

RESEARCH ARTICLE

Phylogeography, genetic diversity and demographic history of the Iranian Kurdish groups based on mtDNA sequences

FATAH ZAREI and HASSAN RAJABI-MAHAM*

Faculty of Biological Sciences, Department of Animal Sciences, Shahid Beheshti University (GC), Tehran 1983963113, Iran

Abstract

Throughout the history of modern humans, the current Kurdish-inhabited area has served as part of a tricontinental crossroad for major human migrations. Also, a significant body of archaeological evidence points to this area as the site of Neolithic transition. To investigate the phylogeography, origins and demographic history, mtDNA D-loop region of individuals representing four Kurdish groups from Iran were analysed. Our data indicated that most of the Kurds mtDNA lineages belong to branches of the haplogroups with the Western Eurasian origin; with small fractions of the Eastern Eurasian and sub-Saharan African lineages. The low level of mtDNA diversity observed in the Havrami group presented a bias towards isolation or increased drift due to small population size; while in the Kurmanji group it indicated a bias towards drift or mass migration events during the 5–18th century AD. The Mantel test showed strong isolation by distance, and AMOVA results for global and regional scales confirmed that the geography had acted as the main driving force in shaping the current pattern of mtDNA diversity, rather than linguistic similarity. The results of demographic analyses, in agreement with archaeological data, revealed a recent expansion of the Kurds (~9,500 years before present) related to the Neolithic transition from hunting and gathering, to farming and cattle breeding in the Near East. Further, the high frequencies of typical haplogroups for early farmers (H; 37.1%) and hunter-gatherers (U; 13.8%) in the Kurds may testify the earlier hunter-gatherers in the Kurdish-inhabited area that adopted and admixed the Kurds ancestors following the Neolithic transition.

[Zarei F. and Rajabi-Maham H. 2016 Phylogeography, genetic diversity and demographic history of the Iranian Kurdish groups based on mtDNA sequences. *J. Genet.* **95**, 767–776]

Introduction

The Kurds are the fourth largest ethnic group in the Middle East (Hassanpour and Mojab 2005). Their original geographic region, the mountainous northern perimeter of the historic Fertile Crescent—the cradle of civilization, farming and domestication—extends between the Black Sea and Mesopotamia on one side, and the Anti-Taurus region and Iranian Plateau on the other (Izady 1992). They speak an Indo-European language from the northwestern branch of the Indo-Iranian subfamily, closely related to Persian. Undoubtedly, they belong to the Aryan stock who were nomadic tribes currently considered as plausible ancestors of most of the present Iranian's, fairly pure, as might be expected from their isolated geographical position (Cumberland 1926). They may represent the descendants of the first shepherds that occupied the Kurdistan highlands since the first Neolithic after the technologies for food production were

discovered (Braidwood and Howe 1960; Comas *et al.* 2000). For the time being, the generally accepted grouping of the Kurds based on the Kurdish dialects is that presented by D. N. Mackenzie (MacKenzie 1962), who distinguished four groups which are quite distinct from each other: (i) Northern Kurdish, also known as Kurmanji, is the most widely spoken variety of Kurdish; (ii) Central Kurdish has one main subgroup, Sorani; (iii) Southern Kurdish known as Kalhori; and (iv) Hawrami Kurdish found in Hawraman region in the Middle Zagros.

Genetic evidence concerning the origins of the Kurds is limited. Using classical markers allele frequencies, Cavalli-Sforza *et al.* (1994) showed that the Kurmanji groups from the Caucasus region are often outside the range observed for European populations, but they showed genetic similarities with other populations from the Middle East (Cavalli-Sforza *et al.* 1994). On the other hand, Comas *et al.* (2000) found that these Kurmanji groups present mitochondrial DNA (mtDNA) lineages that are typical for the European gene pool (Comas *et al.* 2000). MtDNA, due to its high mutation rate,

*For correspondence. E-mail: H_Rajabi@sbu.ac.ir.

Keywords. mtDNA; Iranian Kurds; phylogeography; demographic history; genetic structure.

Table 1. Study populations and additional data about the original and published mtDNA datasets.

Geographic region Population	Abbreviation	N	Latitude N	Longitude E	LF	Source Gen Bank acc. no.
Middle East						
Sorani	So	20	35°98'	46°20'	IE	This study
Hawrami	Ha	21	35°21'	46°35'	IE	This study
Kurmanji	Kur	19	37°25'	44°90'	IE	This study
Kalhari	Kal	19	34°47'	47°34'	IE	This study
Iraqi Kurds	Iq_Ku	15	33°32'	44°42'	IE	Al-Zahery <i>et al.</i> (2013)
Persian*	Per	127	33°83'	53°91'	IE	EU239582–EU239560; Schönberg <i>et al.</i> (2011)
Arab**	Ara	23	31°31'	48°68'	AA	EU239536–EU239655
Central Asia						
Afghan	Af	98	38°27'	67°89'	IE	Irwin <i>et al.</i> (2010)
Kazakh	Kaz	251	41°73'	70°15'	Al	Irwin <i>et al.</i> (2010)
Kyrgyz	Ky	244	41°00'	71°66'	Al	Irwin <i>et al.</i> (2010)
Tajik	Taj	234	39°18'	67°26'	IE	Irwin <i>et al.</i> (2010)
Turkmen	Tum	248	38°06'	66°78'	Al	Irwin <i>et al.</i> (2010)
Uzbek	Uz	319	40°85'	65°51'	Al	Irwin <i>et al.</i> (2010)
South Asia						
Changpa	Ch	50	34°17'	77°58'	ST	HM036578–HM036529
Tharu	Tha	17	27°30'	85°66'	IE	Malyarchuk <i>et al.</i> (2010)
Northeast Asia						
Eskimo	Es	10	65°58'	170°98'	EA	Volodko <i>et al.</i> (2008)
East Asia						
Mongol	Mo	107	40°84'	111°74'	Al	Cheng <i>et al.</i> (2008)
West Asia						
Tatar	Tat	73	54°90'	50°69'	Al	Malyarchuk <i>et al.</i> (2010)
Caucasus						
Armenian	Arm	30	40°18'	44°51'	IE	Schönberg <i>et al.</i> (2011)
Azeri	Az	30	40°43'	49°86'	Al	Schönberg <i>et al.</i> (2011)
Georgian	Ge	28	41°63'	41°63'	SC	Schönberg <i>et al.</i> (2011)
Turk	Tuk	28	39°92'	32°85'	Al	Schönberg <i>et al.</i> (2011)
East Europe						
Slav	Sl	46	53°83'	26°88'	IE	Mielnik-Sikorska <i>et al.</i> (2013)
Sami	Sam	18	62°20'	32°72'	Ur	Ingman and Gyllensten (2006)
West Europe						
Sardinian	Sar	50	39°88'	09°47'	IE	Fraumene <i>et al.</i> (2006)
Basque	Ba	32	42°81'	–01°64'	Un	Cardoso <i>et al.</i> (2013)

