Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v30i18.3311

GENETIC CHARACTERIZATION OF MITOCHONDRIAL DNA CONTROL REGION OF ARAIN ETHNIC GROUP IN PAKISTAN

Farah Ahmad[a], Muhammad Afzal[a*], Muhammad Saqib Shahzad[b]

[a] Department of Basic and Applied Chemistry, University of Central Punjab Lahore - Pakistan.
[b]Government Post Graduate College Model Town Lahore - Pakistan.

***Corresponding Author:** Muhammad Afzal
dr.afzal@ucp.edu.pk

Abstract

Numerous studies have demonstrated that by examining people's genetic makeup, one can learn about the past movements of communities. Due to its significant evolutionary value, mitochondrial DNA is a suitable tool for determining human migration, geographic distribution, and population origin.. A given ethnic group's maternal lineages, as well as its evolutionary and emigrational histories, can be gleaned from the examination of mtDNA control regions. 100 samples were collected (83 samples showed results) and sequenced the entire mtDNA control region of 83 unrelated individuals in Pakistan from different areas of Pakistan. The samples were compared with the revised Cambridge reference sequences. 83 dictinct haplotypes were observed, all of them were unique, and none of them were shared by any individuals. The Arain population's mtDNA diversity was 0.9916, and its power of discrimination was 0.9917. The results showed a strongly admixture mtDNA pool consisting of Indian sub-continent R5a(R5a2) 7%, South Asia M30g(M30g) 5%, Eurasian, South Asian U2b(U2b) 5%, Central and Northeast Europe T1a(T1a1'3) 4.8%, South Asian hyplogroups include U2b(U2b2) 3.0% M34a(M34a1) 2.0% M18(M18'38) 2.0%. The R5a(R5a2) Haplogroups were discovered to clearly dominate other haplogroups. It implies that the different geographic location of the Arain population may be due to the heterogeneous haplogroup makeup of their mitochondrial DNA. It has highlighted mtDNA based genetic characterization of the population, the diversity and evolutionary aspects. The outcome of study is beneficial for medical applications and for human identification. The information is also a contribution to Pakistani population's national mtDNA data, which will be useful in criminal investigations.

Keywords: Mitochondrial DNA, Arain, hyplogroup, ethnic group, maternal lineages

Introduction

DNA is the life's architectural blueprint. The genetic information necessary for eukaryotic and prokaryotic species, including bacteria and viruses, for their development, growth, reproduction, and operations, is carried by the DNA molecule (Lehninger *et al.*, 2000). DNA analysis is a fundamental instrument in biochemical and genetic research that offers vital information about biology, useful details about ancestry, and evidence for legal cases (Jafari & Ansari 2019). Although DNA is found in the chromosomes of the nucleus in a compact form, mitochondrial DNA, also known as mtDNA, is found in the mitochondria. With 16,569 base pairs encoding 37 genes, mitochondrial DNA was the first organelle of the human genome to be sequenced (Vieira *et al.*, 2016). A single cell contains several thousands of copies of the mitochondrial genome. The situation of the presence of identical

copies of mtDNA inside a cell is known as homoplasmy while the presence of two or more different copies of the mtDNA inside a cell is known as heteroplasmy. Mitochondrial DNA is recognized as the dynamic genetic organelles which are capable to divide, self-replicate, move, and fuse (Vieira *et al.,* 2016). In genomic studies, uniparental DNA markers are used for the mtDNA and Y-chromosome study, while autosomal markers are predominant (Cano *et al.*, 2014). These autosomal markers have a huge parameter of variations inside the autosomal genome therefore are considered a strong tool for human identification (Ma *et al.*, 2014). The use of mtDNA in the biological as well as evolutionary studies has attained huge importance (Pilipenko et al., 2018). The mtDNA analysis helps in forensic science for almost completely degraded samples like bones, cigarette butts, nails, teeth, stains, and the hair shafts that have a minute amount of nuclear DNA in them for their genetic analysis (Hoseinzadeh *et al.*, 2016). The circular shape and the location of human mtDNA have proven its significance in the scientific investigation due to self-replication and energy production mechanisms (Cano *et al.*, 2014).

Displacement loop (D-loop) is the control region of the mtDNA covering approximately 7% (1122 bp) of the total genome of mitochondria (Hoseinzadeh *et al.*, 2016). This loop has three highly variable regions (HVS) that are HVSI (342bp), HVSII (268bp), and HVSIII (137bp). From these, HVSI and HVSII are regions with a higher level of polymorphism (Cano *et al.*, 2014). Anderson formed the Cambridge Reference Sequence (CRS) by sequencing the whole mtDNA genome of human (Anderson *et al.*, 1981) which is now being used for standard comparisons. Any distinction when observed between the Anderson and the sequences of the individual sample is referred to as polymorphism (Cano *et al.*, 2014). The haplogroups of the mitochondrial DNA are the genome markers that are inherited maternally. The analysis of mutations in the mtDNA has shown the presence of alike mutations in individuals who share a similar maternal line (Ma *et al.*, 2014). Heteroplasmy is considered the form of polymorphism where it contains two different forms of mitochondrial DNA either in a cell or the individual (Barbieri *et al.*, 2014). There exist seven macro-haplogroups that are L0, L1, L2, L3, L4, L5, and L6. These are specific for different area wise populations like Africans, the L haplogroup was reported in western Africa in 2014 (Barbieri *et al.*, 2014). Other haplogroups like M, N, and R groups are subgroups of the macrohaplotypes (Goto *et al.*, 2011). The studies conducted in Pakistan include mtDNA analysis for determining Pathan lineage (Rakha *et al.*, 2011), Makrani (Siddiqi *et al.*, 2015), Kalash (Ayub *et al.*, 2015), and Saraiki (Hayat *et al.*, 2015).

Arain population has huge importance among the social interactive populations and these have not previously been investigated (Ayub & Skaletsky 1999). In Pakistan's Punjab and Sindh provinces, the Arain (also known as Raeen) are a sizable agricultural tribe with a distinct political identity and high level of organization (Ayub & Skaletsky 1999). The Arain, who at first worshipped Hinduism but eventually converted to Islam in large numbers, are said to be displaced farming communities who relocated to Punjab from Sindh and Multan as Arab Muslim troops encroached. This theory is held by historian and political scientist Christophe Jaffrelot. According to him, the group is connected to the Kamboj Rajput group, which is primarily found in Pakistan and India (Sutherland 2003; Nesheva 2014). Since almost all Arain are and have always been Sunni Muslims, like the early Arabs of Muhammad bin Qasim's expedition, many Arain claim to be of Arab ancestry (Van & Kayser 2009). (Ishtiaq Ahmed 2006).

Despite having a predominance of Muslims, the Arain also includes Sikh and Hindu people, just as the other agricultural castes in the Punjab (Ishtiaq Ahmed 2006). Therefore, rather than being a precise genealogical identity, the Arab origin claim can be mostly understood as a wish to claim an uninterrupted practice of Islam over the years for modern status (Ishtiaq Ahmed 2006). O is the most frequent blood group (across all ethnicities), with the exception of the Arain, where B is most

prevalent, according to a study published in the Pakistan Journal of Medical Sciences, with the difference being statistically significant. It doesn't prove that the Arain have non-local heritage, but it does show how they differ from the other castes who live in the Punjab. For additional information about the caste's migratory origins, obviously more study is necessary (http://www.pjms.com.pk/issues/janmar05/article/article6.html).We collected information for the Hyper Variable Region 1&2 (HVR1&2) of mtDNA from 100 Arain people from various regions of Pakistan in order to examine all potential lineages among different ethnic groups. mtDNA haplogroup affiliations were determined using various servers and software, and we then analyzed the distribution of mtDNA among the different subpopulations, including local ethnic groups from Pakistan and its neighbors.

