REVIEW

Annals of
human genetics WILEY

A comprehensive review of HVS-I mitochondrial DNA variation of 19 Iranian populations

Motahareh Amjadi[1,5] 🆔 | **Zahra Hayatmehr**[2] | **Balázs Egyed**[3] |
Mahmood Tavallaei[4] | **Anna Szécsényi-Nagy**[5]

[1]Department of Genetics, ELTE Doctoral School of Biology, Eötvös Loránd University, Budapest, Hungary

[2]Faculty of Management and Financial Science, Department of Management, Khatam University, Tehran, Iran

[3]Department of Genetics, ELTE Eötvös Loránd University, Budapest, Hungary

[4]Kawsar Human Genetic Research Center, Tehran, Iran

[5]HUN-REN Research Centre for the Humanities, Institute of Archaeogenomics, Budapest, Hungary

Correspondence
Motahareh Amjadi, Department of Genetics, ELTE Doctoral School of Biology, Eötvös Loránd University, 1117 Budapest, Hungary. Email: motahareh.amjadi@abtk.hu

Funding information
Tempus public foundation

Abstract

Iran is located along the Central Asian corridor, a natural artery that has served as a cross-continental route since the first anatomically modern human populations migrated out of Africa. We compiled and reanalyzed the HVS-I (hypervariable segment-I) of 3840 mitochondrial DNA (mtDNA) sequences from 19 Iranian populations and from 26 groups from adjacent countries to give a comprehensive review of the maternal genetic variation and investigate the impact of historical events and cultural factors on the maternal genetic structure of modern Iranians. We conclude that Iranians have a high level of genetic diversity. Thirty-six haplogroups were observed in Iran's populations, and most of them belong to widespread West-Eurasian haplogroups, such as H, HV, J, N, T, and U. In contrast, the predominant haplogroups observed in most of the adjacent countries studied here are H, M, D, R, U, and C haplogroups. Using principal component analysis, clustering, and genetic distance-based calculations, we estimated moderate genetic relationships between Iranian and other Eurasian groups. Further, analyses of molecular variance and comparing geographic and genetic structures indicate that mtDNA HVS-I sequence diversity does not exhibit any sharp geographic structure in the country. Barring a few from some culturally distinct and naturally separated minorities, most Iranian populations have a homogenous maternal genetic structure.

KEYWORDS
genetic diversity, hypervariable segment, Iran, mitochondrial DNA, population genetics

1 | INTRODUCTION

Over the past two decades, various molecular markers have been employed to elucidate the origins, admixture, and migratory patterns of human populations. Mitochondrial DNA (mtDNA) serves as an indispensable haploid marker for tracing maternal lineage history (Galtier et al., 2009).

Within the mtDNA, the D-loop region houses the hypervariable segment (HVS) which exhibits a mutation rate significantly higher than other segments of the mitogenome. Notably, the HVS-I within this region, containing critical haplogroup-defining loci, has been extensively researched (Van Oven & Kayser, 2009). Our study aims to provide an exhaustive review of the extant data concerning mtDNA

FIGURE 1 Geography of the Iranian Plateau: The Elburz Mountains, including Mount Demavend, the Turkmen-Khorasan Mountains, like the Kopeh Dagh, and the Hindu Kush, are located in the northern arc. In contrast, the Zagros Mountains form the southern and southwestern arches. Several mountain ranges alternate with plateaus and large intermountain-closed depressions inside the region, like the Dasht-e Kavir and the Dasht-e Lut. *Source*: The map has been created with the genetic mapping tools http://generic.mapping.tools.org.

variation among the populations residing in the Iranian Plateau, a region encompassed within Central Asia. Geographically, Central Asia runs from the Caspian Sea and extends eastward to China and Mongolia, demarcated by Afghanistan and Iran in the south and bordered by Russia in the north. For the purpose of this study, nations, such as Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan, are also integrated into our definition of Central Asia.

From the early dispersals of anatomically modern humans from Africa, the Iranian territory has occupied a strategic position within the Central Asian corridor. This corridor, serving as a tri-continental juncture, has historically facilitated human migrations (Vallini et al., 2022). The Iranian Plateau (Figure 1) is a prominent geographical and geological expanse within Central Eurasia, and its internal features are distinctly demarcated by surrounding mountain arcs (Bell, 1832).

The Iranian population displays significant ethnic and linguistic heterogeneity, with an estimated population of approximately 83 million, positioning them among the top decile of nations globally in terms of ethnic diversity. The predominant ethnic group in Iran are the Persians (Fars), constituting approximately 51% of the total population. Based on the study by Rashidvash (2012), Iranian Azeris represent 24% of the population, succeeded by Iranian Gilakis and Mazandaranis (8%), Iranian Kurds (7%), Iranian Arabs (3%), Iranian Lurs (2%), Iranian Balochs (2%), and Iranian Turkmens (2%). Additional ethnic minorities cumulatively account for roughly 1% of the population (Figure 2, Rashidvash, 2012).

This figure was designed via Adobe Photoshop 2022 V.23.3. Operating system windows 11 64-bit. The base map originates from https://www.cleanpng.com/free/iran-map.html.

The investigation of genetic variations in Iranian populations, particularly from a matrilineal perspective, has hitherto remained somewhat under-explored. Although the HVS-I region of the mtDNA has been examined (Farjadian et al., 2011), there exist two foundational studies that focus on the entire mtDNA genomes of the Iranian population (Bahmanimehr et al., 2015; Derenko et al., 2013). However, these seminal works did not encompass all the Iranian samples previously scrutinized. The present research endeavors to discern the relationships between HVS-I mtDNA gene pools within the modern Iranian genetic framework and those of neighboring Eurasian nations. This is achieved by an integrative reanalysis of extant data on contemporary Iranian cohorts. To facilitate interpopulation comparisons, HVS-I nucleotide sequences from regional clusters of proximate territories were incorporated. The study seeks to address the subsequent inquiries: (1) To what extent do genetic similarities pervade among Iranian populations across diverse geographical regions? (2) How have the mtDNA gene pools of neighboring populations influenced the Iranian genetic spectrum in the context of transboundary migration within the Iranian Plateau? (3) Are Iranian minor ethnic groups genetically distinct from the rest of the country's population due to specific cultural practices such as endogamy?

2 | MATERIALS AND METHODS

2.1 | Study populations

HVS-I nucleotide sequences from a collective sample of 3840 individuals were amassed, representing 19 Iranian populations ($n = 1498$) and 26 populations from neighboring countries ($n = 2342$) (Figure 3A,B). These sequences were either retrieved from the GenBank database or procured directly from the authors of pertinent studies (Table 1). In this analysis, the populations of three major cities, namely, Tehran, Isfahan, and Shiraz, were treated as distinct entities. This delineation was necessitated due to the absence of detailed ethnic categorizations in available datasets for these urban centers and the recognition of their status as some of Iran's most populous cities, encompassing a diverse spectrum of ethnic affiliations. Although this exhaustive review primarily centered on the extensively studied mtDNA HVS-I, it is noteworthy that certain Eurasian populations, currently devoid of published HVS-I mtDNA data, could not be integrated into this evaluation.

TABLE 1 Description of the studied populations in Iran and in the comparative regions.

Population	Abbreviation	Sample size	References
Iranian Arabs	IR_Ab	91	Farjadian et al. (2011), Nasidze et al. (2008), MZ416977-MZ4116987
Iranian Armenians	IR_Ar	62	Derenko et al. (2013), Farjadian et al. (2011)
Iranian Azeris	IR_Az	75	Derenko et al. (2013), Farjadian et al. (2011), MZ16988-MZ16998
Iranian Bakhtiaris	IR_BA	53	Nasidze et al. (2008)
Iranian Balochs	IR_BL	61	Farjadian et al. (2011)
Iranian Gilakis	IR_GL	75	Derenko et al. (2013), Farjadian et al. (2011), Nasidze et al. (2006), MZ16999-MZ17009
Iranians in Isfahan	IR_IS	44	Nasidze et al. (2004)
Iranian Jews	IR_Je	56	Farjadian et al. (2011)
Iranian Kurds	IR_Ku	80	Derenko et al. (2013), Farjadian et al. (2011), Nasidze et al. (2005), MZ17010-MZ17020
Iranian Lurs	IR_Lu	83	Derenko et al. (2013), Farjadian et al. (2011), Quintana-Murci et a. et al. (2004)
Iranian Mazanderanis	IR_Ma	102	Derenko et al. (2013), Farjadian et al. (2011), Nasidze et al. (2006), MZ17200-MZ17222
Iranian Persians (Fars)	IR_P	179	Derenko et al. (2013)
Iranian Qashqai	IR_QA	161	Derenko et al. (2013), Farjadian et al. (2011)
Iranians in Qeshm	IR_QE	55	Farjadian et al. (2011)
Iranians in Shiraz	IR_SH	58	Farjadian et al. (2011)
Iranian Southern Talyshes	IR_ST	47	Nasidze et al. (2009)
Iranians in Tehran	IR_TE	108	Nasidze et al. (2004), Schönberg et al. (2011)
Iranian Turkmens	IR_TU	59	Derenko et al. (2013), Farjadian et al. (2011)
Iranian Zoroastrians	IR_ZO	49	Farjadian et al. (2011)
Armenians from Armenia	A_O_I	53	Margaryan et al. (2017), Schönberg et al. (2011)
Altai-Kizhi from Russia	AK	77	Derenko et al. (2013)
Ashkenazi Jews	Ash	59	Costa et al. (2013)
Azeris from Azerbaijan	AZ_O_I	71	Nasidze et al. (2004), Schönberg et al. (2011)
Balochis from Afghanistan	Bl_O_I	50	Lippold et al. (2014)
Brahuis from Pakistan	Br	50	Lippold et al. (2014)
Dravidians from India	Drav	41	Derenko et al. (2013), Sylvester, C. et al. (2019)
Georgians from Georgia	Gr	85	Schönberg et al. (2011), Nasidze and Stoneking (2001)
Huis from China	Hui	42	Yao et al. (2004)
Kalash from Pakistan	Kal	143	Lippold et al. (2014), Siddiq et al. (2021)
Karakalpaks from Uzbekistan	Kar	55	Marchi et al. (2017), Irwin et al. (2010)
Kazakhs from Kazakhstan	Kaz	251	Marchi et al. (2017)

TABLE 1 (Continued)

