

MtDNA and Y-chromosome Variation in Kurdish Groups

Ivan Nasidze^{1,*}, Dominique Quinque¹, Murat Ozturk², Nina Bendukidze³ and Mark Stoneking¹

¹Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103, Leipzig, Germany

²Norwegian University of Science and Technology TBS, NO-7491, Trondheim, Norway

³H&I DNA Reference Laboratory, National Blood Service, Southmead Road, Bristol, BS7 9AH, UK

Summary

In order to investigate the origins and relationships of Kurdish-speaking groups, mtDNA HV1 sequences, eleven Y chromosome bi-allelic markers, and 9 Y-STR loci were analyzed among three Kurdish groups: Zazaki and Kurmanji speakers from Turkey, and Kurmanji speakers from Georgia. When compared with published data from other Kurdish groups and from European, Caucasian, and West and Central Asian groups, Kurdish groups are most similar genetically to other West Asian groups, and most distant from Central Asian groups, for both mtDNA and the Y-chromosome. However, Kurdish groups show a closer relationship with European groups than with Caucasian groups based on mtDNA, but the opposite based on the Y-chromosome, indicating some differences in their maternal and paternal histories. The genetic data indicate that the Georgian Kurdish group experienced a bottleneck effect during their migration to the Caucasus, and that they have not had detectable admixture with their geographic neighbours in Georgia. Our results also do not support the hypothesis of the origin of the Zazaki-speaking group being in northern Iran; genetically they are more similar to other Kurdish groups. Genetic analyses of recent events, such as the origins and migrations of Kurdish-speaking groups, can therefore lead to new insights into such migrations.

Keywords: Kurds, Y chromosome, mtDNA

Introduction

Kurds are an Indo-European speaking group that inhabit the highlands in the border area of Turkey, Syria, Iran and Iraq. This region lies astride the Zagros Mountains of Iran and the eastern extension of the Taurus Mountains in Turkey, and extends in the south across the Mesopotamian plain to include the upper reaches of the Tigris and Euphrates rivers (Figure 1). Several languages and/or dialects of Kurdish are recognized, and are classified as belonging to the northwestern branch of Iranian languages (Ethnologue, 2000).

The first mention of the Kurds in historical records is in cuneiform writings from the Sumerians from around 3,000 B.C. (Wixman, 1984), who talked of the “land

of the Karda”. In the 7th century A.D., the Arabs conquered the area and in time converted everyone in it – including the Kurds – to Islam. In the centuries that followed, the Kurds withstood invasions from Central Asia which brought the Turkic peoples as far west as Asia Minor (now Turkey), probably because they occupied an area that was too difficult for outsiders to reach. As the Ottoman Empire rose to power in the 13th through to the 15th centuries, it extended its territory to what is roughly now the border between Iran and Iraq. From then until World War I, the area inhabited by the Kurds was under the dominion of the Ottomans and Persians.

There have been several migration events involving Kurds. An extensive resettlement of Kurds from Turkey and Iran into the Caucasus began during the late 19th century and continued through World War I (Wixman, 1984). Since that time Kurds have formed compact settlements in Georgia and Armenia, and have

* Corresponding author: Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, 04103, Leipzig, Germany. E-mail: nasidze@eva.mpg.de

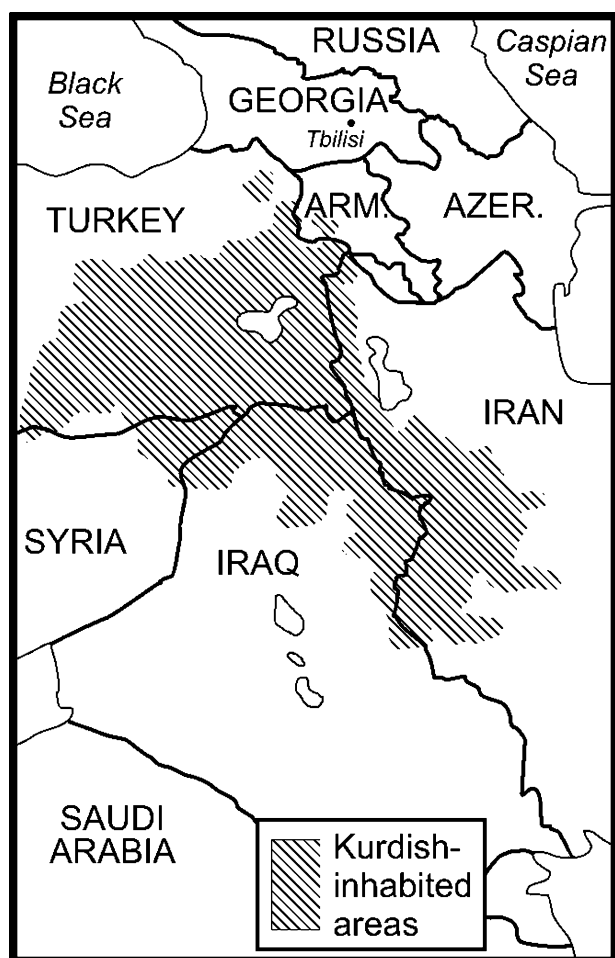


Figure 1 A map of the geographic location of Kurdish-speaking groups.

kept their cultural and linguistic identity. At the time of the 1979 census there were 51,000 Kurds in Armenia and 26,000 in Georgia (USSR Census 1979). Kurds from Central Asia (Turkmenistan and Kazakhstan) speak the Kurmanji language (Ethnologue, 2000). They are also descendants of Kurds from Eastern Anatolia and Iran.