In this study we aim to present the atlas of maternally inherited DNA of Arain living in Pakistan. It will highlight mtDNA based genetic characterization of the population, the diversity and evolutionary aspects. The outcome of study is beneficial medical applications and for human identification. The information is likewise a commitment towards public mtDNA information of Pakistani populace which will be useful in criminal examinations.

Materials and methods

Blood samples from 100 maternally unrelated individuals were taken from different areas of Pakistan from Mughal ethnic group. Mughal people were identified based on their different ethnicity, their settlement, and other details in the different areas of Pakistan. All the details were recorded in the consent form that was filled and signed by the donor. EDTA vacutainers were used for preserving the blood samples. All the ethical aspects were kept in a note while conducting the sampling.

DNA extraction

DNA Mini Kit (Qiagen, Hilden, Germany) was used for the DNA Extraction from the preserved blood samples previously stored in the EDTA tubes. DNA extraction Protocol was adopted that was given by the manufacturer’s instructions.

PCR Amplification and Sequencing

Quantification of the genomic DNA was done with the help of Nano Drop™ 1000 Spectrophotometer (Thermo Scientific, USA). For amplification of the entire mt DNA’s D-loop, forward and reverse primers were used (Rakha et al., 2011). The amount of genomic DNA used was 1-2 ng. In each PCR reaction mix, 0.4 μM was the concentration of the primers used. Ampli Taq Gold® 360 PCR Master Mix (Applied Biosystems, Foster City, CA, USA) of volume 50 μL was used for performing PCR amplification. The overall process of PCR amplification consists of; the first step that is pre-denaturation conducted for 11 minutes at 95°C, the second step that is 35 cycles of denaturation for the 30s at 95°C, the third step of annealing for 30s at 56°C, the fourth step of extension for 90s at 72°C and fifth and final step of final extension for 7 minutes at 72°C. The size and quality of the amplified fragment of mtDNA is shown in **figure 1 & 2.**

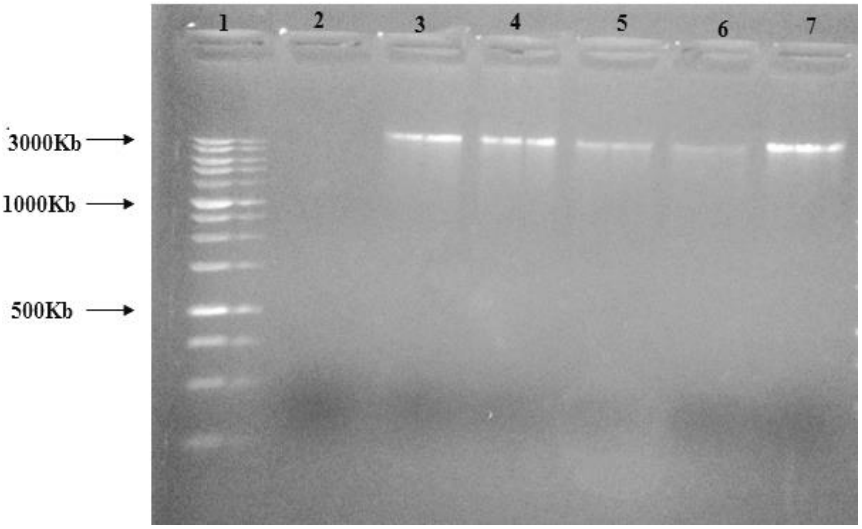

Figure 1: Agarose gel electrophoresis examination of separated genomic DNA from Arain blood tests (1) 1kb ladder (ThermoTM SM # 0323), (2) Negative Control, (3) AR0001, (4) AR0002, (5) AR0003, (6) AR0004, (7) AR0005

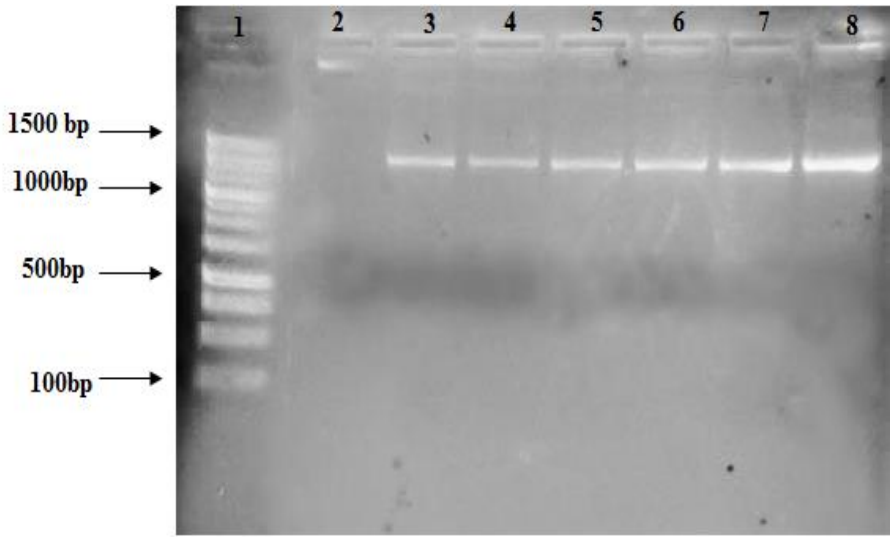

Figure 2: garose gel electrophoresis examination of separated genomic DNA from Arain blood tests (1) 100bp DNA ladder (ThermoTM SM # 0323), (2) Negative Control, (3) AR0005, (4) AR0006, (5) AR0007, (6) AR0008, (7) AR0009, (8) AR00010.

Sequencing of the mtDNA Control Region

Purification of the amplified product of the mtDNA control region was followed by a bidirectional Sanger sequence using the mtDNA control region sequence. The chromatogram was independently cross-checked for each sample after it had been processed twice. The bidirectional chromatograms of Arain (AR0009) for complete mtDNA control locale by utilizing forward groundwork (F15975) and turn around preliminary (R635) are displayed as models in **figures 3a** and 3b separately. Chromatograms for Arain (ARAR0081) got by utilizing forward (F16524) groundwork and opposite (R042) preliminary are displayed in Figures 4a and 4b. For haplotype affirmation, extra sequencing for distinguishing proof of important SNPs was done.

Figure 3 (a): Chromatogram of Arain individual (AR0009) for entire mtDNA control region sequenced by forward primer (F15975).