Population	Abbreviation	Sample size	References
Kyrgyz from Kyrgyzstan	Kyr	244	Marchi et al. (2017)
Mongolians from Mongolia	Mon	85	Derenko et al. (2013), Marchi et al. (2017)
Northern Talyshes from Azerbaijan	NT	77	Nasidze et al. (2009)
Pakistanis from Pakistan	Pak	50	Lippold et al. (2014)
Pathans from Afghanistan	Path	51	Lippold et al. (2014)
Shors from Russia	Sh	39	Marchi et al. (2017)
Sindhis from Pakistan	Sid	44	Lippold et al. (2014)
Turks from Turkey	Tur	106	Nasidze et al. (2004), Schönberg et al. (2011)
Turkmens from Turkmenistan	T_O_I	51	Marchi et al. (2017)
Tajiks from Tajikistan	TJK	67	Chen et al. (2020), Marchi et al. (2017)
Uzbeks from Uzbekistan	Uzb	319	Irwin et al. (2010), Marchi et al. (2017)
Yaghnobis from Tajikistan	Yab	68	Cilli et al. (2019)
Yunnans from China	Yuna	59	Yao and Zhang (2002)
Zoroastrians from India	Z_O_I	105	Morawala-Patell et al. (2021)

2.2 | Genetic analyses

2.2.1 | Alignment of HVS-I sequences

Sequences obtained were aligned employing the UGENE software, version 40.1, with the HVS-I sequences referenced to the Revised Cambridge Reference Sequence (rCRS, Anderson et al., 1981). Subsequent to this, a consistent nucleotide range was established for all datasets, designating the HVS-I region (16,046–16,381 bp). This region was then trimmed using a script written in Python, version 3.10. Notably, all variations in the polycytosine tract within the range 16,184–16,193 bp were excised. Mitochondrial haplogroup determinations were performed by HaploGrep 2 (Weissensteiner, Pacher et al., 2016), which uses Phylotree mtDNA tree Build 17 (Van Oven, 2015; Weissensteiner, Forer et al., 2016).

To categorize aligned FASTA files by individual populations, facilitating the generation of the pertinent Arlequin file (.arp), the FaBo online toolbox was employed, specifically leveraging the "DNA to haplotype collapse and converter" function (Villesen, 2007).

2.2.2 | Statistical analysis

The number of polymorphic (segregating) sites (S), haplotype (gene) diversity (Hd), nucleotide diversity (pi), number of haplotypes (H), and the mean number of differences between all pairs of haplotypes (π) as fundamental indices of the molecular diversity comparisons were calculated by Arlequin v3.5.2.2 software from prepared (.arp) for each population. Based on HVS-I sequences (np 16,046–16,381) from the 45 populations, F_{ST} values were calculated in Arlequin (Excoffier & Lischer, 2010). The Tamura and Nei substitution model (Tamura & Nei, 1993) was used, with a gamma value of 0.325 inferred from the software FindModel based on PAML likelihoods. Ten thousand permutations were used to test significant variations in F_{ST} values (Posada & Crandall, 1998).

First, to conduct a principal component analysis (PCA), the mtDNA haplogroups of all aligned HVS-I FASTA files were checked with HaploGrep 2 V.2.4.0 compared to PhyloTree build 17 (Van Oven, 2015). These haplogroups stand for the primary 36 groups as A, B, C, D, F, G, H, HV, I, J, J1, J2, JT, K, L, M, N, P, R*, R0, R2, T, T1, T2, U*, U1, U2,

FIGURE 2 Iranian ethnic population structure with language information. NA, not applicable; Pop, proportion of the population compared to the whole Iranian population. The language family/language family sublevel/language (dialect) attributions are in green.

U3, U4, U5, U6, U7, W, X, Y, and Z (haplogroup frequencies tables are available in Supplementary Table, sheets 1 and 2). Then, the relative frequencies of all haplogroups for the whole 45 populations were calculated. The PCA scores were calculated from the relative frequencies via the *prcomp* function and for categorical PCA in the R v4.1.3 package.

To measure and to adjoin to certain PCAs, we performed hierarchical clustering analyses employing the Ward type (Ward, 1963) algorithm and the Euclidean similarity measurement method. The first six PCAs were used for the clustering. The Ward clustering results were visualized in R as a dendrogram using the hclust function with Euclidean distances. The significance of each cluster was evaluated by 10,000 bootstrap replicates using the pvclust function in R. Pairwise F_{ST} value and Slatkin's distance matrices were acquired in Arlequin software for the multidimensional scaling (MDS) (Excoffier & Lischer, 2010). Based on PAML likelihoods, the evolution model of Tamura and Nei (1993) with a gamma value of 0.325 was picked by the FindModel program (Posada & Crandall, 1998). Finally, using the metaMDS function based on Euclidean distances implemented in the vegan library of R, MDS was performed to a matrix of linearized Slatkin F_{ST}

values (Slatkin, 1995) and visualized in a two-dimensional space. The standard analysis of the molecular variance (AMOVA) function implemented in Arlequin was used to calculate variances, fixation indices, and *p* values (*p*). Ten thousand permutations were used to test the importance of F_{ST} values. The distance matrix was used to infer the haplotype definition. AMOVA was run for each assembly after dividing the groups into two or three clusters. The highest among-groups and the lowest within-groups variances, F_{CT} and F_{SC} values were selected as the best models for our AMOVA analysis. A Mantel test was conducted to determine any potential correlations between the genetic and geographic distances of the samples under study using an F_{ST} genetic distance matrix (Mantel, 1967).

3 | RESULTS AND DISCUSSION

3.1 | Genetic diversity of the studied populations

In this review, we have for the first time systematically collated and presented all extant mtDNA (full mitogenome and HVS-I) data from contemporary Iran. Utilizing HVS-I
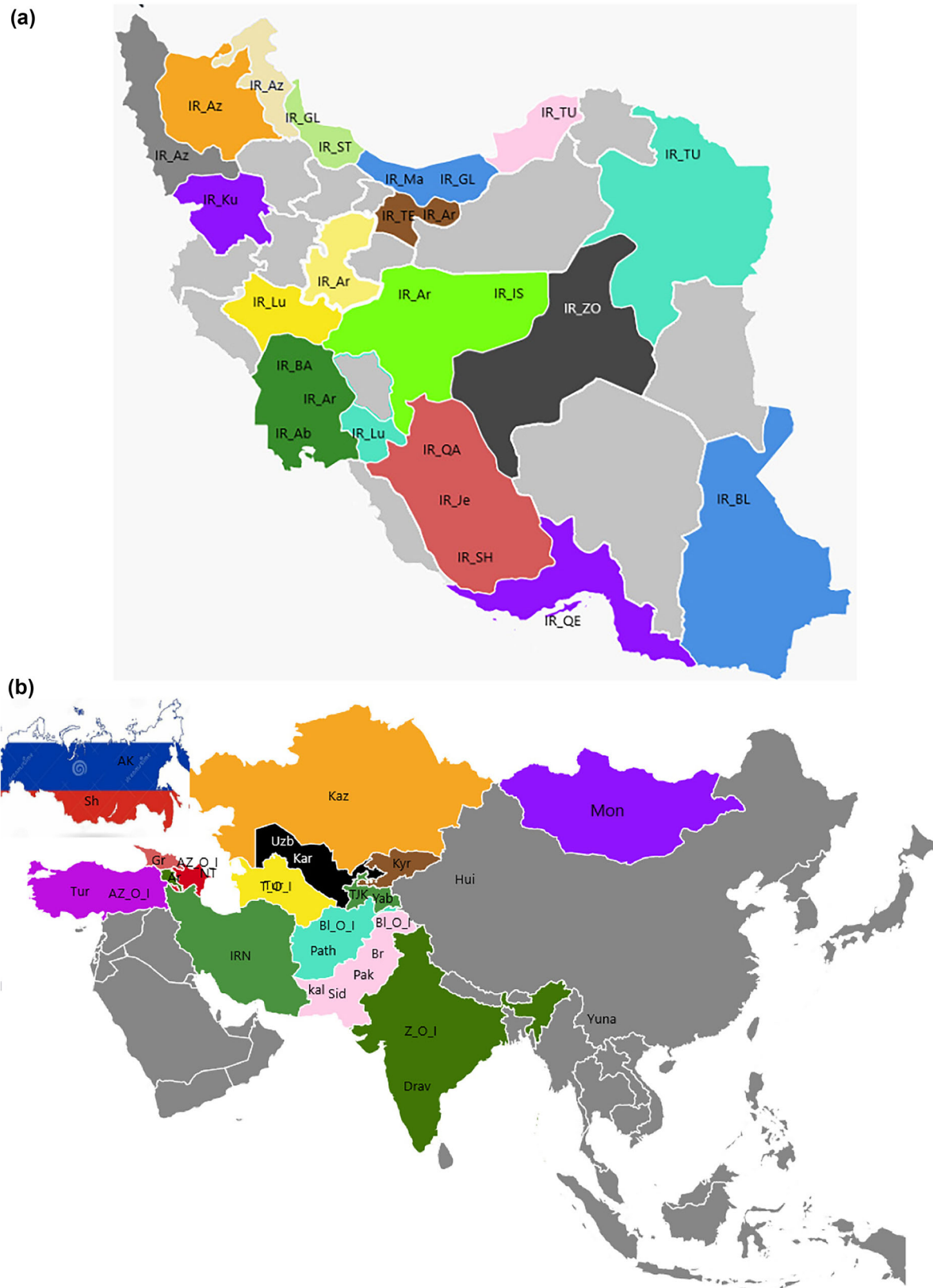
FIGURE 3 (a and b): Geographic distribution of 19 and 26 investigated populations in Iran and Eurasia. The "Persian" and "Ashkenazi" ethnic groups were not specified on the map because they originate from several provinces of Iran and areas of Eurasia. Abbreviations are listed in Table 1. *Source*: The maps were created with the MapChart Tools: https://www.mapchart.net/world.html.

sequences that span a length of 335 bp, specifically covering mtDNA nucleotide positions 16,046–16,381, we identified a total of 2505 distinct haplotypes from 3840 Eurasian individuals. Notably, 1119 of these haplotypes were discerned within the Iranian cohort ($n = 1498$ in total).

In a broader Eurasian perspective, certain populations displayed notably low haplotype diversity: Dravidian (0.2622), Ashkenazi Jews (0.8136), Kalash (0.8227), Iranian Zoroastrians (0.9099), Yunnan (0.9106), Shors (0.9366), Yaghnobi (0.9460), Iranian Jews (0.9435), and Iranian Balochs (0.9508). The range of nucleotide diversity extended from a low of 0.003752 in the Dravidian group to a high of 0.0292 in the Iranian Azeris. Further analysis highlighted that the low-π value observed in Dravidian (1.2195), Ashkenazi (2.5751), Kalash (3.9728), and Yaghnobi (3.7616) populations corroborated a pronounced decline in variability, aligning with observed haplotype diversity metrics in these groups. Conversely, the juxtaposition of low haplotype diversity and elevated π values in populations, such as Iranian Zoroastrians, Jews, and Shors, can be attributed to deep-rooted haplotype genealogies that evolved in consistently sized small populations.

