Pakistan J. Zool., vol. 53(1), pp 133-141, 2021 DOI: https://dx.doi.org/10.17582/journal.pjz/20191024091047

Forensic and Genetic Characterization of mtDNA Lineages of Shin, a Unique Ethnic Group in Pakistan

Muhammad Umer Khan[1,2]*, Muhammad Farooq Sabar[1], Atif Amin Baig[3], Arif-un-Nisa Naqvi[4] and Muhammad Usman Ghani[1]

[1]*Centre for Applied Molecular Biology, University of the Punjab, Lahore, Pakistan*
[2]*Faculty of Allied Health Sciences, The University of Lahore, Lahore, Pakistan*
[3]*Faculty of Medicine, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia*
[4]*Karakoram International University, Gilgit-Baltistan, Pakistan*

ABSTRACT

Mitochondrial DNA has been adopted as a versatile genetic marker all over the globe and provides a unique maternal ancestry portrait of a person's genetic pin code. Overall knowledge of mtDNA profiles of worldwide populations can benefit population genetics and forensic sciences. Consequently, this study was designed to establish the mtDNA profiles of the Shin ethnic group in Gilgit-Baltistan, the northern most territory of Pakistan. Phlebotomy was performed for a total of 79 maternally unrelated Shin volunteers. Genomic DNA was extracted from whole blood samples and subjected to PCR amplification using specific primers for the control region of mtDNA (covering positions 16024–16569 and 1–576), including the three hypervariable segments (HVS1, HVS2, HVS3). The PCR products were subjected to cycle sequencing and further evaluated through computational analysis. A total of 75 different haplotypes were identified in Shin people; among them, 72 were unique and 3 were shared by more than one individual. This study revealed the predominance of West Eurasian lineages in the Shin population (59.49%), followed by South Asian lineages (25.32%) and then East and Southeast Asian lineages (15.19%). Shin population presented a high genetic diversity of 0.9996 and a low random match probability of 0.0129. To the best of our knowledge, this is the first report of mtDNA profiles of the Shin population, providing a complementing dataset for curative generation of future mtDNA databases in Pakistan.

Article Information
Received 24 October 2019
Revised 11 February 2020
Accepted 09 March 2020
Available online 11 December 2020

Authors' Contribution
MUK and MUG wrote the manuscript. MFS supervised the whole project and also helped in sequencing analysis. AAB provided the technical assistance regarding genetic and statistical analysis. ANN provided the facility for sampling.

Key words
Mitochondrial DNA, Shin population, Haplotypes, Haplogroups, Control region

INTRODUCTION

Mitochondrial DNA has emerged as one of the most popular genetic markers to investigate the genetic diversity of human populations (Pugach and Stoneking, 2015). In forensic practice, mtDNA analysis functions as a pivotal tool for human identity testing, population genetics, phylogenetics, anthropology, archaeology, human evolution and migration studies (Gupta *et al*., 2015).

MtDNA typing provides a unique maternal genealogical portrait of a person's genetic code. Its remarkable characteristics, which include a high copy number within the cells, an exclusive maternal inheritance, a high level of variation in its control regions, its size, and a neutral mode of evolution, make it a marker of choice in those circumstantial forensic caseworks where routine nuclear markers are not applicable (Conrad *et al*., 1968; Legros *et al*., 2004; Nilsson *et al*., 2008; Khan, 2013; O'Neill, 2013). Specifically, due to its significantly high copy number per cell, it has an advantage and provides valuable data in the legal scenarios where only degraded DNA is available (Gupta *et al*., 2015). Moreover, its absolute maternal inheritance pattern and absenteeism of recombination events allow specific mtDNA sequences to be well reserved in all maternally-related family members of a family (Conrad *et al*., 1968). This has led to an extraordinary evolutionary consistency of genetic factors across multiple generations through the entire four billion spectrum of years since the birth of Adam and Eve (Zenil, 2017). Consequently, forensic comparisons can be made using a reference sample from multiple generations (Conrad *et al*., 1968).

All these benefits of mtDNA analysis are employed by forensic scientists for multiple purposes such as recognition of the relics of missing persons in disasters or matching evidential DNA recovered from a crime scene to those available in a database (of, e.g., convicted criminal profiles or the database for probable relatives) (Ziętkiewicz *et al*., 2012). Hence an overall knowledge of mtDNA profiles of worldwide populations is imperative to take advantage of mtDNA in a plethora of applications

* Corresponding author: umer.khan685@gmail.com
0030-9923/2021/0001-0133 $ 9.00/0

including forensic genetics and phylogenetic studies (Butler, 2009). Population specificity of mitochondrial genome (mtgenome) is widely reported in literature. MtDNA has been found to be very informative for inference of ethnicity (Prieto *et al*., 2011, Ladoukakis and Zouros, 2017).

Through historical perspective of human population migrations, South Asia comes next to Africa in holding heterogeneity and genetic diversity of populaces. Pakistan is situated in the core region of South Asia and probably was inhabited during primitive human movements (Shi *et al*., 2008). This zone is thus considered as the cradle of multiple civilizations. Currently a number of racial groups and minority units reside in Pakistan (Ayub and Tyler-Smith, 2009).

Gilgit-Baltistan is an important independent territory of Pakistan (Afzal, 2017). It is situated in the northern zone of Pakistan and consists of sight worthy valleys disjointed by some of the globe's highest mountain ranges including Hindu Kush, the Himalayas, Karakoram and the Pamir Mountains. Thus, it embraces a mixture of dynamic cultures and civilizations. Hence it is well-intentioned to study the ethnicity of people residing there (Khan, 2013).

