Following the Trace of HVS II Mitochondrial Region Within the Nine Iranian Ethnic Groups Based on Genetic Population Analysis

Anousheh Shasttiri1 · Misagh Moridi2 · Abbas Safari2 · Sayed Haidar Abbas Raza3 · Mostafa Ghaderi-Zefrehei4 · Massoud Houshmand5 · Ahmad Oryan6 · Mohammad Hossein Sanati7 · Jacqueline Smith8 · Motahareh Amjadi9

Received: 10 April 2021 / Accepted: 7 October 2021 / Published online: 18 October 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
The Iranian gene pool is seen as an important human genetic resource for investigating the region connecting Mesopotamia and the Iranian plateau. The main objective of this study was to explore gene flow in nine Iranian ethnic/subpopulation groups (402 samples) by examining mtDNA HVS2 sequence variations. This then allowed us to detect mtDNA HVS2 sequence mutations in two independent thalassemia and cystic fibrosis patient sample groups. The patient groups did not explicitly belong to any of the aforementioned nine subpopulations. Across all subpopulations, the haplogroups B4a1c3a, H2a2a1, N10b, H2a2a2, and J1 were seen to be predominant. High haplogroup diversities along with admixture of the exotic groups were observed in this study. The Arab subpopulation was shown to be independent from the others. It was revealed that there is a far distant relationship between Arab and Azeri groups. The thalassemia patient group, represented an almost random sample of most Iranian ethnic groups, and revealed few significant differences ($P < 0.05$) in their HVS2 sequence. It turned out that the IVS II-I (G→A) mutation in the thalassemia β-globin gene was highly significant. Since the thalassemia patients in the present study represent many unique haplotypes, we can begin to comprehend the importance of mtDNA with this disease and the necessity for more studies in this context.

Keywords Thalassemia · Cystic fibrosis · Iranian subpopulations · Haplogroups · mtDNA

* Mostafa Ghaderi-Zefrehei
Mghaderi@yu.ac.ir

Extended author information available on the last page of the article
Introduction

Mitochondria have their own genetic material called mitochondrial DNA (mtDNA) (Galanello and Origa 2010). In total, at least 12 features distinguish human mtDNA from its counterpart nuclear genome (Chial and Craig 2008). Human mtDNA, is a maternally inherited circular double-stranded DNA molecule, comprising 16,569 bp, with approximately 103–104 copies present per cell. mtDNA has an almost 5–15 times higher mutation rate than the nuclear genome (Bircan et al. 2018). The D-loop region, which has the highest rate of polymorphisms, consists of the hyper variable-region I (HVRI) [at position 15971 to 16414 in the mtDNA genome], and hyper variable-region II (HVRII) [format position 15 to 389] and has a unique triple-strand characteristic (Ngili et al. 2012; Nicholls and Minczuk 2014; Ubyaan et al.). Therefore, mtDNA alterations can result in a variety of single and multi-system disorders affecting various organs, including brain, heart, kidneys and skeletal muscles (Jamali et al. 2016).

Iran can be seen as a staging area for the Neolithic Agricultural Revolution, and the home to some of the earliest world empires. Iran, centered in the Asian continent, consists of diverse ethnic and linguistic groups, namely the Arabs, Armenians, Assyrians, Azeris, Baluchis, Gilak, Mazandaranis, Kurds, Lurs, Persian, Turkmen, and Zoroastrians. Based on mtDNA haplogroup diversity it is being suggested that in populations living in Iran and its neighboring regions e.g. Turkey, Georgia, and Central Asia, the main influx of gene flow has been from west (the Fertile Crescent) to east (Pakistan), a hypothesis that is also supported by genome wide association studies (GWAS)(Karimi et al. 2018) (Di Cristofaro et al. 2013). To date, in order to reveal the genetic structure, expansion patterns and population movements of Iranian subpopulations using complete mtDNA variation, two studies have been done (Schönberg et al. 2011; Derenko et al. 2013). However, further data will improve significance of genetic values, especially when there is no prior information regarding-specific ethnicity.

Thalassemia, the common single-gene disorder in humans, the most common hereditary anemia, containing more than 300 mutations, can manifest in three primary clinical forms due to mutations in the human β-globin locus, with major, intermediate and minor clinical manifestations (Galanello and Origa 2010). Iran is one of the major centers for the prevalence of thalassemia. Due to the high consanguinity in Iran’s different ethnic subpopulations, it is estimated that there are more than three million thalassemia carriers (4–8%) in Iran (De Sanctis et al. 2017). The thalassemia gene frequency is high and varies considerably among different Iranian geographical locations, with double the country average rate in the Mazandaran, Sistan and Baluchistan, Fars, Hormozgan, and Kerman provinces and half the average rate in Tehran, East Azerbaijan, Khorasan, Hamadan, Yazd, and West Azerbaijan regions. There are more than 47 different β-globin gene mutations, in which, the most predominant one is the IVS II-I (G→A) mutation (Rezaee et al. 2012; Salehi et al. 2010).

Cystic fibrosis (CF) is an autosomal recessive disorder which is considered as the most common cause of pancreatic insufficiency in children and one of the most
important reasons for chronic lung disease (Najafi et al. 2015). This disease is caused by a mutation in a CF trans-membrane conductance regulator gene regulating the activity of other chloride and sodium channels at the cell surface epithelium (Brzezinski et al. 2011; Rafeeq and Murad 2017). A heterogeneous mutation spectrum for the CFTR gene in 60 CF patients has been reported in Iran (Elahi et al. 2006). However, the most pronounced mutation, ΔF508 (p.F508del), represented only 16% of the expected mutated alleles. At a rate of 81.9%, mutation was detected in the CFTR gene, which is the highest recorded mutation detection rate for the CFTR gene (Alibakhshi et al. 2008). Dissection of the genetics underlying Iranian CF patients is appealing as it is reported that most frequent Iranian mutations are not included in a commonly reported set of CF mutations (Elahi et al. 2006). This underlines the importance of identifying geographic/ethnic-specific mutations in Iran. Exploring mutation in Iranian (sub) populations using either mtDNA or nuclear genomes is therefore an appealing route for investigation. The objectives of this study were to decipher the mtDNA relationships among nine different Iranian subpopulations/ethnic groups (Arab, Armani, Azeri, Bandari, Gilak, Jewish, Kurd, Lur, and Fars) e.g. haplotype sharing and nucleotide variation, with further examination of genetic mutation in two independent thalassemia and CF patients sample groups.