For the encompassing Eurasian dataset, the mean haplotype diversity was calculated at 0.9974, with the observed values spanning from a minimum of 0.2622 in the Dravidian population to a maximum of 1. This maximum value was observed in multiple groups: Northern Talyeshes from Azerbaijan, Bakhtiaris, Southern Talyshes, and the population from Isfahan, Iran (as detailed in Table 2).

3.2 | mtDNA haplogroup distribution

In the present study on the Iranian population, 36 distinct haplogroups were identified. A significant proportion of these haplogroups correspond to the more ubiquitous West-Eurasian lineages, including H, HV, J, N, T, and U.

Farjadian et al. (2011) elucidated that a vast majority (90.8%) of HVS-I lineages predominantly traced their origins to Western Eurasian haplogroups, namely, H, HV, I, J, K, N, T, U, V, and W. Conversely, haplogroups typically associated with East-Eurasian (A, C, D, F, G, and Z), South-Asian (M), and Sub-Saharan African origins (L1, L2, and L3) were confined to specific, smaller ethnic subsets. Complementing these findings, Bahmanimehr et al. (2015) undertook an extensive exploration of the entire mtDNA mitogenome within the Iranian populace, revealing the predominant distribution frequencies of West Eurasian lineages, such as H, J, and U. However, this dominance was not observed within the Baloch and Zoroastrian communities. Our findings resonate with the aforementioned studies, particularly highlighting the prominence

of mtDNA haplogroups H (19.69%), N (9.07%), HV and J1 (7.07%), T (5.88%), U7 (5.07%), and M (5.03%) across the Iranian communities with the Zoroastrian ethnicity as a notable exception.

The summary table of our review shows that four haplogroups B (from the Lurs group), Y (Turkmen), JT (Mazanderanis), and P (Tehran) were represented singularly. In contrast, the remaining 32 identified haplogroups each marked their presence at least two instances within the broader Iranian demographics (Supplementary Table, sheet 1).

Bahmanimehr et al. (2015) have explained that the South Asian haplogroup M is a common ancestor found in all of Iran's neighbors to the east. According to our data, the M haplogroup is most prevalent in the Tehran and Baloch groupings among Iranian communities (Supplementary Table, sheet1). The two East-Eurasian mtDNA haplogroups of G and D as a descendants haplogroup of M can be traceable for the Turkmens of this study in which distinctively sets them apart from other Iranian populations.

Shared maternal haplogroups between the Iranian population and populations of other countries, reflecting geographical trends, extend beyond just Iran's terrestrial regions. Sea routes also appear to have facilitated maternal gene flow. For instance, the African haplogroup component (L) is most prevalent in Qeshm and the Arab populations residing along the Persian Gulf in southern Iran. This prevalence suggests an African influence, potentially channeled through the Strait of Hormuz, a key conduit between the two continents (Figure 4).

Derenko et al. (2013) reported that haplogroup R2 likely dates back to the pre-LGM/Late Glacial era, estimated to have an overall coalescence time of 21-31 kya years ago, and possibly originated in southern Iran. Significantly, this haplogroup is most prominent in the Qashqai and Persian populations, accounting for 3.3% and 1.8%, respectively, among the surveyed Iranian datasets. The investigations by Derenko et al. substantiated that communities from Iran, Anatolia, the Caucasus, and the Arabian Peninsula shared a set of maternal lineages. Even with the aggregated data from both Derenko and Farjadian et al. studies about the Qashqai ethnic group, the observed frequency of the R2 maternal haplogroup remains consistent with Derenko et al.'s report from 2013 (2%), resting at approximately 1.86%.

In the majority of the studied neighboring countries of Iran, the haplogroups H (14.17%), M (10.16%), D (9%), R (6.78%), U (4.56%), and C (4%) are predominant. The East-Eurasian haplogroups C and D are notably prevalent among all Turco-Mongol-speaking groups (except for haplogroup C in Kyrgyz, Turkmen, and Karakalpak). Conversely, the South Asian haplogroup M stands out as the

TABLE 2 Measurements of molecular diversity in the current study.

Population	N^1	S^2	Hd^3	$Hd \pm SD^4$	Pi^5	$Pi \pm SD^6$	H^7	π^8	$\pi \pm SD^9$
Dravidians from India	41	6	0.262	0.0831	0.004	0.003	3	1.2196	0.7923
Ashkenazi Jews from outside of Iran	59	31	0.814	0.0506	0.008	0.005	24	2.5752	1.4008
Kalash from Pakistan	143	25	0.823	0.0153	0.012	0.007	15	3.9729	1.9998
Iranian Zoroastrians	49	36	0.91	0.0231	0.019	0.01	19	6.0596	2.9355
Yunnans from China	59	38	0.911	0.0257	0.017	0.009	27	5.5564	2.7079
Shors from Russia	39	38	0.937	0.0219	0.021	0.011	20	6.7625	3.2566
Yaghnobis from Tajikistan	68	38	0.946	0.0127	0.012	0.007	26	3.7617	1.9203
Iranian Jews	56	33	0.944	0.0149	0.018	0.01	24	5.6936	2.7697
Iranian Balochs	61	50	0.951	0.0206	0.014	0.008	40	4.4028	2.2035
Zoroastrians from India	105	58	0.959	0.0071	0.015	0.008	41	4.7731	2.3517
Iranians in Qeshm	55	57	0.964	0.0153	0.017	0.009	36	5.633	2.7441
Iranian Qashqai	161	74	0.969	0.0082	0.015	0.008	79	4.8542	2.3803
Balochi from Afghanistan	50	50	0.976	0.0127	0.014	0.008	36	4.5894	2.292
Tajiks from Tajikistan	67	49	0.977	0.0071	0.017	0.009	41	5.4492	2.6566
Turkmens from Turkmenistan	51	44	0.98	0.0091	0.016	0.009	36	5.2902	2.5977
Altai-Kizhi from Russia	77	51	0.982	0.0052	0.018	0.01	46	5.6955	2.7593
Iranian Lurs	83	76	0.982	0.0075	0.018	0.01	76	5.7644	2.7869
Iranian Persians (Fars)	179	103	0.986	0.0043	0.016	0.008	121	5.0467	2.4623
Brahuis from Pakistan	50	52	0.986	0.0084	0.016	0.009	39	5.1511	2.5378
Iranian Turkmens	59	62	0.99	0.0056	0.018	0.01	46	5.8323	2.828
Pakistanis from Pakistan	50	59	0.99	0.0071	0.017	0.009	42	5.5315	2.704
Kyrgyz from Kyrgyzstan	244	100	0.991	0.0021	0.018	0.01	147	5.9039	2.8287

(Continues)

TABLE 2 (Continued)

Population	N[1]	S[2]	Hd[3]	Hd ± SD[4]	Pi[5]	Pi ± SD[6]	H[7]	π[8]	π ± SD[9]
Karakalpaks from Uzbekistan	55	58	0.991	0.0072	0.016	0.009	48	5.3293	2.6116
Turks from Turkey	106	74	0.991	0.0038	0.018	0.01	84	5.9762	2.8725
Sindhis from Pakistan	44	59	0.993	0.0064	0.021	0.011	37	6.6438	3.1966
Iranians in Shiraz	58	73	0.995	0.0051	0.017	0.009	52	5.5584	2.7094
Iranian Mazanderanis	102	78	0.995	0.0027	0.016	0.009	39	5.1346	2.509
Mongolians from Mongolia	85	76	0.995	0.0027	0.019	0.01	69	6.1211	2.9408
Iranian Azeris	75	177	0.995	0.0035	0.029	0.015	66	9.4905	4.4033
Uzbeks from Uzbekistan	319	123	0.996	0.0009	0.018	0.009	232	5.6575	2.7204
Iranian Armenians	62	68	0.996	0.0037	0.017	0.009	54	5.4332	2.6523
Kazakhs from Kazakhstan	251	111	0.996	0.0011	0.018	0.01	190	5.9246	2.8373
Iranian Kurds	80	80	0.996	0.0031	0.018	0.01	71	5.9763	2.8799
Pathans from Afghanistan	51	53	0.996	0.0048	0.017	0.009	46	5.5867	2.7272
Huis from China	42	55	0.997	0.0061	0.02	0.011	39	6.345	3.0689
Armenians from Armenia	53	55	0.997	0.0047	0.015	0.008	50	4.8905	2.4215
Iranians in Tehran	108	64	0.998	0.002	0.015	0.008	103	4.9886	2.4446
Iranian Arabs	91	85	0.999	0.0019	0.018	0.01	87	5.7336	2.7712
Iranian Gilakis	75	91	0.999	0.0023	0.02	0.011	73	6.3925	3.0624
Georgians from Georgia	85	70	0.999	0.0019	0.015	0.008	83	4.9479	2.432
Azeris from Azerbaijan	71	78	1	0.0025	0.017	0.009	70	5.3453	2.6096
Northern Talyshes from Azerbaijan	77	80	1	0.0021	0.017	0.009	77	5.3845	2.6242
Iranian Bakhtiaris	53	78	1	0.0037	0.02	0.011	53	6.5	3.1236
Iranian Southern Talyshes	47	47	1	0.0044	0.015	0.008	47	4.766	2.3719
Iranians in Isfahan	44	60	1	0.0048	0.017	0.009	44	5.5402	2.7141