N, sample size; *samples from five provinces including Tehran, Esfahan, Yazd, Shiraz and Mashhad; **samples from Khuzestan province of Iran; LF, language families; IE, Indo-European; AA, Afro-Asiatic; Al, Altaic; ST, Sino-Tibetan; EA, Eskimo-Aleut; SC, South Caucasian; Ur, Uralic; Un, unknown.

haploid and strictly maternal mode of inheritance, has been accepted to be a highly informative tool for studying population relationships and tracing past demographic scenarios. Following the early findings, a number of international research studies were performed on the origin and genetic landscape of the Kurds, exclusively on the Kurmanji Kurds and unique to mtDNA (Richards *et al.* 2000; Quintana-Murci *et al.* 2004). Nasidze *et al.* (2005) introduced the first main genetic study on the Kurmanji groups based on the mtDNA and Y-chromosome variation. According to the results based on mtDNA, different Kurmanji groups demonstrated closer genetic relationships with the European groups than with the Caucasian groups; while the opposite results using the Y-chromosome, indicating some mismatch in their maternal and paternal histories (Nasidze *et al.* 2005). However, these studies were focussed only on the Kurmanji groups from Georgia, Eastern Turkey and Northern Iraq. Therefore, they may not reflect properly the possible genetic heterogeneity of the Kurds.

Despite the presence of all four major Kurdish groups in western Iran, so far, no comprehensive study has been conducted in this particular area. In the present study, using phylogeographic and population genetic analyses, we have extended previous observations by analysing the mtDNA D-loop sequence in a larger cohort consisting of individuals representing four Kurdish groups from western Iran, east of the Kurdish-inhabited region, as well as we have compared them to available data from the world populations in order to understand the origins, past demographic scenarios, any specific population structure governing the mtDNA diversity, and genetic relationships between four Kurdish groups from Iran and with other Eurasian populations.

Materials and methods

Population samples

A total of 79 blood samples were collected from unrelated males, representing four Kurdish groups from Iran:

Sorani, Hawrami, Kurmanji and Kalhori (table 1; figure 1). Informed consent of all the donors and information about their ethnic background were obtained. Only native people were included in the sample; the Kurdish ancestry was ascertained for three generations. For population comparison purposes, the mtDNA D-loop sequence data from several Middle Eastern (GenBank acc. no. EU239582–EU239560 and EU239536–EU239655; Schönberg *et al.* 2011; Al-Zahery *et al.* 2013), European (Fraumene *et al.* 2006; Ingman and Gyllensten 2006; Cardoso *et al.* 2013; Mielnik-Sikorska *et al.* 2013), Caucasian (Schönberg *et al.* 2011), South Asian (Malyarchuk *et al.* 2010; GenBank acc. no. EU239560 and EU239536–EU239655), Central Asian (Irwin *et al.* 2010), North Asian (Volodko *et al.* 2008), East Asian (Cheng *et al.* 2008) and West Asian (Malyarchuk *et al.* 2010) populations were taken from NCBI.

DNA extraction and typing

Genomic DNA was extracted using a standard phenol–chloroform method. Approximately 655 bp were amplified from 5' region of the mtDNA D-loop gene using two specific primers (Azimzadeh-Irani 2011): HDL29 (5'-GGTCTATCA CCCTATTAACCACT-3'), and HDH901 (5'-ACTTGGGTT AATCGTGTGA-3'). The 25 μ L PCR reaction mixes included 16 μ L of ultrapure water, 2.5 μ L of MgCl₂ included 10 \times PCR buffer, 1.2 μ L of each primer (0.01 mM), 0.5 μ L of each dNTP (0.05 mM), 0.2 U of Taq polymerase, and 1.4 μ L of DNA template. The thermal cycler (Primus 96 advanced Gradient, PeqLab, Germany) regime for amplification consisted of an initial step of 3 min at 95°C, followed by 30 cycles (30 s at 94°C, 45 s at 57°C, and 45 min

at 72°C), with a final extension of 10 min at 72°C. Best quality products were selected for DNA typing using an ABI Prism™ 3730 Genetic Analyzer (Applied Biosystems, Foster City, USA) by the Macrogen Company of South Korea.

Multiple alignment and data analysis

The BioEdit 7.1 (Hall 1999) was applied to read the electropherograms, whereas ClustalW implemented in the Mega 6.0 (Kumar *et al.* 2008) was used to align all sequences with the revised Cambridge Reference Sequence (rCRS). The mtDNA haplogroups were determined on the basis of diagnostic sites using the PhyloTree 16 (Van Oven and Kayser 2009). To reduce the error rate during haplogroup classification process, we also applied the algorithm implemented in the HaploGrep 2.0 (Kloss-Brandstätter *et al.* 2011). The best-fit model of sequence evolution was detected by the jModelTest 2.1.3 (Posada 2008) based on the corrected Akaike information criterion (AICc); (Hurvich and Tsai 1989). Genetic diversity statistics, pairwise F_{ST} values (with 10,000 permutations), and analyses of molecular variance (AMOVA) based on different criteria were calculated using the Arlequin 3.5 (Excoffier and Lischer 2010). The F_{ST} values were used to make a neighbour-joining (NJ) tree of populations. Historical demography was examined following two different approaches: (i) neutrality tests including Tajima's D (Tajima 1989), Fu's F_s (Fu 1997) and R_2 (Ramos-Onsins and Rozas 2002); and (ii) mismatch distribution (MMD; Harpending 1994). The statistical significance of the observed data to the expected distribution modelled for sudden expansion growth was calculated based on sum of squared deviations (SSD) and Harpending's raggedness index (Harpending 1994). The population parameter τ (post-expansion time in mutational units) was estimated from the model of sudden expansion based on the following equation $\tau = 2\mu t$ (Slatkin and Hudson 1991), where μ is fragment length dependent mutation rate and t is post-expansion time in years. The Mantel test (Mantel 1967) implemented in the Vegan library (Dixon 2003) was applied to evaluate isolation by distance (IBD) using the R 3.1.3 (Team 2012).

