doi: 10.1046/j.1529-8817.2004.00092.x

# Mitochondrial DNA and Y-Chromosome Variation in the Caucasus

I. Nasidze[1,*], E. Y. S. Ling[1,2], D. Quinque[1], I. Dupanloup[3], R. Cordaux[1], S. Rychkov[4], O. Naumova[4], O. Zhukova[4], N. Sarraf-Zadegan[5], G. A. Naderi[5], S. Asgary[5], S. Sardas[6], D. D. Farhud[7], T. Sarkisian[8], C. Asadov[9], A. Kerimov [9] and M. Stoneking[1]

*[1]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, D-04103, Germany; [2]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK (current address: The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK); [3]Ferrara University, Department of Biology, Via L.Borsari 46, 44100, Ferrara, Italy; [4]Vavilov Institute of General Genetics, Gubkin str.3, 119991, Moscow, Russia; [5]Isfahan Cardiovascular Research Center, Isfahan, Iran; [6]Gazi University, Toxicology Department 06330, Ankara, Turkey; [7]Tehran University of Medical Sciences, Department of Human Genetics, Tehran, P.O. Box 14155-6446, Iran; [8]Center of Medical Genetics, National Academy of Sciences of Republic of Armenia, 5/1 Zakyan Str., 375010, Yerevan, Armenia; [9]Scientific-Research Institute of Hematology and Transfusiology, Azerbaijan Republic Ministry of Health, Gashgay Str. 87, Baku, Azerbaijan*

## Summary

We have analyzed mtDNA HVI sequences and Y chromosome haplogroups based on 11 binary markers in 371 individuals, from 11 populations in the Caucasus and the neighbouring countries of Turkey and Iran. Y chromosome haplogroup diversity in the Caucasus was almost as high as in Central Asia and the Near East, and significantly higher than in Europe. More than 27% of the variance in Y-haplogroups can be attributed to differences between populations, whereas mtDNA showed much lower heterogeneity between populations (less then 5%), suggesting a strong influence of patrilocal social structure. Several groups from the highland region of the Caucasus exhibited low diversity and high differentiation for either or both genetic systems, reflecting enhanced genetic drift in these small, isolated populations. Overall, the Caucasus groups showed greater similarity with West Asian than with European groups for both genetic systems, although this similarity was much more pronounced for the Y chromosome than for mtDNA, suggesting that male-mediated migrations from West Asia have influenced the genetic structure of Caucasus populations.

Keywords: Y chromosome, mtDNA, Caucasus

## Introduction

The Caucasus region provides an unparalleled opportunity to investigate the influence of geography and language on the genetic structure of human populations. There are approximately 50 ethnic groups in the Caucasus, speaking languages belonging to four major families (South Caucasian, North Caucasian, Indo-European and Turkic), while the Caucasus Mountains, with peaks up to six thousand meters high, are a potential major geographic barrier that divides the region into the North and South Caucasus.

Previous analyses of a number of classical genetic markers (blood groups, serum proteins, red cell enzymes) showed substantial genetic diversity in the Caucasus (Barbujani *et al.* 1994a,b). The results of these studies indicated a single ancient origin for the Caucasus populations, with subsequent subdivision along linguistic and geographic borders.

More recently, a study of sequence variation in the first hypervariable segment (HV1) of the mtDNA control region in nine populations from the Caucasus showed a high level of diversity, exceeding that found

*Correspondence: Dr. Ivan Nasidze, Phone: (49) 341 3550505. Fax: (49) 341 3550555, e-mail: nasidze@eva.mpg.de

within Europe, and only slightly lower than West Asian mtDNA diversity (Nasidze & Stoneking, 2001). The genetic relationships among Caucasus groups reflected geographic rather than linguistic relationships. In particular, Caucasus groups speaking non-Caucasian languages (i.e., Indo-European speaking Armenians and Turkic-speaking Azerbaijanians) grouped with their geographic neighbours in the Caucasus, and not with their linguistic neighbours elsewhere (i.e., other Indo-European or Turkic groups, respectively). Phylogenetic and principal coordinate analyses placed the Caucasus populations in an intermediate position between West Asian and European groups, albeit much closer genetically to the latter, suggesting a common ancestry and/or admixture of European and Caucasus groups.

Studies of eight *Alu* insertion loci (Nasidze *et al.* 2001) and 11 bi-allelic Y-chromosome markers (Nasidze *et al.* 2003) in largely the same set of Caucasus groups reinforced the conclusions based on mtDNA analysis. Neither geographic nor linguistic factors had a strong influence on the genetic structure of the Caucasus groups. Overall, it appears that the major factor in the region has been primarily genetic drift, caused by small population size and/or isolation.

However, the Y-chromosome results differed from the mtDNA results in that they indicated a much closer relationship of Caucasus groups to West Asian groups, whereas mtDNA indicated a closer relationship of Caucasus groups to Europe than to West Asia. Since the above studies were based on a limited number of Caucasus groups, we have expanded the sampling both of Caucasus groups and of neighbouring West Asian groups, to more fully investigate the influence of geographic and linguistic factors on the genetic structure of the Caucasus groups. We report here the results of analyses of mtDNA and Y-chromosome variation; these results extend our previous studies, but differ in that the mtDNA results now indicate a closer relationship between Caucasus and West Asian groups than was found previously.

## Materials and Methods

### Samples

A total of 371 samples (182 whole blood and 189 cheek cell swabs) from unrelated individuals were obtained from 11 populations in the Caucasus, Turkey and Iran (Table 1). The number of Avarian, Balkarian and Karachaian male samples was insufficient for Y chromosome analyses. Published data from an additional 14 Caucasus groups were included (Table 1); a map of all of the Caucasus and West Asian sampling localities is in Figure 1. Additional published Y-SNP data on 21 Azerbaijanians and 47 Armenians (Wells *et al.* 2001), and 63 Georgians (Semino *et al.* 2000; Wells *et al.* 2001), do not differ from our samples [the new samples did not differ from the published samples in that $F_{st}$ values were not significantly different from zero (data not shown)] and were not included in the analysis because having two samples each of Azerbaijanians, Georgians, and Armenians (all from the South Caucasus) but only one sample each from the north Caucasus groups would introduce an uneven weighting based on the number of samples. Published Y-SNP data (Semino *et al.* 2000; Wells *et al.* 2001) for European, West Asian, and Central Asian populations were also included in some analyses, as were published mtDNA HV1 sequences from 106 Basques (Bertranpetit *et al.* 1995; Corte-Real *et al.* 1996), 101 British (Piercy *et al.* 1993), 69 Sardinians and 42 Middle Easterners (DiRienzo & Wilson, 1991), 72 Spaniards, 37 Poles, 83 Czechs (Corte-Real *et al.* 1996; Richards *et al.* 1996), 102 Russians (Orekhov *et al.* 1999), 18 Slavs (Maliarchuk *et al.* 1995), 101 Estonians (Richards *et al.* 2000), 45 Israeli Drusi (Macaulay *et al.* 1999), and 29 Kurds (Comas *et al.* 2000).