Abbreviations and symbols: 1. Number of samples (N), 2. Number of polymorphic (segregating) sites (S), 3. Haplotype diversity (Hd), 4. Standard deviation for Hd, 5. Nucleotide diversity (Pi), 6. Standard deviation for Pi, 7. Number of haplotypes (H), 8. The mean number of differences between all haplotype pairs (π), 9. Standard deviation for π. The order of the populations in this table is based on haplotype diversity.
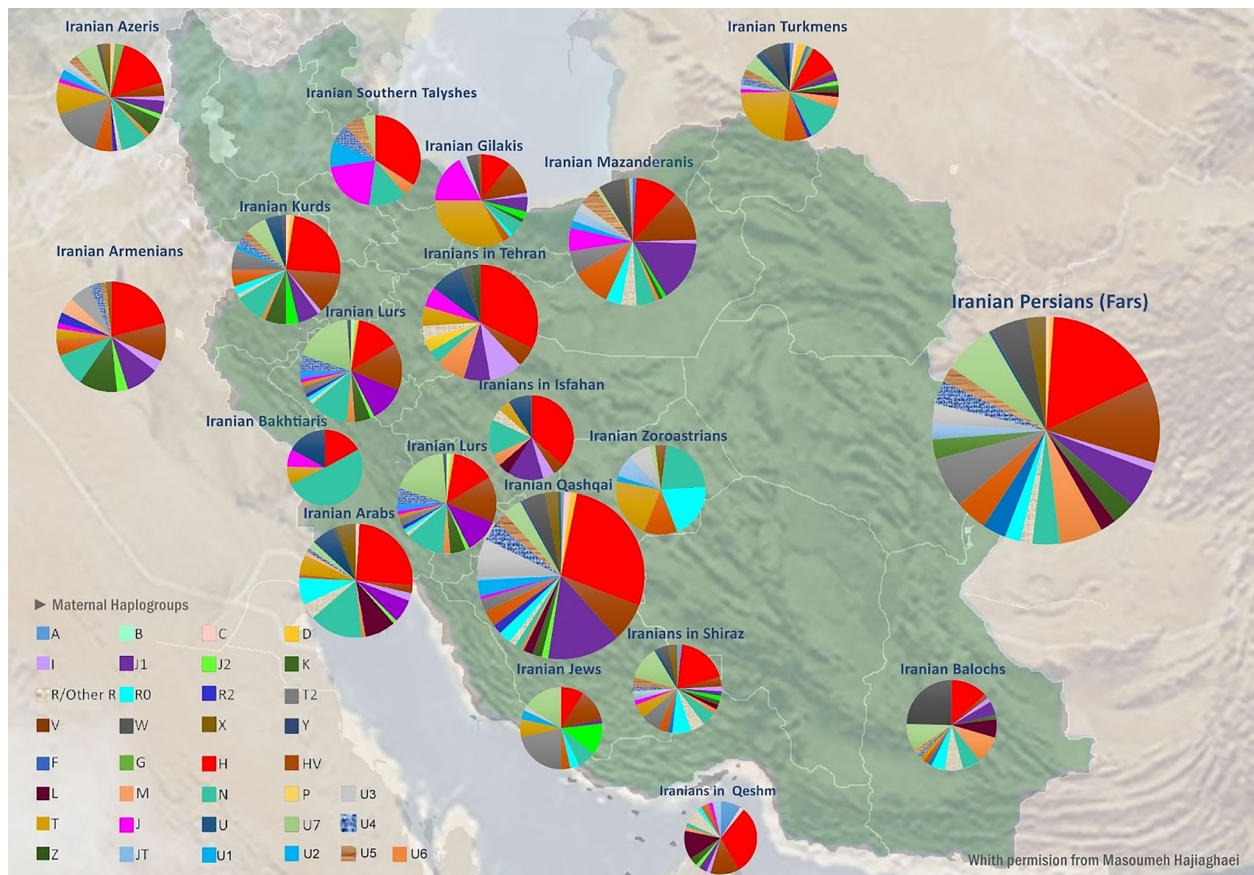
FIGURE 4 Relative mitochondrial DNA (mtDNA) haplogroup frequencies of the Iranian populations. Circle sizes correspond to sample sizes analyzed in this study. This figure was designed via Adobe Photoshop 2022 V.23.3. Operating system windows 11 64-bit. *Source*: The raw map was extracted from https://www.cleanpng.com/free/iran-map.html.

most frequent among most of the surveyed individuals outside of Iran, with exceptions, including the Yaghnobi, Turk, Dravidian, Georgian, and Armenian populations (see Supplementary Table, sheet 2).

The ethnic groups "Persian," "Armenian," and "Azeri" are not delineated on the map as their data are sourced from multiple provinces in Iran, as indicated in Figure 3A. Supplementary Tables, sheets 1 and 2 provide a summary of mtDNA haplogroups distribution for both Iranian and other Eurasian communities studied.

3.3 | Haplogroup frequency-based comparisons of the studied populations

A PCA was conducted on the frequencies of derived haplogroups from a dataset comprising 3840 HVS-I mitochondrial sequences. This was executed to discern the genetic affiliations between Iranian groups and other Eurasian populations (see Figure 5A,B). Drawing from 36 haplogroups, the first two principal components (PCs) encapsulated 25.504% of the total variance.

There is a distinct demarcation along PC1 between populations speaking the Turco-Mongol languages (viz., Mongolia, Turkmen, Shors, Karakalpak, Kyrgyzstan, Kazakhstan, and Uzbekistan) and other groups. Notably, haplogroups Y, G, C, A, and D manifest with higher frequencies in Mongolia, Kyrgyzstan, and Uzbekistan compared to other areas. Parallelly, Yunnan/Altaian–Kizhi and Kyrgyzstan/Uzbekistan exhibit congruent frequencies for the D and C haplogroups.

Counterposing this on the PC1 axis, and with a distinction on the PC2-PC3, the Kalash ethnic group is characterized by pronounced frequencies of haplogroups R0, J2, and specific sub-lineages of the U haplogroup, such as U4 and U2. In a similar PC1 orientation with the Kalash, the Eurasian outlier Dravidian population exhibits the highest frequency of haplogroup R0. The Yaghnobi, marked by the highest HV frequency in the dataset, clusters with Iranian populations, including Persians, Mazanderanis, Lurs, Armenians, Qashqai, Kurds, Shiraz, Azeris, and Jews in Iran.

The ancient U7 lineage was strikingly high at a frequency of 18.07% in the Lur ethnic group within the Iranian
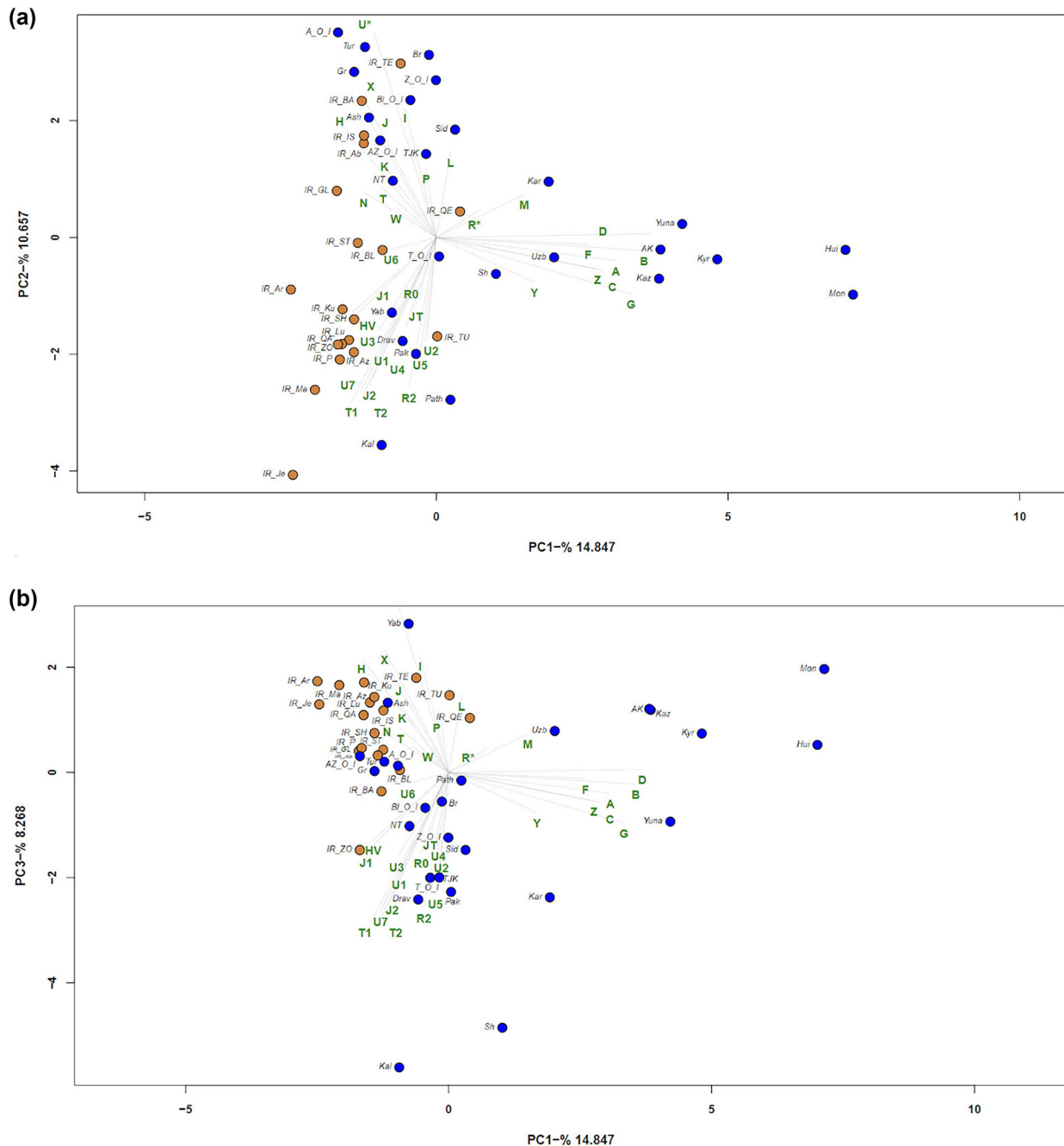
FIGURE 5 (a) Principal component analysis (PCA) is performed based on haplogroup frequencies. Brown and blue dots represent Iranian and other non-Iranian Eurasian populations, respectively. Principal component (PC)1-2 together accounted for 25.50% of the total variance. (b) PCA is performed based on haplogroup frequencies. Brown and blue dots represent Iranian and other non-Iranian Eurasian populations, respectively. PC1-3 together accounted for 23.12% of the total variance.

gene pool of this study. Moreover, uncommon haplogroups such as K, M, and the smaller percentage of H and N were distinguished. All these observations could characterize the reduced mtDNA variability of Lurs among other Iranian populations.

We used the first six PCs' scores for the Ward-type hierarchical clustering with Euclidean distance calculation method.