Results

Haplogroup profile distribution

A total of 580 bp of the mtDNA D-loop region, tRNA^{Phe}, and 12S rRNA sequence comprising nucleotide positions 156 to 736 were determined for individuals. Thirteen major haplogroups including nine Western Eurasian (H, HV, U, J, K, W, T, N*, R*, and V), two Eastern Eurasian (G and M1), and one sub-Saharan African (L3*) were identified in the Kurdish groups (table 2). Three of 13 haplogroups including V (V22 in the Hawrami group), M1 (M1a in the Iraqi Kurds), and G (G3b in the Hawrami group) were singletons and the other haplogroups were shared between two or more

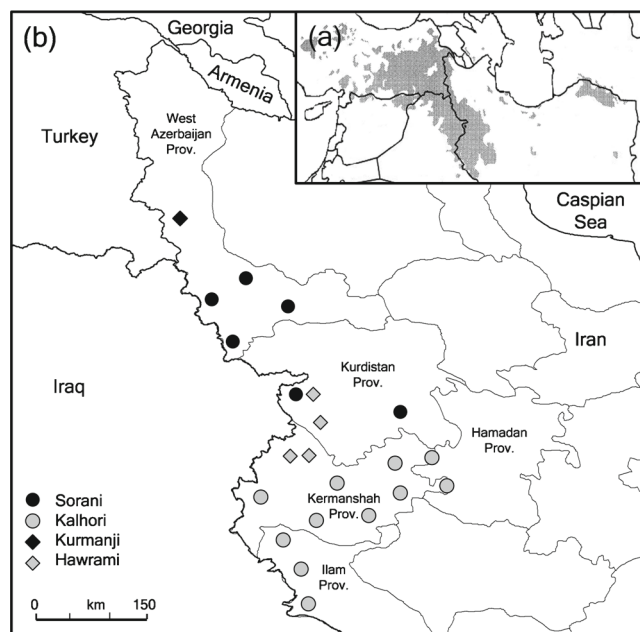


Figure 1. Map of the Kurdish-inhabited areas in the Middle East (a) after Izady 1992; and (b) our sampling localities in western Iran.

Table 2. MtDNA haplogroup frequencies in the Kurdish groups.

Haplogroup	So	Ha	Kur	Kal	Iq_Ku	Total
V		4.76				1.06
V22		4.76				1.06
HV	5	4.76	5.26	5.26		4.24
HV14	5	4.76	5.26	5.26		4.24
H	35	71.4	31.56	31.56	6.66	37.10
H1	5	23.8	10.52	15.78		11.70
H2	25	47.6	15.78	10.52		21.20
H5					6.66	1.06
H6			5.26			1.06
H7	5					1.06
H13				5.26		1.06
U	10		21.04	15.78	26.64	13.80
U1	5		10.52	10.52	6.66	6.36
U2	5					1.06
U3			5.26		6.66	2.12
U4			5.26		6.66	2.12
U6				5.26		1.06
U7					6.66	1.06
K	15	9.52	21.04			9.54
K1a	15	9.52	21.04			9.54
J	10		5.26	10.52	20	8.48
J1				5.26		1.06
J1b	5			5.26	13.32	4.24
J1c	5		5.26		6.66	3.18
T		4.76			13.32	3.18
T1					13.32	2.12
T2		4.76				1.06
R	10		5.26	10.52	13.32	7.42
R0					13.32	2.12
R6	10		5.26	10.52		5.30
W				10.52	6.66	3.18
N*	15		5.26	10.52		6.36
N1b	10			5.26		3.18
N3	5		5.26	5.26		3.18
M1					6.66	1.06
M1a					6.66	1.06
L3*			5.26	5.26	6.66	3.18
L3d			5.26	5.26		2.12
L3e					6.66	1.06
G		4.76				1.06
G3b		4.76				1.06

So, Sorani; Ha, Hawrami; Kur, Kurmanji; Kal, Kalhori; Iq_Ku, Iraqi Kurds.

populations. Overall, the Kurds mtDNA pool was characterized by the high frequencies of Western Eurasian haplogroups: H (37.1%), U (13.8%), J (8.48%), K (9.54%), R* (7.42%), N* (6.36%), and HV (4.24%). The Hawrami group stood out in its lack of two common and high frequency haplogroups U and J, and in a higher than average frequency of haplogroup H (71.42% as compared with 38.27% in the other Kurdish groups). The Kurdish groups showed small fractions of the Eastern Eurasian lineages (M1a in the Iraqi Kurds (6.66%) and G3b in the Hawrami group (4.76%); overall 2.13%). The sub-Saharan African lineages were virtually absent in the analysed Kurdish groups except L3d in the Kurmanji (5.26%) and Kalhori (5.26%) groups, and L3e in the Iraqi Kurds (6.66%).

Genetic diversity

Subsequent analyses were restricted to 418 bp of mtDNA D-loop region (156–573). We recognized 199 haplotypes in 2157 individuals which 24 of them observed in the Iranian Kurdish groups. From these, eight of 24 haplotypes (33.33%) were singletons and 16 (66.67%) were shared between populations. The haplotypes no. 1 and no. 4 showed the first and second highest frequencies in the Iranian Kurds (in 48 and four individuals, respectively). The global haplotype diversity was 0.774 ± 0.007 , ranging from 0.865 ± 0.058 for the Iranian Arabs, to 0.466 ± 0.131 for Eskimos (table 3). The haplotype diversity for the pooled samples of the Kurds was 0.668 ± 0.056 , ranging from 0.847 ± 0.087 for the Iraqi Kurds, to 0.500 ± 0.132 for the Hawrami group. Other Kurdish groups presented haplotype diversities of 0.736 ± 0.111 (Kalhori) to 0.705 ± 0.111 (Sorani) and 0.614 ± 0.130 (Kurmanji). The nucleotide diversity range from 0.0073 ± 0.004 for the Iraqi Kurds to 0.0025 ± 0.002 for the Hawrami group. The high and low levels of mtDNA diversity observed in the Iraqi Kurds and Hawrami group were clearly evident in the high and low levels of mean number of pairwise differences.