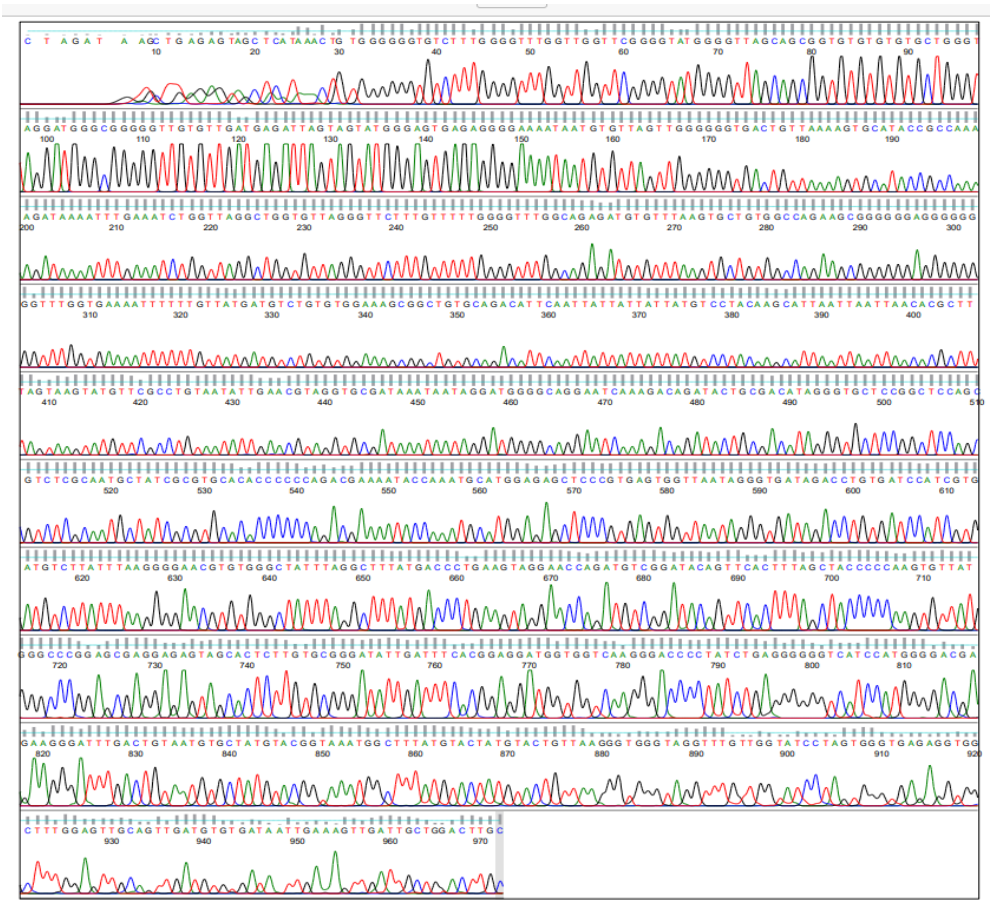

Figure 3 (b): Chromatogram of Arain individual (AR0009) for entire mtDNA control region sequenced by reverse primer (R635).

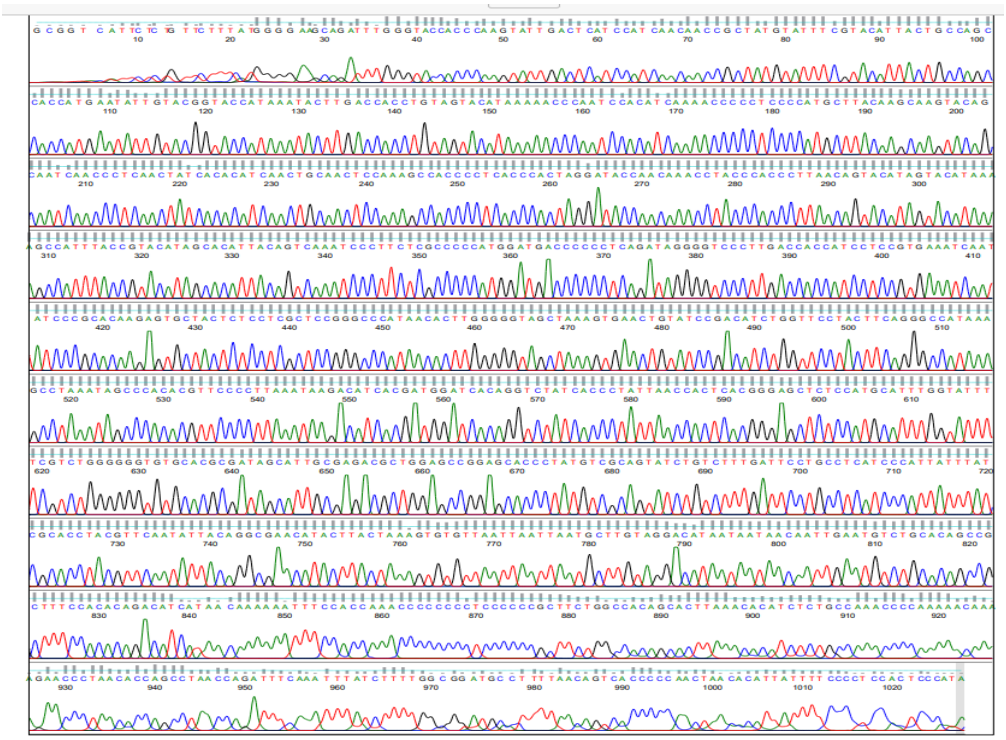

Figure 4 (a): Chromatogram of Arain individual (AR00081) for entire mtDNA control region sequenced by forward primer (F16524).