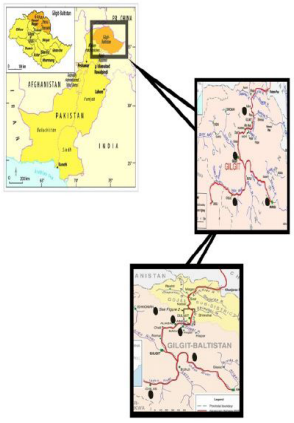
One of the dominant populations of Gilgit-Baltistan is Shin, a Dardic tribe, whose mother tongue is Shina (Radloff, 1992; O'Neill, 2013). Unfortunately, this smaller but significant ethnic group of Pakistan has remained neglected and understudied. Understanding the genetic structure of this population is important, not only from a historic standpoint, but also for effective implementation and interpretation of forensic genetics.

In this regard, the current study was aimed to establish the mitochondrial DNA profiles of the Shin population, residing in Gilgit-Baltistan. The entire mtDNA control region of Shin individuals was sequenced and analyzed (as per recommendations) (Parson *et al*., 2014). This is the first study to report the mtDNA profiles of the Shin population. The main target of this work was to establish the predominant mtDNA lineages of this population to infer their ethnicity and history of their settlements in Pakistan and to compare them with other relevant races. The outcomes of this study will be useful for generating a genetic database of these areas which may be utilized for multipurpose future forensic implications.

MATERIALS AND METHODS

Samples

Blood samples were collected from 79 maternally unrelated Shin individuals, both males and females, living in different regions of Gilgit-Baltistan, Pakistan (Fig. 1). Only individuals who confirmed their Shin origin of at least last three generations on the maternal side were included in the study. Scripted informed consent was taken from all the volunteer participants according to the declarations of Helsinki. Sample collection was performed in different towns and cities of Gilgit-Baltistan to achieve a reliable and complete representation of Shin population. A bioethical clearance certificate was obtained by the Bioethics committee of University of the Punjab. All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.



Sampling Areas	No. of Samples
Aliabad	1
Astor	9
Baseen	2
Bonji	1
Chillas	1
Danyor	9
Diameer	2
Ganj,Skrdu	1
Gilgit	22
Ghizzer	12
Gujal	1
Gulmit	1
Gunji	1
Haramosh	4
Hunza	1
Jager Baseen	1
Juglot	3
Mominabad	1
Naltar	2
Nigar	1
Sultanabad	1
Yalabad	1
Yasin	1

Fig. 1. Geographic location of Gilgit-Baltistan in map of Pakistan, information on right side of the figure depicting the sampling region for Shin population. Samples were collected from these 23 regions of Gilgit-Baltistan.

DNA extraction, amplification and sequencing

Whole blood samples collected in EDTA vials were subjected to DNA extraction via QIAamp DNA Mini Kit as per manufacturer instructions (Qiagen, Hilden, Germany. Cat No./ID: 51304). The quality and purity of extracted DNA samples were determined and adjusted (Nano Drop TM 1000 Spectrophotometer). The amplification of the desired sequences was done by Polymerase Chain Reaction (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA, USA), using specific primers as mentioned in Table I. PCR cyclic reactions were performed in 50µl of reaction mixture containing a total of 25µl 2x PCR hotstart master mix (abm, Canada, Cat. No. G906), a total of 21µl nuclease free H 2 O (Ambion, ThermoFisher Scientific, USA) a total of 2µl forward (10µM) and reverse primers (10µM) and a 2µl of DNA template. The PCR (30 cycles) was engineered to be; initial denaturation at 94°C for 10 min; second denaturation at 94°C for 30 seconds, annealing

Table I. Primers used in this study to amplify the control region of mitochondrial DNA in Shin ethinic group in Pakistan.

Sr. No.	Primer name (Control region)	Primer sequences (5′→3′)	PCR	Sequencing
1	F15975	CTC CAC CAT TAG CAC CCA AA	Yes	Yes
2	F16327	CCG TAC ATA GCA CAT TAC AGT C	No	Yes
3	F155	TAT TTA TCG CAC CTA CGT TC	No	Yes
4	R16419	GAG GAT GGT GGT CAA GGG A	No	Yes
5	R042	AGA GCT CCC GTG AGT GGT TA	No	Yes
6	R635	GAT GTG AGC CCG TCT AAA CA	Yes	Yes
7	F403	CCG CTT CTG GCC ACA GCA CT	No	Yes
8	R389	CTG GTT AGG CTG GTG TTA GG	No	Yes
9	F16524	AAG CCT AAA TAG CCC ACA CG	No	Yes

at 60 C for 30 seconds and extension at 72 °C for 1.5 min. followed by a final extension 72 °C for 5 min.. The amplified PCR products were analyzed for their quality and purity (Nano Drop TM 1000 Spectrophotometer, USA) to be subjected further to cycle sequencing (BigDye Terminator v3.1 Cycle Sequencing Kit, Thermo Fisher Scientific, USA) using manufacturer instructions, followed by DNA sequencing readout through Applied Biosystems 3730xl Genetic Analyzer (Thermo Fisher Scientific, USA).

Data analysis

The mtDNA control region forward and reverse sequences were aligned through the sequence analysis tool Geneious (Version 7.0.3, Biomatters Ltd, New Zealand) (Drummond, 2009) and were then compared to the revised Cambridge Reference Sequence (Andrews *et al*., 1999) using mtDNA profiler (Yang *et al*., 2013). To ensure high quality, two independence evaluations of the raw data were performed as per recommendations (Parson and Bandelt, 2007). The haplogroup assignments were carried out using previously published data (Metspalu *et al*., 2004; Behar *et al*., 2008; van Oven *et al*., 2011; Elmadawy *et al*., 2013), and by using Mitotool (www.mitotool.org) (Fan and Yao, 2011) and Haplogrep (www.haplogrep.uibk.ac.at) online tools (Kloss-Brandstätter *et al*., 2011), based on PhyloTree Build 17 (http://www.phylotree.org) (Van Oven and Kayser, 2009) as classification tree.