Population differentiation and genetic structure

The best-fit nucleotide substitution model determined; TN93 + G; TN93 (Tamura and Nei 1993) distance method and a gamma correction (Meyer *et al.* 1999) value of 0.437 was used to compute pairwise F_{ST} values (table 4). Interestingly, the F_{ST} values between the Hawrami group and other Kurdish groups were much higher than those between the other Kurdish groups. Even its F_{ST} value with the Iraqi Kurds population was significant ($F_{ST} = 0.061$, $P < 0.05$). Overall, the F_{ST} values among four Kurdish groups from Iran were not significantly different from zero, implying that they are genetically inseparable. Thus, all Kurdish samples from Iran were considered as a single population (i.e. the Iranian Kurds population; In_Ku) and the F_{ST} values recomputed (table 4). In most cases, the estimated F_{ST} values were positively significant with the exception of those with its geographic neighbours including the Armenians, Azeris, Iranian Arabs, Iraqi Kurds and Turks from Turkey. The Iranian Kurds population respectively showed the highest F_{ST} values with Tharu (0.341) and Changpa (0.238) from south, Eskimo (0.2) from northeast, Mongol (0.159) from east, and Kirgiz (0.135) from Central Asia. The NJ tree displays the Kurdish, Caucasian, Middle Eastern and European populations at one edge, the South Asian populations (i.e. Tharu and Changpa) at the opposite end, and the Central Asian populations in tree centre (figure 2). The correlation between geographic and pairwise F_{ST} distance matrices was highly positive and significant ($r = 0.631$, $P = 0.001$), implying that geographic distance has limited gene flow and extremely influenced the genetic structure.

The AMOVA result for geographical classification of all populations revealed that 4.73% of total mtDNA variation was related to differences among the geographical groups,

Table 3. Basic statistics of genetic diversity and neutrality tests for the study populations.

Population	<i>s</i>	<i>H</i>	<i>h</i> ± SD	π ± SD	<i>k</i>	Tajima's <i>D</i>	Fu's <i>F_s</i>	<i>R₂</i>
Kurds								
So	22	9	0.705 ± 0.111	0.0058 ± 0.004	2.47	-1.92*	-1.56	0.069*
Ha	19	7	0.500 ± 0.132	0.0025 ± 0.002	1.09	-2.30*	-1.00	0.064*
Kur	18	8	0.614 ± 0.130	0.0052 ± 0.003	2.18	-2.18*	-2.06	0.082*
Kal	25	10	0.736 ± 0.111	0.0071 ± 0.004	3.01	-2.04*	-2.30	0.069*
Iq_Ku	18	9	0.847 ± 0.087	0.0073 ± 0.004	3.07	-1.29	-1.54	0.088*
Total	49	28	0.668 ± 0.056	0.0054 ± 0.003	2.32	-2.09*	-26.18**	0.030*
Others								
Af	24	10	0.530 ± 0.057	0.0030 ± 0.002	1.39			
Arm	20	8	0.554 ± 0.105	0.0052 ± 0.003	2.22			
Az	22	10	0.600 ± 0.103	0.0045 ± 0.003	1.93			
Ba	26	11	0.703 ± 0.082	0.0062 ± 0.004	2.63			
Ch	13	9	0.659 ± 0.067	0.0038 ± 0.002	1.64			
Es	7	2	0.466 ± 0.131	0.0034 ± 0.002	1.45			
Ge	28	11	0.740 ± 0.083	0.0058 ± 0.004	2.45			
Kaz	70	56	0.813 ± 0.018	0.0057 ± 0.003	2.45			
Ara	24	11	0.865 ± 0.058	0.0075 ± 0.004	3.18			
Ky	50	42	0.785 ± 0.020	0.0052 ± 0.003	2.23			
Sl	23	10	0.570 ± 0.085	0.0033 ± 0.002	1.41			
Sam	6	3	0.568 ± 0.070	0.0027 ± 0.002	1.17			
Sar	20	10	0.557 ± 0.081	0.0031 ± 0.002	1.33			
Taj	47	30	0.768 ± 0.022	0.0051 ± 0.003	2.15			
Tat	36	22	0.772 ± 0.046	0.0062 ± 0.004	2.62			
Th	15	8	0.669 ± 0.129	0.0046 ± 0.003	1.96			
Tum	53	33	0.763 ± 0.022	0.0053 ± 0.003	2.26			
Tuk	32	12	0.640 ± 0.106	0.0052 ± 0.003	2.24			
Uz	83	63	0.795 ± 0.018	0.0057 ± 0.003	2.43			
Mo	37	20	0.694 ± 0.035	0.0047 ± 0.003	1.02			
Per	45	30	0.695 ± 0.045	0.0036 ± 0.006	2.56			
Total	158	199	0.774 ± 0.007	0.0077 ± 0.004	3.33			

* $P < 0.05$; ** $P < 0.02$; *s*, number of polymorphic sites; *H*, number of haplotypes; *h*, haplotype diversity; π , nucleotide diversity; SD, standard deviation; *k*, mean pairwise differences.

whereas interpopulation and intrapopulation differences explained 1.50% and 93.76% of variation, respectively (table 5). When population samples were subdivided according to linguistic affiliation, among group's percentage of variance (2.31%) was significantly lower than among populations within groups (3.11%). To further investigation on the patterns of genetic variation in smaller geographic space, we performed a second AMOVA only for the Iranian populations. Compatible with the first analysis, the regional AMOVA analysis also revealed that a geographic classification of the Iranian populations for the observed mtDNA diversity gave a highly better fit than what linguistic classification did.