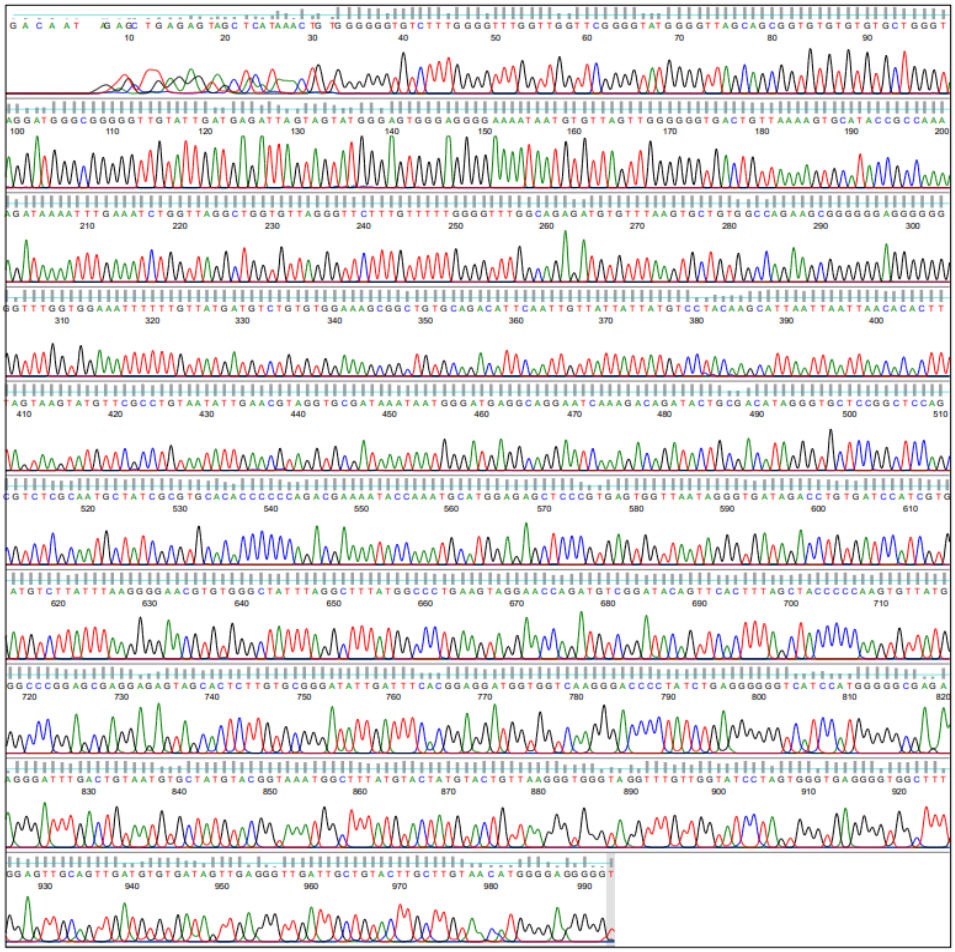

Figure 4 (b): Chromatogram of Arain individual (AR0081) for entire mtDNA control region sequenced by reverse primer (R042).

Reconstruction and Alignment with rCRS

The "mtDNA assembly tool," one of the tools in the mtDNA profiler, was used to reconstruct the anticipated mtDNA sequence of each Arain individual from a number of fragmented sequences (Yang et al., 2013). All Arain samples' reconstructed sequences were aligned with rCRS to determine their differences from rCRS.

Data Analysis

Every one of the examples were sequenced bi-directionally and adjusted utilizing arrangement examination programming Geneius (Form 7.1.5, Biomatters Ltd, New Zealand). A comparison of the sequences and results was done in both the directions with rCRD (Anderson et al., 1981: Andrew et al., 1999). In order to determine the haplotype of each sequence, the sequences were subjected to be analysed by MITOMASTER, a tool present on the MITOMAP platform. The application of MITOMASTER takes each mtDNA control region sequence as input file and compares it with rCRS for the assignment of respective haplotype and haplogroups. The tool utilizes the HaploGrep2 with Phylotree 17 for haplogroup determination. The statistical parameters of the population, like the diversity of haplotypes (GD = $\frac{n}{n-1}1-\sum_i Pi^2$), random match probability (PM = $\sum iGi^2$) and power of discrimination (PD = $1-PM$) were statistically calculated according to the previous studies (Fisher & Garner, 2020).

Results

The study highlights very important genetic an evolutionary dynamic of the Arain population. Out of the 100 samples under the study,83 samples showed results and it turned up with total 83 haplotypes. The 83 haplotypes out of 83 observed haplotypes were found to be as unique. None of the haplotypes were found to be shared with other haplotypes within the population under study. The haplotypes corresponded to respective haplogroups. The haplogroups give the idea of genetic relatedness of population to specific geographic locations. In the Arain population, R5a(R5a2) is the most prevalent haplogroup was found in the 7% samples marking the connection with Indian Sub-continent origins. It was followed by the haplogroup M30g(M30g) found in the 5% samples. This haplogroup (M30g(M30g) is found within South Asian countries but also it was reported to be found in the Eastern Saudi Arabia. The haplogroups U2b(U2b) was found to be 5% defining these samples be related with South Asian as well as Eurasian origins. The majority of the samples from this population were observed to be genetically close with South Asian populations. The haplogroup T1a(T1a1'3) has been observed with frequency of 4.8% within the population pointing out its connection with Central and Northren European haplogroups. The frequency of other occupant haplogroups have been displayed in the graph. The population comes with the higher genetic diversity (GD) 0.9916 signifying hugely diverse population as far as mtDNA control region is concerned. The higher genetic diversity also gives rise to higher Power of discrimination (PD) 0.9917. It also implies the very less random match probability (RMB) as 0.0083 making the persons from this population be easily distinguished from other population without consuming excessive sources. The sites having polymorphism are given as haplotypes in **Table 1**. Data of mtDNA results for its diversity as calculated statistically for the respective population are given in **Table 3.**

Table 1 The assessed haplotypes and haplogroups in Arain populace of Pakistan.

Sequence	Predicted Haplogroup	Total Variants	Variants
mtDNA_AR_0001	H1c(H1c+152)	7	T152C, A263G, C309CC, C325d(=C324dâ€), A361d(=A357dâ€), A376d, T16519C
mtDNA_AR_0003	R5a(R5a2)	15	A73G, C150T, T152C, A263G, A361d(=A357dâ€), C371d(=C369dâ€), A376d, G16036GG, A16038G, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0004	M34a(M34a1)	15	A73G, T146C, A263G, T310TTC, T310C, A361d(=A357dâ€), A376d, C387d(=C386dâ€), T401d(=T398dâ€), G16036GG, C16095T, C16223T, T16249C, T16359C, T16519C