Kurmanji is a North-West Iranian language spoken in Southeast Anatolia in Turkey. This language belongs to the Kurdish linguistic branch, along with the Behdini, Herki, Kurdi, Shikaki and Surchi languages, all of which are spoken in Northern Iraq (Ethnologue, 2000). Zaza (or Dymli) is also a northwest Iranian language spoken in Southeast Anatolia, northwest of the Kurdish speaking regions. Although first thought to be a Kurdish dialect, since the beginning of the 20th century Zazaki has been accepted as a language of its own (Paul, 1998). Virtually

nothing is known about the origin of Zazaki-speaking people, as they do not possess a written language and therefore lack a recorded history. Based on some structural similarities of the Zazaki language with the Talyshi, Gilani and Mazandarani languages spoken in northern Iran, linguists have hypothesized that their origin lies in the mountainous region of the Southern Caspian Sea area (MacKenzie, 1962).

Only a few genetic studies have been carried out on Kurdish groups. Previous genetic studies of classical markers (Cavalli-Sforza *et al.* 1994) indicated an overall genetic similarity of Kurds with other Middle Eastern populations. Comas *et al.* (2000) studied mtDNA HV1 sequence variability among Kurmanji-speaking Kurds living in Georgia (Caucasus), and found close European affinities for Kurdish mtDNA lineages. Richards *et al.* (2000) studied mtDNA HV1 sequence variability among 53 Kurds from Eastern Turkey and found that some mtDNA haplotypes found in Kurdish samples presumably originated in Europe, and were associated with back-migrations from Europe to the Near East. Wells *et al.* (2001) investigated Y chromosome SNP haplogroup distributions among Central Asian groups, including a group of Kurmanji-speaking Kurds living in Turkmenistan, but no specific conclusions were made regarding the history of the Kurdish group. Nebel *et al.* (2001) studied Y chromosome SNP and short tandem repeat (STR) loci among different groups from the Middle East, including a group of 95 Kurds from northern Iraq, and found close affinities for the Kurdish group to other Middle Eastern groups. Finally, Quintana-Murci *et al.* (2004) studied 20 Kurds from Western Iran and 32 Kurds from Turkmenistan, among other groups from Iran, Pakistan and Central Asia, but did not come to any specific conclusions concerning the Kurds.

None of these previous studies were focused specifically on Kurdish groups and their origins or relationships. Rather, these previous studies targeted a larger geographic scale and/or more general questions concerning the genetic relationships of different populations. In order to investigate the genetic relationships between different Kurdish groups, we present here data on mtDNA and Y chromosomal variation in 139 individuals from Zazaki and Kurmanji speaking groups (from Turkey), and from Kurds living in Georgia (Caucasus) who are also Kurmanji speakers.

The new genetic data presented in this study were combined with previously-published data on mtDNA HV1 sequence variation in 29 Kurds living in Georgia (Comas *et al.* 2000) and with other previously-published data from Kurds from Iran, Turkmenistan and Turkey (Richards *et al.* 2000; Quintana-Murci *et al.* 2004) to address the following questions: (1) how closely related genetically are the Kurdish groups living in different areas and/or speaking different languages?; (2) what is the genetic relationship between Indo-European speaking Kurdish groups and other West Asian Indo-European and non-Indo-European speaking populations?; (3) can a source population for the Kurdish group from Georgia be identified?; (4) was there any subsequent genetic exchange between the Kurdish group from Georgia and the surrounding Georgian population?

Materials and Methods

Samples and DNA Extraction

A total of 114 cheek cell samples from unrelated males, representing two Kurdish groups - Zazaki speakers (27 samples) and Kurmanji speakers (87 samples) - were collected in Turkey. The Zazaki language used to be classified as a dialect of Kurmanji, however it is now considered to be separate and not a Kurdish language (Paul, 1998), but rather belonging to the Zaza-Gorani group of northwest Iranian languages (ethnologue, 2000). Since the Zazaki-speakers analyzed here self-identify as Kurds (Donald Stilo, personal communication), we included them in the analyses of the groups speaking Kurdish languages.

Genomic DNA from cheek cell swabs was extracted in Leipzig using a standard salting-out procedure (Miller *et al.* 1988). An additional 25 blood samples from Kurdish males from Georgia (Tbilisi) were collected and genomic DNA was extracted in Bristol using a standard phenol-chloroform method (Maniatis *et al.* 1982). Informed consent and information about birthplace, parents and grandparents was obtained from all donors.

MtDNA HV1 Sequences

The first hypervariable segment (HV1) of the mtDNA control region was sequenced in the samples from

Zazaki and Kurmanji individuals. Primers L15996 and H16410 (Vigilant *et al.* 1989) were used to amplify the first hypervariable segment (HV1) of the mtDNA control region, as described previously (Redd *et al.* 1995). The nested primers L16001 (Cordaux *et al.* 2003) and H16401 (Vigilant *et al.* 1989) were used to determine sequences for both strands of the PCR products with a DNA Sequencing Kit (Perkin-Elmer), following the protocol recommended by the supplier, and an ABI 3700 automated DNA sequencer. Individuals with the "C-stretch" between positions 16184-16193, which is caused by the 16189C substitution, were sequenced again in each direction, so that each base was determined twice.

Published mtDNA HV1 sequences were also used from 29 Kurds from Georgia (Comas *et al.* 2000) and from Caucasian, West and Central Asian, and West and East European groups (Di Rienzo & Wilson, 1991; Maliarchuk *et al.* 1995; Cote-Real *et al.* 1996; Richards *et al.* 1996, 2000; Kivisild *et al.* 1999; Comas *et al.* 2000; Nasidze & Stoneking, 2001; Al-Zahery *et al.* 2003; Quintana-Murci *et al.* 2004; Nasidze *et al.* 2004).