Demographic history

Being strongly negative and significantly different from zero Tajima's *D* and Fu's *F_s*, and small and significant *R₂* values for pooled dataset of the Kurds, these estimates reflect an excess of low-frequency variants resulted from recent demographic expansion in the inspected mtDNA pool (table 3). However, in the population analyses, neutrality

tests represented different results: significant (with the exception of the Iraqi Kurds) and negative Tajima's *D* and significant positive *R₂* values indicated recent expansion of each Kurdish group, while insignificant negative Fu's *F_s* value for each group provided support for rejecting the hypothesis of population growth probably due to the strong effects of drift and / or the small sample size. The MMD analysis when applied to the pooled dataset of the Kurds was unable to reject the model of sudden expansion ($P(\text{sim} \geq \text{obs}) > 0.05$). The MMD curve of haplotypes showed a skewed exponential shape due to excess frequencies of small pairwise differences (figure 3). Altogether, these results implied an excess of recently diverged haplotypes and a deficit of deeper coalescence events. With a mutation rate of 32% per site per Myr (Sigurðardóttir *et al.* 2000), the τ value of 2.57 was obtained by MMD analysis on the pooled dataset of the Kurds translated into an expansion time of ~9,500 YBP. Due to insignificant *F_{ST}* values among the Kurdish groups and the disagreement of neutrality tests in confirmation of population growth for each individual Kurdish group, we were unable to perform the MMD analysis for any Kurdish group, individually.

Table 4. Pairwise F_{ST} values among the study populations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
So		.550	.937	.895	N/A	.047	.531	.292	.306	.000	.000	.110	.002	.497	.482	.000	.072	.010	.104	.006	.082	.000	.015	.720	.010	.000	.214
Ha	-.002		.315	.203	N/A	.066	.096	.061	.0419	.000	.000	.017	.000	.060	.014	.000	.176	.003	.220	.004	.017	.000	.003	.853	.004	.000	.051
Kur	-.023	.002		.996	N/A	.165	.287	.362	.182	.000	.000	.533	.007	.533	.310	.000	.547	.015	.114	.028	.342	.000	.025	.837	.027	.000	.336
Kal	-.022	.011	-.032		N/A	.107	.508	.326	.163	.000	.002	.518	.019	.713	.714	.000	.194	.050	.059	.043	.451	.000	.070	.347	.061	.000	.738
In Ku	N/A	N/A	N/A	N/A		.005	.136	.101	.025	.000	.000	.021	.000	.096	.108	.000	.161	.002	.055	.000	.001	.000	.000	.504	.000	.000	.016
Af	.034	.030	.014	.020	.019		.011	.067	.000	.000	.001	.016	.000	.007	.039	.000	.078	.039	.001	.004	.028	.000	.002	.049	.003	.000	.011
Arm	-.007	.033	.006	-.006	.010	.041		.289	.202	.000	.001	.111	.000	.567	.877	.000	.012	.013	.057	.000	.054	.000	.014	.212	.002	.000	.306
Az	.006	.026	.002	.003	.011	.020	.005		.083	.000	.000	.287	.004	.070	.362	.000	.076	.041	.050	.016	.091	.000	.021	.396	.023	.000	.080
Ba	.003	.043	.013	.016	.027	.071	.010	.024		.000	.003	.004	.000	.114	.262	.000	.001	.004	.007	.000	.004	.000	.000	.127	.000	.000	.006
Ch	.249	.320	.251	.202	.238	.215	.234	.208	.255		.007	.000	.000	.000	.000	.032	.000	.100	.000	.000	.000	.046	.000	.000	.000	.000	.174
Es	.196	.326	.209	.153	.200	.234	.206	.209	.204	.212		.000	.005	.002	.007	.000	.000	.011	.000	.000	.001	.000	.002	.000	.002	.002	.001
Ge	.019	.037	-.003	-.003	.0207	.037	.019	.004	.058	.220	.168		.000	.065	.002	.035	.009	.002	.002	.259	.000	.002	.064	.001	.000	.000	.059
Kaz	.068	.081	.056	.042	.069	.032	.061	.042	.093	.050	.111	.064		.001	.065	.002	.000	.990	.000	.019	.002	.000	.110	.000	.156	.004	.000
Ara	-.004	.033	-.005	-.014	.013	.056	-.008	.024	.021	.217	.147	.027	.067		.790	.000	.003	.012	.008	.004	.023	.000	.011	.275	.003	.000	.307
Iq_Ku	-.006	.061	.005	-.018	.020	.051	-.028	.003	.008	.171	.160	.011	.031	-.021		.076	.000	.127	.021	.061	.348	.000	.274	.175	.132	.005	.634
Ky	.136	.159	.124	.100	.135	.089	.122	.097	.160	.017	.158	.115	.011	.130	.076		.000	.947	.000	.000	.000	.001	.000	.000	.000	.000	.313
Sl	.020	.009	-.003	.010	.006	.013	.039	.016	.058	.266	.244	.023	.061	.048	.061	.130		.003	.017	.000	.008	.000	.000	.114	.000	.000	.007
Sam	.094	.184	.091	.051	.082	.050	.084	.060	.110	.041	.259	.071	-.021	.077	.038	-.019	.111		.000	.723	.262	.007	.770	.003	.944	.592	.088
Sar	.019	.007	.017	.025	.012	.043	.027	.021	.050	.319	.304	.045	.092	.046	.065	.167	.021	.164		.000	.000	.000	.000	.135	.000	.000	.000
Taj	.058	.063	.040	.032	.053	.019	.056	.032	.089	.092	.158	.052	.006	.058	.035	.033	.046	-.012	.076		.008	.000	.012	.002	.084	.000	.000
Tat	.022	.040	.002	-.002	.024	.015	.022	.015	.050	.132	.144	.004	.021	.033	.002	.054	.025	.007	.054	.017		.000	.025	.018	.016	.000	.104
Th	.326	.459	.340	.265	.341	.364	.326	.316	.328	.055	.360	.294	.144	.281	.233	.085	.405	.163	.456	.195	.212		.000	.000	.000	.011	.000
Tum	.046	.067	.041	.025	.050	.023	.035	.029	.071	.082	.140	.052	.002	.044	.006	.027	.050	-.013	.075	.007	.012	.184		.001	.355	.000	.005
Tuk	-.009	-.009	-.013	.002	-.001	.024	.008	.001	.017	.254	.175	.018	.066	.006	.018	.140	.010	.102	.010	.053	.030	.354	.054		.002	.000	.044
Uz	.044	.056	.036	.024	.047	.017	.043	.025	.073	.073	.127	.047	.001	.050	.019	.024	.040	-.017	.067	.002	.012	.166	.001	.049		.000	.000
Mo	.159	.200	.154	.122	.159	.121	.150	.122	.181	.006	.172	.142	.018	.149	.097	.001	.162	-.010	.208	.045	.074	.080	.037	.169	.032		.000
Per	.008	.026	.002	-.011	.014	.015	.002	.015	.040	.148	.150	.019	.031	.003	-.010	.074	.026	.024	.038	.024	.006	.238	.013	.021	.017	.090	

Pairwise F_{ST} values below the diagonal and their P values above the diagonal; significance threshold was $P \leq 0.05$; N/A, not available.