Results

Variant Detection and Frequency

We identified seven characterized mutations including IVS II-I (G→A), codons 36/37 (-T), IVS I-5 (G→C), codons 8/9 (+G), IVS I-110 (G→A), IVS I-6 (T→C) and IVS I-1 (G→A) in the thalassemia samples with 24, 10, 9, 7, 7, 2 and 2 counts, respectively. Using 402 mitochondrial D-loop sequences from nine different Iranian subpopulations and two independent patient groups, 242 haplotypes were revealed (GenBank accession number: MK562093-MK562334). Sequence alignment with the reference sequence (rCRS) resulted in 197 allelic variants with 33 different base variations (Table S1). From 197 variants, 14 allelic variants were detected in all studied subpopulations. The most frequent allele variants were 315C→CCC (91.8%), 311C→T (83.6%) and 73A→G (72.1%) which appeared in all subpopulations. We found 30 novel alleles amongst the 197 allelic variants from all sequences (Table S1). Allele variants of 310 T→CC (11.8%), 62G→CG (7.5%) and 81G→GA (7%) had the highest frequencies amongst all the novel variants. Overall, the most common nucleotide variations were C→T (23.1%) and T→C (22.4%) followed by A→G (13.3%) (Table S1). The 315C→CCC, 311C→T and 73A→G allelic variants were the most frequent in thalassemia (88.5%, 85.2% and 73.8%, respectively) and CF (92.9%, 85.7% and 71.4%, respectively) patients. It was also revealed that 39.3% of the thalassemia patients possessed the IVS II-I (G→A) mutation. The highest number of nucleotide variants in thalassemia patients were seen in the IVS II-I (G→A) mutation (51 variants), followed by IVS I-110 (G→A), IVS I-5 (G→C), codons 36/37 (-T), codons 8/9 (+G), IVS I-1 (G→A) and IVS I-6 (T→C) with frequencies of 45, 35, 22, 21, 10 and 8, respectively (Table S2).
Classification of Haplotypes and Haplogroups

Table 1 displays information about the subpopulation haplogroups. In total, we identified 108 haplogroups from three lineages (L, M, and N). The most common lineage was N (82.3%), followed by M (13.7%) and then L (4.0%). Additionally, the most common haplogroups in the N, M and L lineages were B4a1c3a (16.9%), M56 (18.2%) and L3e3’4’5 (43.8%), respectively. The highest number of haplotypes (78) and haplogroups (47) were found in the Fars subpopulation and the lowest was found within the CF patients (Table 1). Furthermore, the majority of sequences were classified in B4a1c3a (13.9%) and H2a2a1 (13.4%) haplogroups. Other frequent haplogroups were N10b (4%), H2a2a2 (3%) and J1 (3%). The most diverse observed haplogroups were B4a1c3a, H2a2a1, J1b8 and U7a4a1, with appearance in 11, 11, 8 and 7 studied subpopulations, respectively (Table 1). We found 62 unique haplogroups out of 108 detected haplogroups. The B5b1a2, K1a1, L2a2b2, M50a, O1, R1a and W7 groups in the thalassemia patient group, and the F2i, J1b7, L0k2, T1a1 + @152 and U2e1h groups in the CF samples were found to be unique in our study. The highest numbers of unique haplogroups were in the Fars (14) and Bandari (9) subpopulations. The Kurd and Lur subpopulations showed only two unique haplogroups and had the lowest number of unique haplogroups.

Clinical Subpopulation Haplogroups

The IVS II-I (G→A) mutation was the most diverse thalassemia mutation in our study. This illustrates the fact that the IVS II-I (G→A) mutation is most likely an endemic mutation, found in the entire Iranian population (assuming the thalassemia group represents an unbiased random sample of the entire population). In this way, the H (20.8%), B (16.7%), and W (16.7%) groups were the most frequent haplogroups in thalassemia patients with the IVS II-I (G→A) mutation (Table 2). Overall, the H (27.9%)—which is by far the most common mtDNA lineage in West Eurasian human populations—and B (14.8%) haplogroups represented the highest frequencies in the thalassemia group. However, as we addressed already, the H haplogroup has been predominant in the Middle East, Western and Eastern Europe, Caucasus, Central Asia and Africa. The most common haplogroup found in the CF patients was the U haplogroup (21.4%).