Y-Chromosome Bi-Allelic Markers

The same set of 139 samples from three Kurdish groups was genotyped for ten Y chromosomal SNP markers: RPS4Y (M130), M9, M89, M124, M45, M173, M17, M201, M170, and M172 (Underhill *et al.* 2000 and references therein); the YAP *Alu* insertion polymorphism (Hammer & Horai, 1995) was also typed. The markers M9 and RPS4Y were typed by means of PCR-RFLP as described elsewhere (Kayser *et al.* 2000). The markers M17, M124, M170, M172, M173, M45 and M201 were typed using PIRA-PCR (primer introduced restriction analysis) assays (Yoshimoto *et al.* 1993) as described previously (Cordaux *et al.* 2004; Nasidze *et al.* 2004). M89 was typed as described by Ke *et al.* (2001), while the YAP *Alu* insertion was typed as described by Hammer & Horai (1995). The samples were genotyped according to the hierarchical order of the markers as described by Underhill *et al.* (2000). The Y-SNP haplogroup nomenclature used here is according to the recommendations of the Y Chromosome Consortium (2002).

The samples from Georgian Kurds which were used for sequencing mtDNA HV1 (Comas *et al.* 2000) also

have been used for Y-SNP genotyping in this study. All 29 samples were typed for the X- and Y-linked zinc finger protein genes in order to determine the gender of the sample (Wilson & Erlandson, 1998); 25 samples were determined to be from males and were included in this study. Published Y-SNP data for 17 Kurds from Turkmenistan (Wells *et al.* 2001), and from several European, West Asian, and Central Asian groups (Semino *et al.* 2000; Wells *et al.* 2001) were also included in some analyses.

Y-Chromosome STRs

Samples belonging to Y-SNP haplogroups *P1(M124)*, *I*(M170)* and *J2*(M172)* were genotyped for nine Y chromosome short tandem repeat (Y-STR) markers: DYS19 (DYS394), DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393. These loci were amplified in pentaplex and quadraplex PCRs and detected on an ABI PRISM 377 DNA sequencer (Applied Biosystems) as described elsewhere (Kayser *et al.* 1997, 2001). In order to distinguish genotypes at DYS385a and DYS385b, an additional PCR was carried out as described in Kittler *et al.* (2003) and detected on an ABI PRISM 377 DNA sequencer (Applied Biosystems).

Statistical Analysis

Basic parameters of molecular diversity and population genetic structure, including analyses of molecular variance (AMOVA), were calculated using the software package Arlequin 2.000 (Schneider *et al.* 2000). The statistical significance of F_{st} values was estimated by permutation analysis, using 10,000 permutations. The statistical significance of the correlation between genetic distance matrices, based on the mtDNA HV1 sequences and the Y chromosome SNP data, was evaluated by the Mantel test with 10,000 permutations. The STATISTICA package (StatSoft Inc.) was used for multi-dimensional scaling (MDS) analysis (Kruskal, 1964). Network analysis for Y-STR and mtDNA HV1 sequence data was carried out using the software package NETWORK version 3.1 (Bandelt *et al.* 1999).

Results

MtDNA HV1 Sequence Variability

A total of 377 bp of the mtDNA HV1 region, comprising nucleotide positions 16024 to 16400 (Anderson *et al.* 1981), were determined for 78 individuals from the Zazaki speakers and Kurmanji speakers from Turkey (hereafter referred to as Zazaki-T and Kurmanji-T, respectively). MtDNA HV1 sequence data for the Kurmanji speakers from Georgia (Kurmanji-G) and for Kurds from Iran (Kurds-I), Turkmenistan (Kurds-Tm), and Turkey (Kurds-T) were previously published (Comas *et al.* 2000; Richards *et al.* 2000; Quintana-Murci *et al.* 2004). As a check on the accuracy of the HV1 sequences, we used the network method to search for so-called “phantom” mutations (Bandelt *et al.* 2002). No such artifacts were found in the Kurdish HV1 sequences (analysis not shown). The sequences have been deposited in the HVRbase database (www.HVRbase.de) and are also available upon request from the authors.

Parameters summarizing some characteristics of the mtDNA HV1 sequence variability in Kurdish groups are presented in Table 1. The haplotype diversity and mean number of pairwise differences (MPD) were both lower in the Kurmanji-G than in the other samples, but not significantly so. Tajima's D was negative and significantly different from zero in five of the six Kurdish groups, suggesting population expansion (Table 1).

The mean number of pairwise nucleotide differences (Table 1) are in the range of values observed in the Caucasus (4.40–5.87; Nasidze *et al.* 2004), but lower for the Zazaki-T and Kurmanji-G than in West Asia (5.38 – 7.08; Nasidze & Stoneking, 2001). Haplotype diversity in these groups (Table 1) falls within the range observed in the Caucasus and West Asia (0.948–0.999; Nasidze *et al.* 2004).

Pairwise F_{st} comparisons between different Kurdish groups showed that the Kurmanji-T and Kurds-Tm are most different from the other Kurdish groups, whereas the other four Kurdish groups all show lower F_{st} values with each other (Table 2).

Overall, all six Kurdish groups show basically the same pattern: the smallest F_{st} values are for West Asia, next smallest are Europe [combining West and East European groups, as they do not differ significantly with respect

Table 1 MtDNA HV1 sequence variability among Kurdish populations

Population	N	no. of haplotypes	Haplotype diversity and SE	MPD	Tajima's D	Source
Kurmanji-T	51	43	0.988+/-0.008	6.95	-2.05**	present study
Zazaki-T	27	23	0.986+/-0.015	4.95	-1.81*	present study
Kurmanji-G	29	22	0.958+/-0.028	4.29	-2.15**	Comas <i>et al.</i> 2000
Kurds-I	20	19	0.995+/-0.018	6.13	-1.57*	Quintana-Murci <i>et al.</i> 2004
Kurds-Tm	32	20	0.970+/-0.015	6.55	-1.20	Quintana-Murci <i>et al.</i> 2004
Kurds-T	53	40	0.979+/-0.012	5.06	-1.94**	Richards <i>et al.</i> 2000

*P < 0.05; **P < 0.01

to pairwise F_{st} values with Kurdish groups (t-test; data not shown)], then the Caucasus, and then Central Asia (Table 2).