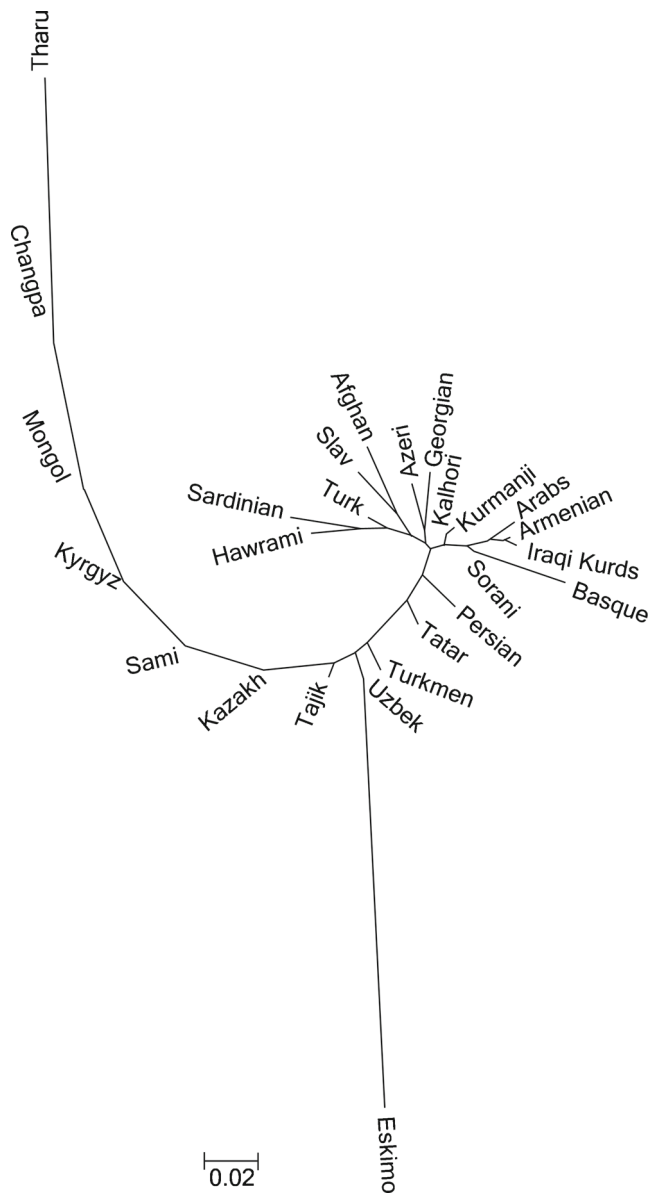


Figure 2. NJ tree of populations based on pairwise F_{ST} values.

Discussion

Despite the high linguistic heterogeneity, our results revealed close genetic relationships among the Kurdish groups. Prior to this study, based on a linguistic tree of the Iranian ethnic groups (Zarei 2013), we hypothesized the significant genetic differentiation between the Hawrami group and other Kurdish groups. However, despite its greater F_{ST} values, the only statistically significant F_{ST} value belonged to the Hawrami group / Iraqi Kurds. The Iranian Kurds population presented the insignificant low F_{ST} values with its geographic neighbours, the significant low F_{ST} values with other Western Eurasian Indo-European populations and the highest F_{ST} values with South and Eastern Eurasian populations. Our results suggested that linguistic classification of the populations do not provide an infallible guide to ancestral

Table 5. AMOVA based on different criteria.

Classification	Source of variation (%)		
	Among groups	Among populations within groups	Within populations
Geography			
Total**	4.73*	1.50*	93.76*
Iran†	2.38*	-0.70	98.33*
Linguistic			
Total‡	2.31*	3.11*	94.58*
Iran§	-1.23*	1.49*	99.74

* $P < 0.05$; **Middle East, Central Asia, Caucasus, West Europe, East Europe, West Asia, South Asia, northeast Asia and East Asia; †Northwest Iran, West Iran, Central Iran, South Iran and East Iran; ‡Indo-European, Altaic, South Caucasian, Uralic, Sino-Tibetan, Afro-Asiatic and Basque; §Indo-European and Afro-Asiatic.

relationships. Presence of a sharp geographical structure both on global and regional scales for mtDNA D-loop diversity, in agreement with study of Nasidze *et al.* (2005), was supported by the AMOVA analyses. Also, the Mantel test, in confirmation of isolation by distance, revealed a high and significant correlation indicating that the level of genetic resemblance between populations is strongly depend on geographic distance (Mantel 1967). These results suggested that the real affinities of the Kurdish groups with other Indo-European populations is partly masked by the effects of gene flow from neighbouring non Indo-European populations, as represented by the NJ tree based on pairwise F_{ST} values.

The Kurdish groups were characterized by the presence of substantial western Eurasian mtDNA lineages. This testifies to a hypothesis that ancestors of the modern Kurds originally diverged from the proto-Indo-European population. The prevailing haplogroups in the East Eurasia such as M1a and G3b, and the sub-Saharan African haplogroups such as L3d and L3e were rare or even absent in the Kurdish groups. However, the role of recent gene flow from the east Eurasian (probably during the Mongol and Turkish invasions (Meri 2005), or trade along the Silk Road until the 16th century AD (Yang *et al.* 2008)), and the West-Central African groups in shaping the current mtDNA pool of the Kurds cannot be ignored.