Indices of forensics and population genetics; e.g., genetic diversity, random match probability and power of discrimination were analyzed as explained previously (Tajima, 1989; Prieto *et al*., 2011). The current study strictly adhered to the guidelines and recommendations from the International Society for Forensic Genetics (ISFG) (Parson and Bandelt, 2007).

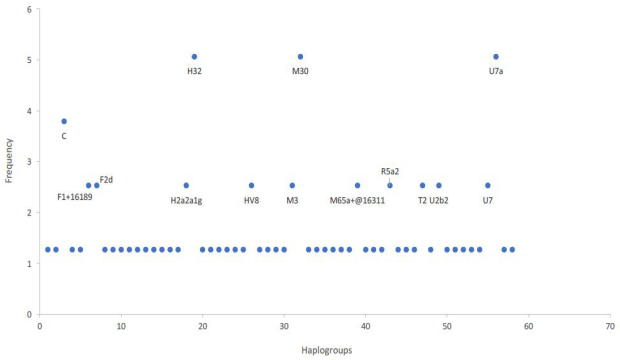


Fig. 2. Frequency of different haplogroups in Shin population; following are the haplogroups which showed the frequency of 1.27%. A+152+16362, A2v, D4g2a, F, G2c, H1+152, H106, H14a, H15a1b, H1e1a1, H29, H2a+152 16311, H2a2a, H3ak, H3b6, H3s, H6, HV14, HV2a, J1, J1b1a1, J2b1a, K1b2, M30+16234, M30c1, M35b+16304, M35b1, M3d, M65, M9, N5, P2, R8, T1a+152, T1a1'3, T2b5a1, U3a2a, U4b2, U5a1f1, U5a2a, U5b2a1a2, W+194, Z+152.

RESULTS AND DISCUSSION

The present study generated population data for the complete mtDNA control region (16,024–576) of 79 subjects from the Shin ethnic group. A total of 75 haplotypes were observed including 72 unique and 3 shared haplotypes. The most frequent haplotype (16129A 16223T 16298C 16327T 16519C 73G 249d 263G 315.1C 315.2C 489C) was found in 3.79 % of the sampled population (Table II). In 1122 positions analyzed, 174 variable sites were found in the mtDNA control region of the Shin population.

MtDNA analysis of the subjects revealed that the Shin population exhibited mtDNA genetic diversity of 0.9996, random match probability of 0.0129 and

Table II. Polymorphism in control region of mitochondrial DNA in Shin ethnicity in Pakistan.