An MDS plot (Figure 2a) based on the pairwise F_{st} values illustrates these patterns. Most of the Kurdish groups are located near each other, whereas the Kurmanji-T and Kurds-Tm are further away. Moreover, the Kurdish groups all tend to fall within West Asian groups and near European groups, with the Caucasus and Central Asian groups being further away.

Y-SNP Haplogroups

Overall, eleven Y-SNP haplogroups were found (Table 3). The number of haplogroups varied from four in the Kurds-Tm to eleven in the Kurmanji-T. Haplogroup $P1(M124)$ was found at an unusually high frequency in the Kurmanji-G, as was haplogroup $J2(M172)$; together, these two haplogroups account for 75% of the Kurmanji-G Y-chromosomes, but less than 22% of the Y-chromosomes in any other Kurdish group. Haplogroup $I^*(M170)$ was found in substantial frequencies in the Zazaki-T and Kurmanji-T, whereas this haplogroup is absent in Kurds from Georgia or Turkmenistan. Similarly, haplogroup $E^*(YAP)$ was found at frequencies close to 0.10 in the Zazaki-T and Kurmanji-T, but was absent in Kurds from Georgia or Turkmenistan. Only haplogroup $F^*(M89)$ was found in all four groups. Haplogroup $G^*(M201)$ was not typed in the sample from Turkmenistan (Table 1) by Wells *et al.* (2001). We therefore used the same procedure as described elsewhere (Nasidze *et al.* 2003) in order to include this group in the analysis; namely, we classified

all $G^*(M201)$ individuals as the haplogroup they would belong to if M201 had not been typed, i.e. $F^*(M89)$. Haplogroup diversity varied substantially (Table 3); in most of these groups it falls within the range (0.15 – 0.86) observed in the Caucasus and West Asia (Nasidze *et al.* 2004).

All pairwise F_{st} values between Kurdish groups are significantly different from 0, except that between the Kurmanji-T and Zazaki-T. With respect to non-Kurdish populations, average F_{st} values show that all four Kurdish groups are most similar to West Asians and most distant from Europeans. Two groups (Kurmanji-T and Kurds-Tm) are next most similar to Caucasians and then to Central Asians, while the other two groups have the reverse order, being next most similar to Central Asians and then to Caucasians (Table 2). The Kurmanji-G group has the highest F_{st} values with both other Kurdish groups and with non-Kurdish groups.

The MDS analysis (Figure 2b) illustrates these patterns. The Kurmanji-G group falls outside a major cluster consisting of the other Kurdish groups and groups from the Caucasus and West Asia. Eastern and Western European groups, and Central Asian groups, fall outside this cluster.

Y-Chromosome STRs

Because of the unusually high frequency of haplogroups $P1(M124)$ and $J2^*(M172)$ in the Kurmanji-G, we typed nine Y-STR loci in individuals with these Y-SNP haplogroups. Since haplogroup $I^*(M170)$ was found at relatively high frequencies in the Zazaki-T and Kurmanji-T, we also typed the same set of Y-STR loci in

Table 2 Pairwise F_{st} values between three Kurdish groups, and Caucasian, European, and West and Central Asian populations. Below diagonal – pairwise F_{st} values based on Y-SNP haplogroups; above diagonal – pairwise F_{st} values based on mtDNA HVI sequences

	Kurmanji-T	Zazaki-T	Kurmanji-G	Kurds_I	Kurds_T	Kurds_Tm	Caucasus	West Europe	East Europe	Centr. Asia	West Asia
Kurmanji-T		0.010	0.017	0.013	0.008	0.019*	0.032*	0.011	0.013	0.076**	0.012
Zazaki-T	0.014		0.001	0.004	0.001	0.015	0.032*	0.015	0.012	0.078**	0.008
Kurmanji-G	0.091**	0.228**		0.001	0.002	0.018*	0.030*	0.019*	0.020*	0.070**	0.018
Kurds_I	N/A	N/A	N/A		0.002	0.011	0.030*	0.029*	0.030*	0.059**	0.009
Kurds_T	N/A	N/A	N/A	N/A		0.012	0.021*	0.010	0.013	0.053**	0.010
Kurds_Tm	0.073*	0.137**	0.189**	N/A	N/A		0.034*	0.020*	0.025*	0.037**	0.012
Caucasus	0.087	0.145	0.217	N/A	N/A	0.127		0.023*	0.026*	0.021*	0.020*
West Europe	0.221	0.224	0.362	N/A	N/A	0.218	0.278*		0.009	0.092**	0.025*
East Europe	0.137	0.081	0.333	N/A	N/A	0.234	0.259*	0.364**		0.076**	0.025*
Centr. Asia	0.094**	0.120	0.205	N/A	N/A	0.160	0.225*	0.376**	0.153		0.056**
West Asia	0.019	0.089	0.144	N/A	N/A	0.074**	0.251*	0.227*	0.169*	0.117	

* $p < 0.05$; ** $p < 0.01$

individuals with this Y-SNP haplogroup (Y-STR haplotypes are available from the authors upon request). We compared the results with the same set of loci on the same Y-SNP background, typed in other groups from the Caucasus, Iran and Turkey (Nasidze et al. 2003b). The average variance in number of repeats per locus on the background of haplogroup $P1(M124)$ is reduced in Kurmanji-G compared with Kurmanji-T and groups from the Caucasus, but higher than in Iran, although the latter is based on just two individuals (Table 4). The average variance in number of repeats per locus on the background of haplogroup $J2^*(M172)$ is also reduced in Kurmanji-G compared with Kurmanji-T, and roughly equal to those for their geographic neighbours, groups from the Caucasus and Iran. The average variance in number of repeats per locus on the background of haplogroup $I^*(M170)$ is roughly equal in Zazaki-T and Kurmanji-T, as well as groups from the Caucasus, Iran and Turkey (Table 4).