The dendrogram on Figure 6 helps verifying the PCA results. Out of the two major branches, the first left one contains more than half of the Iranian dataset (63.15%). Interestingly, the majority of the population who live in southern parts of the Zagros mountains stand together with the strongest probability (Shiraz, Lurs, Kurds, and Qashqai). On the other hand, the high ratio of the autochthonous haplogroup U7 among the population who
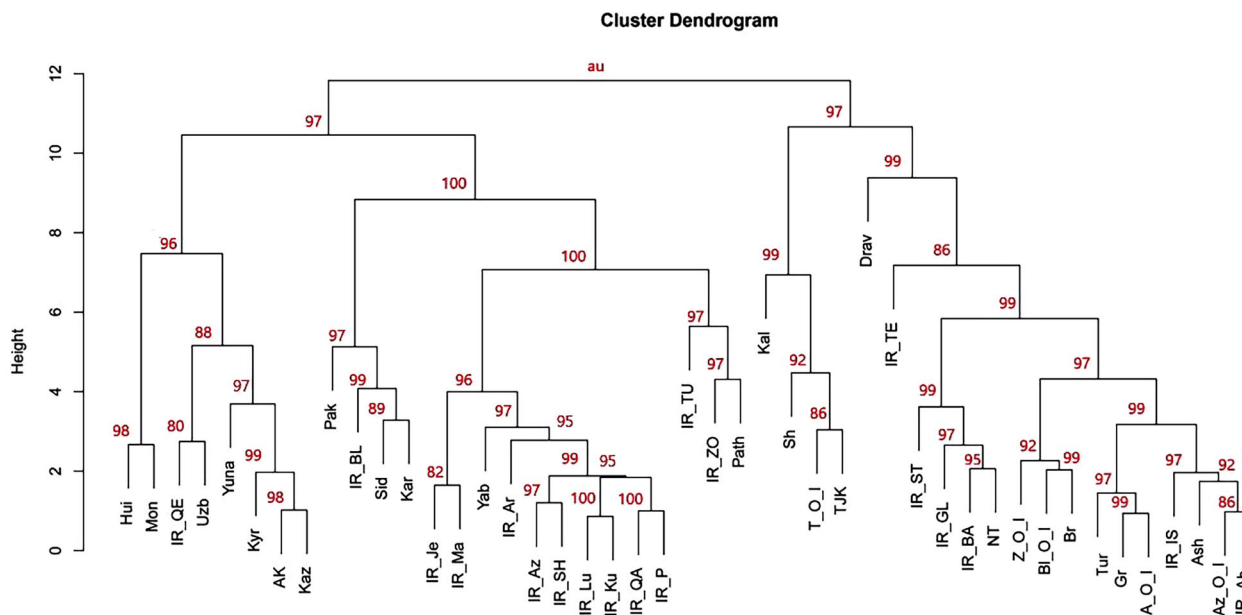
Cluster Dendrogram

FIGURE 6 Dendrogram based on Ward clustering with bootstrap probability values (*p*).

lived in the southern part of the Zagros mountain can represent the persistence of this ancient haplogroup from a long back period up to now for the modern population. It is worth mentioning that Iranian Zoroastrians and Pathans from Afghanistan are joined with high probability values (97%). This can link back to the historical root and cultural practices of the Pashtun ethnic groups from Zoroastrianism (Wynbrandt, 2009).

Another subbranch system composed of the majority of Turco-Mongol language speaking groups like Mongolians, Uzbeks, Kyrgyz, Kazakhs, and Karakalpaks with high probability value (96%), showing their common roots in their genetic structure. More than half the adjacent countries of Iran stand for the second major cluster (97%). It is shown on the dendrogram that within the second major cluster, Turks, Georgians and Armenians; Brahuis, Balochs and Zoroastrians; Turkmens and Tajiks that are nearby non-Iranian populations stand within the same branch system. Interestingly, the Talyeshes from Azerbaijan and Iran cluster with each other at 99% probability. The predominance of haplogroups H, N, J, and U in both Talyeshes populations connect them genetically across the border.

3.4 | Sequence-based genetic differentiation discussion

Based on the HVS-I mtDNA sequence data, pairwise genetic distances (F_{ST}) were calculated. F_{ST} range of eight populations (Dravidian, Yunna, Kalash, Ashkenazi, Iranian Jews and Zoroastrians, Shors, and Yaghnobi) was notably higher compared to other groups, designating these populations as outliers in the MDS plot (Supplementary Table, sheet 3). Additionally, all F_{ST} *p* values were significant except the following pairs: Yaghnobi and Iranian Azeris, Shors and Iranian Azeris, Shors and Turks, Iranian Zoroastrians and Iranian Azeris, Ashkenazi and Iranian Azeris, and Iranian Jews and Iranian Azeris (Supplementary Table, sheet 3).

The pairwise F_{ST} values were close to zero in some identical ethnonyms living in Iran and beyond its borders, such as Armenians, Turkmens, Azeris, Balochs, and Talyshes, showing high genetic affinity between them. An exception to this was the observed significant F_{ST} of 0.12 ($p < 0.05$) between Iranian Zoroastrians and their counterparts in India.

The pairwise F_{ST} measured between the Persian-speaking urban populations (Tehran, Isfahan, and Shiraz) was low and nonsignificant, correlating with the high maternal gene flow and the linguistic evidence of the modern Persian language's origin in these populations (Renfrew, 1990). MDS and heatmap plots were created from pairwise F_{ST} values to depict the relationships between the investigated Iranians and other populations from Eurasia (Figures 7 and 8).

Generally, the amounts of pairwise F_{ST} displayed considerable similarity among Kazakhs, Karakalpaks, Mongolians, Kyrgyz, Hui, Altai-Kizhi, Sindhi, and Zoroastrians from India. These populations are located next to each other on the MDS plot. However, they had quite different F_{ST} from the rest of the populations, leading to a separate cluster in the MDS and heatmap plots.

The Gilaki and Mazandarani ethnic groups, inhabitants of the South Caspian region in Iran, predominantly speak the North-Western branch of Iranian languages. Linguistic
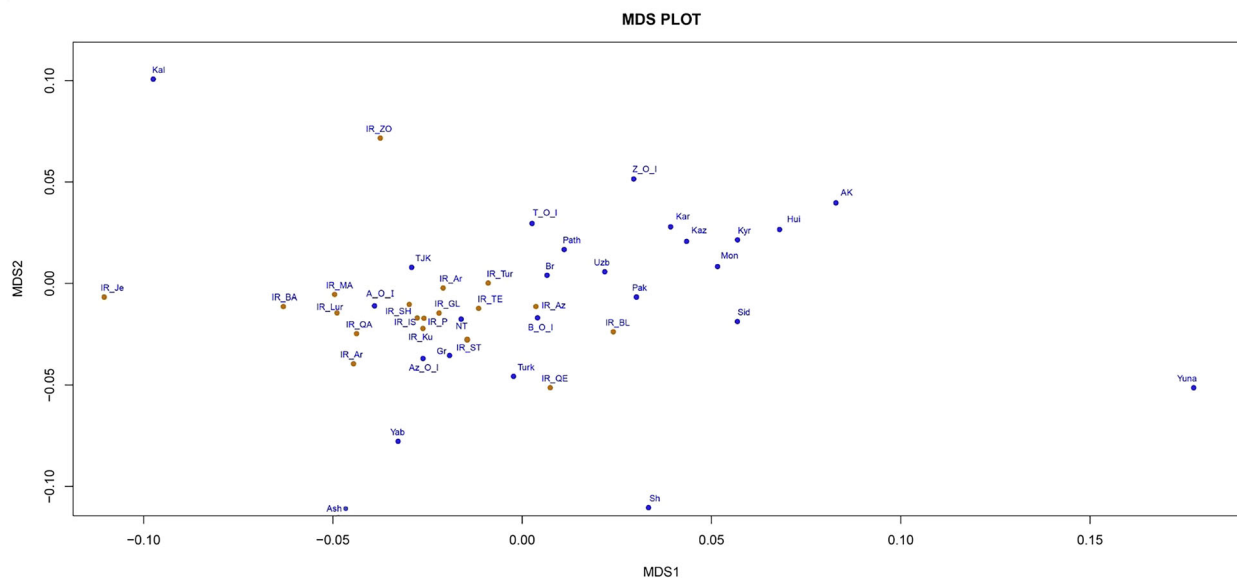
FIGURE 7 Multidimensional scaling plot based on hypervariable segment (HVS-I) mitochondrial DNA (mtDNA) sequences. Brown and blue dots represent Iranian and other non-Iranian Eurasian populations, respectively.

evidence, highlighted by the shared typological characteristics among Gilaki, Mazandarani, and Caucasian languages, supports the hypothesis that these groups may have originated in the Caucasus, potentially displacing a prior South Caspian population (Sana'ati et al., 2017). Our genetic findings, derived from sequence-level analysis, indicate that the Iranian Gilakis have genetic ties closely resembling the Azeris and Northern Talysh people from Azerbaijan. Meanwhile, the Iranian Mazanderanis appear genetically similar to Armenians from Armenia. Furthermore, both groups display genetic proximity to the broader Iranian gene pool. These results are corroborated by PCA and Ward's method of hierarchical clustering, reinforcing the genetic affinities observed among these populations.

In 1813, the Golestan agreement between Russia and Iran separated the North Talysh area from Iran, dividing Astara city into two parts. The Astrachai River was used as the border, and the area to the north was ceded to the Russian government (which became the Republic of Azerbaijan in 1991 CE), and the southern area remained with Iran (Fard et al., 2019). The MDS and heatmap plots of the current study show that the Talyeshes from Azerbaijan are the most similar in their maternal lineages to the Talyeshes from Iran (Figures 7 and 8).

Historically, the Qashqai migrated to Iran at different periods, eventually settling in the Fars province, where they intermingled with local Lur, Kurds, and Turks. A significant event during the Qajar dynasty around 1870 CE saw this tribe grappling with famine, leading to its eventual disintegration and the subsequent absorption of approximately 5000 Qashqai families into the Bakhtiari tribe (Oberling, 1974). Various distributions of haplogroups and

sub-haplogroups can be found in this ethnic group. In fact, Qashqai stands the highest variable position (77.8%) for haplogroups within the Iranian gene pool. Qashqai people from Iran have a stronger affinity to Kurds, Lur, and Bakhtiari populations than to other Iranian ones based on our F_{ST} result.

The genetic distance between the Lur and Bakhtiari ethnic groups was negligible, as reflected by the fact that they are close to each other on the MDS plot. Notably, the Lur exhibits a more diverse array of mtDNA haplogroups compared to the Bakhtiaris. Out of the 19 haplogroups identified in Lurs, they share only four (H, T, N, and J) with the Bakhtiaris. Such genetic patterns might be attributed to the Bakhtiaris' pronounced endogamous cultural practices and their nomadic way of life, compared to the Lurs (Baigi & Sadeghi, 2022). The historical distinction between Lur-i Buzurg (Greater Lur) and Lur-i Kuchak (Lesser Lur) can be traced back to 913 CE within the confines of Lurestan (Fragner, 1987). The Bakhtiaris, a subset of the Greater Lur tribe, predominantly inhabit hilly terrains in Khuzestan and Isfahan, located in southwestern Iran. A significant portion of the Bakhtiari population maintains a nomadic lifestyle, alternating between the northern and southern territories of Iran, during summer and winter (in Farsi, called *sardsir* and *garmsir* camps). The close location of these two ethnic groups in the MDS and current F_{ST} results indicates Bakhtiaris and Lurs do not have significant maternal gene pool differences from each other among the Iranian populations.

