Full Length Research Article
Advancements in Life Sciences – International Quarterly Journal of Biological Sciences

ARTICLE INFO

Date Received: 23/10/2019;
Date Revised: 21/11/2021;
Date Published Online: 25/05/2022;

Authors' Affiliation:
1. Faculty of Life Sciences, University of Central Punjab, Lahore - Pakistan
2. Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore - Pakistan
3. Faisalabad Medical University, Faisalabad – Pakistan
4. Centre for Applied Molecular Biology (CAMB), The University of Punjab, Lahore – Pakistan
5. Soil and Water Testing Lab Jhelum – Pakistan

***Corresponding Author:**
Azam Ali
Email: azam.ali@imbb.uol.edu.pk

How to Cite:
Jaffer M, Shahzad MS, Rahman Z, Ali A, Kalsoom S et al., (2022). Genetic Characterization of Wakhi People from Hunza Valley of Pakistan by employing Mitochondrial DNA Control Region. Adv. Life Sci. 9(1): 13-17.

Keywords:
Forensic Science; Control region; Haplogroup; Genetic diversity

Open Access

Genetic Characterization of Wakhi People from Hunza Valley of Pakistan by employing Mitochondrial DNA Control Region

Marriam Jaffer[1], Muhammad Saqib Shahzad[1], Zia Ur Rahman[1], Azam Ali[2*], Saeeda Kalsoom[2], Javed Iqbal Bajwa[3], Muhammad Farooq Sabar[4], Alamgir Alvi[5]

Abstract

Background: Study of populations at genetic level marks high importance in terms of determination of population specific patterns. The study and analysis of population specific genetic patterns have wide ranging applications from medicine to forensic science. The study of mtDNA control regions gives the insight of maternal lineages of particular ethnic group and their evolutionary as well emigrational history.

Methods: This research gives information of mitochondrial DNA haplotypes data of CR (control region) covering the area from 16024bp to 576bp of mitochondrial DNA of Wakhi population of Hunza valley from Pakistan. Samples of 40 unrelated Wakhi from upper Hunza were sequenced and their sequences encapsulating mtDNA control region was compared to rCRSs (revised Cambridge reference sequence) to see maternally inherited DNA variation at genetic level in this population.

Results: The results showed that all forty (40) haplotypes are unique. The haplotypes corresponded to 67.9% West Eurasian haplogroups followed by the Middle East and variety of Asian haplogroups exhibiting admixed maternal genetics of this population. Wakhi population comes with high genetic diversity (0.998) in turn lowest random match probability (0.026) and high power of discrimination (0.974).

Conclusion: This study gives interesting highlights important aspect of uni-parental genetics of Wakhi population and is also a contribution to mtDNA control region data of Pakistani populations for applications in criminal investigations.

Introduction

The mtDNA control region assists to find out the immigration trail of a population throughout the history along with other related uni parental genetic constitution [1]. Analysis of mtDNA haplogroups has been proven valuable in determining not only the evolutionary account of human populations but also the turns the trick in favor of crime investigators to identify a human profile where conventional STR typing comes with limitation to find a hit [2]. Pakistan (Figure 1) happened to be in the most recurrently followed expedition road, which attracted people from out of Africa and referred as one of the ancient regions where contemporary human being chosen to reside [3,4]. By the language and civilization, Pakistani population is mixture of 16 cultural groups of assorted ancestries [5]. Hunza valley is located in the highest Pakistani northern zone in the Karakoram Mountains [6]. Upper part of Hunza is inhabited by Wakhi people believed to be evolved from prehistoric Iranian people. The Wakhi identify them by their Eastern Iranian language which belongs to Pamirian group. Wakhi language is very different from Tajik language, and it is also very different from other Pamiri languages [7]. No genetic study has ever been carried out to dissect the genetic structure of this population. Therefore, we conducted uniparental marker (mtDNA control region) study to find out the maternally inherited DNA pattern of Wakhi population for the usefulness in criminal investigations and snapshot of evolutionary as well as emigrational track record.