Sr. no.	Sample ID	Sampling area	Haplotypes	Haplogroups
1	SHN-009	Ghizer	16114A 16192T 16256T 16270T 16294T 16526A 73G 263G 309.1C 361d	U5a2a
2	SHN-010	Ghizer	16129A 16242T 16356C 73G 200G 263G 309.1C 315.1C550.1C	H3b6
3	SHN-011	Astore	16166d 16309G 16318T 16519C 73G 151T 152C 263G 315.1C 523d 524d 550.1C 573.1C	U7a
4	SHN-012	Astore	16126C 16223T 16519C 73G 263G 315.1C 482C 489C 523d 524d 550.1C 573.1C	M3
5	SHN-013	Sultan Abad	16224C 16311C 16320T 16519C 73G 89C 146C 195C 263G 309.1C 315.1C 524.1A 524.2C 524.3A 5 24.4C 573.1C	K1b2
6	SHN-014	Haramosh	16150T 16166d 16223T 16519C 73G 146C 152C 195A 263G 315.1C 489C 523d 524d 549T 573.1C	M30c1
7	SHN-015	Astore	16223T 16304C 16519C 73G 151T 199C 204C 263G 309.1C 315.1C 489C 550.1C 573.1C	M35b1
8	SHN-016	Astore	16223T 16519C 73G 152C 195A 263G 272G 309.1C 315.1C489C 523d 524d 573.1C	M30
9	SHN-017	Astore	152C 200G 235G 263G 315.1C 523d 524d 573.1C	H1+152
10	SHN-018	Astore	73G 152C 263G 315.1C 523d 524d 573.1C	H32
11	SHN-019	Danyor	16126C 16163G 16186T 16189C 16294T 16519C 73G 152C 263G 315.1C 573.1C 573.2C 573.3C573 .4C	T1a+152
12	SHN-020	Gilgit City	16223T 16234T 16274A 16519C 73G 195A 263G 309.1C 309.2C 315.1C 489C 523d 524d 573.1C	M30+16234
13	SHN-021	Gunji	263G 315.1C 573.1C	H29
14	SHN-022	Danyor	16129A 16223T 16298C 16327T 16519C 73G 249d263G 315.1C 315.2C 489C	C
15	SHN-023	Gultari	16344T 16519C 73G 263G 315.1C 482C 489C 523d 524d549.1C 550.1C 573.1C	M3d
16	SHN-024	Gilgit City	16126C 16294T 16296T 16519C 73G 263G 309.1C315.1C 550.1C 573.1C	T2b5a1
17	SHN-025	Gilgit City	16049.1G 16182C 16183C 16189C 16194C 16195A 16196.1G 16197G 205.1G 206G 220A 221T 230G 237T 240T 250A 253T 256G 257C 260A	H2a2a1g
18	SHN-026	Jalalabad	16145A 16192T 16256T 16270T 16304C 16311C 16399G 73G 263G	U5a1f1
19	SHN-027	Gilgit City	16111T 16223T 16391A 16519C 73G 195A 263G	N5
20	SHN-028	Gilgit City	16223T 16519C 73G 152C 195A 263G 272G 309.1C315.1C 489C 523d 524d 573.1C	M30
21	SHN-029	Gilgit City	73G 152C 263G 315.1C 573.1C	H32
22	SHN-030	Jager Basen	16086C 16185T 16223T 16260T 16519C 73G 152C 249d 263G 309.1C 315.1C 489C 573.1C	Z+152
23	SHN-031	Astore	16166G 16256T 16352C 16519C 200G 263G 309.1C309.2C 315.1C 573.1C	H14a
24	SHN-032	Haramosh	16126C 16294T 16296T 16325C 16519C 16527T 73G 263G 315.1C 523d 524d 573.1C	T2
25	SHN-033	Chilas	63C 64T 73G 105G 114T 141T 146C 151T 167T 186G 194T 242T 253T 263G 295T 309.1C 315.1C 462T 489C 552.1T 573.1C	J1
26	SHN-034	Gilgit City	16309G 16318T 16519C 73G 152C 263G 309.1C315.1C 523d 524d 573.1C	U7
27	SHN-035	Diamir	16223T 16304C 16519C 73G 146T 199C 263G 309.1C 309.2C 315.1C 489C 573.1C	M35b+16304
28	SHN-036	Jaglot	16309G 16318T 16343G 16519C 73G 151T 152C 185A 263G 315.1C 368.1A 368.2G 368.3A 368.4A 573.1C	U7a
29	SHN-037	Ganj Skrdu	16092C 16223T 16290T 16319A 16362C 73G 152C 235G 263G 309.1C 315.1C 573.1C	A+152+16362
30	SHN-038	Naltar	16126C 16294T 16296T 16325C 16519C 16527T 73G 263G 315.1C 523d 524d 550.1C 573.1C	T2
31	SHN-039	Gilgit City	16309G 16318T 16519C 73G 152C 263G 309.1C315.1C 523d 524d 573.1C	U7
32	SHN-040	Haramosh	73G 194T 263G 315.1C	P2
33	SHN-041	Gulmit	16223T 16289G 16290T 16360T 16519C 73G 198T 263G 315.1C 489C 511T 524.1A 524.2C 573.1C	M65a+@16311
34	SHN-042	Nagar	16309G 16318T 16519C 64T 73G 151T 152C 263G 309.1C 315.1C 523d 524d 573.1C	U7a
35	SHN-043	Jaglot	16266T 16304C 16311C 16356C 16524G 73G 146C 152C 263G 315.1C 523d 524d 549.1C 550.1C	R5a2
36	SHN-044	Baseen	16309G 16318C 16368C 16519C 73G 152C 200G 263G 315.1C 523d 524d 573.1C	H32
37	SHN-045	Bunji	16086C 16311C 263G 309.1C 315.1C 480C 573.1C	HV14

Continued on next page.....