The Zoroastrians from Iran and Iranian Jews exhibit a marked divergence from other Iranian populations on the MDS plot, potentially stemming from their distinct
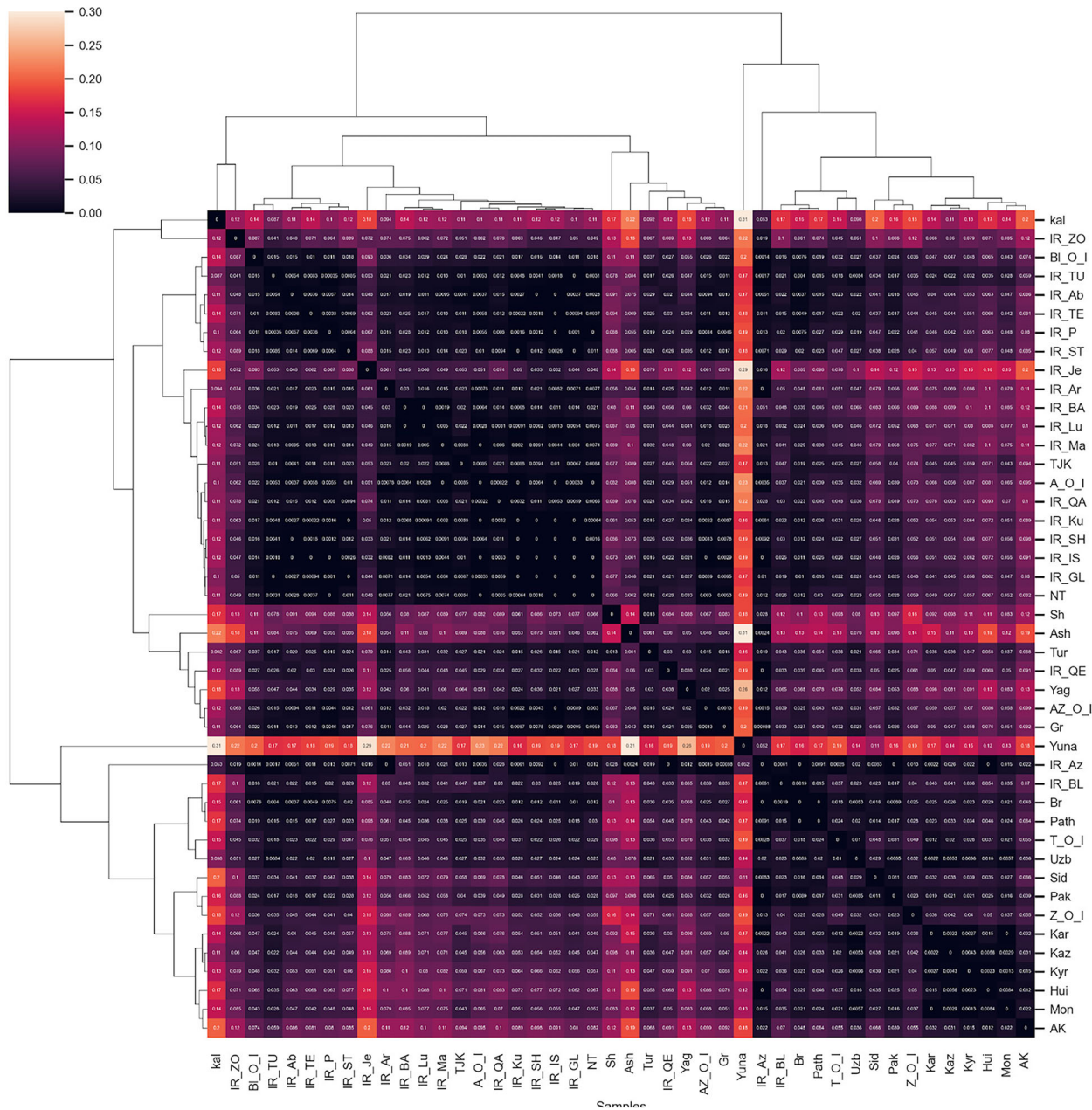
FIGURE 8 Heatmap plot and clustering based on Slatkin's F_{ST} distance matrix.

endogamous traditions and cultures (Baigi & Sadeghi, 2022).

The Balochs had the second lowest mtDNA genetic diversity after Zoroastrians in the Iranian populations of the current research. This can be explained with long term geographical isolation (Bittles & Black, 2010). In fact that they enhanced by geographical separation in a southern part of the Zagros mountainous area and the Lut and Kavir deserts, their gene flow restricted between northern and western Iranian districts for a longer time.

The location of the Iranian Azeris on PCA and MDS which locates as a border between Iranian and other non-Iranian population can be explained by the highest number of exogamous marriages with Iranian and non-

Iranian populations (Arakelova, 2015; Baigi & Sadeghi, 2022). The Iranian Azeris have low genetic distances to Brahui, Pakistan, and Iranians in Isfahan and Qeshm (Figures 7 and 8).

The MDS plot generally distinguishes an explicitly separate cluster of Iranians from other Eastern adjacent countries' populations. Iranian Azeris, Turkmens, and Balochs groups are placed between the two independent clusters of Iranians and Eastern Eurasian people.

The Kalash are a group of Indo-Iranian people who live in Pakistan's Khyber Pakhtunkhwa Province's Chitral area (Ayub et al., 2015). Kalash culture and belief systems differ from the various ethnic groups surrounding them but are close to practices of the Vedic and

pre-Zoroastrians (Mela-Athanasopoulou, 2011). They form a divergent group compared to the other Central Eurasian population on the MDS plot of this research, which shows the lack of South Asian mtDNA lineages. Interestingly, the spatial location of the Kalash on the MDS plot was too far from the Eastern Eurasian populations and stood in line with Zoroastrians from Iran. A similar pattern was observed on PC1-2 (Figure 5A), and the haplogroup-based dendrogram also shows them further from other Pakistani ethnic groups and stands nearer to Iranian population in which cluster with some of them like Tehran, Southern Talyshes, Gilakis, Bakhtiaris, Isfahan, and Arab groups (Figure 6).

The Parsis are Zoroastrian Iranian people who live primarily in South Asia, particularly along the western coast of India. Various historical documents support their Iranian descent. They can be called as the first emigrant community of Iranians as their ancestors fled from Iran to India (Gujarat Province) during the Arab regimes of Iran in the 7th century CE (Fox, 1967). In the Iranian Zoroastrian population, most of the Western-Eurasian haplogroups (H, I, J, and K) are missing entirely, which implies that a considerable genetic drift and founder effects should have occurred in this population following the Muslim expansion in the 7th–8th centuries (Moulton, 1917). In this article, we compared the Zoroastrian-Parsi mtDNA pool from India with the Iranian Zoroastrians. About 35% of Zoroastrian-Parsi's maternal gene pool comprises two Asian haplogroups (M, D) and U, which is recognized as a typical European macro haplogroup. These haplogroups were predominant among the ethnic groups in India, and the genetic structure can also show signs of intermarriages with local women in India. As historical evidence supports, their migration and admixture could have led to their shifted genetic structure and the intermediate location of the Zoroastrian-Parsi between the Eastern Eurasian population and other Iranian populations on the MDS plot. Meanwhile, the high number of endogamous marriages can be considered a potential reason for the small diversity and the distinctiveness of the Iranian Zoroastrian population (Baigi & Sadeghi, 2022).

3.5 | Analyses of the population structure

In order to reveal possible HVS-I sequence-based differences within the dataset compared to geography, we tested a locus-by-locus AMOVA at the country (only the Iranian population) and Eurasia (the whole dataset) scale. Furthermore, a series of AMOVA was conducted, in which we tested a different number of population groups, building geographically distinct clusters (Supplementary Table, sheet 4). On the other hand, identifying clusters that were

geographically homogeneous and genetically maximally distinctive from one another enabled us to detect the possible genetic barriers between the Iranian groups. The lack of a sharp geographical structure in mtDNA HVS-I sequence diversity was statistically supported by the performed tests.

The majority of haplotype variance for the Iranian population was explained within-population level (93.23%, $p < 0.001$). It should be noted that the geographic subdivision of the Iranian samples (North-West and South-Central of Iran) consistently produced low and nonsignificant among-groups variance (F_{CT}) (Supplementary Table, sheet 4). Under both grouping criteria, the F_{CT}, F_{ST}, and F_{SC} values decreased when the outlier groups (Zoroastrians, Jews) were removed from the analyses. As a result of sequential AMOVA analyses for the Iranian population, the highest genetic differentiation between groups occurred when three distinct clusters were identified: The first two clusters represented the Zoroastrian and Jewish communities, and the third cluster represented the remaining Iranian ethnic groups. This clustering configuration displayed the highest among-groups percentage of variance (4.16%) with a significant but low among-groups fixation index ($F_{CT} = 0.041$, $p = 0.009$), thus indicating genetic discontinuity between Zoroastrians, Jews, and the majority of Iranian samples (Supplementary Table, sheet 4).

We performed the AMOVA analysis on the whole Eurasian dataset, in which Dravidians were not included, as their genetic difference from the other groups is so large that they would unbalance the results. Corresponding to the MDS, our dataset fit into two large groups rather than three, as two-group clustering gave a higher range for among group-genetic variance (Table 3). Most of the haplotype variation in this constellation accounts for significant within-population differences (95.43%, $p < 0.05$).

We also performed AMOVA by linguistic groups (Afro-Asiatic, Altaic, and Indo-European) and assess what contribution linguistic differences may have to genetic differentiation in Iran. It seems that a random mtDNA differentiation pattern with respect to linguistics can be invoked, suggesting that linguistic factors did not actually represent insurmountable barriers to matrilineal gene flow among the Iranian population.

Constellations 1 and 2 consist of two and three subdivisions, respectively. Model 1, Subgroup 1: Karakalpak, Mongolia, Kazakhstan, Kyrgyz, Altai-Kizhi, Hui-china, Sindhi, and Zoroastrian from India. Subgroup 2: Iranian Arabs, Iranian Azeris, Iranian Balochs, Iranian Kurds, Iranian Lur, Iranian Mazanderanis, Iranian Persians, Shiraz, Tehran, Turkmen, Pathan, Georgia, Baloch, Azerbaijan, Armenian, Turk, Turkmen, Tajik, Iranian Qashqai, Isfahan, Iranian Bakhtiaris, Pakistan, Brahui, Iranian Gilakis, Uzbekistan, Southern Talyshes, Northern Talyshes from Azerbaijan, Iranian Armenians, and Iranians in Qeshm

TABLE 3 Analysis of molecular variance (AMOVA) based on the Eurasian mitochondrial DNA (mtDNA) hypervariable segment I (HVS-I) sequence dataset.