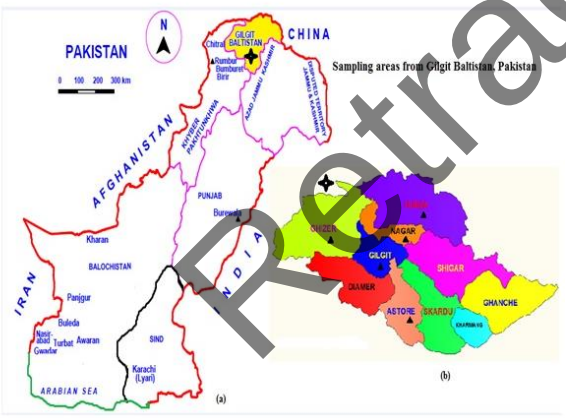

Figure 1: Map of Pakistan showing major ethnic groups and sampling area.

Methods

Collection of Samples

3-5 ml of blood samples were collected from 40 maternally unrelated individuals of Wakhi population in Hunza Valley of Pakistan. Sampling was done from different areas of Hunza Valley (Figure 2) with the written and oral consent following the according to the declarations of Helsinki [25]. Approval for sample collection was obtained from ethical committee of (UCP) University of Central Punjab, Lahore.

DNA Extraction, Quantification, Amplification

DNA of all samples was extracted by organic method. The genomic DNA was quantified by the real time PCR (7500 SDS Real-Time PCR System).

Entire mtDNA (1122bp approximately) control region was amplified by utilizing the primer set from (http://forensic.yonsei.ac.kr/protocol/mtDNA-CR.pdf). The primer sequence has been given in detail in (Table 1). Amplification was performed in a volume of 25 µl consisting of 1ng of genomic DNA, 0.4 µM of both forward and reverse primers and adding 0.4 µl Taq Polymerase (Thermo scientific #EP0402). The PCR procedure involves the denaturation at 94°C, annealing at 54°C for 30 seconds following by extension at 72°C for 60 seconds. While final extension process was done at 72°Cfor 60 seconds. After the amplification, PCR products were treated with ExoSAP-IT® (USB, Cleveland, OH, USA) to get rid of unused primers and dNTPs.

Primer name (Control region)	Primer Sequences (5´→3´)	Concentration of each primer (µM)	Melting Temperature (°C)
Amplification and sequencing primer-F15975	CTC CAC CAT TAG CAC CCA AA	0.2	55.1
Amplification and sequencing primer-R635	GAT GTG AGC CCG TCT AAA CA	0.2	54.7

Table 1: Sequences of primers used in current study

Sequencing

Bidirectional sequencing of entire mtDNA control region (covering nucleotide position 16,024–16,569 and 1–576) was performed using the Big Dye Terminator Cycle Sequencing v3.1 Ready Reaction Kit (Applied Biosystems) according to the manufacturer's instructions.

Data Analysis

All bidirectionally sequenced samples were assessed twice by independent researchers as per commendation of [8] utilizing the Geneious (Version 7.0.3, Biomatters Ltd, New Zealand) as sequencing analysis tool. The eminence of mtDNA data was determined by the MitoTool [9] and mtDNA Profiler [10].

Haplogroup assignment was done by using HaploGrep [11] by taking the PhyloTree [12] Build 17 into consideration. The statistical parameters like Genetic diversity, power of discrimination and random match probability were calculated following the [13, 14] to

determine richness of gene pool and in formativeness of mtDNA as marker for forensic applications for this population.

Results

All of forty (40) samples were sequenced and compared to rCRSs for detection of haplotypes. After the sequence analysis, all samples were found to be different on genetic level marking the rich gene pool of population. Haplogroup frequency of each individual was calculated. The haplogroup H14a+146 were more frequent (5%) in Wakhi population as compared to other haplogroups. The haplogroup frequency of mitochondrial DNA of Wakhi population is presented in table 2.