Sr. no.	Sample ID	Sampling area	Haplotypes	Haplogroups
38	SHN-046	Jalal abad	73G 152C 235G 263G 309.1C 315.1C 573.1C	F2d
39	SHN-047	Danyor	93G 152C 263G 309.1C 309.2C 315.1C 524.1A 524.2C	H1e1a1
40	SHN-048	Ghizer	16189C 150T 263G 315.1C 550.1C 573.1C	H1e1a6
41	SHN-049	Gilgit City	73G 152C 235G 263G 309.1C 315.1C 573.1C	F2d
42	SHN-050	Baseen	16037G 16039.1G 16266T 16304C 16311C 16356C 16524G 16526A 73G 150T 152C 263G 309.1C 315.1C 523d 524d 573.1C	R5a2
43	SHN-052	Ghizer	16126C 16163G 16186T 16189C 16294T 16519C 73G 152C 195C 263G 309.1C 309.2C 315.1C 524.1A 524.2C 573.1C	T1a1'3
44	SHN-053	Jalal abad	16136C 16356C 73G 195C 263G 309.1C 315.1C 499A 524.1A 524.2C 524.3A 524.4C 573.1C	U4b2
45	SHN-054	Ghizer	16069T 16126C 16145A 16172C 16222T 16261T 16519C 73G 242T 263G 295T 315.1C 462T 489C 550.1C 573.1C	J1b1a1
46	SHN-055	Ghizer	16519C 143A 263G 309.1C 309.2C 315.1C 572.1G	H3ak
47	SHN-056	Gilgit City	73G 150T 200G 263G 309.1C 315.1C 550.1C	U5b2a1a2
48	SHN-057	Danyor	16209C 16239T 16311C 16352C 16353T 73G 146C 152C 153G 234G 263G 309.1C 315.1C 573.1C	U2b2
49	SHN-058	Danyor	16126C 16147T 16223T 16519C 73G 195C 263G 309.1C 315.1C482C 489C	M3
50	SHN-059	Astore	16223T 16274A 16362C 16519C 73G 263G 298T 309.1C 309.2C 315.1C 489C 550.1C	D4g2a
51	SHN-060	Ghizer	16069T 16126C 16193T 16278T 16519C 73G 150T 152C 235G 263G 295T 315.1C 489C 550.1C	J2b1a
52	SHN-061	Danyor	16111T 16239T 16362C 16482G 239C 263G 309.1C 309.2C315.1C 549.1C	H6
53	SHN-062	Danyor	16183C 16189C 16304C 16519C 73G 249d 263G 309.1C 315.1C 523d 524d	F1+16189
54	SHN-063	Momin abad	16129A 16223T 16298C 16327T 16519C 73G 249d263G 315.1C 315.2C 489C	C
55	SHN-064	Gilgit City	16360T 16519C 73G 198T 263G 315.1C 489C 511T 524.1A524.2C	M65
56	SHN-065	Gilgit City	16093C 16223T 16519C 73G 195A 263G 315.1C 352C 489C 523d 524d	M30
57	SHN-066	Ghizer	16311C 16354T 263G 315.1C 550.1C	HV8
58	SHN-067	Gilgit City	16124C 16184T 16311C 44.1C 55C 57C 146C 263G 309.1C 315.1C 550.1C	H15a1b
59	SHN-068	Diamer	16311C 16354T 263G 315.1C	HV8
60	SHN-069	Yasin	16209C 16239T 16311C 16352C 16353T 73G 146C 152C 153G 234G 263G 309.1C 315.1C	U2b2
61	SHN-070	Hunza	263G 309.1C 315.1C	H2a2a
62	SHN-071	Gujal	16049.1G 16183d 16189d 16194d 16195C 16196C 16519C 73G 153G 195C 263G 309.1C 315.1C 550.1C	R8
63	SHN-072	Juglot	16049.1G 16183C 16189C 16277C 16304C 16519C 73G 249d 263G 309.1C 315.1C 523d 524d	F1+16189
64	SHN-073	Astore	16183C 16189C	H2a2a1g
65	SHN-074	Astore	16179T 16223T 16292T 16295T 16519C 73G 146C 189G 194T 195C 204C 207A 263G 309.1C 315.1C 550.1C	W+194
66	SHN-075	Gilgit City	16519C 72C 73G 146C 152C 195C 263G 315.1C 550.1C	HV2a
67	SHN-076	Danyor	16129A 16223T 16298C 16327T 16519C 73G 249d263G 315.1C 315.2C 489C	C
68	SHN-077	Astore	16223T 16302G 16519C 73G 143A 195A 263G 315.1C 489C 523d 524d 550.1C	M30
69	SHN-078	Jalal abad	16311C 152C 263G 309.1C 315.1C 548.1A 550.1C	H2a+152 16311
70	SHN-079	Gilgit	16223T 16289G 16519C 73G 263G 315.1C 489C 511T	M65a+@16311
71	SHN-080	Juglot	16245T 16309G 16318T 16356C 16519C 73G 151T 152C 263G 315.1C 523d 524d	U7a
72	SHN-081	Diamer	16223T 16245T 103A 235G 263G 309.1C 315.1C	H106
73	SHN-082	Yasin	16270T 16343G 73G 150T 200G 263G 309.1C 315.1C550.1C	U3a2a
74	SHN-083	Hunza	16093C 16145A 16223T 16362C 73G 152C 195C 263G 309.1C 315.1C 489C	G2c
75	SHN-084	Gujal	194T 200G 263G 309.1C 309.2C 315.1C 550.1C	H3s
76	SHN-085	Duglot	73G 152C 263G 309.1C 315.1C 524.1A 524.2C	H32
77	SHN-086	Astore	16223T 16362C 16519C 73G 263G 309.1C 315.1C489C	M9
78	SHN-087	Astore	16183d 16189d 16194d 16195C 16196C 16519C 73G 249d 263G 309.1C 315.1C 523d 524d	F
79	SHN-088	Gilgit City	16111T 16239T 16362C 239T263G 309.1C 315.1C 523d 524d	A2v

Table III. Forensic and genetics parameter indices in Shin population.

Total number of samples	79
No. of haplotypes	75
Unique haplotypes	72
Polymorphic positions	178
Random match probability	0.0129
Power of discrimination	0.9871
Genetic diversity	0.9996

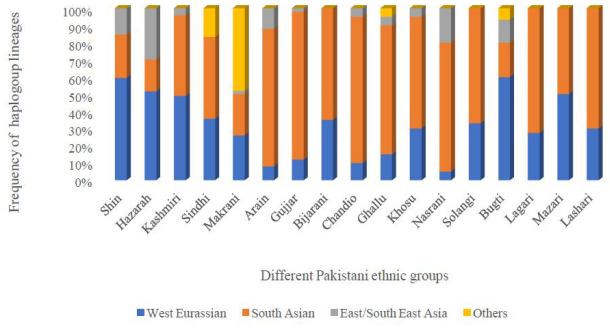


Fig. 3. Percentage association of phylogeographical roots in different Pakistani populations.

power of discrimination of 0.9871, as presented in Table III. We compared the forensic and population genetics parameters including no. of haplotypes, unique haplotypes, genetic diversity, random match probability and power of discrimination of the Shin population with the other reported indigenous populations of Pakistan such as Saraiki, Sindhi, Makrani, Pathan, Kashmiri and Hazara, and found that the Shin population had the highest proportion of unique haplotypes reflecting high population heterogeneity in Shins. The large proportion of unique haplotypes in Shin population also corresponded well with their greatest genetic diversity (0.9996) when compared to other ethnic groups of Pakistan, i.e. Pathan (0.9978), Kashmiri (0.9977) Hazara (0.9945), Sindhi (0.9924), Makrani (0.9905) and Saraiki (0.9570) (Rakha *et al*., 2011, 2016, 2017; Hayat *et al*., 2015; Siddiqi *et al*., 2015; Yasmin *et al*., 2017) (Table IV).