### DNA Extraction

Genomic DNA from blood samples was extracted using a conventional phenol-chloroform method (Maniatis *et al.* 1982). DNA from cheek cell swabs was extracted using a salting-out procedure (Miller *et al.* 1988).

### MtDNA HV1 Sequencing

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

**Table 1** Population information

| Population | mtDNA* | source | Y chromosome* | source | language |
|---|---|---|---|---|---|
| *North Caucasus* | | | | | |
| Abazinians | 23 | Nasidze & Stoneking, 2001 | 14 | Nasidze *et al*. 2003 | N. Caucasian |
| Adygheians | 50 | Macaulay *et al*. 1999 | N/A | N/A | N. Caucasian |
| Balkarians | 16 | present study | N/A | N/A | Turkic |
| Avarians | 32 | present study | N/A | N/A | N. Caucasian |
| Chechenians | 23 | Nasidze & Stoneking, 2001 | 19 | Nasidze *et al*. 2003 | N. Caucasian |
| Cherkessians | 44 | Nasidze & Stoneking, 2001 | N/A | N/A | N. Caucasian |
| Darginians | 37 | Nasidze & Stoneking, 2001 | 26 | Nasidze *et al*. 2003 | N. Caucasian |
| Ingushians | 35 | Nasidze & Stoneking, 2001 | 22 | Nasidze *et al*. 2003 | N. Caucasian |
| Kabardinians | 51 | Nasidze & Stoneking, 2001 | 59 | Nasidze *et al*. 2003 | N. Caucasian |
| Karachaians | 13 | present study | N/A | N/A | Turkic |
| Lezginians | 45 | present study | 25 | present study | N. Caucasian |
| Ossetians (Ardon) | 26 | present study | 28 | present study | Indo-European |
| Ossetians (Digora) | 30 | present study | 31 | present study | Indo-European |
| Rutulians | 31 | present study | 24 | present study | N. Caucasian |
| *South Caucasus* | | | | | |
| Abkhazians | 27 | present study | 12 | present study | N. Caucasian |
| Armenians | 42 | Nasidze & Stoneking, 2001 | 100 | Nasidze *et al*. 2003 | Indo-European |
| Azerbaijanians | 41 | Nasidze & Stoneking, 2001 | 72 | Nasidze *et al*. 2003 | Turkic |
| Georgians | 57 | Nasidze & Stoneking, 2001 | 77 | Nasidze *et al*. 2003 | S. Caucasian |
| Kazbegi | N/A | N/A | 25 | Wells *et al*. 2001 | S. Caucasian |
| Lezginians (Azerbaijan) | N/A | N/A | 12 | Wells *et al*. 2001 | N. Caucasian |
| S. Ossetians | 201 | Kivisild *et al*. 1999 | 17 | Wells *et al*. 2001 | Indo-European |
| Svans | N/A | N/A | 25 | Wells *et al*. 2001 | S. Caucasian |
| *West Asia* | | | | | |
| (Tehran) Iranians | 79 | present study | 80 | present study | Indo-European |
| (Isfahan) Iranians | 46 | present study | 50 | present study | Indo-European |
| Turks | 39 | present study | 39 | present study | Turkic |
| **Total** | **991** | | **757** | | |

*number of individuals studied.

(Cordaux *et al*. 2003) and H16401 (Vigilant *et al*. 1989) were used to determine sequences for both strands of the PCR products with the Big Dye DNA Sequencing Kit (Perkin-Elmer), following the protocol recommended by the supplier, and an ABI 3700 automated DNA sequencer. Those sequences with a C at position 16189 (Anderson *et al*. 1981) terminate prematurely in each direction at the "C-stretch" region between positions 16184-16193; sequences with 16189C were sequenced again in each direction, so that each base was determined twice. To verify the accuracy of this procedure, we cloned PCR products from 3 different individuals with 16189C using the TOPO TA cloning kit (Invitrogen) and the protocol recommended by the supplier. We sequenced 2 to 4 clones for each PCR product as described above, using vector specific M13 reverse and M13(-20) forward primers.

MtDNA HVI sequences have been submitted to the mtDNA sequence database (www.HVRbase.de).

## Y Chromosome Markers

Ten Y chromosomal SNP markers previously reported to be polymorphic in Europe and the Near East (Semino *et al*. 2000) were analyzed: 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. All SNP markers (except M130) were typed by means of Taqman® (Applied Biosystems) assays as described previously (Nasidze *et al*. 2003). M130