A median network (Figure 3) based on Y chromosome STR haplotypes on the background of haplogroup $P1(M124)$ shows close relationships between haplotypes found in the Kurmanji-G and Kurmanji-T, although there was no haplotype sharing between these two groups. The median network shows a clear distinction of Kurdish Y-STR haplotypes and haplotypes found in the Caucasus and Iran. Furthermore, clustering of the latter is much more compact, suggesting that haplogroup $P1(M124)$ may have arisen in a population ancestral to present-day Kurds. For the median network of Y-STR haplotypes on the background of haplogroup $J2^*(M172)$, individuals from different groups are intermixed (data not shown). The median network based on Y-STR haplotypes for haplogroup $I^*(M170)$ similarly did not reveal any differences between the Zazaki and Kurmanji and their geographic neighbours (data not shown).

Comparison of Mitochondrial and Y-Chromosome Data

The geographic and linguistic structure of the Kurdish, Caucasian, European, and West and Central Asian groups, as assessed by mtDNA and Y chromosome variation, was investigated by the AMOVA procedure (Table 5). As is typically seen in human populations,

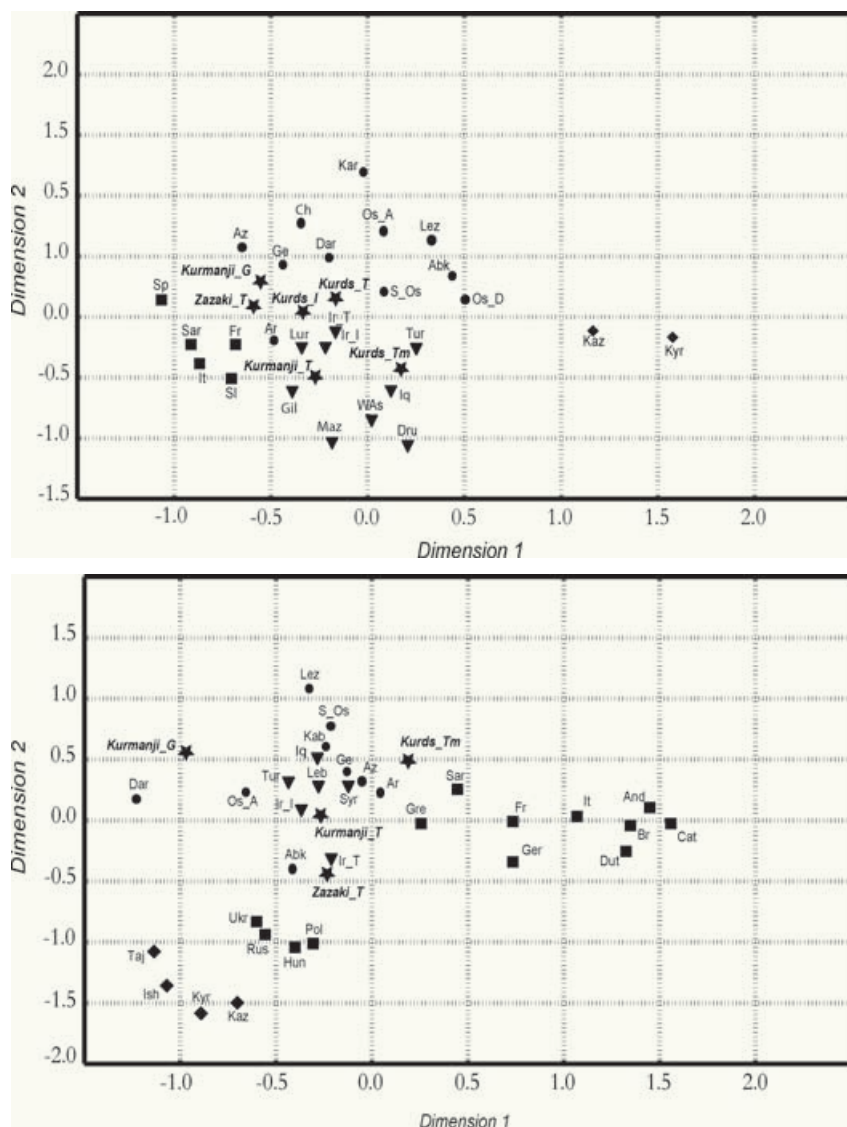


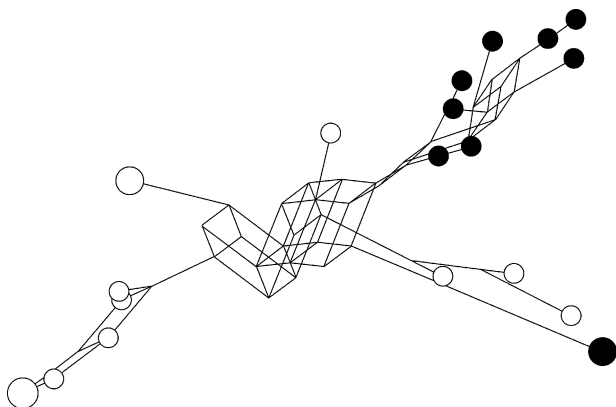
Figure 2 MDS plots based on pairwise F_{st} values, showing relationships among the Kurdish groups (stars) and Caucasian (circles), European (squares), Central (diamonds) and West Asian (triangles) groups. **A.** Based on mtDNA HV1 sequence data. The stress value for the MDS plot is 0.117. **B.** Based on Y chromosome SNP data. The stress value for the MDS plot is 0.152. The names of the populations are given using the following abbreviations: Kurmanji-G – Kurmanji speakers from Georgia, Zazaki-T – Zazaki speakers, Kurmanji-T – Kurmanji speakers, Kurds-I – Kurds from Iran, Kurds-T – Kurds from Eastern Turkey, Kurds-Tm – Kurds from Tuirkmenistan, Os-D – Ossetians from Digora, Os-A – Ossetians from Ardon, S-Os – South Ossetians, Lez – Lezginians, Ge – Georgians, Az – Azerbaijanians, Ar – Armenians, Abk – Abkhazians, Kar – Karachaians, Bal – Balkarians, Kab – Kabardinians, Ir-I – Iranians from Isfahan, Ir-T – Iranians from Tehran, Lu- Lur, Maz – Mazandarani, Gil – Gilaki, Leb – Lebanese, Syr – Syrians, Tur – Turks, Iq – Iraqi, Rus - Russians, Ukr – Ukrainians, Pol – Polish, Hung – Hungarians, Gre – Greeks, Sar – Sardinians, Fr – French, Ger – Germans, It – Italians, Br – British, And – Andalusians, Dut – Dutch, Cat – Catalans, Kyr – Kyrgyz, Ishk – Ishkinasi, Kaz – Kazakh, Taj - Tajik.