Haplogroup affiliations

The haplogroups observed in the Shin population showed affiliations with different phylogenetic lineages. They were mainly assigned into three continental groups, namely the West Eurasian (59.41%), South Asian (25.32%) and East and Southeast Asian (15.19%) groups. Thus, a high degree of genetic association with West Eurasian lineage was observed as compared to South Asian and South East Asian lineages. The most frequent haplogroups identified in the Shin population were U7a (5.06%), M30 (5.06%) and H32 (5.06%), carried by 15.19% of the population. The rest of the haplogroups observed in the Shin population were C (3.79%), U5a2a (1.27%), H3b6 (1.27%), M3 (2.53%), K1b2 (1.27%), M30c1 (1.27%), M35b1 (1.27%), H1+152(1.27%), T1a+152 (1.27%), M30+16234 (1.27%), H29 (1.27%), M3d (1.27%), T2b5a1 (1.27%), U5a1f1 (1.27%), N5 (1.27%), Z+152 (1.27%), H14a (1.27%), T2 (1.27%), U7 (2.53%), M35b+16304 (1.27%), A+152+16362 (1.27%), P2 (1.27%), M65a+@16311 (2.53%), R5a2 (2.53%), HV14 (1.27%), F2d (2.53%), H1e1a1 (1.27%), H1e1a6 (1.27%), T1a1'3 (1.27%), U4b2 (1.27%), J1b1a1 (1.27%), H3ak (1.27%), U5b2a1a2 (1.27%), U2b2 (2.53%), D4g2a (1.27%), J2b1a (1.27%), H6 (1.27%), F1+16189 (2.53%), M65 (1.27%), HV8 (2.53%), H15a1b (1.27%), H2a2a (1.27%), R8 (1.27%), W+194 (1.27%), HV2a (1.27%), H2a+152 16311 (1.27%), H106 (1.27%), U3a2a (1.27%), G2c (1.27%), H3s (1.27%), M9 (1.27%), F (1.27%) and A2v (1.27%) (Fig. 2, Table V).

The current study revealed that the majority of the haplogroups of the Shin population indicated affiliation with West Eurasian lineage. A similar pattern was observed in other studies conducted on other Pakistani ethnic groups such as the Pathan, Hazara and Kashmiri, Bugti and Laghari, where maximum frequencies of West Eurassian haplogroups were reported. However, the rest of the Pakistani ethnic groups, such as the Gujjar, Araiyn, Bijrani, Chandio, Ghallu, Khosu, Nasrani, Solangi, Laghari, Lashari, Makrani, Saraiki and Sindhi, represented quite contrasting genetic structure and affiliations (Rakha *et al*., 2011, 2016, 2017; Hayat *et al*., 2015; Siddiqi *et al*., 2015; Yasmin *et al*., 2017; Bhatti *et al*., 2017, 2018a, b) (Fig. 3). The pronounced prevalence of West Eurasian matrilineal lineages may root back to great historical movements from Europe and Central Asia such as the invasion by the soldiers of Alexander the Great, the Arab and Muslim takeovers, and the era of the British Indian Empire (McElreavey and Quintana-Murci, 2005).

CONCLUSION

To the best of our knowledge, this is the first report regarding a forensic dataset of the Shin population including entire mtDNA control region sequences. The results reveal high genetic diversity and low random match probability, predicting the worth of mtDNA profiles of the Shin population for exploring maternal genetic lineages and routine forensic investigations in Pakistan. The outcomes of this study show the West Eurasian

Table IV. Comparison of forensic and genetic diversity indices of mtDNA control region of main ethnic groups of Pakistan.

Parameters	Shin	Saraiki	Sindhi	Makrani	Pathan	Kashmiri	Hazara
No. of samples	79	85	88	99	230	317	319
No. of haplotypes	75	63	66	71	192	251	189
No. of unique haplotypes	72	58	50	54	128	201	124
Genetic diversity	0.9996	0.957	0.9924	0.9905	0.9978	0.9977	0.9945
Power of discrimination	0.9871	0.9458	0.9811	0.7172	0.8348	0.7918	0.5925
Random match probability	0.0129	0.0542	0.0188	0.0195	0.0066	0.0054	0.0085

haplogroups to be predominant in the Shin population. The data reported in this study will contribute in generation of mtDNA databases in Pakistan and will be beneficial for multipurpose future forensic implications.

Table V. Haplogroup frequencies of 79 Shins from Gilgit Baltistan, Pakistan.