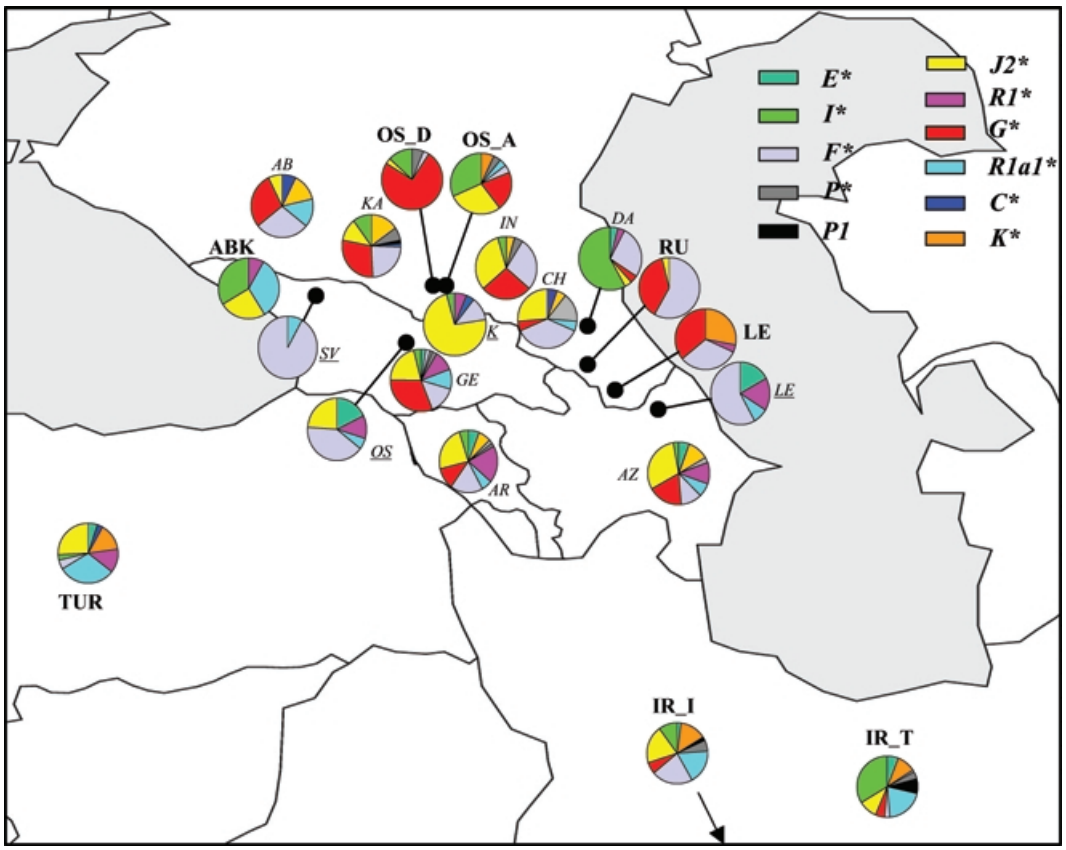


**Figure 1** Map of the Caucasus showing the Y-SNP haplogroup frequencies. AB - Abazinians, Abk – Abkhazians, Ar – Armenians, Az – Azerbaijanians, Ch – Chechenians, Da – Darginians, Ge – Georgians, In – Ingushians, K – Georgians from Kazbegi, Ka – Kabardinians, Le_Az – Lezginians from Azerbaujan, Le_Dag – Lezginians from Dagestan, Os – South Ossetians, Os_A – Ossetians (Ardon), Os_D – Ossetians (Digora), Ru – Rutulians, Sv – Svans, Tur – Turks, Ir_I – Iranians (Isfahan), Ir_T – Iranians (Tehran). Population names in boldface are from the present study, in italic from Nasidze *et al.* (2003), and in underlined italic from Wells *et al.* (2001).

was typed using the PCR-RFLP method described elsewhere (Kayser *et al.* 2000), while the YAP *Alu* insertion was typed as described previously (Hammer & Horai, 1995) The samples were genotyped according to the hierarchical order of the markers (Underhill *et al.* 2000). YAP and M130 were first typed in all individuals, then M89 was typed in individuals with the ancestral state at both YAP and M130. M9 was typed in individuals with the mutant state at M89, and M45 was typed in individuals with the mutant state at M9. M124 was typed in individuals with the mutant state at M45, while M173 was typed in individuals with the mutant state at M45 and the ancestral state at M124. M17 was typed only in individuals with the mutant state at M173. M170, M172 and M201 were typed in individuals with the mutant state at M89 and the ancestral state at M9. The Y-SNP haplogroup nomenclature used here is according to the recommendations of the Y Chromosome Consortium (2002).

## Statistical Analyses

Basic parameters of molecular diversity and population genetic structure, including analyses of molecular variance (AMOVA), were calculated using the package Arlequin 2.000 (Schneider *et al.* 2000). The statistical significance of $F_{st}$ values was estimated by permutation analysis, using 10,000 permutations. $F_{st}$ values were calculated based on the number of pairwise differences between sequences (for mtDNA) or

 

haplogroups (for the Y chromosome). The statistical significance of the correlation between geographic distance and genetic distance matrices, as well as between genetic distance matrices based on mtDNA vs. Y chromosome data, was evaluated by the Mantel test with 10,000 permutations. The STATISTICA package (StatSoft Inc.) was used for multidimensional scaling (MDS) analysis (Kruskal, 1964). Since data for both mtDNA and the Y chromosome are not available from all populations from the Caucasus, two datasets were used to compare the mtDNA and Y chromosome results. The complete data set includes all available information for mtDNA HVI sequences and Y chromosome SNP genotypes, while the reduced data set consists only of those populations for which both mtDNA and Y chromosome data were available.

In order to define geographic regions of large genetic changes (or genetic barriers), spatial analysis of molecular variance (SAMOVA) was used (Dupanloup *et al.* 2002). In contrast to other tests of genetic structure (Barbujani *et al.* 1989; Barbujani & Sokal, 1990), in which groups are defined *a priori* based on geographical, linguistic, or ecological characters, SAMOVA determines groups of populations that are geographically and genetically homogeneous and maximally differentiated from each other, based solely on genetic data. The method is based on a simulated annealing procedure that aims at maximizing the proportion of the total genetic variance due to differences between groups of populations; the final number of groups is based on the largest value for this variance component. As a by-product, SAMOVA also leads to the identification of genetic barriers between these groups.

## Results

### MtDNA HVI Sequence Validation and Variability

A total of 377 bp of the mtDNA HV1 region, comprising nucleotide positions 16024 to 16400 (Anderson *et al*. 1981), were determined for 368 individuals from the seven Caucasus groups and three neighbouring groups from Turkey and Iran. As a check on the accuracy of the HV1 sequences, we used the network method of Bandelt *et al.* (2002) to search for so-called "phantom" mutations in both the new HV1 sequences reported here and in the Caucasus HV1 sequences that we published previously (Nasidze & Stoneking, 2001). This method eliminates presumed rapidly-evolving sites from the data and then looks for excessive reticulations (indicating parallel mutations) in networks based on the remaining slowly-evolving sites; since slowly-evolving sites are not expected to undergo parallel mutations, any reticulations may indicate sequence errors. No excess reticulations were observed in the new data, but a large number of reticulations were observed in the network for the previously-published Caucasus sequences (analysis not shown). These reticulations were due to just one sequence, as removal of this sequence from the analysis removed the excess reticulations. We resequenced this sample and found that the previous sequence for this sample was not verified; we therefore included the corrected sequence for this sample (from a Cherkessian, CH579) in the analyses in the present paper.

It has also been claimed that sequences with 16189C are particularly error-prone (Bandelt *et al*. 2002), since the sequences terminate prematurely at the C-stretch region in both directions. We routinely sequence each sample with 16189C twice in each direction, so that each base is determined twice. However, since each base is determined twice from the same strand, in theory it is possible for strand-specific artifacts to be reproduced by this method, leading to sequence errors. To check on the accuracy of our method, we cloned the PCR products from 3 samples with 16189C and sequenced several clones. Clear sequences could be obtained from the clones, in accordance with the view that the problem with sequencing through the C-stretch lies with polymerase slippage during PCR amplification that then generates templates of different lengths, not with the sequencing itself. The clone sequences did not differ from the sequences obtained directly from the PCR products, indicating that sequencing each strand twice is a reliable method for determining HV1 sequences from samples with 16189C.