Typically, the low genetic diversity is the characteristic of small or bottlenecked population. Such populations show a rapid increase in the number of identical haplotypes (Hoelzel *et al.* 1993; Weber *et al.* 2000). The very high frequency of H derived lineages and the lack of two common and high frequency haplogroups U and J in the Hawrami group are indicative of this idea. The low mtDNA diversity observed in the Hawrami group indicated a bias towards isolation or increased drift due to small population size. This group living in the Hawraman region of the Middle Zagros is considered as the smallest and most isolated Kurdish group. The well-known local Kurdish sentiment that ‘they have no friends but the mountains’ demonstrates their ethnic and geographical isolation (Bulloch and Morris 1992). Surprisingly, the

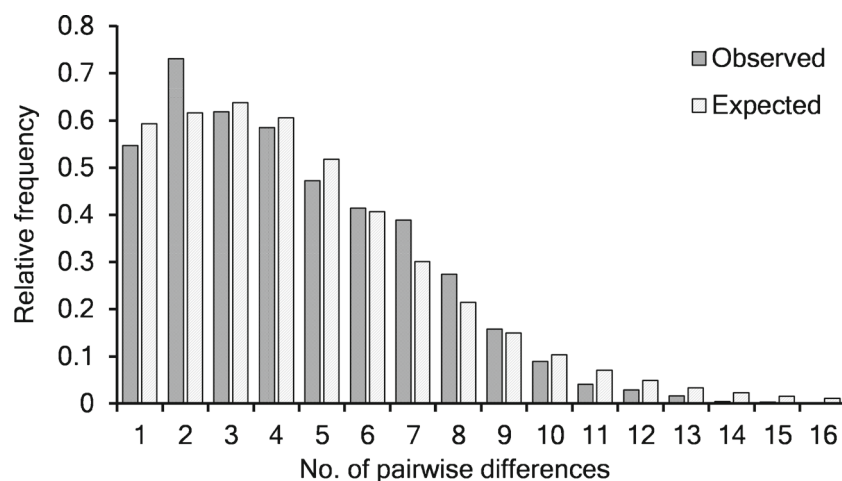


Figure 3. MMD for the pooled dataset of the Kurds.

low mtDNA diversity observed in the Kurmanji group can be explained by drift or mass migration events during its contemporary history (i.e. 5–12th century AD; Izady 1992). Later, at the beginning of the 16th century AD, particularly after the Chaldiran battle in 1514, the Safavid dynasty (16–18th century AD) initiated the mass migration of the Kurmanji Kurds from Northwestern Iran to the North Khorasan province (figure 1), which was primarily aimed at dispersing the compact Kurdish population in the border zone with the Ottomans, as well as at making a defense line in the northeast frontiers of the country against the constant inroads of the Turkmen and Uzbek nomads (Madih 2007).

Here, to detect the population size growth, we used three neutrality tests which differ somewhat in their approach. However, Ramos-Onsins and Rozas (2002) showed that the most powerful tests are Fu's F_s and the recently developed R_2 . Fu's F_s did not detect significant deviations from neutrality for any individual Kurdish group, but this may be due to the lack of power to detect deviation in small datasets (Ramos-Onsins and Rozas 2002). On the other hand, the behaviour of the R_2 test is superior for small sample sizes. In population analyses, the significant positive R_2 estimates were consistent with Tajima's D results, indicating that populations were not in genetic equilibrium or decline. However, the Kurdish groups, as a whole, had significant negative D and F_s , and significant positive R_2 values, which reflect an excess of low-frequency variants in the surveyed mtDNA pool resulted from recent demographic expansion. The MMD curve of the total Kurds mtDNA haplotypes showed a skewed exponential shape, but smooth unimodal distribution with the low raggedness value indicating that a single expansion occurred at some time in the past, ~9,500 YBP, when farming and cattle breeding in the modern societies appeared first in the Fertile Crescent (Diamond and Bellwood 2003; Gupta 2004; Bellwood and Oxenham 2008). The introduction of agriculture and domestication to pre-Neolithic hunter-gatherer societies significantly impacted on the establishment and distribution of ancestral Kurdish

population, more or less parallel with the expansion of the first farmers into Europe from the Fertile Crescent (Diamond and Bellwood 2003). Our results are largely in agreement with the time estimates based on the zoological and botanical remains of farming and domestication which have been found at three main archaeological sites in the Kurdish-inhabited region: (i) Çayönü (Diyarbakır, Turkey; dating to 9,000 YBC; (Braidwood *et al.* 1971)), (ii) Ganj Dara (Kermanshah, Iran; dating to 10,000 YBP; (Zeder and Hesse 2000)), and (iii) Jarmo (Sulaymaniyah, Iraq; dating to 8,000 YBP; Braidwood and Howe 1960)). Also, the signature of recent expansion of the Kurds is well represented in the high frequency of H derived lineages, as a recent study by Fu *et al.* (2012) suggested that the H-type in the European mtDNA genome shows a population expansion related to the prevalence of animal domestication and cultivation around 9,000 YBP and continuing to the present. In contrast, the U-type mtDNAs show a population growth between 20,000 and 10,000 YBP seeming to represent earlier hunter-gatherers that initiated farming practices and admixed with immigrant populations (Fu *et al.* 2012). In agreement with this scenario, the presence of relatively high frequency haplogroup U in the Kurds may testify the earlier hunter-gatherers in the Kurdish-inhabited area that adopted and admixed the Kurds ancestors, following the expansion.

Overall, our mtDNA data has supported the scenario of recent demographic expansion of the Kurds related to the Neolithic transition in the Near East, a scenario that is largely in agreement with a significant body of archaeological evidence. However, to provide more details about the peopling of the current Kurdish-inhabited region, it is also necessary to employ the nonrecombining Y-chromosomal (NRY) markers.

Acknowledgements

We are grateful to all the original donors for providing blood samples, Esmail Majidi and Maryam Maali for assisting in the sampling, and Hassan Nazari for linguistic editing. We appreciate the

anonymous reviewers for their valuable suggestions on the manuscript. This work which was extracted from a master thesis, supported in part by Faculty of Biological Sciences, Shahid Beheshti University.