Broad haplogroup	Number	Proportion %	Haplogroup	Number	Proportion %	Possible origin
A	2	3.4	A+152+16362	1	1.27	East Asia
			A2v	1	1.27	East Asia
C	1	1.7	C	3	3.79	East Asia
D	1	1.7	D4g2a	1	1.27	East Asia
F	3	5.2	F	1	1.27	East Asia
			F1+16189	2	2.53	East Asia
			F2d	2	2.53	East Asia
G	1	1.7	G2c	1	1.27	East Asia
H	15	25.7	H1+152	1	1.27	West Eurasian
			H106	1	1.27	West Eurasian
			H14a	1	1.27	West Eurasian
			H15a1b	1	1.27	West Eurasian
			H1e1a1	1	1.27	West Eurasian
			H1e1a6	1	1.27	West Eurasian
			H29	1	1.27	West Eurasian
			H2a+152 16311	1	1.27	West Eurasian
			H2a2a	1	1.27	West Eurasian
			H2a2a1g	2	2.53	West Eurasian
			H32	4	5.06	West Eurasian
			H3ak	1	1.27	West Eurasian
			H3b6	1	1.27	West Eurasian
			H3s	1	1.27	West Eurasian
			H6	1	1.27	West Eurasian
HV	3	5.2	HV14	1	1.27	West Eurasian
			HV2a	1	1.27	West Eurasian
			HV8	2	2.53	West Eurasian
J	3	5.2	J1	1	1.27	West Eurasian
			J1b1a1	1	1.27	West Eurasian
			J2b1a	1	1.27	West Eurasian
K	1	1.7	K1b2	1	1.27	West Eurasian
M	10	17.2	M3	2	2.53	South Asian
			M30	4	5.06	South Asian
			M30+16234	1	1.27	South Asian
			M30c1	1	1.27	South Asian
			M35b+16304	1	1.27	South Asian
			M35b1	1	1.27	South Asian
			M3d	1	1.27	South Asian
			M65	1	1.27	South Asian
			M65a+@16311	2	2.53	South Asian
			M9	1	1.27	South Asian
N	1	1.7	N5	1	1.27	West Eurasian
P	1	1.7	P2	1	1.27	East Asian
R	2	3.4	R5a2	2	2.53	West Eurasian
			R8	1	1.27	West Eurasian
T	4	6.9	T1a+152	1	1.27	West Eurasian
			T1a1'3	1	1.27	West Eurasian
			T2	2	2.53	West Eurasian
			T2b5a1	1	1.27	West Eurasian
U	8	13.79	U2b2	2	2.53	South Asian
			U3a2a	1	1.27	West Eurasian
			U4b2	1	1.27	South Asian
			U5a1f1	1	1.27	West Eurasian
			U5a2a	1	1.27	West Eurasian
			U5b2a1a2	1	1.27	West Eurasian
			U7	2	2.53	West Eurasian
			U7a	4	5.06	West Eurasian
W	1	1.7	W+194	1	1.27	West Eurasian
Z	1	1.7	Z+152	1	1.27	East Asian

Continued on next column.....

ACKNOWLEDGEMENTS

We are grateful to Dr. Muhammad Hassan Siddiqui for his support in sample collection and Dr. Emmanuel William Smith for English language editing.

Accession number

The mtDNA control region sequences of Shin population reported in the current study have been submitted to GenBank and are available under MK032930 to MK033007 accession number.

Statement of conflict of interest

The authors have declared no conflict of interests regarding the publication of this article.

REFERENCES

Afzal, 2017. Dawn newspaper. https://www.dawn.com/news/1365854.

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

Ayub, Q. and Tyler-Smith, C., 2009. Genetic variation in South Asia: Assessing the influences of geography, language and ethnicity for understanding history and disease risk. *Brief. Funct. Genom. Proteom*., **8**: 395-404. https://doi.org/10.1093/bfgp/elp015

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S. and Comas, D., 2008. The dawn of human matrilineal diversity. *Am. J. Human Genet*., **82**: 1130-1140. https://doi.org/10.1016/j.ajhg.2008.04.002

Bhatti, S., Abbas, S., Aslamkhan, M., Attimonelli, M., Trinidad, M.S., Aydin, H.H., de Souza, E.M.S. and Gonzalez, G.R., 2018a. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochond. DNA Part A*, **29**: 714-726. https://doi.org/10.1080/24701394.2017.1350951

Bhatti, S., Khan, M.A., Abbas, S., Attimonelli, M., Gonzalez, G.R., Aydin, H.H. and de Souza, E.M.S., 2018b. Problems in Mitochondrial DNA forensics: While interpreting length heteroplasmy conundrum of various Sindhi and Baluchi ethnic groups of Pakistan. *Mitochond. DNA Part A,* **29**: 501-510. https://doi.org/10.1080/24701394.2017.1310853

Bhatti, S., Aslamkhan, M., Attimonelli, M., Abbas, S. and Aydin, H.H., 2017. Mitochondrial DNA variation in the Sindh population of Pakistan. *Aust. J. Forensic Sci*., **49**: 201-216. https://doi.org/10.1080/00450618.2016.1144788

Butler, J.M., 2009. *Fundamentals of forensic DNA typing*. Academic Press.

Radloff, C.F., 1992. The dialects of Shina. In: *Sociolinguistic Surv. Northern Pakistan* (eds. P.C. Backstromm and C.F. Radloff). International Academic Bookstore, TX 75236, USA. pp. 89-201.

Conrad, G., Mehl, E., Moussine-Pouchkine, A. and Conrad, J., 1968. *Région de la Gara Azzel Matti, Sahara central*. Paris, Éd. du Centre national de la recherche scientifique.

Drummond, A.J., Cheung, A.B.M., Heled, J., Kearse, M. and Moir, R., 2009. Geneious. v4.7. 2009 from http://www.geneious.com.

Elmadawy, M.A., Nagai, A., Gomaa, G.M., Hegazy, H.M., Shaaban, F.E., and Bunai, Y., 2013. Investigation of mtDNA control region sequences in an Egyptian population sample. *Leg. Med*., **15**: 338-341. https://doi.org/10.1016/j.legalmed.2013.06.006

Fan, L. and Yao, Y.-G., 2011. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion,* **11**: 351-356. https://doi.org/10.1016/j.mito.2010.09.013

Gupta, A., Bhardwaj, A., Sharma, P. and Pal, Y., 2015. Mitochondrial DNA-a tool for phylogenetic and biodiversity search in equines. *J. Biodiv. Endang. Species*, **2015**: 1-8. https://doi.org/10.4172/2167-1206.S1-006

Hayat, S., Akhtar, T., Siddiqi, M.H., Rakha, A., Haider, N., Tayyab, M., Abbas, G., Ali, A., Bokhari, S.Y.A. and Tariq, M.A., 2015. Mitochondrial DNA control region sequences study in Saraiki population from Pakistan. *Leg. Med*., **17**: 140-144. https://doi.org/10.1016/j.legalmed.2014.10.010