For the purposes of comparing the sequences reported here with published data, further analyses were

**Table 2** MtDNA HV1 sequence variability among the Caucasus and West Asian populations

| Population | N | No. of Haplotypes | Nucleotide Diversity | Haplotype Diversity | MPD | Tajima's D |
|---|---|---|---|---|---|---|
| Abazinians | 23 | 19 | 0.014 | 0.980 | 5.19 | −2.02** |
| Adyghe | 50 | 32 | 0.014 | 0.954 | 4.98 | −1.55* |
| Balkarians | 16 | 13 | 0.018 | 0.975 | 5.87 | −1.15 |
| Avarians | 32 | 26 | 0.016 | 0.988 | 5.48 | −1.87* |
| Chechenians | 23 | 18 | 0.012 | 0.972 | 4.40 | −1.67* |
| Cherkessians | 44 | 37 | 0.015 | 0.986 | 5.35 | −1.98** |
| Darginians | 37 | 27 | 0.014 | 0.975 | 5.10 | −2.04** |
| Ingushians | 35 | 26 | 0.013 | 0.970 | 4.75 | −1.57* |
| Kabardinians | 51 | 36 | 0.013 | 0.975 | 4.88 | −2.23** |
| Karachaians | 13 | 10 | 0.015 | 0.949 | 5.31 | −1.23 |
| Lezginians | 45 | 34 | 0.016 | 0.985 | 5.67 | −1.52* |
| Ossetians, Ardon | 26 | 19 | 0.013 | 0.948 | 4.65 | −1.66* |
| Ossetians, Digora | 30 | 21 | 0.016 | 0.977 | 5.46 | −1.26 |
| Rutulians | 31 | 25 | 0.015 | 0.985 | 5.33 | −1.90* |
| Abkhazians | 27 | 19 | 0.016 | 0.969 | 5.86 | −1.25 |
| Armenians | 42 | 35 | 0.014 | 0.980 | 5.22 | −2.18** |
| Azerbaijanians | 41 | 37 | 0.014 | 0.995 | 5.17 | −2.13** |
| Georgians | 57 | 40 | 0.014 | 0.971 | 5.16 | −1.99** |
| S. Ossetians | 201 | 65 | 0.018 | 0.969 | 6.44 | −1.40* |
| Iranians, Tehran | 79 | 63 | 0.015 | 0.984 | 5.53 | −2.05** |
| Iranians, Isfahan | 46 | 42 | 0.017 | 0.996 | 6.17 | −2.13** |
| Turks, Ankara | 39 | 38 | 0.015 | 0.999 | 5.51 | −1.91** |

*$P < 0.05$; **$P < 0.01$

restricted to 365 bp (nucleotide positions 16024-16388) of HV1. The nucleotide diversity ranged from 0.012 to 0.018 in the different Caucasus groups, while the haplotype diversity ranged from 0.947 to 0.995 (Table 2). The mean number of pairwise nucleotide differences was fairly uniform across the different Caucasus groups, ranging from 4.40-6.44. These estimates are towards the upper limit of the range of mean pairwise differences found in European populations (3.15-5.03; Comas *et al*. 1997) but lower than those for Middle Eastern groups (5.38–7.08) (Nasidze & Stoneking, 2001). The mismatch distributions for the Caucasus groups are all roughly bell-shaped (data not shown), suggesting prehistoric population expansions. This demographic scenario is reinforced by Tajima's D statistic (Tajima, 1989), which is negative in all of the Caucasus groups, and significantly so (after Bonferroni correction for multiple tests) in all but the Abkhazians, Balkarians, Karachaians, and Ossetians from Digora (Table 2); negative values of D, together with bell-shaped mismatch distributions, are signatures of population expansions (Aris-Brosou & Excoffier, 1996).

An MDS plot (Figure 2a) based on $F_{st}$ values placed the Caucasus groups in an intermediate position between West Asian and European groups, except that the two Iranian groups are much closer to the Caucasus groups than to the other West Asian groups. The Karachaians are separated from the major Caucasus cluster in the second dimension; neither geographic nor linguistic clusters are apparent in the plot.

The correlation between the geographic and genetic (pairwise $F_{st}$) distances separating pairs of Caucasus populations was not statistically significant (Mantel test, $Z = -0.177$, $P = 0.839$). Removing the Karachaians and other possible outliers (Ingushians, Lezginians, Abkhasians, Balkarians, North Ossetians (Digora) and South Ossetians) resulted in a correlation that was slightly higher but still non-significant ($Z = -0.110$, $P = 0.717$).

## Y-SNP Haplogroups in the Caucasus

Eleven Y-SNP haplogroups were found in the Caucasus (Table 3, Figure 1). The most frequent haplogroups

 

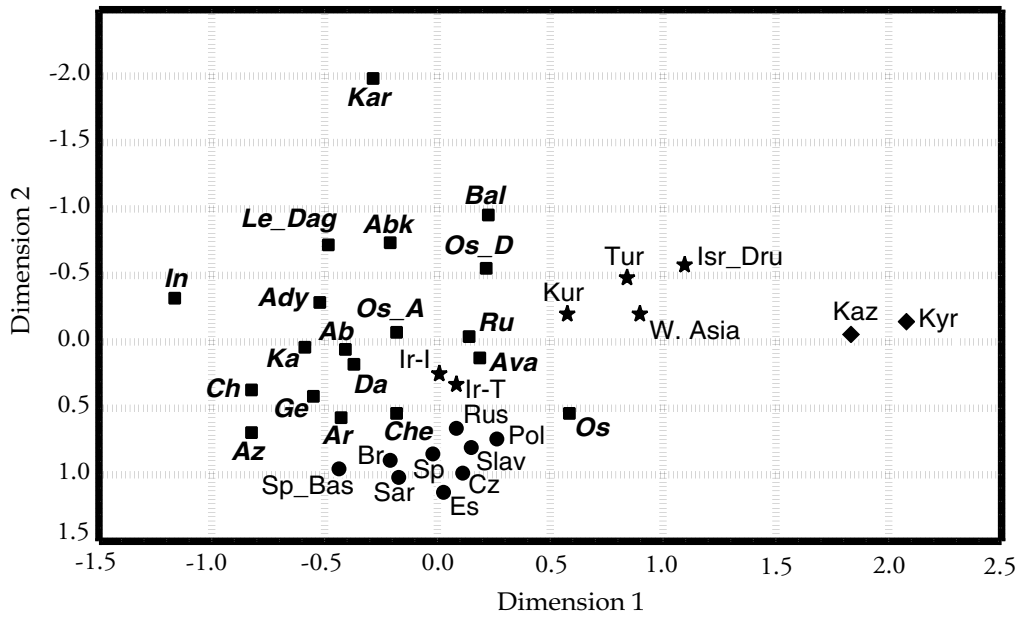


**Figure 2** MDS plot based on pairwise $F_{st}$ values, showing relationships among the Caucasus, European, West Asian, and Central Asian populations. The names of the Caucasus and some neighbouring populations are given the same abbreviations as in Figure 1. Additional populations from the Caucasus studied for mtDNA are as follows: Ady – Adyghe, Bal – Balkarians, Ava – Avarians, Kar – Karachaians, Kur – Kurds. European, West and Central Asian populations are abbreviated as follows: Sp_Bas – Spanish Basques, Fr_Bas – French Basques, Sar – Sardinians, And – Andalusians, Br – British, Sp – Spanish, Cat – Catalans, Dut – Dutch, It – Italians, Fr- French, Gre – Greek, Ger – Germans, Cz & Sl – Czech and Slovak, Kar – Karakalpak, Bar – Bachtiar, Uz – Uzbek, Kyr – Kyrgiz, Kaz – Kazakh, Dun – Dungan, Taj – Tajik, Isr_Dru – Israeli Drusi, W.Asia – West Asians, Ukr – Ukrainians, Rus – Russians, Slav – Slavs, Pol – Polish, Hun – Hungarians, Es – Estonians, Leb – Lebanese, Syr – Syrians. A. Based on mtDNA data. Squares correspond to Caucasus groups (population names are in boldface italic), circles to European groups, diamonds to Central Asian groups, and stars to West Asian groups. The stress value for the MDS plot is 0.113. B. Based on Y chromosome data. The same symbols are used to designate groups as in Figure 2A, except that West European groups are presented by circles and East European groups by triangles. The stress value for the MDS plot is 0.122.