Pattern	Source of variation	Variance components	%	F_{SC}	F_{ST}	F_{CT}
	Among groups	0.0982	2.73			
Model 1	Among populations within group	0.0661	1.84	*0.0189	*0.0457	*0.0273
	Within population	3.4321	95.43			
Model 2	Among groups	0.0865	2.42			
	Among populations within group	0.0509	1.42	*0.0146	*0.0385	0.0242*
	Within population	3.4393	96.16			
Language family+	Among groups	−0.00619	−0.21	*0.01886	*0.01683	−0.00207
	Among populations within group	0.05651	1.89			
	Within population	2.94067	98.32			

Note: We considered the whole dataset for geographical criteria and linguistic groups within the Iranian population.

were considered. For model 2, Subgroup 1: Iranian population excluded Jew and Zoroastrians, Tajik, Armenian, Azerbaijan, Georgia, and Northern Talyshes. Subgroup 2: Turkmen, Pathan, Brahui, Uzbekistan, and Baluch. Subgroup 3: Karakalpak, Mongolia, Kazakhstan, Kyrgyz, Altai-Kizhi, Hui-china, Sindhi, and Zoroastrian from India are the representatives of each subdivision. *Stars denote *p* values <0.05. There are eight ethnic groups (Shors, Ashkenazi, Yaghnobi, Dravidian, Iranian Zoroastrians, Iranian Jews, Kalash, and Yunnan-China) considered outlier populations in both models of AMOVA based on their high genetic differentiations from the other studied groups.

Language family for Iranian populations, based on: Afro-Asiatic, Altaic, and Indo-European

In order to examine further the relationship between geographical features and mtDNA variance, we conducted a Mantel test. We found that genetic and geographic distances were significantly correlated when considering the full dataset of Eurasian populations ($r = 0.65$, $p < 0.001$) (Supplementary Table, sheet 6). In contrast, when analyzing only Iranian samples, correlations were minor and nonsignificant ($r = 0.12$, $p = 0.22$) (Supplementary Table, sheet 5). On the other hand, As a result of considering the entire dataset of Eurasian populations a positive spatial autocorrelation among geographic distances and pairwise F_{ST} were observed (Supplementary Table, sheet 6).

4 | CONCLUSIONS

Comprehensive studies delving into the maternal genetic variations within Iranian communities have been limited. In this study, by merging and reanalyzing previously published mtDNA data from the gene pool of the modern Iranian population, we sought to shed light on the HVS-I mtDNA genetic makeup of modern Iranians and neighboring populations across Eurasia. Furthermore, we endeavored to identify potential external influences on the Iranian mtDNA pool and to map out the genetic interactions among various Iranian ethnic groups spanning the Iranian Plateau.

The observed high levels of haplotype and nucleotide diversity underscore the vast genetic diversity within the Iranian gene pool. The matrilineal genetic structures revealed a pronounced genetic affinity among the majority of Iranian populations. Notably, low pairwise F_{ST} values were evident among the Persian-speaking urban populations of Tehran, Isfahan, and Shiraz, suggesting extensive maternal gene flow between these groups. This genetic cohesion is in line with linguistic evidence pointing to the origins of the modern Persian language in Iran's core regions. Further, Iranian Lurs and Bakhtiaris exhibited a particular pattern of high genetic congruence, underscored by both F_{ST} and MDS analyses, which is consistent with the Bakhtiaris' historical subdivision from the greater Lur tribes. In addition the close maternal genetic similarity of Qashqai people with Lurs, Bakhtiaris, and Iranian Kurds corresponds to their admixture with Qashqai after the later population's collapse during the late Qajar dynasty.

Despite this similarity among Iranians, genetic trends can be observed in the country's periphery. The high frequency of South Asian components in the Balochi ethnic group corresponds to the geographical position of the Balochistan province in the southeast of Iran. The Qeshm minority living in the south accumulated some African components of the L haplogroup, which can imply the African inflow through the Strait of Hormuz, also serving as a choke point and a maritime passage between the two continents. Besides geographic, cultural and political factors seem to have acted as obstacles to maternal gene flow between Iranian populations, such as Balochi, Jews, Zoroastrians, and Turkmens, corresponding to long centuries of reproductive isolation caused by language, religion, and other cultural barriers.

In this study, we also considered particular populations with identical ethnonyms within and beyond the

Iranian borders and investigated the degree of genetic affinity between them. The relatively recent separation of the Talysh population into two groups in Iran and Azerbaijan did not lead to significant maternal genetic differences between them. Iranian and Indian Zoroastrians, on the contrary, for whom there is also evidence of long isolation for cultural and religious reasons after the Muslim expansion in the 7–8th centuries, showed a significant haplogroup and sequence-based maternal genetic structure.

Our findings suggest that barring a few culturally unique and geographically isolated groups, the broader Iranian populace exhibits a cohesive maternal genetic structure. This pattern can be extrapolated to ethnic groups residing in regions historically under Iranian governance prior to 1828 CE (encompassing Turkmenistan, Azerbaijan, Georgia, and Armenia).

As the field progresses, the use of mitochondrial and whole genome sequencing is becoming increasingly pivotal, especially with advances in molecular biology. Deploying complete mitogenomes in phylogenetic analyses of Iranian populations will enhance genetic resolution for both inter- and intrapopulation assessments.

One limitation of our study is the inherent imbalance in the mtDNA dataset, due to either smaller sample sizes for certain Iranian groups or asymmetrical sampling of specific minority groups (e.g., Assyrian, Tat, Sistani, Bandari, Iranian Georgians, Mandaean, Tabari, Yarsan, and Chaldean). These samples may not holistically represent the diverse Iranian society. By consolidating extant mtDNA data, we aimed to offer a clearer depiction of the Iranian maternal genetic landscape. Future research, concentrating on comprehensive mitochondrial genomes and previously overlooked Iranian populations, especially with balanced urban and rural sampling, will further enrich this domain.

AUTHOR CONTRIBUTIONS

Motahareh Amjadi and Mahmood Tavallaei recruited data from reviewed articles and their authors for this study. Motahareh Amjadi and Zahra Hayatmehr performed data analysis. Motahareh Amjadi performed literature review and drafted the manuscript with inputs from all authors. Anna Szécsényi-Nagy, Mahmood Tavallaei, and Balázs Egyed supervised and edited the article. All authors revised and approved the final manuscript.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Motahareh Amjadi https://orcid.org/0009-0002-1613-3548

REFERENCES

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465. https://doi.org/10.1038/290457a0

Arakelova, V. (2015). On the number of Iranian turkophones. *Iran and the Caucasus*, *19*(3), 279–282. https://doi.org/10.1163/1573384X-20150306

Asgharzadeh, M., Samadi Kafil, H., Ranjbar, A., Rahimipour, A., Najati, K., & Rahbani Nobar, M. (2011). Molecular diversity of mitochondrial DNA in Iranian Azeri ethnicities vis-à-vis other Azeris in Asia. *Iranian Journal of Biotechnology*, *9*(2), 120–125.

Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S. Q., & Tyler-Smith, C. (2015). The Kalash genetic isolate: Ancient divergence, drift, and selection. *American Journal of Human Genetics*, *96*(5), 775–783. https://doi.org/10.1016/j.ajhg.2015.03.012

Bahmanimehr, A., Eskandari, G., & Nikmanesh, F. (2015). The congruence between matrilineal genetic (mtDNA) and geographic diversity of Iranians and the territorial populations. *Iranian Journal of Basic Medical Sciences*, *18*(1), 30.

Baigi, M., & Sadeghi, R. (2022). Marriage patterns of Iranian ethnic groups during the last two decades: Convergence or continuity of differences? *Scientific Quarterly Journal of Social and Cultural Strategy, Strategic Research Institute*, *11*(45), 437–472.

Bell, J. (1832). *A system of geography, popular and scientific, or, a physical, political and statistical account of the world and its various divisions*. Archibald Fullarton.

Bennett, C., & Kaestle, F. A. (2010). Investigation of ancient DNA from Western Siberia and the Sargat culture. *Human Biology*, *82*(2), 143–156. https://doi.org/10.3378/027.082.0202

Bittles, A. H., & Black, M. L. (2010). Consanguineous marriage and human evolution. *Annual Review of Anthropology*, *39*, 193–207. https://doi.org/10.1146/annurev.anthro.012809.105051

Blackwell, C. (2001). *Tradition and society in Turkmenistan: Gender, oral culture and song*. Curzon.

Chen, Y., Gong, L., Liu, X., Chen, X., Yang, S., & Luo, Y. (2020). Mitochondrial DNA genomes revealed different patterns of high-altitude adaptation in high-altitude Tajiks compared with Tibetans and Sherpas. *Scientific Reports*, *10*(1), 1–9.

Cilli, E., Sarno, S., Gnecchi Ruscone, G. A., Serventi, P., De Fanti, S., Delaini, P., & Gruppioni, G. (2019). The genetic legacy of the Yaghnobis: A witness of an ancient Eurasian ancestry in the historically reshuffled central Asian gene pool. *American Journal of Physical Anthropology*, *168*(4), 717–728. https://doi.org/10.1002/ajpa.23789

Costa, M. D., Pereira, J. B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., & Richards, M. B. (2013). A substantial prehistoric

European ancestry amongst Ashkenazi maternal lineages. *Nature Communications*, *4*(1), 1–10. https://doi.org/10.1038/ncomms-3543

Cummings, S. N. (2013). *Understanding central Asia: Politics and contested transformations*. Routledge.

Derenko, M., Malyarchuk, B., Bahmanimehr, A., Denisova, G., Perkova, M., Farjadian, S., & Yepiskoposyan, L. (2013). Complete mitochondrial DNA diversity in Iranians. *PLoS ONE*, *8*(11), e80673. https://doi.org/10.1371/journal.pone.0080673

Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Rogalla, U., Grzybowski, T., & Zakharov, I. (2012). Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. *PLoS ONE*, *7*(2), e32179. https://doi.org/10.1371/journal.pone.0032179

Elfenbein, J. (2019). Seever. In B. Sanford (Ed.), *The Dravidian languages* (2nd ed., p. 495). Routledge.

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

Fard, F. K., Basiri, M. A., & Yazdani, E. (2019). Ethnic bargaining and separatism in the South Caucasus. *Region*, *8*(2), 173–196. https://doi.org/10.1353/reg.2019.0011

Farjadian, S., Sazzini, M., Tofanelli, S., Castrì, L., Taglioli, L., Pettener, D., & Luiselli, D. (2011). Discordant patterns of mtDNA and ethno-linguistic variation in 14 Iranian ethnic groups. *Human Heredity*, *72*(2), 73–84. https://doi.org/10.1159/000330166

Fisher, W. B., Avery, P., Hambly, G. R. G., & Melville, C. (1991). *The Cambridge history of Iran* (Vol. 7). Cambridge University Press.