Wakhi population has an admixed mitochondrial DNA pool. 67.9% West Eurasian haplogroups (H101, H15a1b, H14a+146, H1+16239, H14a+146, H4a1a1a3, H15a1b, H2a1, U4b1a1a1, U4b2, U2b2, U4'9, k1a13, k1b2, T1a, T2d1b, T, T1a1'3,).Wakhi population also exhibited the haplogroup J: 7.5% the haplogroup having origin of Middle East (J1b1b, JT) and found in Middle East with the percentage as high as 12%. 7.5 % of East African haplogroup (N9a1), 5% South Asian haplogroup (M31a1, M3), 2.5% East Asian haplogroup (A+152+16362), 2.5% of African haplogroup, 2.5% of Southeastern haplogroup (B4), 2.5% of Central Asian haplogroup (C4a1), 2.5% of Southeast Asian haplogroup (R2).

Sr. No.	Sample ID	Haplotype ID	Differences to rCRS (309.1C,315.1C,524,16519C,309.2C,523,524,16030d are disregarded)	Haplogroup
1	HW35	M1	16047T, 16050d, 16051d, 16053A, 16054.1G, 16054.2G, 16060A, 16126C, 16163G, 16186T, 16189C, 16220C 16294T, 73G, 263G, 489C, 511T	T1a
2	HW37	M2	16027.1C, 16030T, 16038d, 16126C, 16163G, 16186T, 16220C, 16294T, 73G, 152C, 195C, 263G	T1a1'3
3	HW38	M3	16223T, 16290T, 16319A, 16362C, 73G, 152C, 235G 263G, 533G	A+152+16362
4	HW39	M4	16026G, 16028C, 16031T, 16033d, 16039d, 16052d, 16127C, 16148T, 16224T, 16262.1C, 73G, 195C, 263G, 284.1A, 482C, 489C	M3
5	HW41	M5	16026G, 16028C, 16031T, 16033d, 16039d, 16052G, 16210C, 16240T, 16254.1A, 16311C, 16352C, 16353T, 73G, 146C, 52C, 153G, 234G, 263G, 284.1A	U2b2
6	HW42	M6	16124C, 16184T, 16311C, 44.1C, 55C, 57C, 146C, 263G, 297C	H15a1b
7	HW43	M7	16203G, 16256T, 16352C, 146C, 263G	H14a+146
8	HW44	M8	16126C, 16294T, 73G, 194T, 200G, 263G	T
9	HW45	M9	16184T, 16263C, 16304C, 16357C, 146C, 195C, 263G, 497T	H101
10	HW46	M10	16027.1C, 16030T, 16038d, 16224C, 16311C, 73G, 146C, 195C, 263G, 497T	K1a13
11	HW47	M11	16111T, 16239T, 16362C, 16482G, 239C,263G	H1+16239
12	HW49	M12	16356C, 16362C, 73G, 195C, 263G, 524.1A, 524.2C, 524.3A, 524.4C	U4b1a1a1
13	HW50	M13	16126C, 16294T, 73G, 194T, 200G, 263G	T2d1b
14	HW51	M14	16126C, 16294T, 73G, 194T, 200G, 263G, 282.1T	T2d1b
15	HW52	M15	16136C, 16356C, 73G, 195C, 263G, 284.1A, 294.1T, 499A, 524.1A, 524.2C, 524.3A, 524.4C	U4b2
16	HW53	M16	16224C, 16311C, 16320T, 73G, 146C, 195C 263G, 524.1A, 524.2C, 524.3A, 524.4C	K1b2
17	HW54	M17	16036d, 16037d, 16069T, 16126C, 16145A, 16261T 16311C, 73G, 185A, 263G, 271T, 295T, 458.1T, 462d, 463d, 489C	J1b1b