were F*, G* and J2*; together the frequency of these three haplogroups was 0.53-1.00 in almost all groups except for the Darginians and Abkhazians (where the frequencies were 0.35 and 0.25 respectively). North Ossetians from Digora had the highest frequency of the haplogroup G* (0.74). Three populations from the highland region of the Caucasus – Rutulians (present study), and Lezgi and Svans reported by Wells *et al.* (2001) - had a high frequency of haplogroup F* (0.58, 0.58 and 0.92 respectively). Haplogroup I* was at high frequency in Darginians (0.58), Abkhazians (0.33), and North Ossetians from Ardon (0.32). This haplogroup was found elsewhere in the Caucasus at a frequency of only 0.13 or less, although it was also at high frequency in the Turks (0.26) and Iranians from Tehran (0.34). The Georgian population from Kazbegi had a high frequency of haplogroup J2* (0.72) (Wells *et al.* 2001).

The North Ossetians from Digora, Darginians, Rutulians, Lezgi (from the S. Caucasus), Svans, and Kazbegi had the lowest haplogroup diversities (range 0.15-0.65), while for the other groups the haplogroup diversity was 0.72–0.86. Almost all of the groups with low haplogroup diversities are small in size, isolated, and inhabit the highland Caucasus region. The overall Y-haplogroup diversity in the Caucasus was equal to 0.688. When we excluded those groups with extremely low diversity valeus (Darginian, Svan, Abkhazian, Ossetian (Digora), Rutulian and Kazbegi), the Y-haplogroup diversity became almost as high (average value 0.790) as in Central Asia (average value 0.824) and West Asia (average value 0.769), and was significantly higher (t-test,

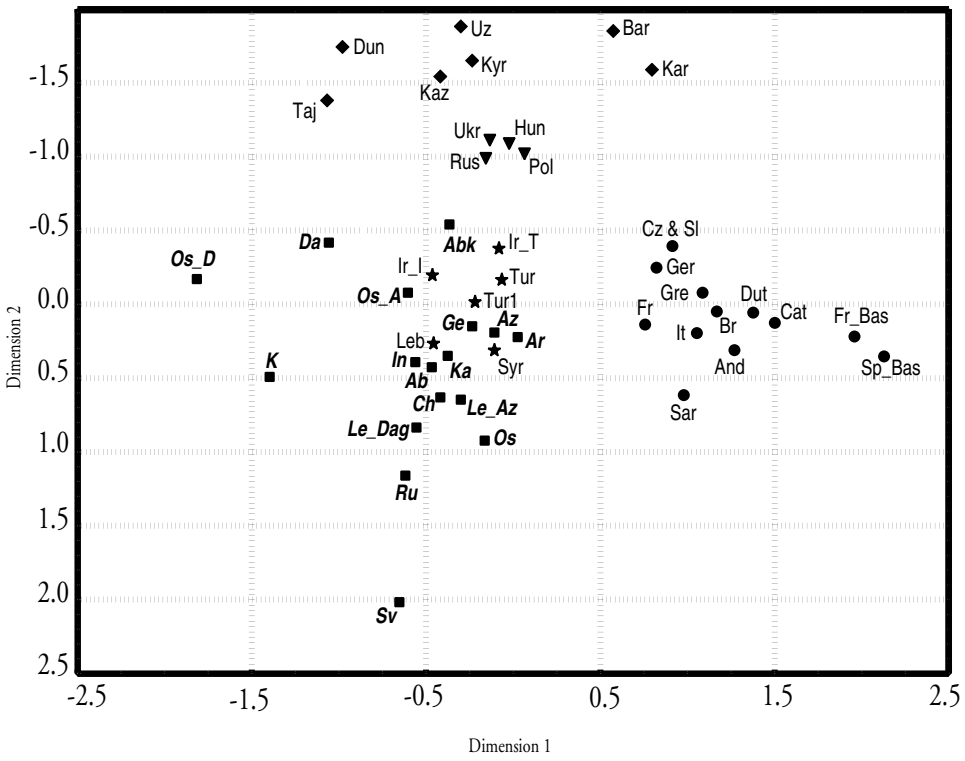


**Figure 2** (Continued)

P = 0.024) than the haplogroup diversity in Europe (average value 0.633).

When compared with other populations, the common Caucasus haplogroup, G*, is rare or absent in Europe and in the Turkish and Lebanese groups (present study; Semino *et al*. 2000), but not in populations from Tehran and Isfahan (frequency of 0.1 and 0.2 respectively). Two other common Caucasus haplogroups (F* and J2*) are also common in some Near Eastern populations, with frequencies between 0.03-0.40 and 0.03-0.30 respectively, but are also found in lower frequencies in Europe, with average frequencies of 0.021 and 0.074 respectively (present study; Semino *et al*. 2000). Haplogroup R1*, which is common in Western and Central Europe, is observed mostly in the South Caucasus. The other Caucasus Y-haplogroups occurred predominantly at low frequencies (Table 3).

The MDS and $F_{st}$ analyses included some groups from Wells *et al*. (2001) in which the M201 marker, which distinguishes haplogroup G* from haplogroup F*, was not analyzed (Table 3). In the above analyses, these individuals were classified as haplogroup F*, although some unknown proportion could belong to haplogroup G*. To determine the impact of this classification on the results, we followed our previous procedure (Nasidze *et al*. 2003) and classified the haplogroup G* individuals from all populations as haplogroup F*, and repeated the analyses. The results (not shown) were essentially identical; thus the inability to distinguish between haplogroups F* and G* in some groups does not influence our conclusions. An MDS plot (Figure 2b) based on $F_{st}$ values splits European populations into Western and Eastern groups, as has been observed previously (Semino *et al*. 2000). The Caucasus populations are intermingled with West Asian populations. Svans, Ossetians (Digora), Kazbegi, and Darginians are fairly well separated from the remaining Caucasus groups. The pairwise $F_{st}$ value was highest between Svans and the other Caucasus groups (average $F_{st}$ = 0.380), followed by the Kazbegi (average $F_{st}$ = 0.281), Darginians (average $F_{st}$ = 0.248), Abkhazians (average $F_{st}$ = 0.164), Rutulians (average $F_{st}$ = 0.154), and North Ossetians from Digora (average $F_{st}$ = 0.153); the average pairwise $F_{st}$ value among the remaining Caucasus groups was 0.070. These high $F_{st}$ values, coupled with the lower haplogroup diversity and reduced number of haplogroups in these groups, are

