area with the conquest of Mahmud of Ghazni on this tribe. Tamil from South India and Kalash from Pakistan are clustered separately (Fig.4). Kalash have retained its outlying position (Qamar *et al.*, 2002; Rosenberg *et al.*, 2006) even with the addition of some more comparative population datasets.

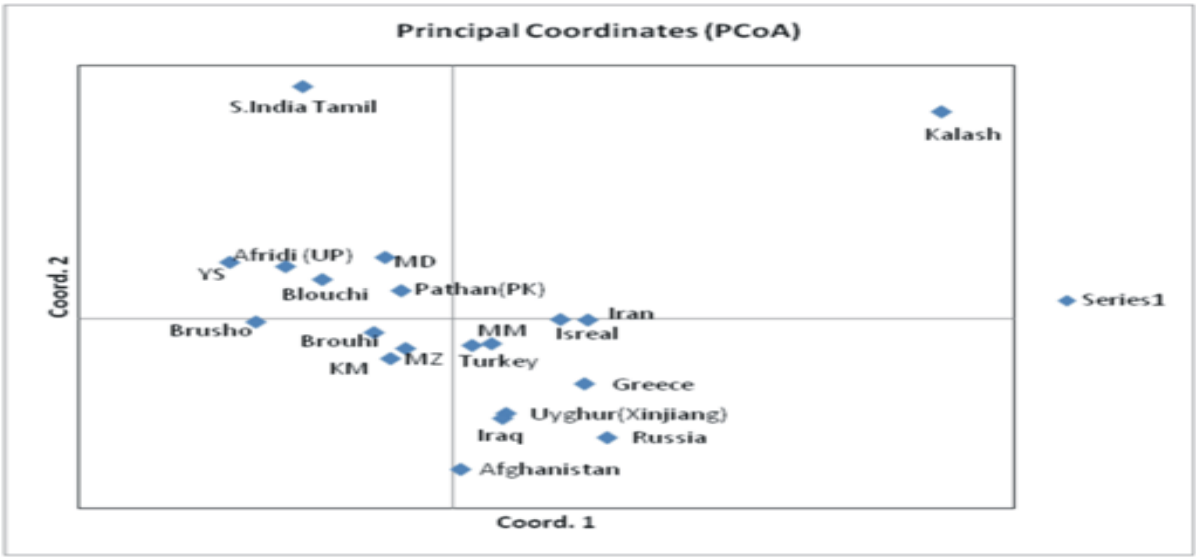


**Fig. 4.** mtDNA-$F_{ST}$ based PCoA plot of five populations of Charsada and Mardan and comparative indigenous Pakistani and global populations.

*Spatial correlation of data*

The results reveal that there is no significant correlation between genetic distance based on mtDNA-$F_{ST}$ and geographic distance (Table: 5).

## Conclusion

On the basis of these investigations and findings it can be concluded that maternal gene pool of Pathan population is heterogeneous comprising haplogroups of South Asian, Southwest Asian and central Asian origin with little or limited contribution from East Asia. Pathans sub-tribes should not be pooled as a single population. Different sub-tribes exhibit varying genetic diversities and some sub-tribes like Kakakhel Mian (KM) have retained their genetic differentiation and this status accounts for being strictly endogamous while Mohmand from Charsada (MM) and Mohmand from Mardan (MD) being same tribe exhibit different genetic structure and it may be contributed by geographic isolation of the MM and encompassing significant Turkish genetic influence. It is therefore recommended to investigate Mohmand population extensively by sequencing whole mitochondrial genome as well as Y_STR amplification.

## Acknowledgements

We acknowledge the contribution of all sample donors and district administrative staff and school heads in sample collection. Further experimentation was funded by Higher Education Commission of Pakistan under International Research and Support Initiative Program (IRSIP). Authors are also grateful to the members of Malhi's Molecular Anthropology Lab for their cooperation and moral and logistic support.

## References

**Abu-Amero KK, José ML, Vicente MC, Ana MG.** 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evolutionary Biology **8,** 45.
http://dx.doi.org/10.1186/1471-2148-8-45

**Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N.** 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics **23(2),** 147.
http://dx.doi.org/10.1038/13779

**Caroe O.** 1984. The Pathans: 500 B.C.-A.D. 1957. (Oxford in Asia Historical Reprints). Oxford University Press.1976.

**Caroe O.** 1984. The Pathans: 500 B.C.-A.D. 1957 (Oxford in Asia Historical Reprints). Oxford University Press.1976.

**Comas D, Calafell F, Mateu E, Pe'rez-Lezaun A, Bertranpetit J.** 1996. Geographic Variation in Human Mitochondrial DNA Control Region Sequence: The Population History of Turkey and its Relationship to the European Populations. Molecular Biology and Evolution **13(8),** 1067-1077.

**Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, Yepiskoposyan L.** 2013. Complete Mitochondrial DNA Diversity in Iranians. PLOS One **8(11),** e80673. http://dx.doi.org/10.1371/journal.pone.0080673

**Elahi S.** 1996. Sayyid. Encyclopedia of world cultures. Encyclopedia.com. 29 Mar. 2016 http://www.encyclopedia.com/topic/Sayyid.aspx#1

**Excoffier L, Lischer HEL.** 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources **10,** 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x

**Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT Nucleic Acids Symposium Series **41,** 95-98.

**Hayat S, Akhter T, Siddiqi MH, Rakha A, Haider N, Tayyab M, Abbas G, Ali A, Bokhari SYA, Tariq MA, Khan FM.** 2015. Mitochondrial DNA Control region sequences study in Saraiki population from Pakistan. Legal Medicine **17,** 140-144.

**Hunley KL, Healy ME, Long JC.** 2009. The Global Pattern of Gene Identity Variation Reveals a History of Long-Range Migrations, Bottlenecks, and Local Mate Exchange: Implications for Biological Race. American Journal of physical Anthropology **139,** 35– 46. http://dx.doi.org/10.1002/ajpa.20932

**Irwin J, Saunier J, StroussK, Paintner C, Diegoli T, Sturk K, Kovatsi L, Brandstatter A, Cariolou MA, Parson W, Parsons TJ.** 2008. Mitochondrial control region sequences from northern Greece and Greek Cypriots. International Journal of Legal Medicine **122(1),** 87-8. http://dx.doi.org/10.1016/j.fsigen.2014.06.013

**Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M.** 2003. Reduced Y-chromosome, but not mtDNA, diversity in human populations from West New Guinea. American Journal of Human Genetics **72,** 281–302. http://dx.doi.org/10.1086/346065

**Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA, Villems R.** 2003. The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations. American Journal of Human Genetics **72(2),** 313–332. http://dx.doi.org/10.1086/346068

**Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC.** 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER. Current Protocols in Bioinformatics **1(123),** 1-26. http://dx.doi.org/10.1002/0471250953

**Malhi RS, Cybulski JS, Tito RY, Johnson J, Harry H, Dan C.** 2010. Brief communication: mitochondrial haplotype C4c confirmed as a founding genome in the Americas. American Journal of Physical Anthropology **141(3),** 494-7. http://dx.doi.org/10.1002/ajpa.21238

**Marisi AIDAR, Sergio Roberto Peres LINE.** 2007. A simple and cost–effective protocol for DNA isolation from buccal epithelial cells. Brazilian Dental Journal **18(2),** 148-152. http://dx.doi.org/10.1590/S010364402007000200012

**McElreavey K, Quintana-Murci L.** 2005. A

population genetics perspective of the Indus Valley through uniparentally-inherited markers. Annals of Human Biology **32,** 154–162.

**Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M,Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R.** 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 Genetics **5(1),** 26. http://dx.doi.org/10.1186/1471-2156-5-26

**Nei M.** 1987. Molecular Evolutionary Genetics, Columbia University Press, New York.

**Peakall R, Smouse PE.** 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6,** 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

**Qamar R, Ayub Q, Khaliq S, Mansoor A, Karafet T, Mehdi SQ, Hammer MF.** 1999. African and Levantine origins of Pakistani YAP+ Y chromosomes. Human Biology **71,** 745–755.

**Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ.** 2002. Y-Chromosomal DNA Variation in Pakistan. American Journal of Human Genetics **70(5),** 1107–24. http://dx.doi.org/10.1086/339929

**Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al Zahery N, Semino O, Santachiara Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K.** 2004. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. American Journal of Human Genetics **74,** 827–845.

**Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, Kashyap VK.** 2005. Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages, BMC Evolutionary Biology **5,** 26. http://dx.doi.org/10.1186/1471-2148-5-26

**Rakha A, Shin KJ, Yoon JA, Kim NY, Siddique MH, Yang IS, Yang WI, Lee HY.** 2011. Forensic and genetic characterization of mtDNA from Pathans of Pakistan. International Journal of Legal Medicine **125(6),** 841–8. http://dx.doi.org/10.1007/s00414-010-0540-7

**Rosenberg NA, Mahajan S, Gonzalez-Quevedo C, Blum MG, Nino-Rosales L, Ninis V, Das P, Hegde M, Molinari L, Zapata G, Weber JL, Belmont JW, Patel PI.** 2006. Low levels of genetic divergence across geographically and linguistically diverse populations from India. PLOS Genet **2,** e215 http://dx.doi.org/ 10.1371/journal.pgen.0020215

**Siddiqi HS, Akhter T, Rakha A, Abbas G, Ali A, Haider N, Ali A, Hayat S, Masooma S, Ahmad J, Tariq MA, Oven M, Khan FM.** 2015. Genetic Characterization of Makrani People of Pakistan from mitochondrial DNA control-region data. Legal Medicine **17,** 134-139.

**Van Oven M, Kayser M.** 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutatation **30(2),** E386-E394. http://www.phylotree.org.

**Wolpert S.** 2000. A new history of India. New York: Oxford University Press.

**Ye Y, Gao J, Fan G, Liao L, Hou Y.** 2015. Population genetics for 23 Y-STR loci in Tibetan in China and confirmation of DYS448 null allele. Forensic Science International Genetics **16,** e7–e10. http://dx.doi.org/10.1016/j.fsigen.2014.11.018