Table 3 Y chromosome haplogroup frequencies in Kurdish populations. Kurds from Turkmenistan were studied by Wells *et al.* HD, haplogroup diversity

Population	N	Haplogroups											HD
		E* YAP	C* RPS4Y	K* M9	P1 M124	P* M45	R1* M173	R1a1* M17	F* M89	G* M201	J2* M172	I* M170	
Zazaki_T	27	0.111	0.037	0	0	0.037	0.111	0.259	0.074	0.037	0	0.333	0.818
Kurmanji_T	87	0.115	0.011	0.127	0.080	0.057	0.046	0.127	0.115	0.023	0.138	0.161	0.920
Kurmanji_G	25	0	0	0.080	0.440	0.04	0	0	0.120	0	0.320	0	0.710
Kurds_Tm	17	0	0	0	0	0	0.290	0.120	0.410	-	0.180	0	0.742

Table 4 Number of Y-STR haplotypes (*n*) and average variance in number of repeats per locus (*avn*) on the background of Y-SNP haplogroups P1(M124), J2* (M172) and I* (M170) observed among Kurdish, Caucasian, Iranian and Turkish individuals. Data from Caucasus, Iran and Turkey are from Nasidze *et al.* (2003) (2004)

population	haplogroups					
	P1		J2*		I*	
	<i>n</i>	<i>avn</i>	<i>n</i>	<i>avn</i>	<i>n</i>	<i>avn</i>
Kurmanji-G	10	0.99	8	0.72	0	-
Kurmanji-T	4	1.24	15	1.23	9	1.16
Zazaki-T	0	-	0	-	19	1.26
Caucasus	8	1.27	77	0.75	42	1.15
Iran	2	0.61	17	0.72	32	1.04
Turkey	0	-	1	-	10	1.15

**Figure 3** Median network constructed for nine Y chromosome STR loci on the background of Y-SNP haplogroup P1(M124). Open circles correspond to haplotypes from Kurdish groups, while filled circles correspond to haplotypes from West Asian and Caucasian groups.

the within-populations proportion of the variance was much higher for mtDNA (about 96%) than for the Y chromosome (about 79%). For both the mtDNA and the Y-chromosome data, a geographic classification of populations gave a slightly better fit (in terms of higher

among-group variance and lower among-populations-within-groups variance) than a linguistic classification did (Table 5).

The correlation between pairwise F_{st} values based on mtDNA and Y-SNP data in the three Kurdish groups was not significant (Mantel test, $Z = -0.686$, $p = 0.825$). When the geographic and linguistic neighbours of the Kurdish groups (Iranians from Tehran and Isfahan) were included in the Mantel test, the correlation between pairwise F_{st} values based on mtDNA and Y-SNP data was still non-significant ($Z = -0.096$, $p = 0.568$).

Discussion

The Kurdish groups in this study come from different regions and speak different languages: the Kurmanji-T and Zazaki-T reside in geographical proximity, but speak different languages, whereas the Kurmanji-G and Kurds-Tm speak the Kurmanji language (Ethnologue, 2000) but are located in different geographic areas; the original publications from which the data on the Kurds-I and Kurds-T were obtained do not specify the languages spoken by these groups. How closely related genetically are the Kurdish groups living in different areas and/or speaking different languages? MtDNA data indicates close relationships among four Kurdish groups (Zazaki-T, Kurmanji-G, Kurds-I, and Kurds-T), while the Kurmanji-T and Kurds-Tm differ from these and from each other (Figure 2a). No clear pattern in terms of either geography or language is evident from our results (i.e., two Kurdish groups from Turkey are different from each other, and Kurmanji speakers from Turkey are not similar to Kurmanji speakers from Georgia). However, the Y chromosome shows somewhat different patterns, indicating some effect of geography. The

Table 5 AMOVA results according to different classifications

Classifications	mtDNA			Y-SNP		
	Among groups	Among populations within groups	Within populations	Among groups	Among populations within groups	Within populations
Geography *	2.11	1.79	96.10	7.73	13.20	79.07
Linguistic**	1.24	2.34	96.43	5.13	15.24	79.63

*West Asia, Caucasus, Europe, Central Asia.

**Caucasian, Indo-European and Turkic.

Kurmanji-T and Zazaki-T, who are geographic neighbours in Turkey, are closer to each other compared with the Kurmanji-G and Kurds-Tm (Figure 2b). The latter two groups also showed significantly reduced haplogroup diversity (Table 3), suggesting a bottleneck effect in these groups.

The Zazaki (or Dymli) language used to be classified as a Kurdish dialect, but is now considered to be a separate Iranian language (Paul, 1998). It has been suggested, on linguistic grounds, that the origins of the Zazaki people may be in northern Iran (MacKenzie, 1962). However, the Zaza people self-define themselves as Kurds (Donald Stilo, personal communication), so it was of interest to examine the genetic relationships of the Zazaki speakers with Kurdish-speaking groups. Genetically, it turned out that Zazaki speakers overall group quite closely with other Kurdish groups, and with their geographic neighbours from the South Caucasus for mtDNA, and with Kurmanji-T for the Y-SNP haplogroups. The previous hypothesis of a close relationship of the Zaza people to populations from northern Iran (MacKenzie, 1962) therefore does not gain genetic support, although the genetic evidence of course does not preclude a northern Iranian origin for the Zazaki language itself.