Comparisons of Subpopulation Diversity

Haplotype and nucleotide diversity of all the subpopulations are represented in Table 1. Total haplotype (0.9916 ± 0.0014) and nucleotide (0.00793 ± 0.00023) diversities were high in this study. Generally, all subpopulations revealed high diversity, with the highest and lowest genetic diversity being observed in the Arab
| Geographical location          | n* | k* | Haplogroup (number of samples)                                                                 | Haplotype diversity | Nucleotide diversity           |
|-------------------------------|----|----|-----------------------------------------------------------------------------------------------|---------------------|--------------------------------|
| Arab Khuzestan (Ahvaz)        | 24 | 20 | B4a1c3a (2), H2a2al (3), J1 (2), J1b8 (1), J1d1a1a (1), L2’3’4’5’6 (1), M53 (1), M56 (1), N10a (1), N3 (2), R0a‘b (2), R1a (1), U1 (1), U3a2al (1), U6a8a (1), U7a4al (2), X1’3 (1) | 1 ± 0.016          | 0.01063 ± 0.00129               |
| Armani Tehran (Aramaneh center)| 26 | 24 | B4a1c3a (7), H2a2al (4), H2a2alf (1), J1 (1), J1b5a (1), J1b8 (1), K1a (1), K1alc (1), M56 (1), N10 (1), R0a1a (1), R1a (1), U1a1a (2), U1a1a3 (1), U1a1a3 (1), W (1), X2+225 (1) | 1 ± 0.012          | 0.00756 ± 0.00074               |
| Azeri East Azerbaijan (Tabriz) | 25 | 22 | B4a1c3a (3), C4a2b2 (1), H2a2al (1), H2a2a2 (2), J1b8 (1), L0 (1), L3e3’4’5 (1), M33a1a (2), M3a1+204 (1), M56 (2), M62a (1), N1o (1), N10b (1), R2+13500+195 (1), U5a2e (3), U7a4a1 (1), W+194 (1) | 1 ± 0.014          | 0.00831 ± 0.00108               |
| Bandari Bushehr (Bushehr)     | 40 | 34 | A15 (1), B4a1c3a (8), C4a2b2 (1), D2c (2), H2a2al (4), H2a2a1f (1), H2a2a2 (1), H6 (1), HV2a3 (1), J1b1b (1), J1b8 (2), J2a1a1a (2), J2a2 (1), K1a1b2a (1), L0a2a2 (2), L3e3’4’5 (1), L3f1b (1), M3 (1), M33a2a (1), M56 (1), M7c1 (1), N22a (1), W+194 (1), W4 (1), W5a2 (2) | 1 ± 0.007          | 0.00918 ± 0.00083               |
| Gilak Gilan (Rasht)           | 24 | 21 | B4a1c3a (3), H2a2al (2), J1b1a (1), J1b1a1+146 (1), J1b8 (1), J1c1 (1), K1a (2), M27c (1), M33a1a (1), M56 (1), N10 (2), N1a1b (1), N1a3 (1), R2+13500+195 (1), R8b (1), U4a2 (1), U6a7b (1), U7a4a1 (2) | 1 ± 0.015          | 0.00856 ± 0.00095               |
| Johoud Tehran (Johoud center)  | 40 | 33 | B4a1c3a (4), D2c (1), D4b4a (1), H15a1b (1), H2a2al (6), H2a2a1f (2), H2a2a2 (1), J1b1b (3), J1c (1), J1c16 (1), J2 (1), J2a1a1a (1), K1a (1), K1b1c (1), M29a (1), M36b (1), M62a (3), N10 (1), N10b (2), N2a (1), R2+13500+195 (1), U4a2 (1), U5a2e (3), U5b1b (2) | 1 ± 0.007          | 0.00821 ± 0.00071               |
| Geographical location                                      | \( n^* \) | \( k^* \) | Haplogroup (number of samples) | Haplotypic diversity | Nucleotide diversity       |
|-----------------------------------------------------------|----------|----------|--------------------------------|----------------------|----------------------------|
| Kurd Kurdistan (Sanandij)                                 | 25       | 23       | B4a1c3a (5), H2a2a1 (1), H2a2a2 (2), J1 (1), J1b1b (2), J1b8 (2), J1c (1), K1a+150 (1), K1a4j1 (1), M3a1a (2), M62a (2), N10 (1), N10b (1), U1a1a (1), U7a4a1 (1), W4 (1) | 1 ± 0.013              | 0.00807 ± 0.00093          |
| Lur Lurestan (Khorramabad) and Chaharmahal and Bakhtiari (Shahr-e Kurd) | 25       | 20       | B4a1c3a (2), H2a2a1 (4), H2a2a2 (2), J1b1a1+146 (1), J1b1b (1), J1b1b1c (2), J1b1a (2), K1a (1), L3c (1), L3e3'4'5 (2), N10 (1), N1a1b (1), N3 (1), U4a1d (1), U7a4a1 (1) | 1 ± 0.016              | 0.00759 ± 0.00070          |
| Fars Esfahan (Isfahan), Razavi Khorasan (Mashhad), Fars (Shiraz), Yazd (Yazd) | 98       | 78       | B4a1c3a (11), B4b (1), C4a2b2 (1), D2c (3), D5a (2), Fl1a1c (1), G2b1a2 (1), Hl5a1b (1), H2a2a1 (13), H2a2a2 (4), H6a1a7 (1), I1 (2), J1 (2), J1b1a (2), J1b1b1a+146 (1), J1b1b (1), J1b1b1c (1), J1b8 (2), J2a1a1a (2), J2a1a1al (1), L2d (1), L3e3'4'5 (2), M29a (1), M2a1a3 (1), M3 (1), M30b (1), M56 (4), M62a (2), M7b1a1a3 (1), N10b (9), N1a1b (1), N1a3 (1), N3 (2), R2+13500+195 (1), R6b (1), R8b (1), T1a7 (1), T2b34 (1), U1 (1), U1a1b (2), U4a2 (1), U5a2e (1), U7a4a1 (2), W+194 (3), W+1194 (1), W4 (1), W4a (1) | 1 ± 0.002              | 0.00838 ± 0.00043          |
| CF Random samples from different populations in National Institute of Genetic Engineering and Biotechnology Clinic in Tehran | 14       | 14       | B4a1c3a (1), F2i (1), H2a2a1 (2), H2a2a2 (2), J1b7 (1), J1b8 (1), K1a (1), L0k2 (1), L3e3'4'5 (1), M29a (1), T1a1a+152 (1), U2e1h (1), U5a2e (1), U5b1b2 (1) | 1 ± 0.027              | 0.00717 ± 0.00077          |
| Geographical location | n | Haplogroup (number of samples) | Haplotype diversity | Nucleotide diversity |
|-----------------------|---|--------------------------------|---------------------|---------------------|
| Thalassemia Random samples from different populations in Medical Genetic Laboratory (Special Medical Center in Tehran) | 61 | 50 | B4a1c3a (10), B5b1a2 (1), D2a (2), H2a2a1 (14), H2a2a1f (2), H7i1 (2), J1 (3), J1b1a (2), J1b1b (1), J1b5a (1), J1c (1), J2a1a (2), K1a (2), K1a1 (1), L2a2b2 (1), M33a1a (1), M50a (1), M62a (1), N10b (3), N1a1b (1), O1 (1), O1a1a (1), U4a2 (1), U5a2 (1), U5b1b2 (1), U7a4a1 (1), W+194 (1), W7 (1) | 1 ± 0.004 | 0.00870 ± 0.00075 |