mtDNA_AR_0005	U2e(U2e2)	76	G16036GG, A16051G, T16092C, G16129C, A16182C, A16183C, T16189C, C16197G, G16204A, C16205A, G16208A, T16209A, C16214G, T16217A, T16224C, C16228A, C16232A, C16245G, A16258C, T16263C, A16265C, A16269C, G16273A, T16276A, C16282A, A16293C, C16301A, T16304G, G16310A, T16311A, A16322C, T16330G, C16332A, G16336A, C16337A, C16344A, T16347G, T16352A, C16358T, G16361C, T16362C, A16367C, A16371G, T16372G, T16381C, G16384A, T16386A, T16392G, T16409C, T16413G, T16418A, T16422A, G16428C, G16434A, G16436A, T16437G, C16439G, C16442A, T16448C, G16450C, T16462A, C16465A, T16475G, G16477A, T16484G, T16491G, C16498A, T16505G, T16509C, T16519G, C16527G, T16533A, G16535A, A16539C, A16541C, T16550C
mtDNA_AR_0006	U7a(U7a3b)	26	A73G, T125A, C151T, T152C, T204G, T206G, T208G, T216A, A263G, C269d(=C268dâ€), C315CC, C324G, A326C, A328d, C330A, C338d, C343T, C345T, A361d(=A357dâ€), G16036GG, T16140C, A16207G, A16309G, A16318T, T16362C, T16519C
mtDNA_AR_0007	M34a(M34a1)	17	A73G, T146C, A263G, C309CC, C324G, C330A, C362d, A374C, A376d, C387d(=C386dâ€), T393d(=T391dâ€), T401d(=T398dâ€), C16095T, C16223T, T16249C, T16359C, T16519C
mtDNA_AR_0008	G1a(G1a)	16	A73G, C150T, A263G, C324G, A376d, C387d(=C386dâ€), C394A, AT397-, T414d, G16036GG, C16223T, C16234T, C16295G, T16311C, T16362C, T16519C
mtDNA_AR_0009	W(W+194)	12	A73G, G143A, A189G, C194T, T195C, T204C, G207A, A263G, G16036GG, C16223T, C16292T, T16519C
mtDNA_AR_0010	M52a(M52a)	9	A73G, T195C, A263G, G16036GG, C16223T, A16275G, G16390A, C16400T, T16519C
mtDNA_AR_0011	T1a(T1a1'3)	15	A73G, T152C, T195C, A263G, T310C, T310TTC, C325d(=C324dâ€), G16036GG, T16126C, A16163G, C16186T, T16189C, C16193T, C16294T, T16519C
mtDNA_AR_0012	M3(M3)	14	A73G, T195C, C198T, G207A, A263G, C325d(=C324dâ€), G16036GG, T16126C, A16162G, C16223T, T16342C, T16359C, G16390A, T16519C
mtDNA_AR_0013	U7a(U7a)	8	A73G, C151T, T152C, A263G, G16036GG, A16309G, A16318T, T16519C
mtDNA_AR_0015	U2b(U2b1)	12	A73G, T146C, A215G, A263G, G16036GG, A16051G, C16168T, C16201T, A16258C, T16263C, G16390A, T16519C
mtDNA_AR_0016	H5(H5)	4	A263G, C303A, G16036GG, T16304C
mtDNA_AR_0017	HV2(HV2)	8	A73G, T152C, A263G, G16036GG, C16214T, T16217C, T16311C, T16519C
mtDNA_AR_0018	M30g(M30g)	10	A73G, T195A, T204C, A263G, C309T, C315CC, A352d(=A350dâ€), G16036GG, C16223T, T16519C
mtDNA_AR_0019	M2a(M2a1a)	14	A73G, T195C, T204C, A263G, A291d(=A286dâ€), T310C, T310TTC, C325d(=C324dâ€), G16036GG, C16223T, C16270T, G16319A, T16352C, T16519C
mtDNA_AR_0021	U2a(U2a1a)	10	A73G, A263G, A365d(=A363dâ€), G16036GG, A16051G, T16154C, A16206C, A16230G, T16311C, T16519C
mtDNA_AR_0022	B6a(B6a1a)	6	G16023GG, G16036GG, G16049GG, A16051G, A16183C, T16189C
mtDNA_AR_0023	M30(M30)	10	A73G, T139C, T195A, A263G, C324G, A365d(=A363dâ€), G16036GG, G16145A, C16223T, T16519C
mtDNA_AR_0025	J1d(J1d)	10	A73G, T152C, A263G, C295T, C315CC, G16036GG, C16069T, T16126C, C16193T, T16519C
mtDNA_AR_0026	J1b(J1b8)	12	A73G, T152C, A263G, C295T, C315CC, A326d, G16036GG, C16069T, T16126C, G16145A, C16260T, C16261T
mtDNA_AR_0027	M34a(M34a)	52	G16023GG, G16036GG, G16049GG, A16051G, A16183C, T16189C, C16193CC, T16195G, C16197T, C16201A, C16205A, T16209A, A16212G, C16214A, C16223T, C16228T, A16233C, A16235C, A16237T, C16242T, T16243G, C16245G, T16249C, A16258C, T16263C, A16265C, T16271A, T16276A, C16282A,

			T16288A, C16294T, C16301A, T16304A, T16308A, T16311A, A16314T, A16322T, T16330A, A16333T, A16338C, A16340T, C16344A, T16347C, C16358T, T16359C, T16362C, T16368G, T16372G, T16381C, G16384A, T16386A, T16392C
mtDNA_AR_0030	R30a(R30a1c)	8	A73G, A263G, C315CC, G16036GG, T16172C, C16278T, C16355T, T16519C
mtDNA_AR_0031	T1a(T1a1'3)	13	A73G, T152C, T195C, A263G, T310TTC, T310C, G16036GG, T16126C, A16163G, C16186T, T16189C, C16294T, T16519C
mtDNA_AR_0032	H6(H6)	5	T239C, A263G, G16036GG, T16362C, A16482G
mtDNA_AR_0033	M30g(M30g)	10	A73G, T195A, T204C, A263G, C308T, C309d(=C303d†), G16036GG, C16223T, T16311C, T16519C
mtDNA_AR_0034	M70(M70)	11	A73G, T236C, A263G, A352C, G16036GG, C16214T, C16223T, T16297C, T16298d(=T16297d†), T16342C, T16381C
mtDNA_AR_0035	H5(H5)	5	A263G, C315CC, C324G, G16036GG, T16304C
mtDNA_AR_0036	M30g(M30g)	10	A73G, T195A, T204C, A263G, C309T, C315CC, G16036GG, C16223T, T16356C, T16519C
mtDNA_AR_0038	U2b(U2b2)	12	A73G, T146C, T152C, A234G, A263G, C315CC, G16036GG, A16051G, T16086C, G16129A, C16353T, T16519C
mtDNA_AR_0039	M33a(M33a3)	12	A73G, T146C, C151T, A257G, A263G, T310TTC, T310C, G16036GG, T16172C, C16221T, C16223T, T16519C
mtDNA_AR_0040	U2b(U2b)	11	A73G, T146C, A234G, A263G, C315CC, G16036GG, A16051G, T16086C, G16129A, C16353T, T16519C
mtDNA_AR_0041	R8(R8)	6	A73G, T152C, T195C, A263G, G16036GG, C16355T
mtDNA_AR_0042	U2b(U2b2)	12	A73G, T146C, T152C, A234G, A263G, G16036GG, A16051G, T16209C, C16239T, T16352C, C16353T, T16519C
mtDNA_AR_0043	H5(H5)	11	A263G, T310C, T310TTC, C324G, C330G, C348G, C362A, G366A, G16036GG, A16241T, T16304C
mtDNA_AR_0044	B2b(B2b+152)	21	T55G, A73G, T152C, T195C, A263G, C269d(=C268d†), C309CC, C317G, C320T, C324G, A326C, C330A, C338d, C343T, C345T, C348G, G16036GG, G16129A, T16189C, T16217C, G16558A
mtDNA_AR_0045	R5a(R5a2)	15	A73G, C150T, T152C, A263G, C309CC, A365d(=A363d†), G16036GG, A16038G, G16129A, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0046	M35b(M35b+16304)	13	A56d, T58A, G71GG, A73G, G143A, T199C, A263G, C315CC, G16036GG, C16192T, C16223T, T16304C, T16519C
mtDNA_AR_0047	T1a(T1a1'3)	12	A73G, T152C, T195C, A263G, G16036GG, T16126C, A16163G, C16186T, T16189C, C16193T, C16294T, T16519C
mtDNA_AR_0048	R5a(R5a2)	14	A73G, C150T, T152C, A263G, T310TTC, T310C, G16036GG, A16038G, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0049	R5a(R5a2a)	9	A73G, T152C, A263G, G16036GG, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0050	H1c(H1c+152)	10	T152C, T199C, A263G, T310C, T310TTC, C324G, A361d(=A357d†), G16036GG, G16319A, C16344T
mtDNA_AR_0051	M30g(M30g)	12	A73G, T195A, T204C, A263G, C308T, C309d(=C303d†), A376d, G389A, T401d(=T398d†), G16036GG, C16223T, T16519C
mtDNA_AR_0052	R5a(R5a2)	15	A73G, C150T, T152C, A263G, T310C, T310TTC, C362d, G16036GG, A16038G, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0053	HV2(HV2)	11	A73G, T152C, A263G, T310C, T310TTC, A361d(=A357d†), G389A, G16036GG, T16217C, T16311C, T16519C
mtDNA_AR_0054	R5a(R5a2)	16	A73G, C150T, T152C, A263G, C309CC, C324G, C362d, G16036GG, A16038G, G16129A, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0056	U2b(U2b)	15	A73G, T146C, A234G, A263G, T310C, T310TTC, A352C, G16036GG, A16051G, T16086C, C16259A, C16267T, C16291T, A16326G, C16353T
mtDNA_AR_0	R5a(R5a2)	15	A73G, C150T, T152C, A263G, T310TTC, T310C, C324G,