 

**Table 3** Y chromosomal haplogroup frequencies and haplogroup diversities in the Caucasus, Iran and Turkey. Haplogroup designations are according to the nomenclature proposed by the Y chromosome consortium (2002)

| | | Haplogroups | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | *N* | **E*** | **C*** | **K*** | **P1** | **P*** | **R1*** | **R1a1*** | **F*** | **G*** | **J2*** | **I*** | Haplogroup Diversity |
| *North Caucasus* | | | | | | | | | | | | | |
| Abazinians | 14 | 0 | 0.07 | 0.14 | 0 | 0 | 0 | 0.14 | 0.29 | 0.29 | 0.07 | 0 | 0.85 |
| Chechenians | 19 | 0 | 0.05 | 0.11 | 0.16 | 0 | 0 | 0.05 | 0.32 | 0.05 | 0.26 | 0 | 0.83 |
| Darginians | 26 | 0.04 | 0 | 0 | 0 | 0 | 0.04 | 0 | 0.27 | 0.04 | 0.04 | 0.58 | 0.61 |
| Ingushians | 22 | 0 | 0 | 0.05 | 0 | 0.05 | 0 | 0 | 0.27 | 0.27 | 0.32 | 0.05 | 0.78 |
| Kabardinians | 59 | 0 | 0 | 0.15 | 0 | 0.07 | 0.02 | 0.02 | 0.24 | 0.29 | 0.12 | 0.10 | 0.82 |
| Lezgi (Dagestan) | 25 | 0 | 0 | 0.28 | 0 | 0 | 0.04 | 0 | 0.32 | 0.36 | 0 | 0 | 0.72 |
| Ossetians (Ardon) | 28 | 0 | 0 | 0.07 | 0 | 0.04 | 0 | 0.04 | 0.04 | 0.21 | 0.29 | 0.32 | 0.79 |
| Oseetians (Digora) | 31 | 0 | 0 | 0 | 0 | 0.06 | 0 | 0 | 0.03 | 0.74 | 0.03 | 0.13 | 0.44 |
| Rutulians | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.58 | 0.38 | 0.04 | 0 | 0.54 |
| *South Caucasus* | | | | | | | | | | | | | |
| Abkhazians | 12 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0.33 | 0 | 0 | 0.25 | 0.333 | 0.77 |
| Armenians | 100 | 0.06 | 0 | 0.07 | 0.02 | 0.02 | 0.19 | 0.06 | 0.18 | 0.11 | 0.24 | 0.05 | 0.86 |
| Azerbaijanians | 72 | 0.06 | 0 | 0.11 | 0.03 | 0 | 0.11 | 0.07 | 0.11 | 0.18 | 0.31 | 0.03 | 0.84 |
| Georgians | 77 | 0.03 | 0 | 0.03 | 0.01 | 0.03 | 0.10 | 0.10 | 0.14 | 0.31 | 0.21 | 0.04 | 0.83 |
| Kazbegi* | 25 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0.04 | 0.12 | – | 0.72 | 0.04 | 0.48 |
| Lezgi* (S. Caucasus) | 12 | 0.17 | 0 | 0 | 0 | 0 | 0.17 | 0.08 | 0.58 | – | 0 | 0 | 0.65 |
| S. Ossetians* | 17 | 0.18 | 0 | 0 | 0 | 0 | 0.12 | 0.06 | 0.41 | – | 0.24 | 0 | 0.77 |
| Svans* | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0.92 | – | 0 | 0 | 0.15 |
| *West Asia* | | | | | | | | | | | | | |
| Turks | 39 | 0.05 | 0.03 | 0.15 | 0 | 0 | 0.31 | 0.13 | 0.05 | 0 | 0.03 | 0.26 | 0.81 |
| Iran (Tehran) | 80 | 0.06 | 0 | 0.1 | 0.01 | 0.04 | 0.08 | 0.2 | 0.03 | 0.05 | 0.1 | 0.34 | 0.82 |
| Iran (Isfahan) | 50 | 0.02 | 0 | 0.14 | 0.02 | 0.06 | 0 | 0.18 | 0.22 | 0.06 | 0.2 | 0.1 | 0.86 |

*Not typed for M201 (haplogroup G*)

most likely the result of genetic drift operating in small, isolated populations. As with the MDS plot based on mtDNA sequences (Fig. 2A), there is no clear clustering of Caucasus populations according to their geographical proximity or linguistic affiliation.

The correlation between the geographic and genetic (pairwise $F_{st}$) distances separating pairs of Caucasus populations was not statistically significant (Mantel test, $Z = -0.113$, $P = 0.697$). Removing the presumed outliers (Darginian, Svan, Abkhazian, Ossetian (Digora), Rutulian and Kazbegi groups) resulted in a correlation that was slightly higher but still non-significant ($Z = -0.088$, $P = 0.625$).

## Comparison of Mitochondrial and Y-Chromosome Data

West Asian populations exhibit the highest level of Y-haplogroup diversity (average value 0.824), followed by the Caucasus and Central Asia (average value 0.790 and 0.769 respectively). The lowest level of Y-haplogroup diversity is found in Europe (average value 0.633). MtDNA haplotype diversity ranged from 0.953 to 0.995 in the Caucasus, within the range observed in Europe and the Near East. A correlation analysis between mtDNA and Y-haplogroup diversities shows an absence of concordant patterns between both genetic components ($r = 0.062$, $P > 0.05$). Similarly, the correlation between genetic ($F_{st}$) distances among pairs of Caucasus populations based on mtDNA and Y-haplogroups was not statistically significant (Mantel test, $Z = 0.058$, $P = 0.345$), suggesting differences in the genetic structure of these groups based on mtDNA and the Y chromosome.