Fox, D. A. (1967). Darkness and light: The Zoroastrian view. *Journal of the American Academy of Religion*, *35*(2), 129–137. https://doi.org/10.1093/jaarel/XXXV.2.129

Fragner, B. G., & Garthwaite, G. R. (1987). Khans and Shahs: A documentary analysis of the Bakhtiari in Iran. *Journal of Iranian Studies*, *22*(2–3), 131–133.

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Molecular Ecology*, *18*(22), 4541–4550. https://doi.org/10.1111/j.1365-294X.2009.04380.x

Garnik, A., & Habib, B. (2005). Talish and the Talyshis (state of research). *Iran & the Caucasus*, *9*(1), 43–72.

Heyer, E., Balaresque, P., Jobling, M. A., Quintana-Murci, L., Chaix, R., Segurel, L., Aldashev, A., & Hegay, T. (2009). Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genetics*, *10*, 49. https://doi.org/10.1186/1471-2156-10-49

Irwin, J. A., Ikramov, A., Saunier, J., Bodner, M., Amory, S., Röck, A., & Parsons, T. J. (2010). The mtDNA composition of Uzbekistan: A microcosm of Central Asian patterns. *International Journal of Legal Medicine*, *124*(3), 195–204. https://doi.org/10.1007/s00414-009-0406-z

Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., & Stoneking, M. (2014). Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genetics*, *5*(1), 1–17. https://doi.org/10.1186/2041-2223-5-13

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Marchi, N., Hegay, T., Mennecier, P., Georges, M., Laurent, R., Whitten, M., & Heyer, E. (2017). Sex-specific genetic diversity is shaped by cultural factors in inner Asian human populations. *American Journal of Physical Anthropology*, *162*(4), 627–640. https://doi.org/10.1002/ajpa.23151

Margaryan, A., Derenko, M., Hovhannisyan, H., Malyarchuk, B., Heller, R., Khachatryan, Z., & Allentoft, M. E. (2017). Eight millennia of matrilineal genetic continuity in the South Caucasus. *Current Biology*, *27*(13), 2023–2028. https://doi.org/10.1016/j.cub.2017.05.087

Mela-Athanasopoulou, E. (2011). A morphophonological description of Kalasha as an Indo-Aryan language with Greek roots. *US-China Foreign Language*, *9*(7), 2–17.

Meri, J. W. (2018). *Medieval Islamic civilization: An encyclopedia*. Routledge.

Morawala-Patell, V., Pasha, N., Krishnasamy, K., Mittal, B., Gopalakrishnan, C., Mugasimangalam, R., & Project, T. A. (2021). The first complete Zoroastrian-Parsi mitochondrial reference genome and genetic signatures of an endogamous non-smoking population. *Meta Gene*, *28*, 100882. https://doi.org/10.1016/j.mgene.2021.100882

Moulton, J. H. (1917). *The treasure of the magi: A study of modern Zoroastrianism* (Vol. 3). H. Milford, Oxford University Press.

Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., & Reich, D. (2019). The formation of human populations in south and central Asia. *Science*, *365*(6457), eaat7487. https://doi.org/10.1126/science.aat7487

Nasidze, I., Ling, E. Y. S., Quinque, D., Dupanloup, I., Cordaux, R., Rychkov, S., & Stoneking, M. (2004). Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Annals of Human Genetics*, *68*(3), 205–221. https://doi.org/10.1046/j.1529-8817.2004.00092.x

Nasidze, I., Quinque, D., Ozturk, M., Bendukidze, N., & Stoneking, M. (2005). MtDNA and Y-chromosome variation in Kurdish groups. *Annals of Human Genetics*, *69*(4), 401–412. https://doi.org/10.1046/j.1529-8817.2005.00174.x

Nasidze, I., Quinque, D., Rahmani, M., Alemohamad, S. A., Asadova, P., Zhukova, O., & Stoneking, M. (2009). mtDNA and Y-chromosome variation in the Talysh of Iran and Azerbaijan. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *138*(1), 82–89. https://doi.org/10.1002/ajpa.20903

Nasidze, I., Quinque, D., Rahmani, M., Alemohamad, S. A., & Stoneking, M. (2006). Concomitant replacement of language and mtDNA in South Caspian populations of Iran. *Current Biology*, *16*(7), 668–673. https://doi.org/10.1016/j.cub.2006.02.021

Nasidze, I., Quinque, D., Rahmani, M., Alemohamad, S. A., & Stoneking, M. (2008). Close genetic relationship between semitic-speaking and Indo-European-speaking groups in Iran. *Annals of Human Genetics*, *72*(2), 241–252. https://doi.org/10.1111/j.1469-1809.2007.00413.x

Nasidze, I., & Stoneking, M. (2001). Mitochondrial DNA variation and language replacements in the Caucasus. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1472), 1197–1206.

Oberling, P. (1974). *The Qashqai nomads of Fars*. Walter De Gruyter Inc.

Olivieri, A., Pala, M., Gandini, F., Kashani, B. H., Perego, U. A., Woodward, S. R., Grugni, V., Battaglia, V., Semino, O., Achilli, A., Richards, M. B., & Torroni, A. (2013). Mitogenomes from Two uncommon haplogroups mark late glacial/postglacial expansions from the near east and Neolithic dispersals within Europe. *PLoS ONE*, *8*(7), e70492. https://doi.org/10.1371/journal.pone.0070492

Paul, D. P., Abbess, E., Müller, K., Tiessen, C., & Tiessen, G. (2009). The ethnolinguistic vitality of Yaghnobi(PDF). In *SIL Electronic Survey Report 2010–017, May 201*. SIL International.

Poorolajal, J. (2017). Resistance economy and new population policy in Iran. *Journal of Research in Health Sciences*, *17*(1), 367.

Posada, D., & Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics (Oxford, England)*, *14*(9), 817–818.

Pradhan, S. (2016). *China's maritime silk route and Indonesia's global maritime fulcrum: Complements and contradictions*. Institute of Chinese Studies Occasional Paper.

Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., & McElreavey, K. (2004). Where west meets east: The complex mtDNA landscape of the southwest and Central Asian corridor. *The American Journal of Human Genetics*, *74*(5), 827–845. https://doi.org/10.1086/383236

Rahman, Z. U., Li, Y. C., Tian, J. Y., & Kong, Q. P. (2020). Exploring European ancestry among the Kalash population: A mitogenomic perspective. *Zoological Research*, *41*(5), 552. https://doi.org/10.24272/j.issn.2095-8137.2020.052

Rashidvash, V. (2012). Iranian people and the race of people settled in the Iranian plateau. *Research Journal of Humanities and Social Sciences*, *3*(4), 426–435.

Rastorgueva, V. S., Kerimova, A. A., Mamedzade, A. K., Pireiko, L. A., Edel'man, D. I., & Lockwood, R. M. (2012). *The Gilaki Language*. Acta Universitatis Upsaliensis.

Renfrew, C. (1990). *Archaeology and language: The puzzle of Indo-European origins*. CUP Archive.

Sana'ati, M. K., Pasandyar, F., & Heidar, D. M. (2017). A study of common expressions in English and Farsi languages and their equivalents in Gilaki dialect in North of Iran. *Language in India*, *17*(4), 17.

Schönberg, A., Theunert, C., Li, M., Stoneking, M., & Nasidze, I. (2011). High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: High diversity and demographic inferences. *European Journal of Human Genetics*, *19*(9), 988–994. https://doi.org/10.1038/ejhg.2011.62

Shamoon-Pour, M., Li, M., & Merriwether, D. A. (2019). Rarehuman mitochondrial HV lineages spread from the near east and Caucasus during post-LGM and Neolithic expansions. *Scientific Reports*, *9*(1), 14751. https://doi.org/10.1038/s41598-019-48596-1

Siddiqi, M. H., Rakha, A., Khan, K., & Akhtar, T. (2021). Current pool of ultimate collection of mitochondrial DNA from remnants of Kalash. *Mitochondrial DNA Part B*, *6*(8), 2410–2414. https://doi.org/10.1080/23802359.2021.1952119

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462. https://doi.org/10.1093/genetics/139.1.457

Sylvester, C., Krishna, M. S., Rao, J. S., & Chandrasekar, A. (2019). Maternal genetic link of a south Dravidian Tribe with native Iranians indicating bidirectional migration. *Annals of Human Biology*, *46*(2), 175–180. https://doi.org/10.1080/03014460.2019.1599067

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526.

Terreros, M. C., Rowold, D. J., Mirabal, S., & Herrera, R. J (2011). Mitochondrial DNA and Y-chromosomal stratification in Iran: Relationship between Iran and the Arabian Peninsula. *Journal of*

Human Genetics, *56*(3), 235–246. https://doi.org/10.1038/jhg.2010.174

Vallini, L., Marciani, G., Aneli, S., Bortolini, E., Benazzi, S., Pievani, T., & Pagani, L. (2022). Genetics and material culture support repeated expansions into Paleolithic Eurasia from a population hub out of Africa. *Genome Biology and Evolution*, *14*(4), evac045. https://doi.org/10.1093/gbe/evac045

Van Oven, M. (2015). PhyloTree build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series*, *5*, e392–e394.

Van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386–E394. https://doi.org/10.1002/humu.20921

Villesen, P. (2007). FaBox: An online toolbox for fasta sequences. *Molecular Ecology Notes*, *7*(6), 965–968. https://doi.org/10.1111/j.1471-8286.2007.01821.x

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244. https://doi.org/10.1111/j.1471-8286.2007.01821.x

Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F., & Schönherr, S. (2016). mtDNA-server: Next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Research*, *44*(W1), W64–W69. https://doi.org/10.1093/nar/gkw247

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

Wynbrandt, J. (2009). *A brief history of Pakistan*. Infobase Publishing.

Yao, Y. G., Kong, Q. P., Wang, C. Y., Zhu, C. L., & Zhang, Y. P. (2004). Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Molecular Biology and Evolution*, *21*(12), 2265–2280. https://doi.org/10.1093/molbev/msh238

Yao, Y. G., & Zhang, Y. P. (2002). Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan province: New data and a reappraisal. *Journal of Human Genetics*, *47*(6), 311–318. https://doi.org/10.1007/s100380200042

Zhukovskaia, N. L. (2007). Алтайцы [Altaians]. In L. M. Mints (Ed.), *Народы мира: Энциклопедия* [Peoples of the world: An Encyclopedia] (in Russian) (pp. 34–35). OLMA Media Group.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.