* n = number of samples an k = number of haplotyp
### Table 2 Distribution pattern of mtDNA haplogroups

| Subpopulation  | Africa | Asia | Europe | Total |
|----------------|--------|------|--------|-------|
|                | L0     | L2   | L3     | Total | H     | HV    | I     | J     | K     | R     | T     | U     | W     | X     | Total |
| Iranian populations |       |      |        |       |       |       |       |       |       |       |       |       |       |       |       |
| Arab           | 1      | 1    | 2      | 4     | 3     | 5     | 1     | 16    |       |       |       |       |       |       |       |       |
| Armani         | 0      | 7    | 1      | 9     | 5     | 3     | 2     | 2     | 3     | 1     | 1     | 17    |       |       |       |
| Azeri          | 1      | 1    | 2      | 3     | 6     | 2     | 12    | 1     | 1     | 1     | 1     | 4     | 1     | 11    |       |
| Bandari        | 2      | 2    | 4      | 1     | 8     | 1     | 2     | 17    | 7     | 1     | 6     | 1     | 4     | 19    |       |
| Gilak          | 0      | 3    | 3      | 4     | 10    | 2     | 4     | 2     | 4     | 14    |       |       |       |       |       |
| Johoud         | 0      | 4    | 2      | 5     | 4     | 15    | 10    | 7     | 2     | 1     | 5     | 25    |       |       |       |
| Kurd           | 0      | 5    | 4      | 2     | 11    | 3     | 6     | 2     | 2     | 1     | 14    |       |       |       |       |
| Lur            | 3      | 3    | 2      | 3     | 5     | 4     | 1     | 9     | 1     | 1     | 2     | 17    |       |       |       |
| Fars           | 1      | 2    | 3      | 12    | 1     | 5     | 1     | 11    | 13    | 44    | 19    | 2     | 12    | 3     | 2     | 7     | 6     | 51    |
| Total          | 3      | 2    | 8      | 13    | 1     | 46    | 3     | 9     | 1     | 1     | 36    | 33    | 0     | 130   | 56    | 1     | 4     | 52    | 10    | 12    | 32    | 13    | 2     | 184   |
| Cystic fibrosis| 1      | 1    | 2      | 1     | 1     | 1     | 3     | 2     | 2     | 1     | 1     | 3     | 9     |       |       |       |       |       |
| Thalassemia    |        |      |        |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| IVS II-1 (G → A) | 1      | 1    | 4      | 2     | 3     | 9     | 5     | 2     | 1     | 2     | 4     | 14    |       |       |       |       |       |
| Codons 36/37 (-T) | 0      | 2    | 1      | 1     | 1     | 5     | 1     | 1     | 1     | 2     | 5     |       |       |       |       |       |       |       |
| IVS I-5 (G → C) | 0      | 1    | 1      | 1     | 3     | 4     | 1     | 1     | 1     | 6     |       |       |       |       |       |       |       |
| Codons 8/9 (+G) | 0      | 1    | 1      | 1     | 2     | 3     | 1     | 1     | 5     |       |       |       |       |       |       |       |
| IVS I-110 (G → A)| 0      | 1    | 1      |       | 1     | 6     |       |       |       |       |       |       |       |       |       |       |       |
| IVS I-6 (T → C) | 0      | 0    | 1      |       | 1     | 2     |       |       |       |       |       |       |       |       |       |       |
| IVS I-1 (G → A) | 0      | 0    | 2      |       | 2     |       |       |       |       |       |       |       |       |       |       |       |
| Total          | 0      | 1    | 0      | 1     | 0     | 9     | 0     | 2     | 0     | 0     | 3     | 5     | 1     | 20    | 17    | 0     | 0     | 8     | 2     | 0     | 6     | 5     | 0     | 40    |

mtDNA haplogroup counts in each subpopulation. The haplogroup counts are hierarchically clustered along both axes. The y-axis corresponds to the subpopulation group, and the x-axis corresponds to the previously characterized mtDNA macro-haplogroups. The continental origins of the mtDNA macro-haplogroups are also shown.
(0.01063±0.00129) and CF (0.00717±0.00077) subpopulations, respectively (Table 1). Table 3 shows subpopulation diversity comparisons. In total, 46 out of 91 comparisons were significantly different ($P < 0.05$). The majority of significant comparisons were observed in the African, East Asian and European populations ($P < 0.05$). The highest significant genetic comparisons among Iranian subpopulations in this study, belonged to the Azeri subpopulation which highlights the genetic isolation and high consanguinity in this Iranian subpopulation. The lowest significant genetic comparisons were seen in the Kurd and CF subpopulations. Both the Kurd and CF subpopulations did not show significant different genetic structure from the European population ($P < 0.05$). However, the thalassemia patients displayed six significant differences with the Arab, Azeri, Bandari, African, East Asian and European populations ($P < 0.05$). Additionally, the Bandari and Lur subpopulation revealed six significant differences with the other populations ($P < 0.05$), and the Arab, Johoud and Fars subpopulation displayed five significant differences with the other populations ($P < 0.05$). The Gilak subpopulation showed four significant differences with the other subpopulations ($P < 0.05$). We used Multidimensional Scaling (MDS) approach to compare genetic distances of the Iranian subpopulations (Fig. 1). To do this, a matrix of mtDNA genetic distances were used to compute the MDS plot. The Arab population was found to be distant from Azeri and Kurd subpopulations, with Bandari and Johoud subpopulations also showing a remote genetic relationship.

Discussion

Despite the larger sample size of Fars subpopulation, the highest number of unique variants belonged to the thalassemia patients (23 variants), of which 8 were novel. This might indicate that the thalassemia patient group can be treated as an independent group here, constituting people from different ethnic groups. The Fars and Bandari subpopulations showed 17 and 14 unique variants, respectively. However, only 4 unique allelic variants were observed in the independent CF group. Previous studies in Iranian subpopulations have shown that the thalassemia patients had the highest frequency of the IVS II-I (G→A) mutation, especially in the northern, central, and some areas of southern Iran (Karimi et al. 2000, 2002; Rezaee et al. 2012). In essence, there are some pieces of evidence indicating that the IVS II-I (G→A) mutation has entered from Iraq and then expanded in Iran (Al-Allawi et al. 2006; Rezaee et al. 2012). Additionally, in some of Iran’s neighboring countries including Kuwait, Iraq, the eastern province of Saudi Arabia, Azerbaijan, Jordan, and Syria, it is reported that the IVS II-I (G→A) mutation is the most frequent one seen in thalassemia patients (De Sanctis et al. 2017). It is most likely that the IVS II-I (G→A) mutation in thalassemia patients is not only the most frequent, but also the most diverse.