18	HW55	M18	16051G, 16209C, 16239T, 16311C, 16352C, 16353T, 73G 146C, 152C, 153G, 234G, 263G	U2b2
19	HW56	M19	16036d, 16037d, 16224C, 16311C, 16320T, 16519C, 73G 146C, 195C, 263G,524.1A, 524.2C, 524.3A, 524.4C	K1b2
20	HW57	M20	16224C, 16311C, 16320T, 73G, 89C, 146C, 195C 263G, 524.1A, 524.2C, 524.3A, 524.4C	K1b2
21	HW60	M21	16126C, 16147T, 16223T, 73G, 150T, 263G	L3e'i'k'x
22	HW61	M22	16224C, 16311C, 73G, 146C, 195C, 263G, 497T	K1a13
23	HW63	M7	16203G, 16256T, 16352C, 146C, 263G	H14a+146
24	HW67	M23	16111T, 16129A, 16223T, 16247G, 16257A, 16261T, 73G 194T, 200G, 263G	N9a1
25	HW68	M24	16289G, 44.1C, 55C, 57C, 146C, 263G	H15a1b
26	HW69	M25	16069T, 16126C, 16145A,16261T, 16311C, 73G, 152C 263G, 524.1A, 524.2C	JT
27	HW72	M26	16051G, 16209C, 16239T, 16311C, 16352C, 16353T, 73G, 150T, 263G	U2b2
28	HW74	M27	16111T, 16129A, 16223T, 16247G, 16257A, 16261T, 73G 150T, 263G	N9a1
29	HW75	M28	16182C, 16183C, 16189C, 16194C, 16217d, 16219C 16228T, 16230T, 16234d, 16258C, 16269C 16287T, 16289d, 16293C, 16305T, 16309G, 16311C, 16312C, 16326C, 16338C, 16355T, 16358T, 16360G, 16402C, 16444T, 25C, 73G, 146C, 152C, 181C, 214G, 241C 263G, 368G 418T	H4a1a1a3
30	HW76	M29	16182C, 16183C, 16189C, 16197G, 16204A, 16217C 16245G, 16258C, 16265C, 16276A, 16293C, 16310A 16311A, 16330G, 16332A, 16336A, 16344A, 16384A 16386A, 16413G, 16422A, 16434A, 16436A, 16437G 16465A, 16475G, 16484G, 16498A, 16505G, 16519G 16535A, 16558A, 16560A, 16564C, 10G, 73G, 146C, 152C, 214G, 263G, 368G	B4
31	HW78	M30	16111T, 16129A, 16223T, 16247G, 16257A, 16261T, 73G 195C, 263G, 499A, 524.1A, 524.2C, 524.3A, 524.4C	N9a1
32	HW80	M31	16129A, 16223T, 16298C, 16327T, 73G, 263G	C4a1
33	HW87	M32	16224C, 16311C, 16320T, 73G, 195C, 263G, 499A	U4'9
34	HW89	M33	16028.1C, 16028.2T, 16030T, 16036d, 16037d, 16136C 16356C, 73G, 195C, 263G, 499A, 524.1A, 524.2C, 524.3A, 524.4C	U4b2
35	HW91	M34	16069T, 16126C, 16145A, 16261T, 16311C, 73G, 185A, 263G, 271T, 295T, 458.1T, 462d, 463d, 489C	J1b1b
36	HW92	M35	16136C, 16356C, 73G, 249d, 263G, 489C	M31a1
37	HW107	M36	16025d, 16224C, 16311C, 73G, 146C, 195C, 263G, 497T	K1a13
38	HW110	M37	16071T, 16362C, 73G, 152C, 263G, 524.1A, 524.2C	R2
39	HW115	M38	16309G, 16354T, 93G, 263G	H2a1
40	HW122	M39	16027.1C, 16030T, 16038d, 16224C, 16311C, 16519C, 73G 146C, 195C, 263G, 497T	T1a