057			G16036GG, A16038G, C16266T, T16304C, T16311C, T16356C, A16524G, G16526A
mtDNA_AR_0059	H1f(H1f+16093)	2	G16036GG, T16093C
mtDNA_AR_0060	M71(M71)	12	A73G, T146C, T152C, G207A, A263G, C315CC, G16036GG, G16047GT, C16223T, T16271C, A16399G, T16519C
mtDNA_AR_0061	H2a(H2a2a)	9	A263G, C325d(=C324d†), A361d(=A357d†), A376d, G389A, G16036GG, T16092C, T16381C, A16399G
mtDNA_AR_0062	HV2(HV2)	10	A73G, T152C, A210G, A263G, C309CC, C325d(=C324d†), G389A, G16036GG, T16217C, T16519C
mtDNA_AR_0063	H2a(H2a2a)	15	A263G, T310C, T310TTC, C325d(=C324d†), C330G, A339d, A361d(=A357d†), C371T, T372A, A376d, G389A, G16036GG, T16092C, T16381C, A16399G
mtDNA_AR_0064	U2b(U2b)	13	A73G, T146C, A234G, A263G, C315CC, G389A, G16036GG, A16051G, T16086C, G16129A, C16291T, C16353T, T16519C
mtDNA_AR_0065	U2b(U2b2)	14	A73G, T146C, T152C, A234G, A263G, C315CC, C381G, G16036GG, A16051G, T16209C, C16239T, T16352C, C16353T, T16519C
mtDNA_AR_0066	M30(M30)	10	A73G, T195A, A263G, T310TTC, T310C, C325d(=C324d†), G16036GG, G16145A, C16223T, T16519C
mtDNA_AR_0067	D4e(D4e4a1)	11	A73G, A263G, T310TTC, T310C, C325d(=C324d†), G16036GG, C16223T, C16256T, T16311C, T16362C, T16519C
mtDNA_AR_0068	M30c(M30c1)	12	A73G, T146C, T195A, A263G, C309d(=C303d†), C324G, G16036GG, T16093C, A16146G, A16166d(=A16162d†), C16223T, T16519C
mtDNA_AR_0069	M33a(M33a3)	10	A73G, T146C, A263G, C315CC, G389A, G16036GG, G16129A, T16172C, C16223T, T16519C
mtDNA_AR_0070	M3a(M3a1+204)	11	A73G, T146C, T204C, A263G, C315CC, A361d(=A357d†), G16036GG, T16126C, A16220G, C16223T, T16519C
mtDNA_AR_0071	M4a(M4a)	12	A73G, A263G, C315CC, A365d(=A363d†), C381G, G16036GG, G16145A, C16176T, C16223T, C16261T, T16311C, T16519C
mtDNA_AR_0072	M52a(M52a1)	13	A73G, A189G, T195C, G207A, A263G, T310C, G16036GG, C16223T, A16253G, A16275G, C16327A, G16390A, T16519C
mtDNA_AR_0074	H6(H6)	11	T239C, A263G, T310C, C311CTCC, A326d, C362A, A365AA, G16036GG, G16153A, T16362C, A16482G
mtDNA_AR_0075	U7a(U7a3b)	10	A73G, C151T, T152C, A263G, C315CC, G16036GG, A16207G, A16309G, A16318T, T16519C
mtDNA_AR_0076	M30b(M30b)	23	T55A, A73G, T125A, T152C, T195A, T206G, A263G, C309T, T310C, C325d(=C324d†), C343T, A352C, G16036GG, T16189C, C16192CT, C16223T, C16278T, G16384A, T16413G, G16434A, T16484G, T16519C, G16558A
mtDNA_AR_0077	U7a(U7a)	11	A73G, C151T, T152C, A263G, T292C, C315CC, A361d(=A357d†), G16036GG, C16169T, A16318T, T16519C
mtDNA_AR_0078	U2b(U2b)	17	A73G, T146C, A234G, A263G, C315CC, C324G, A339d, A352d(=A350d†), A365d(=A363d†), G16036GG, A16051G, T16086C, C16259A, C16267T, C16291T, A16326G, C16353T
mtDNA_AR_0079	J1b(J1b5)	13	A73G, A263G, C295T, T310TTC, T310C, A361d(=A357d†), G16036GG, C16069T, T16126C, G16145A, C16222T, C16261T, G16274A
mtDNA_AR_0080	M18(M18'38)	18	A73G, A93G, C194T, T217C, T246C, A263G, T310C, T310TTC, C325d(=C324d†), A365d(=A363d†), A376d, C381G, G16036GG, T16093C, C16134T, C16223T, A16318C, T16519C
mtDNA_AR_AR0081	U7(U7)	10	16210C 16222T 16223T 16228T 16261T 73G 146C 204C 242T 263G 295T 309.1C 315.1C 462T 489C 584d 589d 592d 595d 599d 600d 602.1T 603T 606C 607T 609G
mtDNA_AR_0082	M2a(M2a1a)	20	A73G, T195C, T204C, A263G, C309CC, G316C, C317G, C320T, C324G, A326C, C330G, C338d, C343T, C348G, G16036GG, C16223T, C16270T, G16319A, T16352C, T16519C
mtDNA_AR_0083	U2e(U2e2)	4	G16036GG, A16051G, T16092C, G16129C
mtDNA_AR_0	H1b(H1ba)	6	A263G, C315CC, A365d(=A363d†), G16036GG, C16270T,

084			T16519C
mtDNA_AR_0085	U2(U2+152)	12	A73G, C150T, T152C, A263G, C315CC, A361d(=A357d†), G389A, G16036GG, G16049GG, A16051G, A16247G, A16254G
mtDNA_AR_0086	T1a(T1a1'3)	12	A73G, T152C, T195C, A263G, C315CC, G16036GG, T16093C, T16126C, A16163G, C16186T, T16189C, C16294T
mtDNA_AR_0087	U2b(U2b)	12	A73G, T146C, A234G, A263G, C315CC, A376d, G16036GG, A16051G, T16086C, G16129A, C16353T, T16519C
mtDNA_AR_0088	R30a(R30a1b1)	7	A73G, A263G, C315CC, A376d, G16036GG, T16209C, C16256T
mtDNA_AR_0089	M18(M18'38)	18	A73G, A93G, C194T, T217C, T246C, A263G, T310C, T310TTC, C324G, C362d, A376d, G389A, G16036GG, T16093C, C16134T, C16223T, A16318C, T16519C
mtDNA_AR_0090	HV2(HV2)	9	A73G, T152C, A263G, T310TTC, T310C, G16036GG, T16217C, T16311C, T16519C
mtDNA_AR_0091	M30(M30+16234)	9	A73G, T195A, A263G, T310TTC, T310C, G16036GG, C16223T, C16234T, T16519C
mtDNA_AR_0093	M30g(M30g)	10	A73G, T195A, T204C, A263G, C308T, C309d(=C303d†), A357C, G16036GG, C16223T, T16519C
mtDNA_AR_0095	U2e(U2e)	50	G16023GG, G16036GG, G16049GG, A16051G, G16129C, A16183C, T16189C, C16193CC, T16195G, C16197T, C16205A, T16209A, C16214A, T16224C, C16228A, T16229A, A16235C, C16242T, T16243G, C16245G, A16258C, A16265C, A16269C, T16271A, T16276A, C16282A, T16288A, A16293C, C16301A, T16304A, T16308A, T16311A, A16322T, T16330A, T16334A, A16340T, C16344A, T16347C, C16358T, G16361C, T16362C, T16368G, T16372G, T16381C, G16384A, T16386A, T16392C, T16409C, T16413G, T16422A

Table 2. Haplogroups, their sample size and occurrence in Geographical areas of world.