The observed mtDNA haplogroup distribution in our study confirmed the previously characterized global distribution. For instance, the ancestral L, H, and U haplogroups are predominant in Africa, H, U and J haplogroups in Middle East, the D, B, M, F and N haplogroups in East Asia and H, D, C and U haplogroups are
### Table 3  Fst values of subpopulation comparisons (below diagonal)

|        | Arab | Armani | Azeri | Bandari | Gilak | Johoud | Kurd | Lur | Fars | CF    | Thalasemia | African | East Asian | European |
|--------|------|--------|-------|---------|-------|--------|------|-----|------|-------|------------|---------|-----------|----------|
| Arab   | –    | +      | –     | –       | –     | –      | –    | –   | –    | +     | +          | +       | +         | +        |
| Armani | 0.011| +      | –     | –       | –     | –      | –    | –   | –    | –     | +          | +       | +         | +        |
| Azeri  | 0.041| 0.056  | –     | +       | +     | –      | +    | –   | +    | –     | +          | +       | +         | +        |
| Bandari| 0.021| 0.011  | 0.016 | –       | +     | –      | +    | –   | –    | –     | +          | +       | +         | –        |
| Gilak  | –0.013| 0.011  | 0.031 | 0.003   | –     | –      | –    | –   | –    | –     | +          | +       | +         | +        |
| Johoud | 0.011| 0.006  | 0.028 | 0.018   | 0.015 | –      | –    | –   | –    | –     | +          | +       | +         | +        |
| Kurd   | 0.007| –0.004 | 0.019 | –0.002  | –0.015| –0.015 | –    | –   | –    | –     | +          | +       | +         | –        |
| Lur    | 0.003| 0.012  | 0.068 | 0.028   | –0.008| 0.011  | –0.008| +   | –    | –     | +          | +       | +         | +        |
| Fars   | 0.019| 0.002  | 0.024 | 0.002   | 0.005 | 0.011  | –0.010| 0.028| –    | –     | +          | +       | +         | +        |
| CF     | 0.013| 0.003  | –0.012| –0.007  | 0.016 | –0.019 | –0.017| 0.020| –0.003| –     | +          | +       | +         | –        |
| Thalasemia | 0.030 | –0.006 | 0.055 | 0.017   | 0.017 | 0.006  | –0.012| 0.017| 0.001| –0.000| +          | +       | +         | +        |
| African| 0.120| 0.141  | 0.087 | 0.096   | 0.103 | 0.142  | 0.122 | 0.145| 0.135| 0.092 | 0.162    | +       | +         | +        |
| East Asian | 0.117| 0.141  | 0.142 | 0.120   | 0.120 | 0.114  | 0.109 | 0.107| 0.123| 0.116 | 0.123    | 0.171   | +         | +        |
| European| 0.048| 0.019  | 0.049 | 0.016   | 0.028 | 0.043  | 0.017 | 0.043| 0.013| 0.018| 0.014    | 0.173   | 0.146     | +        |

Pluses (+) and minuses (−) display the existence and non-existence of significant difference amongst subpopulations (above diagonal) ($P$ value < 0.05)
Fig. 1 MDS of genetic distances among studied subpopulations
most frequently found in Central Asia (Chen et al. 1995; Comas et al. 2004; Lott et al. 2013; Rishishwar and Jordan 2017). Also, H, U, T and J haplogroups are the predominant ones in Europe (Lott et al. 2013; Rishishwar and Jordan 2017; Torroni et al. 1996). However, A, B, C and D constitute the majority of haplogroups in North, South and Central American populations, with N and B haplogroups the most predominant in Australia and Oceania (Lott et al. 2013; Rishishwar and Jordan

### Table 4 Major haplogroups of Iranian ethnic subpopulations and their geographical distribution pattern

| Subpopulation       | Major haplogroups and distribution                                                                                                                                 |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Armani and Bandari  | We found the B (26.9%) and H (19.2%) haplogroups the most prevalent ones in the Armani subpopulation (Table 2). The Armani subpopulation is scattered all over the country and mostly are inhabitant of the Tehran, Gazvin, Shahrekord, Abadan and Bushehr cities. In addition, the B (20%) and H (17.5%) haplogroups were the profound ones in Bandari subpopulation. The Bandari subpopulation is inhabitant of the Southern regions of Iran. The most predominate haplogroup, B, in the Armani and Bandari populations had been shown to be one the most frequent haplogroups in the South and Central Americas (27%), North Americas (23%) and Oceania (23%) populations. |
| Azeri               | Geographically, the Azeri subpopulation is mainly inhabitant of the northwest and northern regions of Iran. Our results displayed that the M (24%) and U (16%) haplogroups were the prevalent ones in this subpopulation (Table 2). The M haplogroup is one of the predominant haplogroups in East Asia. Moreover, the northwest regions of Iran are close to the Caucasus and East Europe regions in which the U haplogroup is very common, so the high frequency of this haplogroup in the Azeri subpopulation could be expected. |
| Gilak               | The Gilak subpopulation, located in the northern regions of Iran, revealed diverse N (16.7%), J (16.7%) and U (16.7%) dominant haplogroups (Table 2). This can give us the possibility of high admixture of Gilak subpopulation, owing to the fact that there are different haplogroups with high frequencies in this subpopulation. |
| Arab                | The Arab subpopulation lives in the South and Southwest regions of Iran. The U (20.8%) and J (16.7%) haplogroups represented the most frequent ones in the Arab subpopulation (Table 2). Interestingly though, previous investigations in Caucasus and East Europe disclosed that the U (with 22% frequency) haplogroup were the most common one in these regions. |
| Johoud              | The Johoud subpopulation is scattered all over cities of Iran and mostly are the inhabitant of the Shiraz, Isfahan, Hamadan, Yazd, Kerman, Rafsanjan, Sirjan and Borujerd cities. The two profoundly appeared haplogroups in Middle East, i.e. H (25%) and J (17.5%) haplogroups, displayed high frequencies in Johoud subpopulation in our study (Table 2). |
| Kurd and Lur        | The Kurd and Lur subpopulations are predominantly residents of western regions of Iran. In our study, the J (24% for Kurd and 36% for Lur populations) haplogroup was the most common one in both populations. As the matter of fact, the J haplogroup proven to be one of the most prevalent haplogroups in Middle East populations. |
| Fars                | The Fars subpopulation, the most frequent population of Iran, are mostly the inhabitants of East (especially northeast) through center to lower frequencies in southern regions of Iran. Originally, the Fars subpopulation are mainly located in Mashhad, Isfahan, Kerman, Shiraz and Yazd cities. The H (19.4%) and N (13.3%) haplogroups were the most frequent ones in the Fars populations. As aforementioned, the H haplogroup is the most frequent one in Middle East population. |
Despite low number of CF haplotypes, we identified 10 out of 22 macro-haplogroups that were also seen in various main African, Middle Eastern, Caucasus and East European haplogroups (Table 2). Interestingly, our results displayed high haplogroup diversities along with admixture of the European, American and East Asian populations. In essence, the secular trade along the great Silk Road which extends from Xian in China through the Indian subcontinent to Iran and the Eastern Mediterranean, along with the invasions of the Mongols (1220 A.D.) and the Tatars (1380–87 A.D.) could be historical events that have resulted in such diverse and admixed Iranian haplogroups (De Sanctis et al. 2017; Rezaee et al. 2012). Previous study in the African, East Asian and European populations showed a tight coherent patterns of mtDNA haplogroup distributions, whereas the Indian and American population haplogroups represented more divergent haplogroups consistent with their admixed origins (Rishishwar and Jordan 2017). The Indian populations, like Iranian subpopulations in our investigation, has been shown to be formed of a combination of European and Asian mtDNA haplogroups, along with relatively ancient human immigration and admixture events (Kivisild et al. 1999; Moorjani et al. 2013). As Table 2 shows, we found some ancient African L haplogroups in our haplotypes which would likely reflect the facts of expansion of these haplogroups in the Near East, and also, immigration between Africa and the Near East (Maca-Meyer et al. 2003). However, along with Iranian ethnic populations, the tremendous haplogroup diversities in thalassemia and CF patients could be the result of admixture of different global human populations due to multiple historical events and the existence of high consanguinity in different subpopulations.