The concordance between the geographic, linguistic and genetic structure of the Caucasus and neighbouring populations was further investigated by the AMOVA procedure. The rationale behind this procedure is that a grouping of populations that accurately reflects their genetic relationships should allocate a higher proportion of

**Table 4** 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 1 | 2.53 | 1.66 | 95.81 | 8.54 | 13.74 | 77.72 |
| Geography 2 | 2.39 | 2.00 | 95.61 | 8.78 | 13.96 | 77.27 |
| Geography 3 | 1.37 | 1.86 | 96.78 | 12.59 | 10.31 | 77.1 |
| Geography 4 | 1.57 | 1.54 | 96.89 | 12.31 | 9.96 | 77.74 |
| Linguistic 1 | 0.87 | 2.33 | 96.81 | 5.51 | 17.41 | 77.07 |
| Linguistic 2 | 0.79 | 2.36 | 96.85 | 4.81 | 17.81 | 77.38 |

The following classification were used to group populations:
Geography 1- North Caucasus, South Caucasus, Europe,West Asia, Central Asia
Geography 2- Caucasus, Europe, West Asia, Central Asia
Geography 3- Caucasus, West Europe, East Europe, West Asia, Central Asia
Geography 4- North Caucasus, South Caucasus, West Europe, East Europe, West Asia, Central Asia
Linguistic 1- Caucasian, Indo-European, Turkic
Linguistic 2- South Caucasian, North Caucasian, Indo-European, Turkic

the genetic variance between groups, and a lower proportion among populations within groups. Using geographic or linguistic criteria, we thus defined several groups of samples and compared the results obtained by AMOVA. The within population component of the genetic variance was about 96% for mtDNA vs. only 77% for the Y-chromosome, confirming that genetic distances between populations were on average much larger for the Y-chromosome than for mtDNA (Table 4). For both mtDNA and the Y-chromosome, a higher between group component of genetic variance is found when groups are defined using a geographic criterion. MtDNA and Y-chromosome both suggest a division between North and South Caucasus groups; but the separation of West and East Europe gave better results for the Y-chromosome than for mtDNA (Table 4). Linguistic groupings consistently resulted in a much higher proportion of the genetic variance among populations within groups than among groups for both mtDNA and the Y-chromosome, indicating a poor fit between linguistic classification and the genetic structure of these groups (Table 4).

We also carried out a spatial analysis of molecular variance (SAMOVA) to look for regions of large genetic changes in Europe and Asia that might define genetic barriers. This analysis follows the same principle as used in the AMOVA analysis and attempts to define the groups of samples that are best supported by the genetic data. Figures 3A and 3B show the location of the populations typed for mtDNA and Y-chromosome polymorphisms and the composition of the groups of samples, as defined by SAMOVA. As observed previously, a stronger genetic structure is observed for Y chromosome data than for mtDNA. But, both genetic structures correspond to the groups depicted in the MDS plots shown in Figure 2. SAMOVA applied to mtDNA data suggests the isolation of populations which are clearly outliers on the MDS plot and which are found in Central Asia, Middle-East and in the Southern Caucasus area. The other populations are grouped into a single cluster which is also easily recognizable in Figure 2A. The Y-chromosome data underline also the genetic peculiarities of samples from the South Caucasus region, as well as the Kirghiz sample in Central Asia. We also observe a strong genetic structure in Europe, with a clear separation between West and East European samples as observed previously.

For the reduced dataset (Table 5), the SAMOVA procedure indicates the isolation of the North Ossetians and Abkhazians for both loci. The Lezginians (from Dagestan) and Ingushians are also distinct based on mtDNA; while Darginians and Rutulians are also distinct based on the Y chromosome. These results are again in good agreement with the MDS plots of Figure 2.

The Y-haplogroups contrast sharply with the mtDNA sequences in terms of the overall relationships of the Caucasus with Europe and West Asia. For the Y-chromosome, the Caucasus groups are more closely

 

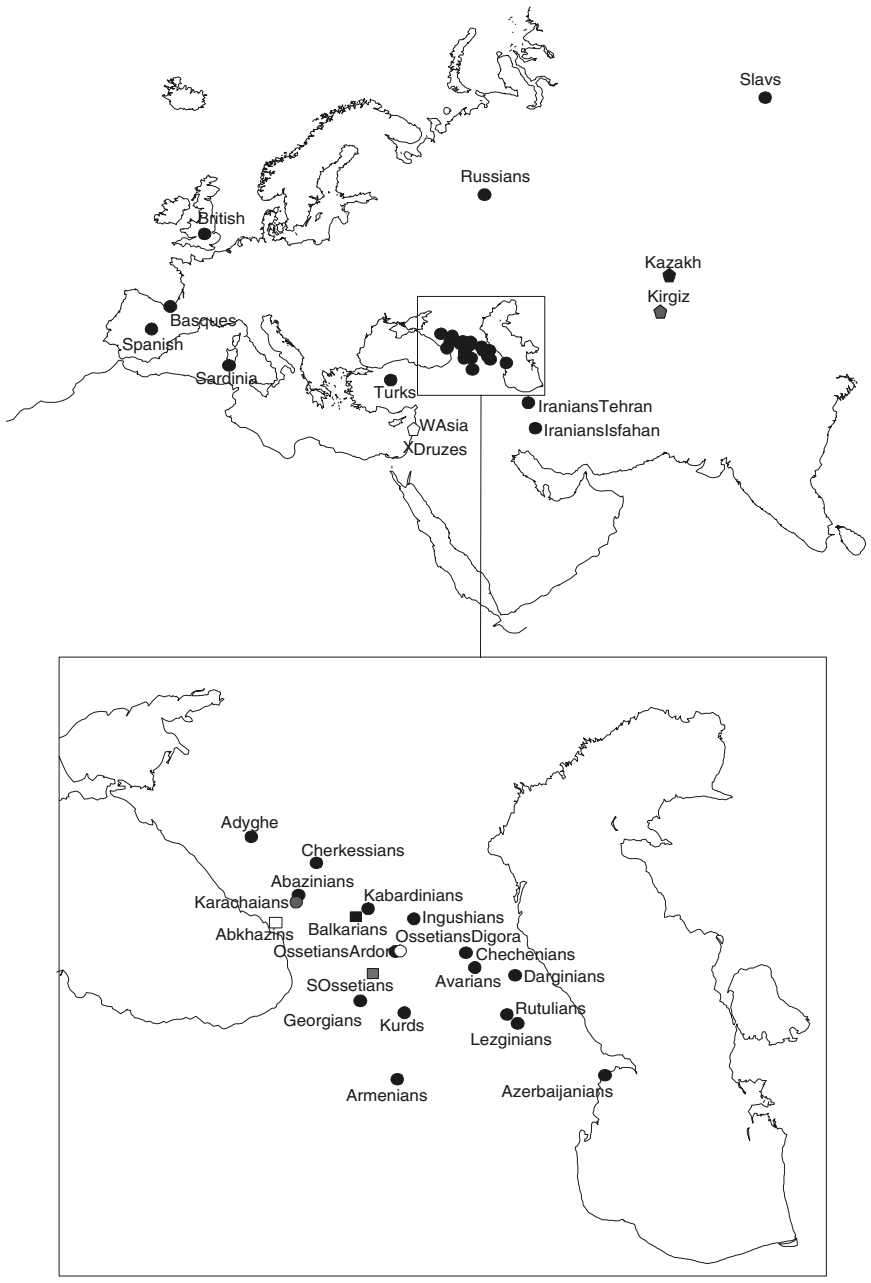


**Figure 3A** A. Distribution of the 34 populations analyzed for mtDNA HV1 sequences and the structure as defined by SAMOVA (samples with the same symbol belong to the same group). Corresponding fixation indices: $F_{SC} = 0.012$ ($p < 0.001$), $F_{ST} = 0.052$ ($p < 0.001$), $F_{CT} = 0.040$ ($p < 0.001$).

related to West Asia (average $F_{st} = 0.110$) than to West or East Europe (average $F_{st} = 0.278$ and 0.259 respectively), whereas for mtDNA, the Caucasus groups are approximately equally closely-related to West or East Europe (average $F_{st} = 0.023$ and 0.026 respectively), and West Asia (average $F_{st} = 0.019$). These relationships are also evident in the MDS plots (Figure 2). The observed trend remained the same when we calculated pairwise $F_{st}$ values using the reduced data set (results not shown). Comparing North and South Caucasus groups separately does not change these conclusions (Table 6).