Sr. #	Haplogroups	Frequency (%)	Corresponding Geographic Region
1	R5a(R5a2)	7%	Indian sub cont
2	M30g(M30g)	5%	South Asia
3	U2b(U2b)	5%	Eurasian, south Asian
4	T1a(T1a1'3)	4.8%	Central and Northeast Europe
5	U2b(U2b2)	3.0%	South Asian
6	M34a(M34a1)	2.0%	South Asian
7	M18(M18'38)	2.0%	South Asian

Table 3. Genetic Properties of Haplogroups.

Total Number of samples	83
Total number of Haplotypes	83
Unique haplotypes	83
Shared haplotypes	0
Genetic Diversity	0.9916
Random Match Probability	0.0083
Power of discrimination	0.9917

Table 4: Comparative Analysis of mtDNA sequence and genetic findings of Other Ethnic Group Studies with Arain Population (current study).

Populations	Wakhi (Current study)	Makrani [1]	Saraiki [2]	Pathan [4]	Four KPK Tribes[19]	Sindhi [20]	Punjabi [15]	Kashmiri [21]	Hazara [23]	Kho (Chitral) [22]	Gujar (HVR1 & 2) [24]	Arain (This study)
Sample size	40	100	85	230	100	88	100	317	319	16	73	83
Random match probability	0.026	0.0408	0.0542	0.0065		0.018	0.0085	0.0054	0.0085		0.0903	0.0083
Power of discrimination	0.974	0.9592	0.9458	0.9978		0.981	0.8819	0.7918	0.9915	0.202	0.9097	0.9917
Average Genetic diversity	0.998	0.9688	0.957	0.993	0.945	0.992	0.9633	0.9977	0.9945	0.215	0.9223	0.9916

Table 7: Percentage occurrence of the Mutations found.

No.	Mutations	Percentage occurrence
1	Transversion	17.23
2	Transition	64.55
3	Insertion	11.6
4	Deletion	6.6

Discussion

Populace investigation of Arains have shown extraordinary variety in correlation with the other revealed populace of Pakistan like Mughal and Kalash (Nesheva, 2014). Unique haplotype distribution is the cause of genetic diversity. The current population under study had 83 distinct haplotypes, which were found to be somewhat consistent with Burusho 78%, Hazara 76%, Makrani 76%, Baluchi 69%, and Brahui 68% among Pakistan's other reported populations, but moderately lower than Saraiki 92%, Sindhi 90%, and Pathan 81%. The population of Gujars showed a high percentage of South Asian ancestry (42%). The extent of South Asian heredities in the other revealed Pakistani populaces were 48% in Sindhi, 39.1% in Pathan, 36% Pashtun, 29.4% in Saraiki and 24% in Makrani (Metspalu et al., 2004). Low recurrence of South Asian genealogies among the significant ethnic gatherings of Afghanistan have additionally been accounted for with the commonness of 15% in Hazara, 13.3% in Baluch and 7.1% in Pashtun, while missing in Tajik (Whale, 2012). The presence of mtDNA haplogroups from south Asia, specifically the Indian subcontinent, in the current study population demonstrated that the actual inhabitants of this region have been reshaped by local demographic events in the past. South Asian haplogroup was the second most predominant haplogroup representing (4.8%) in the people of the current review populace. Its recurrence among the Pathans of Pakistan was accounted for 55% and 26% in Makranis. In addition, the West Eurasian haplogroup was found in Indian Punjabis (40-50%), Kashmiris (30%) and Gujrathis (30%), and in Indian Uttar Pradesh and West Bengal (less than 1%). According to Ayub et al., the major ethnic groups of Afghanistan also had a higher frequency of West Eurasian lineages, with frequencies of 40% for the Hazara, 89% for the Tajik, 74% for the Baluch, and 64% for the Pashtun. 2015). The presence of these lineages demonstrated that, prior to Alexander, the Arabians, Muslims, and the British's invasion, gene flow to this region may have come from the north through Central Asia or from the west through Iran. According to Akbar et al., the megahaplogroups R, M, and N that have been found in the Gujars population are thought to have originated in South Asia and date back between 60.000 and 75.000 years (Akbar et al., 2014), pointing to a South Asian origin for their maternal gene pool.

Genetic characterization of different populations of Pakistan is done for determining the frequency of various haplogroup among those populations. Pathans After studying the Pathan population’s mtDNA control region the macrohaplotype magnitude was 8.7% M3, 10.4% HV, 11.3% U7, 30.09% M, 7.8% N, and 61.3 R. The results showed the ancestral relation of Pathans with South Asia, East Asia and West Eurasia (Rakha et al., 2011). According to the Makrani people's mtDNA analysis, they have African haplotypes such L1a, L2a, L3b, and L3d. (Siddiqi et al., 2015). Saraiki In Saraiki people, haplotypes of South Asia are most important (Hayat et al., 2015). Because haplotype U4 is present in Kalashi people 34% more frequently than in other populations, Kalash Studies have shown that Western Eurasian haplogroups are found to be predominant in this population (Ayub et al., 2015). Studies on Gujars in Khyber Pakhtunkhwa (KPKSwat)'s district revealed that R, M, and N mega haplotypes predominate with 48%, 45%, and 7% of the total population, respectively (Ullah et al., 2017).

Conclusion

After examination with the rCRS, noticed transformations in populace tests under review were seen 723 changes (64.55%) 193 transversions (17.23%), 130 additions (11.6%) and 74 cancellations (6.6%). From 100 samples with 1120 polymorphic sites, 83 distinct haplotypes were identified in 83 positive samples. Each of the 83 individuals is unique, and neither one nor more than one person

shared any of the haplotypes. The haplotypes R5a(R5a2), which account for 7% of the population, were the most frequently observed in this population. In this study, the South Asian haplogroups clearly dominate, with M30g (M30g) accounting for 5%, U2b (U2b) accounting for 5% and 3%, M34a (M34a1) accounting for 2%, and M18 (M18'38) accounting for 2%. The genetic diversity of this Arain population's mtDNA control region is compared to that of other Pakistani ethnic groups that have been studied. It is found that among every one of the examinations ethnic gathering Pathans populace was the most expanded and the Kalash populace was least broadened having haplotype variety 0.993 and 0.851 separately (Rakha et al., 2011). Here, it is tracked down that R5a(R5a2) Haplogroups have clear predominance over other haplogroups. According to Sounier et al., heterogeneous mitochondrial DNA haplogroups may be the cause of the Arain population's distinct geographic location. 2009). Biochemical comparisons, genetics, population studies, and forensic sciences will all benefit from this investigation into the analysis of mitochondrial DNA.