The high haplotype and nucleotide diversities could shed some light on the existence of high genetic diversity in all subpopulations sequences. Previous investigations in Iranian subpopulations revealed high frequencies of thalassemia in the Bandari, Gilak and Fars subpopulations and a low frequency in the Azeri population (De Sanctis et al. 2017). Other studies have revealed a probable ancient and recent gene flow between Iranian populations and the Indian sub-continent and the Arabian Peninsula (Derenko et al. 2013). However, certain mtDNA haplogroups indicate barriers to gene flow due to two major Iranian deserts and the Zagros mountain range (Derenko et al. 2013). Additionally, the results from MDS indicates their evolutionary genetic distance. With unprecedented increasing whole genome data and reliable computational power, genomic simulations and statistical inference in lieu of model-based approaches; a finer scale view of the mtDNA of nine group would be possible. We are expecting using these tools and more well representative ethic group sample, a better spectacular views of their genome proximity could be grasped. It is crucial to say that these methods may help to integrate data from non-genetic factors and sources, a matter that shall help to make accurate inferences from genetic data and to improve our interpretations of ancient events. Interdisciplinary approaches will be essential as we continue to move forward in disentangling human evolutionary history. The interdisciplinary nature of genetic anthropology places us in an ideal position to take advantage of these approaches.
This study has its limitations. We did not calculate the statistical power of our mtDNA data (i.e. beta, the probability of rejecting the null hypothesis when actually it is true). As we were constrained by the small number of mtDNA sequences per ethnic group, this may have affected the final results. Although the Iranian population is made up of geographically, ethnically and linguistically diverse groups, likely created from prehistoric post-glacial expansions, the common underlying gene pool and gene flow among them is expected. In our study this was clearly observed. Assuming a common underlying gene pool and gene flow, there were no genetic differences among the studied subpopulations. Due to some issues like sampling schemes, retracing gene flow of the present Iranian populations has been strongly restricted. To further elaborate on these results, a comparative analyzes of paternal (Y-Chromosome) and maternal (mtDNA) lineages in these subpopulations needs to be done in the future. The normalized ratio of the Y-Chromosome FST distance with respect to the total distance (Y-Chromosome RST + mtDNA) could also help shed some more light in this context. However, a contrast between the results of mtDNA and Y-Chromosome has been observed (Badro et al. 2013).

In summary, the subpopulations displayed high genetic diversity along with diverse haplogroups. The IVS II-I (G→A) mutation was not only the most common thalassemia mutation, but also the most diverse one across all subpopulations. The most significant haplogroups were B4a1c3a, H2a2a1, N10b, H2a2a2 and J1. Our results showed high haplogroup diversity along with admixture of the European, American and East Asian common haplogroups which could be as a result of the secular trade along the great Silk Road and invasion of the Mongols (1220 A.D.) and the Tatars (1380–87 A.D.) as historical events. In fact, we could say that the multi-ethnic population context and complex historical events of Iran resulted in the mtDNA heterogeneity. Few subpopulations especially the e.g. Azeri, and Arab groups are shown to be conserved and display signs of genetic isolation. The thalassemia group which, like the CF group, is assumed to constitute random sampling of the total Iranian population, revealed differences with the Arab, Azeri, and Bandari subpopulations, indicating that this disease is more common in the other subpopulations (i.e. Armani, Gilak, Johoud, Kurd, Lur, Fars). In this way, due to high number of unique and novel nucleotide variations in the thalassemia group, we could claim that these variations are likely associated with this disease in our study.

Thalassemia, heterogeneous at the molecular level, is derived by many mutations (more than 300) leading to a failure of hemoglobin and has relatively high prevalence in Iran. Also, for CF investigated in this study, which is the most common autosomal recessive disorders around the world, the finding indicates even though ΔF508 is the most frequent mutation in populations like Caucasian, but it has not a pivotal role in Iranian CF groups. This study had no mean to associate prevalence of aforementioned diseases with mtDNA sequence information; but tried to pinpoint the matrilineal lineage of CF and Thalassemia groups with other ethical groups in terms of sharing mtDNA.
Methods

Ethical Considerations

The study was approved and monitored by the ethics committee of the National Institute of Genetic Engineering and Biotechnology (IR.NIGEB.EC.1398.12.3.E). All methods were carried out in accordance with relevant guidelines and regulations. All participants gave written informed consent to participate in the study and for their anonymized data to be used for statistical analysis and dissemination, also none of them were under 18 years of age.