Kurdish languages belong to the Iranian branch of the Indo-European language family. What is the genetic relationship between Indo-European speaking Kurdish groups and other West Asian Indo-European and non-Indo-European speaking groups? For both mtDNA and the Y-chromosome, all Kurdish groups are more similar to West Asians than to Central Asian, Caucasian, or European groups, and these differences are significant in most cases. However, for mtDNA, Kurdish groups are all most similar to European groups (after West Asians), whereas for the Y-chromosome Kurds are more simi-

lar to Caucasians and Central Asians (after West Asians) than to Europeans. Richards *et al.* (2000) suggested that some Near Eastern mtDNA haplotypes, among them Kurdish ones from east Turkey, presumably originated in Europe and were associated with back-migrations from Europe to the Near East, which may explain the close relationship of Kurdish and European groups with respect to mtDNA. Subsequent migrations involving the Caucasus and Central Asia, that were largely male-mediated, could explain the closer relationship of Kurdish Y-chromosomes to Caucasian/Central Asian Y-chromosomes than to European Y-chromosomes.

Kurds migrated into the Caucasus at the end of the 19th and beginning of the 20th centuries from Turkey and/or Iran (Wixman, 1984). However, the source population for these migrants is unknown; does the genetic evidence suggest a source population for the Kurds from Georgia? Both the mtDNA and Y-chromosome data indicate reduced diversity in the Georgian Kurds, and the consequently increased effect of drift makes it more difficult to infer the origins of this group. The mtDNA data are equivocal, in that the smallest pairwise F_{st} values involving the Georgian Kurds are with Kurds from Turkey and from Iran. Interestingly, the Kurmanji speakers from Turkey exhibit a larger pairwise F_{st} value with the Georgian Kurds than the Zazaki speakers from Turkey, even though the former speak the same language as the Georgian Kurds. The Y-chromosome data do suggest that the Kurdish group in Georgia was founded by Kurmanji speakers from Turkey, although the presence of haplogroups *P1* and *J2* are at unusually high frequencies in the Georgian Kurds, which is an indication of genetic drift, making conclusions based on the Y-chromosome suspect. However, analysis of Y-STR haplotypes on the background of Y-SNP haplogroups *P1* and *J2* provided some additional insights. Y-STR haplotypes found in

the Kurmanji-G are closely related to those found in the Kurmanji-T, whereas Y-STR haplotypes found in other Turkish or Iranian groups are distinct, and more distantly related to the Y-STR haplotypes found in the Kurds. Furthermore, haplogroups *P1* and *J2* are absent in the Zazaki-T (Nasidze *et al.* 2004), indicating an even more distant relatedness with the Georgian Kurds.

The Georgian Kurds have lived in the Caucasus for more than 100 years since their migration from Turkey and/or Iran. Was there any subsequent genetic exchange between the Kurds from Georgia and the surrounding Georgian population? MtDNA data suggest that there was very little, if any, gene exchange between these two groups, as the only HV1 sequence type shared between these two groups is the Cambridge reference sequence (Anderson *et al.* 1981), which is widespread in this region (Nasidze *et al.* 2004). The absence of any other shared mtDNA sequences between Georgians and the Kurmanji-G suggests very little (if any) maternal admixture between these groups. For the Y-chromosome, the strong founder effect in the Kurmanji-G, as evidenced by the unusually high frequencies of haplogroups *P1* and *J2*, makes it difficult to draw conclusions. However, analysis of the relatedness of Y-STR haplotypes on the background of these Y-SNP haplogroups shed some additional light on this question. A median network (Figure 3) clearly shows the distinctiveness of Y-STR haplotypes in the Kurmanji-G from those in the Georgians, suggesting an absence of genetic exchange between these two groups. On the contrary, Kurmanji-G Y-STR haplotypes are closely related with Kurmanji-T haplotypes, suggesting their common origin.

In conclusion, the genetic results indicate a close relationship between the Kurmanji-G and Kurmanji-T groups, and a clear distinction of the former from Georgians, their geographic neighbours. Moreover, the genetic results indicate that during the Kurdish migration into the Caucasus they experienced a bottleneck effect, and since that time Kurds have not undergone detectable admixture with Georgians. Our results also do not support a hypothesized northern Iranian origin for the Zazaki people; instead, genetically they are closely related to Kurds. Thus, genetic studies of even such recent events as the origins and migrations of Kurds can provide additional insights into the circumstances surrounding such migrations.

Acknowledgments

We are grateful to the original donors for providing DNA samples. We thank Donald Stilo for useful discussion. This research was supported by funding from the Max Planck Society, Germany.

References

- Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A. & Sanachiara-Benerecetti, A. S. (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a cross-road of the early human dispersal and of post-Neolithic migrations. *Mol Phylog Evol* **28**, 458–472.
- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- Bakaev, C. (1973) Language of Kurds from USSR. Comparative characteristics of dialects. Moscow: Acad. Nauk (in Russian).
- Bandelt, H.-J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48.
- Bandelt, H.-J., Quintana-Murci, L., Salas, A. & Macaulay, V. (2002) The Fingerprint of Phantom Mutations in Mitochondrial DNA Data. *Am J Hum Genet* **71**, 1150–1160.
- Bertorelle, G. & Excoffier, L. (1998) Inferring admixture proportion from molecular data. *Mol Biol Evol* **15**, 1298–1311.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) History and geography of human genes. Princeton, NJ: Princeton University Press.
- Asian Populations. *Am J Hum Genet* **63**, 1824–1838.
- Comas, D., Calafell, F., Bendukidze, N., Fananas, L. & 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.
- Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E., Martínez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D. & Bertranpetit, J. (1998) Trading Genes along the Silk Road: mtDNA Sequences and the Origin of Central Asian Populations. *Am J Hum Genet* **63**, 1824–1838.
- Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Saha, N., Sirajuddin, S. M. & Stoneking, M. (2004) Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* **14**, 231–235.
- Cordaux, R., Saha, N., Bentley, G. R., Aunger, R., Sirajuddin, S. M. & Stoneking, M. (2003) Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* **11**, 253–264.