Acknowledgments: Financial assistance provided by the Higher Education Commission (HEC indigenous section) Islamabad is gratefully acknowledged.

Statements & Declarations

Funding

Financial assistance provided by the Higher Education Commission (HEC indigenous section) Islamabad is gratefully acknowledged.

Competing Interests

"The authors have no relevant financial or non-financial interests to disclose."

Non-financial interests: Author B and C have served as supervisor of the project for university of central Punjab Lahore.

Author Contributions

"Auther A and C contributed to the study conception and design. Material preparation, data collection and analysis were performed by Farah Ahmad, DR. M. Afzal and Dr. M. Saqib Shahzad. The first draft of the manuscript was written by Farah Ahmad and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript."

Conceptualization: Dr. Muhammad Saqib Shahzad Methodology: Dr. Muhammad Saqib Shahzad, …; Formal analysis and investigation: Farah Ahmad …; Writing - original draft preparation: Farah Ahmad. Supervision: Dr. Muhammad Afzal and Dr. Muhammad Saqib Shahzad.

Ethics approval

Authors of research involving human or animal subjects should include a statement that confirms that the study was approved (or granted exemption) by the appropriate institutional and/or national research ethics committee (including the name of the ethics committee and reference number, if available). For research involving animals, their data or biological material, authors should supply detailed information on the ethical treatment of their animals in their submission. If a study was granted exemption or did not require ethics approval, this should also be detailed in the manuscript.

"This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Human Ethics Committee of University of Central Punjab,

Consent to participate

"Informed consent was obtained from all individual participants included in the study and a signed copy was saved for future reference. must be obtained from participants.*"*

Consent to publish

Individuals may consent to participate in a study, but object to having their data published in a journal article. If your manuscript contains any individual person's data in any form (including any individual details, images or videos), consent for publication must be obtained from that person, or

in the case of children, their parent or legal guardian. This is in particular applicable to case studies. A statement confirming that consent to publish has been received from all participants should appear in the manuscript.

Availability of data and materials

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. As the publications of the raw data are related with the condition of Article publication before it is submitted to some public forum or some DNA databases.

Reference list

1. Stanley A. Wolpert 2000. A new History of India. Oxford University Press, New York.
2. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., ... & Mehdi, S. Q. (2002). Y-chromosomal DNA variation in Pakistan. The American Journal of Human Genetics, 70(5), 1107-1124.
3. Rozen, S., & Skaletsky, H. (1999). Primer3 on the WWW for general users and for biologist programmers. Bioinformatics methods and protocols, 365-386.
4. Ayub, Q., & Tyler-Smith, C. (2009). Genetic variation in South Asia: assessing the influences of geography, language and ethnicity for understanding history and disease risk. Briefings in Functional Genomics and Proteomics, 8(5), 395-404.
5. Sutherland, W. J. (2003). Parallel extinction risk and global distribution of languages and species. Nature, 423(6937), 276-279.
6. Nesheva D. Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. science, (2014); 1(5): 5-14.
7. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human mutation, (2009); 30(2): E386-E394.
8. The strongest argument in favour of this claim is that Arains are always Muslims and almost entirely Sunnis as were the early Arabs who came with Muhammad bin Qasim.", There is many a slip betwixt cup and lip, Ishtiaq Ahmed, Daily Times, Pakistan, 18/04/2006.
9. at http://www.joshuaproject.net/index.php
10. "On the other hand, agricultural castes of Jats, Gujjars and Rajputs are divided into Muslims, Hindus and Sikhs.", There is many a slip betwixt cup and lip, Ishtiaq Ahmed, Daily Times, Pakistan, 18/04/2006.
11. The Arab-origin claim can be simply a re-orientation towards a (perceived) greater status as conquerors and “original” Muslims.", There is many a slip betwixt cup and lip, Ishtiaq Ahmed, Daily Times, Pakistan, 18/04/2006.
12. In our study we found blood group B is the most frequent blood group in Arain (40.5%), and blood group O is less common (24.7%) while it is the most frequent blood group among Awans, Rajputs, miscellaneous Punjabi sub casts, Balochs, Sindhis, and Pathans..." (http://www.pjms.com.pk/issues/janmar05/article/article6.html).
13. There is no statistically significant difference of ABO and Rh-D distribution among various ethnic groups and casts, O is the most common blood group except in Arains where B is the most common and O is less common."
    (http://www.pjms.com.pk/issues/janmar05/article/article6.html).
    (http://www.pjms.com.pk/issues/janmar05/article/article6.html
14. Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. De Bruijn, A. R. Coulson, and J. Drouin. 1981. Sequence and organization of the human mitochondrial genome. Nature. 290:457-465.
15. Andrew, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, N. Howell. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999; 23:147.
16. Fisher, M. C., & Garner, T. W. (2020). Chytrid fungi and global amphibian declines. Nature Reviews Microbiology, 18(6), 332-343.

17. Nesheva, D. V. (2014). Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. Balkan journal of medical genetics, 17(1), 5-14.
18. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., ... & 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), 1-25.
19. Whale, J. W. (2012). Mitochondrial DNA analysis of four ethnic groups of Afghanistan (Doctoral dissertation, University of Portsmouth).
20. Ayub, Q., M. Massimo, L. Pagani, M. Haber, A. Mohyuddin and K. Shagufta. 2015. The Kalash genetic isolate: ancient divergence, drift, and selection. Am. J. Hum. Genet. 96(5):775-783.
21. Akbar, N., Ahmad, H., Nadeem, M. S., Ali, N., & Saadiq, M. (2015). An efficient procedure for DNA isolation and profiling of the hyper variable MtDNA sequences. Journal of Life Sciences, 9, 530-534.
22. Rakha, A., K. J. Shin, J. A. Yoon, N. Y. Kim, M. H. Siddique and I. S. Yang. 2011. Forensic and genetic characterization of mtDNA from Pathans of Pakistan. Int. J. Leg. Med. 125(6):841-848.
23. Siddiqi, M. H., T. Akhtar, A. Rakha, G. Abbas, A. Ali and N. Haider. 2015. Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. Leg. Med. 17(2):134-139.
24. Hayat, S., T. Akhtar, M. H. Siddiqi, A. Rakha, N. Haider and M. Tayyab. 2015. Mitochondrial DNA control region sequences study in Saraiki population from Pakistan. Legal Medicine. 17(2): 140-144.
25. Ullah, I., H. Ahmad, B. E. Hemphill, M. S. Nadeem, M. Tariq and S. Tabassum. 2017. Mitochondrial genetic characterization of Gujar population living in the Northwest areas of Pakistan. Advancements Lif. Sci. 4(3): 84-91.
26. Ayala, I., Sounier, R., Usé, N., Gans, P., & Boisbouvier, J. (2009). An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. Journal of biomolecular NMR, 43, 111-119.
27. Yang, In Seok, Lee, Hwan Young, Yang, Woo Ick, & Shin, Kyoung-Jin. (2013). mtDNAprofiler: a Web application for the nomenclature and comparison of human mitochondrial DNA sequences. *Journal of forensic sciences*, *58*(4), 972-980.