Sample Collection and DNA Extraction

Figure 2 displays the flowchart of our survey, including data collection, laboratory procedures and statistical analysis. Geographical location and number of each subpopulation samples are shown in Table 1. The study was conducted in accordance with the methodology from previous studies (Akbari et al. 2008; Derakhshandeh et al. 2008; Najmabadi et al. 2001; Rezaee et al. 2012) e.g. sample collection, DNA extraction, primers, PCR amplification and sequencing to extra HVSI and HVSII D-loop data. Briefly, total DNA samples were extracted from blood using a QIAamp DNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer’s instruction. The quality and purity of the extracted DNA were measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

PCR Amplification

The mtDNA was amplified using a set of specific primers (numbering according to NCBI Accession No NC_0122920.1): PF, 5′-ATCATTGGACAAGTAGCATC-3′ (15,791–15,810 bp) and PR, 5′-GAGCTGCATTGCTGCTGCT-3′ (780–761 bp). Polymerase chain reaction (PCR) amplification was carried out using TEMPase Hot Start 2 × Master Mix A BLUE (Ampliqon, Odense, Denmark) in a final reaction volume of 50 μL, containing 100 ng of DNA, 0.32 μL of each primer (10 pmol), 25 μL of TEMPase 2 × Master Mix, and 23.2 μL RNase-free water. PCR amplification was performed with the following program: pre-PCR incubation at 95 °C for 15 min, 35 cycles of 95 °C for 20 s, annealing at 60 °C for 45 s, and extension at 72 °C for 30 s, with a final extension at 72 °C for 5 min. The specific amplification of a 1550 bp fragment was confirmed by 1.5% agarose gel electrophoresis.

Sequencing

PCR products were sequenced by direct DNA sequencing (Bioneer, South Korea). The sequencing results were analyzed, using Codon Code Aligner 6.0.2 software (Codon Code, Centerville, MA, USA, https://www.codoncode.com/aligner/new60.htm). The sequences were compared to the revised Cambridge Reference Sequence (rCRS) (Accession No NC_012920.1), using the BLAST sequence analysis tool.
(NCBI, Bethesda, MD, USA). The sampling scheme of the thalassemia and CF patient groups were quite different than from the nine subpopulations mentioned the above, in a way that individuals constituting these samples were assumed to be an independent random sample of the Iranian population rather than explicitly from a given ethnic group. Therefore, we could say, eleven groups were genetically compared e.g. nine subpopulations with two patient groups.

**Statistical Analyses**

To compare all the sequenced mtDNA D-loop samples to other datasets and reference sequence, we edited and reduced the sequences size to leave the region spanning nucleotides 47 to 766 of the reference sequence. This segment includes
the HVSII region. Editing was done using BioEdit software 7.0.9.0 (https://bioedit.software.informer.com/7.2/) (Hall 1999). In addition, we used GenBank mtDNA data from the African (EF184580-EF184641 and FJ236978-FJ236983), East Asian (DQ826448 and DQ834253-DQ834261, EF446784 and EF488201 and DQ418488, DQ437577, DQ462232-DQ462234 and DQ519035) and European (DQ523619-DQ523681 and AY339402-AY339593) populations (Yao et al. 2009) to assess our population sequences. Sequence alignment was performed with a reference sequence (NC_012920.1) using the CLUSTALW package (Thompson et al. 1994). The polymorphic sites in the aligned sequences were excluded from the analyses, using DNASP 5.10 software (http://www.ub.edu/dnasp/) (Librado and Rozas 2009). Similar sequences were considered as the same haplotype. The haplotype diversity (h) and nucleotide diversity (π) for subpopulations were estimated by DNASP 5.10 software (http://www.ub.edu/dnasp/) (Librado and Rozas 2009). The haplotypes were deposited in GenBank under accession numbers MK562093 to MK562334. In order to evaluate the allelic variants we used Mitomap (https://www.mitomap.org/MITOMAP) and to elucidate the subpopulation haplogroups we submitted the haplotypes to HaploGrep 2 (v2.1.19) software (https://haplogrep.uibk.ac.at). In addition, to compare the genetic structure of subpopulations, we performed Analysis of Molecular Variance (AMOVA), using Arlequin 3.5.2.2 software (http://cmpg.unibe.ch/software/arlequin35/) with the Tamura and Nei method at 0.05 significance level (Excoffier and Lischer 2010). Additionally, XLSTAT, a Microsoft Excel add-on, was used to create multidimensional scaling (MDS) figures. In general, MDS which is a family of different algorithms, each strive to arrive at optimal low-dimensional structures can be accomplished as Classical MDS, Metric MDS and Non-metric MDS. It is patently clear that we resorted on Metric MDS, as genetic distance is in essence a metric measure. However, applying different versions of Metric MDS e.g. Absolute, Ratio, Interval, and Polynomial, different results turned out, even though, the stress function values of MDS were almost the same. However, we used the one with the lowest stress function value.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s10528-021-10141-z.

Acknowledgements All the authors acknowledge and thank their respective Institutes and Universities.

Author Contributions AS and MH provided and sequenced the data, MM, was heavily involved in statistical analysis of the data and paper preparation, MG-Z, was the principal investigator of this study and developed the project proposal, AS, AO and JS edited paper and provided intellectual input. MA, and SHAR provided constructive suggestions for the discussion. MHS was the head of the Human genome diversity project in Iran and helped to collect the samples. All authors revised and approved the final manuscript.

Funding Not applicable.

Data Availability The dataset supporting the conclusions of this article is available in the GenBank public archive under accession numbers MK562093-MK562334.
Declarations

Conflict of interest The authors declare that they have no competing interests. All authors are working as academic staff and research associates in universities and are not employed by the government, do not represent the government and the article is not submitted on behalf of the government.