- Côrte -Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–50.
- Di Benedetto, G., Erguven, A., Stenico, M., Castri, L., Bertorelle, G., Togan, I. & Barbujani, G. (2001) DNA diversity and population admixture in Anatolia. *Am J Phys Anthropol* **115**, 144–156.
- Di Rienzo, A. & Wilson, A. C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* **88**, 1597–1601.
- Dupanloup, I. & Bertorelle, G. (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol.* **18**, 672–675.
- Ethnologue (2000) (www.ethnologue.com)
- Hammer, M. & Horai, S. (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* **56**, 951–962.
- Kayser, M., Brauer, S., Weiss, G., Underhill, P. A., Roewer, L., Schiefenhover, W. & Stoneking, M. (2000) Melanesian origin of Polynesian Y chromosome. *Curr Biol* **10**, 1237–1246.
- Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F. et al (1997) Evaluation of Y chromosomal STRs: a multicenter study. *Int J Legal Med* **110**, 125–133.
- Kayser, M., Krawczak, M., Excoffier, L., Deiltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L. F., Jespersen, J., Bakker, E., Rower, L. & Kniff, P. (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* **68**, 990–1018.
- Ke, Y., Su, B., Song, X., Lu, D., Chen, L., Li, H., Qi, C., Marzuki, S., Deka, R., Underhill, P., Xiao, C., Shriver, M., Lell, J., Wallace, D., Wells, R. S., Seielstad, M., Oefner, P., Zhu, D., Jin, J., Huang, W., Chakraborty, R., Chen, Z. & Jin, L. (2001) African Origin of Modern Humans in East Asia: A Tale of 12,000 Y Chromosomes. *Science* **292**, 1151–1153.
- Kittler, R., Erler, A., Brauer, S., Stoneking, M. & Kayser, M. (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur J Hum Genet* **11**, 304–14.
- Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W. S., Dixon, M. E., Papiha, S. S., Mastana, S. S., Mir, M. R., Ferak, V. & Villems, R. (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* **9**, 1331–1334.
- Kruskal, J. B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.
- MacKenzie, D. N. (1962) Kurdish dialect studies.
- Maliarchuk, B. A., Derenko, M. V. & Solovenchuk, L. L. (1995) Types of regulatory regions in mitochondrial DNA in eastern Slavs. *Genetika (Russia)* **31**, 846–51.
- Maniatis, T., Fritsh, E. F. & Sambrook, J. (1982) Molecular cloning. A laboratory manual. New York: Cold Spring Laboratory.
- Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215.
- Nasidze, I., Ling, E. Y. S., Quinque, D., Dupanloup, I., Cordaux, R., Rychkov, S., Naumova, O., Zhukova, O., Sarraf-Zadegan, N., Naderi, G. A., Asgary, S., Sardas, S., Farhud, D. D., Sarkisian, T., Asadov, C., Kerimov, A. & Stoneking, M. (2004) Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* **68**, 205–221.
- Nasidze, I., Sarkisian, T., Kerimov, A. & Stoneking, M. (2003a) Testing hypotheses of language replacement in the Caucasus: Evidence from the Y-chromosome. *Hum Genet* **112**, 255–261.
- Nasidze, I., Schadlich, H. & Stoneking, M. (2003b) Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Sci Int* **137**, 85–93.
- Nasidze, I. & Stoneking, M. (2001) Mitochondrial DNA variation and language replacements in the Caucasus. *Proc R Soc Lond B* **268**, 1197–1206.
- Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., Faerman, M. & Oppenheim, A. (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* **69**, 1095–1112.
- Paul, L. (1998) Zazaki: Grammatik und Versuch einer Dialektologie. Beiträge zur Iranistik. Reichert verlage, Wiesbaden.
- Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A. S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Qasim Mehdi, S., Torroni, A. & McElreavey, K. (2004) Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. *Am J Hum Genet* **74**, 827–845.
- Redd, A. J., Takezaki, N., Sherry, S. T., McGarvey, S. T., Sofro, A. S. M. & Stoneking, M. (1995) Evolutionary history of the COII/tRNA-lys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* **12**, 604–615.
- Richards, M., Côrte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. & Sykes, B. (1996) Paleolithic and

- Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**, 185–203.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., *et al.* (2000) Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *Am J Hum Genet* **67**, 1251–1276.
- Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikia, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, S., L. Cavalli-Sforza, L. & Underhill, P. A. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver 2.000: A software for population genetics data analysis*. University of Geneva, Switzerland: Genetics and Biometry Laboratory.
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Qasim Mehdi, S., Seielstad, M. T., Wells, S., Piazza, A., Davis, R. W., Feldman, M. W., L. Cavalli-Sforza, L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.
- Vigilant, L., Pennington, R., Harpending, H., Kocher, T. D. & Wilson, A. C. (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* **86**, 9350–9354.
- Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shannmugalakshimi, S., Balakrishnan, K. *et al.* (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* **98**, 10244–10249.
- Wilson, J. & Erlandsson, R. (1998) Sexing of human and other primate DNA. *Biol Chem* **379**, 1287–1288.
- Wixman, R. (1984) The peoples of the USSR. *An ethnographic handbook*. Armonk, NY: M.E. Sharpe, Inc.
- Y chromosome Consortium (2002) A nomenclature system for the tree of Human Y-chromosomal binary haplogroups. Genome Research. *Genome Res* **2**, 339–348.
- Yoshimoto, K., Iwahana, H., Fukuda, A., Sano, T. & Itakura, M. (1993) Rare mutations of the Gs alpha subunit gene in human endocrine tumors. Mutation detection by polymerase chain reaction-primer-introduced restriction analysis. *Cancer* **72**, 1386–1393.

Received: 27 July 2004

Accepted: 19 November 2004