Table 2: Estimated haplotypes and respective haplogroups in Wakhi population.

The genetic diversity of Wakhi population was high (0.998) with lower random match probability is (0.026) and in turn high power of discrimination (0.974) indicating rich gene pool (Table 3).

Sample size	40
Number of different haplotypes	40 (40 Unique)
Random match probability	0.026
Power of discrimination	0.974
Genetic diversity	0.998

Table 3: Statistical parameters of Wakhi people.

Discussion

Since the age of mankind, has utilized plants for the This study revealed the haplotype data of hypervariable regions of the control region of mitochondrial DNA in Wakhi population living in upper Hunza (Gojal), Pakistan. Each individual's mitochondrial DNA control region was sequenced and haplogroup assignment was done. Haplogroups U, T, A, M, H, K, J, L, N, B, C and R were observed in Wakhi population. The haplogroups H, U, T and K were frequently found in samples of this population.

Thirty nine (39) unique haplotypes were observed in Wakhi population, hence this population has a very high genetic diversity (0.998) and lowest random match probability is 0.026. The results of this study have been compared to other populations being studied in Pakistan e.g. especially genetic diversities of Makrani (0.97) [1], Saraiki (0.96) [2], Pakhtun (0.99) [4], Baluchi including three other ethnic groups from Sindh and Baluchistan province (0.97) [19], Sindhi (0.992) [20], Punjabi (0.963) [15], Kashmiri (0.997) [21], Hazara (0.994) [23], Kho from Northwest region (0.0215) [22] and Gujar from Northwest region of Pakistan (0.922) [24] population (Table 4). This population was found to be the most diverse, genetically, than other so far studied populations from Pakistan. The haplogroup diversity observed in present study is comparable with the haplogroup diversity of Pashton population in which predominant haplogroups were of West Eurasian origin (55.6%) as it is common as far as Wakhi are concerned. In this study of Wakhi population, West Eurasian H haplogroup was found in 8 individuals (20%). Haplogroup U of West Eurasian nature was second most prominent with 17.5% frequency after H haplogroup. Different subgroups of super haplogroups-H have been observed in various other ethnic groups from Pakistan like Makrani, Pathan and Kashmiri people. The most prominent mtDNA super haplogroup-U, is frequently found in Pathan (17%) and in Baluch (74%) in Pakistan. The haplogroup U2 is considered to be the most ancient lineage of super Haplogroup U. The haplogroup U2 and other sub haplogroups of haplogroup U are found most frequently in Kashmiri and Kalash people from Pakistan. The haplogroup H is considered to be instigated in Southwest Asia around 20,000 to 25,000 years ago [16]. The haplogroup was also observed among the Tryptillians. The clade was found to be in the DNA of ancient mummies which were discovered in the middle Egypt at the archeological site of Abusir el-Meleq dating from the pre Ptolemaic/ late new kingdom and Ptolemaic interlude [17].

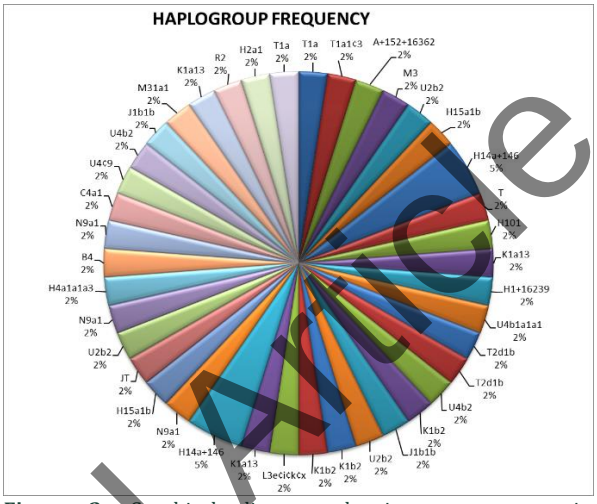

Figure 2: Graphical diagram showing percentage wise frequencies of mtDNA haplogroups in Wakhi population with H14a+146 as most frequent haplogroup up to 5%.

The haplogroup U was discovered from human skeletal remnants discovered in western Siberia which has been marked as most ancient DNA belonging to this clade, was dated to 45000 years ago [18]. Around 11 % of native Europeans bearing this haplogroup in their DNA and haplogroup are categorized as most old prehistoric maternally inherited haplogroup found in the region. Likewise, haplogroup H, it was also found in Egyptian mummies at the archeological site of Abusir el-Meleq dated to the 1st millennium BC [17]. In this context, further studies on this ethnic group could shed light on the interesting genetic aspects which could lead to its possible association with southwest Asian and Europe also emigrational history of this ethnic group to or from these continents.

Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contribution

Marriam Jafar: Study design
Muhammad Saqib Shahzad: Refinement of study design and supervision
Zia Ur Rahman: Technical support

Azam Ali: Data analysis
Rahat Abdul Rehman: Monitoring of research
Saeeda Kalsoom: Drafting
Javed Iqbal Bajwa: Manuscript scanning
Muhammad Farooq Sabar: Manuscript scanning
Alamgir Alvi: Data evaluation

References

1. Siddiqi H, Akhtar T, Rakha A, Abbas G, Ali A, et al. .Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. Legal Medicine, (2015); 2: 134–139.
2. Hayat S, Akhtar T, Siddiqi H, Rakha, A, Haider N, et al. Mitochondrial DNA control region sequences study in Saraiki population from Pakistan. Legal Medicine, (2015);(2): 140–144.
3. Qamar R , Ayub Q, Khaliq S, Mansoor A, Karafet T, et al. African and Levantine origins of Pakistani YAP+ Y chromosomes. Human Biology, (1999);5: 745–755.
4. Rakha A, Shin J, Yoon A, Kim Y, Siddique H, et al. Forensic and genetic characterization of mtDNA from Pathans of Pakistan. International Journal of Legal Medicine, (2011); 6: 841–848
5. Wells S, Mohyuddin A, Santachiara-Benerecetti S, Tyler-Smith C, et al. Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. The American Journal of Human Genetics, (2004); 5: 827–845.
6. Dani, A. (2001). History of Northern Areas of Pakistan: Upto 2000 AD.
7. Shafiq, N. The Wakhi Community Settlements in Northern Pakistan. Journal of Political Studies, 2011; (4)
8. Parson W, Bandelt J. Extended guidelines for mtDNA typing of population data in forensic science. Forensic Science International: Genetics, (2007); 1:13–19.
9. Fan L ,Yao G. MitoTool: A web server for the analysis and retrieval of human mitochondrial DNA sequence variations. Mitochondrion, (2011); 2: 351–356
10. Yang S, Lee Y, Yang I, Shin J. mtDNAprofiler: A Web Application for the Nomenclature and Comparison of Human Mitochondrial DNA Sequences,. Journal of Forensic Sciences, (2013);4: 972–980
11. Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R., et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Human Mutation, (2011); 1: 25–32.
12. Oven V, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation, (2009); 2: E386–E394
13. F Tajima. The effect of change in population size on DNA polymorphism. GENETICS,(1989);3: 597-601.
14. Prieto L, Zimmermann B, Goios A, Rodriguez-Monge A, Paneto G, Alves C, Parson W. The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework. Forensic Science International: Genetics, (2011);2: 146–151.
15. Bhatti S, Abbas S, Aslamkhan M, Attimonelli M, Trinidad S, et al. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. Mitochondrial DNA Part A, (2018); 5: 714–726
16. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. American Journal of Human Genetics, (2004); 5: 910–918.
17. Schuenemann J, Peltzer A, Welte B, Pelt V, Molak P, et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. Nature Communications, (2017); 8: 15694
18. Larruga M , Marrero P, Abu-Amero K, Golubenko M V, Cabrera V M. Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area. BMC Evolutionary Biology, (2017);1: 115.
19. Bhatti S, Khan A, Abbas S, Attimonelli M, Gonzalez R, et al .Problems in Mitochondrial DNA forensics: while interpreting length heteroplasmy conundrum of various Sindhi and Baluchi ethnic groups of Pakistan. Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis, (2018); 4: 501–510.
20. Yasmin M, Rakha A, Noreen S, Salahuddin Z. Mitochondrial control region diversity in Sindhi ethnic group of Pakistan. Legal Medicine, (2017); 26: 11–13.
21. Rakha A, Peng S, Bi R, Song J, Salahudin Z, Adan A, et al. EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu & Kashmir, Pakistan. Forensic Science International: Genetics, (2016); 25: 125–131.
22. Aziz S, Nawaz M, Afridi S G, Khan A. Genetic structure of Kho population from north-western Pakistan based on mtDNA control region sequences. Genetica, (2019);2:177-183
23. Rakha A, Peng F, Adnan A, Bi R, Yasmin M, et al. mtDNA sequence diversity of Hazara ethnic group from Pakistan. Forensic Science International: Genetics, (2017);30: e1–e5
24. Ullah I, Ahmad H, Hemphill E, Nadeem S, Tariq M, Tabassum S. Mitochondrial genetic characterization of Gujar population living in the Northwest areas of Pakistan. Advancements in Life Sciences,(2017);3:84-91
25. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects.. Bulletin of the World Health Organization, (2001);4:373 - 374.