References

Akbari MT, Izadi P, Izadyar M, Kyriacou K, Kleanthous M (2008) Molecular basis of thalassemia intermedia in Iran. Hemoglobin 32:462–470
Al-Allawi NA, Jubrael JM,Hughson M (2006) Molecular characterization of β-thalassemia in the Dohuk region of Iraq. Hemoglobin 30:479–486
Alibakhshi R, Kianishirazi R, Cassiman J-J, Zamani M, Cuppens H (2008) Analysis of the CFTR gene in Iranian cystic fibrosis patients: identification of eight novel mutations. J Cyst Fibros 7:102–109
Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghashibe-Sabbagh M, Johnsrud B, Khaszen G, Matissoo-Smith E, Soria-Hernanz DF, Wells RS (2013) Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern Middle Eastern populations with European and African populations. PLoS ONE 8:e54616
Bircan R, Gözü HI, Esra U, Sarikaya Ş, Gül AE, Şirin DY, Özçelik S, Aral C (2018) The mitochondrial DNA control region might have useful diagnostic and prognostic biomarkers for thyroid tumors. bioRxiv 291435
Brzezinski LXC, Riedi CA, Kussek P, Souza HHMD, Rosário N (2011) Nebulizers in cystic fibrosis: a source of bacterial contamination in cystic fibrosis patients? J Bras Pneumol 37:341–347
Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133
Chial H, Craig J (2008) mtDNA and mitochondrial diseases. Nat Educ 1:217
Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504
De Sanctis V, Kattamis C, Canatan D, Soliman AT, Elsedfy H, Karimi M, Daar S, Wali Y, Yassin M, Soliman N, Sobi P (2017) β-thalassemia distribution in the old world: an ancient disease seen from a historical standpoint. Medit J Hematol Infect Dis 9:e2017018
Derakhshandeh PP, Hour FH, Heydari M, Mir YM (2008) The spectrum of beta-thalassemia mutations in Isfahan Province of Iran. Iran J Public Health 37(2):106–111
Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, Yekipkosposyan L (2013) Complete mitochondrial DNA diversity in Iranians. PLoS ONE 8:e80673
Di Cristofaro J, Pennarun E, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA (2013) Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. PLoS ONE 8:e76748
Elahi E, Khodadad A, Kupershmidt I, Ghasemi F, Alinasab B, Naghizadeh R, Eason RG, Amini M, Esmaili M, Dooki M, Sanati MH (2006) A haplotype framework for cystic fibrosis mutations in Iran. J Mol Diagn 8:119–127
Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567
Galanello R, Origa R (2010) Beta-thalassemia. Orphanet J Rare Dis 5:11
Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Paper presented at the Nucleic acids symposium series. 41, 95-98
Jamali L, Banoei MM, Khalili E, Dadgar S, Houshmand M (2016) Association of genetic variations in the mitochondrial D-loop with β-thalassemia. Mitochondrial DNA Part A 27:1693–1696
Karimi M, Kadivar M, Alaviani GA (2000) Regional mapping of the gene frequency of B-Thalassemia in Fars Province, Iran during 1997–1998. Iran J Med Sci 25:134–137
Karimi M, Yarmohammad H, Farjadian S, Zeinali S, Moghaddam Z, Cappellini MD, Giordano PC (2002) β-Thalassemia intermedia from southern Iran: IVS-II-1 (G→A) is the prevalent thalassemia intermedia allele. Hemoglobin 26:147–154
Karimi N, Alibakhshi R, Almasi S (2018) CFTR mutation analysis in western Iran: identification of two novel mutations. J Reprod Fertil 19:3–9
Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papila SS (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC (2013) mtDNA variation and analysis using MITOMAP and MITOMASTER. Curr Protoc Bioinform 44:1–23
Maca-Meyer N, González AM, Pestano J, Flores C, Larruga JM, Cabrera VM (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genet 4:15
Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P, Berger B, Reich D, Singh L (2013) Genetic evidence for recent population mixture in India. Am J Hum Genet 93:422–438
Najafi M, Alimadadi H, Rouhani P, Kiani MA, Khodadad A, Motamed F, Moraveji A, Hooshmand M, Haghi Ashtiani MT, Rezaei N (2015) Genotype-phenotype relationship in Iranian patients with cystic fibrosis. Turk J Gastroenterol 26:241–243
Najmabadi H, Karimi-Nejad R, Sahebjam S, Pourfarzad F, Teimourian S, Sahebjam F, Amirizadeh N, Karimi-Nejad MH (2001) The β-thalassemia mutation spectrum in the Iranian population. Hemoglobin 25:285–296
Ngili Y, Ubyaan R, Palit EI, Bolly HM, Noer AS (2012) Nucleotide mutation variants on D-Loop HVS1/HVS2 mitochondrial DNA region: studies on Papuan population, Indonesians. Eur J Sci Res 72:64–73
Nicholls TJ, Miniczuk M (2014) In D-loop: 40 years of mitochondrial 7S DNA. Exp Gerontol 56:175–181
Rafeeq MM, Murad HAS (2017) Cystic fibrosis: current therapeutic targets and future approaches. J Transl Med 15:1–9
Rezaee AR, Banoei MM, Khalili E, Houshmand M (2012) Beta-Thalassemia in Iran: new insight into the role of genetic admixture and migration. Sci World J. https://doi.org/10.1100/2012/635183
Rishishwar L, Jordan IK (2017) Implications of human evolution and admixture for mitochondrial replacement therapy. BMC Genom 18:1–11
Salehi R, Fisher CA, Bignell PA, Eslami G, Old JM (2010) Identification of three novel mutations [− 41 (A>C), codon 24 (−G), and IVS-I-109 (− T)], in a study of β-Thalassemia Alleles in the Isfahan Region of Iran. Hemoglobin 34:115–120
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. Eur J Hum Genet 19:988–994
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
Torroni A, Huoponen K, Franchalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savonanta ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850
Ubyaan R, Palit EI, Noer AS, Ngili Y, Point mutations in D-loop region at 3 and 7 generations in one mother lineage and 4 individual of Bali bomb victims
Yao YG, Salas A, Logan I, Bandelt HJ (2009) mtDNA data mining in GenBank needs surveying. Am J Hum Genet 85:929–933

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Authors and Affiliations

Anousheh Shasttiri1 · Misagh Moridi2 · Abbas Safari2 · Sayed Haidar Abbas Raza3 · Mostafa Ghaderi-Zefrehei4 · Massoud Houshmand5 · Ahmad Oryan6 · Mohammad Hossein Sanati7 · Jacqueline Smith8 · Motahareh Amjadi9

1 Department of Cellular Biotechnology, Cell Science Research Center, Royan Institute for Biotechnology, ACECR, Isfahan, Iran
2 Faculty of Agriculture, Guilan University, Rasht, Guilan, Iran
3 College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China
4 Department of Animal Science, Yasouj University, Yasouj, Iran
5 Department of Medical Genetic, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran
6 Department of Pathology, School of Veterinary Medicine, Shiraz University, Shiraz, Iran
7 Department of Medical Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran
8 The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK
9 Department of Genetics, Eötvös Loránd University, Budapest, Hungary