References

- Al-Zahery N., Saunier J., Ellingson K., Parson W., Parsons T. and Irwin J. 2013 Characterization of mitochondrial DNA control region lineages in Iraq. *Int. J. Legal Med.* **127**, 373–375.
- Azimizadeh-Irani M. 2011 Phylogeography and phylogenetic-linguistic relationship of the Iranian Tat groups using mtDNA D-loop region. MSc thesis, Faculty of Biological Sciences, Shahid Beheshti University, Tehran, Iran.
- Bellwood P. and Oxenham M. 2008 The expansions of farming societies and the role of the Neolithic demographic transition in *The Neolithic demographic transition and its consequences* (eds. O. Bar-Yosef and J. P. Bocquet-Appel), pp. 13–34. Springer, Netherland.
- Braidwood R. F. and Howe B. 1960 *Prehistoric investigations in Iraq Kurdistan*, 1st edition. University of Chicago Press, Illinois, USA.
- Braidwood R. J., Cambel H., Redman C. L. and Watson P. J. 1971 Beginnings of village-farming communities in southeastern Turkey. *Proc. Natl. Acad. Sci. USA* **68**, 1236–1240.
- Bulloch J. and Morris H. 1992 *No friends but the mountains*, 1st edition. Viking Penguin, New York, USA.
- Cardoso S., Valverde L., Alfonso-Sánchez M. A., Palencia-Madrid L., Elcoroaristizabal X., Algorta J. *et al.* 2013 The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. *PLoS One* **8**, e67835.
- Cavalli-Sforza L. L., Menozzi P. and Piazza A. 1994 *The history and geography of human genes*, 1st edition. Princeton University Press, Princeton, New Jersey, USA.
- Cheng B., Tang W., He L., Dong Y., Lu J., Lei Y. *et al.* 2008 Genetic imprint of the Mongol: signal from phylogeographic analysis of mitochondrial DNA. *J. Hum. Genet.* **53**, 905–913.
- Comas D., Calafell F., Bendukidze N., Fananas L. and Bertranpetit J. 2000 Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am. J. Phys. Anthropol.* **112**, 5–16.
- Cumberland R. C. 1926 The Kurds. *Muslim World* **16**, 150–157.
- Diamond J. and Bellwood P. 2003 Farmers and their languages: the first expansions. *Science* **300**, 597–603.
- Dixon P. 2003 VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930.
- Excoffier L. and Lischer H. E. 2010 Arlequin, ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567.
- Fraumene C., Belle E. M., Castri L., Sanna S., Mancosu G., Cosso M. *et al.* 2006 High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. *Mol. Biol. Evol.* **23**, 2101–2111.
- Fu Y. X. 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.
- Fu Q., Rudan P., Pääbo S. and Krause J. 2012 Complete mitochondrial genomes reveal Neolithic expansion into Europe. *PLoS One* **7**, e32473.
- Gupta A. K. 2004 Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. *Curr. Sci.* **87**, 54–59.
- Hall T. A. 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
- Harpending H. 1994 Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591–600.
- Hassanpour A. and Mojab S. 2005 *Kurdish diaspora in encyclopedia of diasporas* (ed. C. R. Ember, M. Ember and I. A. Skoggard), pp. 214–224. Springer, New York, USA.
- Hoelzel A., Halley J., O'Brien S., Campagna C., Arnborn T., Le Boeuf B. *et al.* 1993 Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *J. Hered.* **84**, 443–449.
- Hurvich C. M. and Tsai C.-L. 1989 Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.
- Ingman M. and Gyllensten U. 2006 A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur. J. Hum. Genet.* **15**, 115–120.
- Irwin J. A., Ikramov A., Saunier J., Bodner M., Amory S., Röck A. *et al.* 2010 The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns. *Int. J. Legal Med.* **124**, 195–204.
- Izady M. R. 1992 *The Kurds: A concise handbook*, 1st edition. Taylor and Francis, London, UK.
- Kloss-Brandstätter A., Pacher D., Schönherr S., Weissensteiner H., Binna R., Specht G. *et al.* 2011 HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32.
- Kumar S., Nei M., Dudley J. and Tamura K. 2008 MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* **9**, 299–306.
- MacKenzie D. N. 1962 *Kurdish dialect: studies*, 1st edition. Oxford University Press, London, UK.
- Madih A. A. 2007 The Kurds of Khorasan. *Iran Caucasus* **11**, 11–31.
- Malyarchuk B., Derenko M., Denisova G. and Kravtsova O. 2010 Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol. Biol. Evol.* **27**, 2220–2226.
- Mantel N. 1967 The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.
- Meri J. W. 2005 *Medieval Islamic civilization: an encyclopedia*, 1st edition. Routledge, New York, USA.
- Meyer S., Weiss G. and von Haeseler A. 1999 Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**, 1103–1110.
- Mielnik-Sikorska M., Daca P., Malyarchuk B., Derenko M., Skonieczna K., Perkova M. *et al.* 2013 The history of Slavs inferred from complete mitochondrial genome sequences. *PLoS One* **8**, e54360.
- Nasidze I., Quinque D., Ozturk M., Bendukidze N. and Stoneking M. 2005 MtDNA and Y-chromosome variation in Kurdish groups. *Ann. Hum. Genet.* **69**, 401–412.
- Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256.
- Quintana-Murci L., Chaix R., Wells R. S., Behar D. M., Sayar H., Scozzari R. *et al.* 2004 Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am. J. Hum. Genet.* **74**, 827–845.
- Ramos-Onsins S. E. and Rozas J. 2002 Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100.
- Richards M., Macaulay V., Hickey E., Vega E., Sykes B., Guida V. *et al.* 2000 Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276.
- Schönberg A., Theunert C., Li M., Stoneking M. and 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.
- Sigurðardóttir S., Helgason A., Gulcher J. R., Stefansson K. and Donnelly P. 2000 The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.* **66**, 1599–1609.

- Slatkin M. and Hudson R. R. 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.
- Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tamura K. and Nei M. 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526.
- Team R. C. 2012 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Van Oven M. and Kayser M. 2009 Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394.
- Volodko N. V., Starikovskaya E. B., Mazunin I. O., Eltsov N. P., Naidenko P. V., Wallace D. C. *et al.* 2008 Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am. J. Hum. Genet.* **82**, 1084–1100.
- Weber D., Stewart B. S., Garza J. C. and Lehman N. 2000 An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Curr. Biol.* **10**, 1287–1290.
- Yang L., Tan S., Yu H., Zheng B., Qiao E., Dong Y. *et al.* 2008 Gene admixture in ethnic populations in upper part of Silk Road revealed by mtDNA polymorphism. *Sci. China C Life Sci.* **51**, 435–444.
- Zarei F. 2013 Phylogeography and phylogenetic-linguistic relationships of the Iranian Kurds using mtDNA D-Loop region. MSc thesis, Faculty of Biological Sciences, Shahid Beheshti University, Tehran, Iran.
- Zeder M. A. and Hesse B. 2000 The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago. *Science* **287**, 2254–2257.

Received 13 November 2015, in final revised form 14 February 2016; accepted 16 February 2016

Unedited version published online: 22 February 2016

Final version published online: 20 October 2016

Corresponding editor: RAJIVA RAMAN