Khan, A.H., 2013. Constitutional Status of Gilgit Baltistan, Chitral and Shenaki Kohistan. *Himalayan Central Asian Stud*., **17**: 44-51.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G. and Kronenberg, F., 2011. HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat*., **32**: 25-32. https://doi.org/10.1002/humu.21382

Ladoukakis, E.D. and Zouros, E., 2017. Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *J. Biol. Res. Thessaloniki*, **24**: 2-7. https://doi.org/10.1186/s40709-017-0060-4

Legros, F., Malka, F., Frachon, P., Lombès, A. and Rojo, M., 2004. Organization and dynamics of human mitochondrial DNA. *J. Cell Sci*., **117**: 2653-2662. https://doi.org/10.1242/jcs.01134

McElreavey, K. and Quintana-Murci, L., 2005. A population genetics perspective of the Indus Valley through uniparentally-inherited markers. *Ann. Human Biol*., **32**: 154-162. https://doi.org/10.1080/03014460500076223

Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M. and Gilbert, M.T.P., 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**: 1-26. https://doi.org/10.1186/1471-2156-5-26

Nilsson, M., Andréasson-Jansson, H., Ingman, M. and Allen, M., 2008. Evaluation of mitochondrial DNA coding region assays for increased discrimination in forensic analysis. *Forensic Sci. Int. Genet*., **2**: 1-8. https://doi.org/10.1016/j.fsigen.2007.07.004

o'Neill, P.W., 2013. *Tracing shina and wakhi origins: are ethnic classifications commonly used in demographic studies biologically meaningful?* Masters thesis, California State University.

Parson, W. and Bandelt, H.-J., 2007. Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Sci. Int. Genet*., **1**: 13-19. https://doi.org/10.1016/j.fsigen.2006.11.003

Parson, W., Gusmao, L., Hares, D., Irwin, J., Mayr, W., Morling, N., Pokorak, E., Prinz, M., Salas, A. and Schneider, P., 2014. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci. Int. Genet*., **13**: 134-142. https://doi.org/10.1016/j.fsigen.2014.07.010

Prieto, L., Zimmermann, B., Goios, A., Rodriguez-Monge, A., Paneto, G., Alves, C., Alonso, A., Fridman, C., Cardoso, S. and Lima, G., 2011. The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework. *Forensic Sci. Int. Genet*., **5**: 146-151. https://doi.org/10.1016/j.fsigen.2010.10.013

Pugach, I. and Stoneking, M., 2015. Genome-wide insights into the genetic history of human populations. *Invest. Genet.,* **6**: Article number: 6. https://doi.org/10.1186/s13323-015-0024-0

Rakha, A., Peng, M.-S., Adan, A., Bi, R., Yasmin, M. and Yao, Y.-G., 2017. mtDNA sequence diversity of Hazara ethnic group from Pakistan. *Forensic Sci. Int. Genet*., **30**: e1-e5. https://doi.org/10.1016/j.fsigen.2017.07.004

Rakha, A., Peng, M.-S., Bi, R., Song, J.-J., Salahudin, Z., Adan, A., Israr, M. and Yao, Y.-G., 2016. EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu and Kashmir, Pakistan. *Forensic Sci. Int. Genet*., **25**: 125-131. https://doi.org/10.1016/j.fsigen.2016.08.009

Rakha, A., Shin, K.-J., Yoon, J.A., Kim, N.Y., Siddique, M.H., Yang, I.S., Yang, W.I. and Lee, H.Y., 2011. Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int. J. Leg. Med*., **125**: 841-848. https://doi.org/10.1007/s00414-010-0540-7

Shi, H., Zhong, H., Peng, Y., Dong, Y.-L., Qi, X.-B., Zhang, F., Liu, L.-F., Tan, S.-J., Ma, R. Z. and Xiao, C.-J., 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol*., **6**: Article number: 45. https://doi.org/10.1186/1741-7007-6-45

Siddiqi, M.H., Akhtar T., Rakha, A., Abbas, G., Ali, A., Haider, N., Ali, A., Hayat, S., Masooma, S. and Ahmad, J., 2015. Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. *Leg. Med*., **17**: 134-139. https://doi.org/10.1016/j.legalmed.2014.09.007

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**: 585-595.

Van Oven, M. and Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutat*., **30**: E386-E394. https://doi.org/10.1002/humu.20921

van Oven, M., Vermeulen, M. and Kayser, M., 2011. Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. *Invest. Genet*., **2**: 6. https://doi.org/10.1186/2041-2223-2-6

Yang, I.S., Lee, H.Y., Yang, W.I. and Shin, K.J., 2013. mt DNA profiler: A web application for the nomenclature and comparison of human mitochondrial DNA sequences. *J. Forensic Sci*., **58**: 972-980. https://doi.org/10.1111/1556-4029.12139

Yasmin, M., Rakha, A., Noreen, S. and Salahuddin, Z., 2017. Mitochondrial control region diversity in Sindhi ethnic group of Pakistan. *Leg. Med*., **26**: 11-13. https://doi.org/10.1016/j.legalmed.2017.02.001

Zenil, H., 2017. *Reprogramming matter, life, and purpose*. arXiv preprint arXiv:1704.00725.

Ziętkiewicz, E., Witt, M., Daca, P., Żebracka-Gala, J., Goniewicz, M., Jarząb, B. and Witt, M., 2012. Current genetic methodologies in the identification of disaster victims and in forensic analysis. *J. appl. Genet*., **53**: 41-60. https://doi.org/10.1007/s13353-011-0068-7