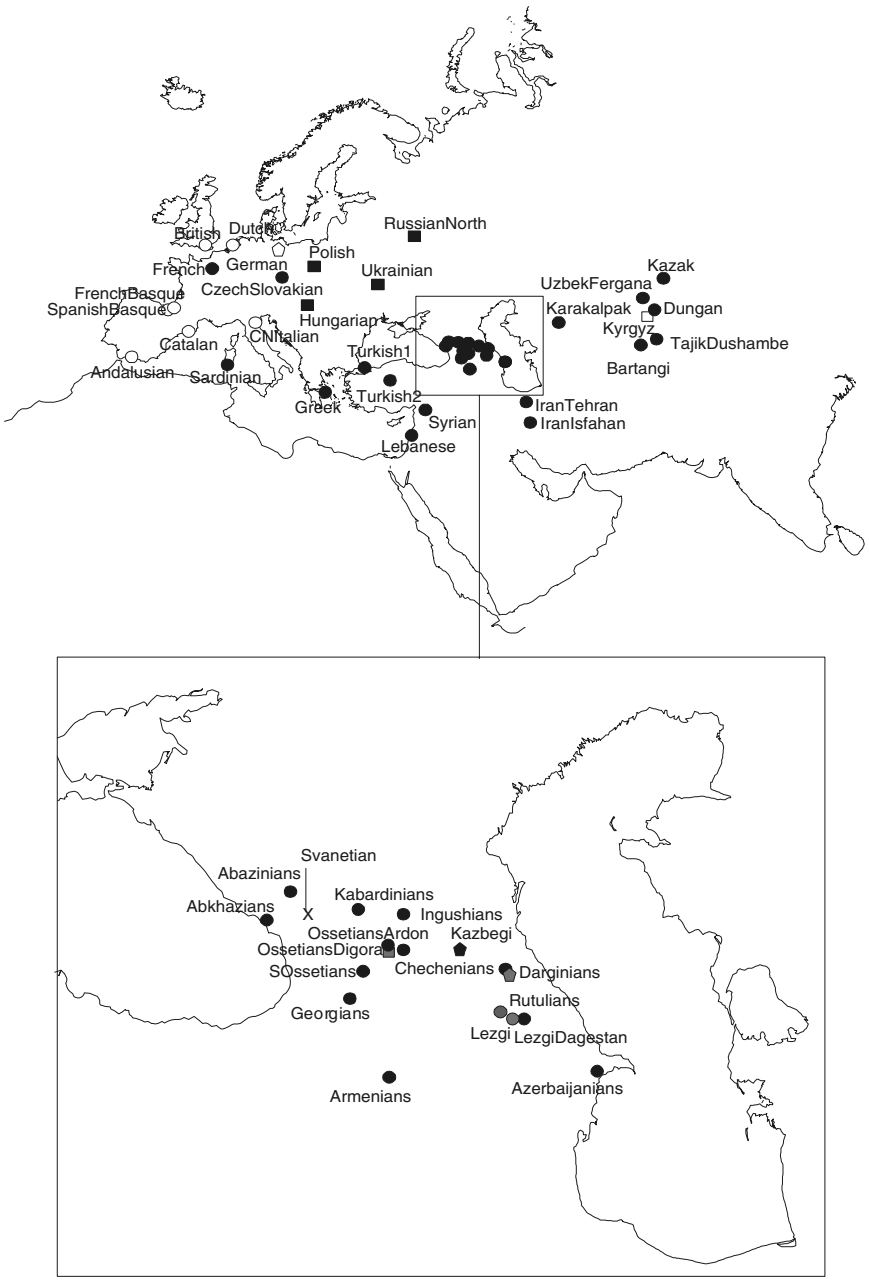


**Figure 3B** Distribution of the 50 populations analyzed for YSNPs haplogroups and 10 groups structure as defined by SAMOVA (samples with the same symbol belong to the same group). Corresponding fixation indices: $F_{SC} = 0.077$ ($p < 0.001$), $F_{ST} = 0.279$ ($p < 0.001$), $F_{CT} = 0.219$ ($p < 0.001$).

## Discussion

This expanded study of mtDNA and Y-chromosome variation in the Caucasus extends previous results (Nasidze *et al.* 2001, 2003). To increase the accuracy of measurements of genetic diversity among populations, it is desirable to increase the number of populations rather than the number of individuals per population (Nei, 1978; Pons & Petit, 1995). Therefore, we added 11 new populations to the 14 previously studied groups from the Caucasus, with overall sample sizes adequate for estimating genetic diversity (Nei, 1978; Pons & Petit, 1995). Overall, the Caucasus groups are genetically heterogeneous, suggesting large genetic differences among

 

**Table 5** Fixation indices corresponding to the groups of populations as inferred by SAMOVA for the reduced dataset (samples from the Caucasus region typed for both mtDNA and Y chromosome polymorphisms)

| Locus | Group composition | $F_{SC}$ | $F_{ST}$ | $F_{CT}$ |
|---|---|---|---|---|
| mtDNA | 1. Abkhazins<br>2. North Ossetians (Digora)<br>3. Ingushians<br>4. Lezginians<br>5. Other samples | 0.008*** | 0.025*** | 0.018** |
| Y chromosome | 1. Abkhazins<br>2. North Ossetians (Digora)<br>3. Darginians<br>4. Rutulians<br>5. Other samples | 0.060*** | 0.174*** | 0.121*** |

***p < 0.001, **P < 0.005

**Table 6** Pairwise Fst values between regions calculated based on reduced data-sets

| Regions | NC | SC | W. Europe | E. Europe | W. Asia | C. Asia |
|---|---|---|---|---|---|---|
| NC | | 0.121 | 0.313 | 0.296 | 0.135 | 0.323 |
| SC | 0.019 | | 0.209 | 0.191 | 0.066 | 0.225 |
| W Europe | 0.022 | 0.025 | | 0.326 | 0.192 | 0.339 |
| E Europe | 0.026 | 0.027 | 0.013 | | 0.122 | 0.158 |
| W Asia | 0.018 | 0.020 | 0.025 | 0.025 | | 0.182 |
| C Asia | 0.023 | 0.021 | 0.092 | 0.076 | 0.053 | |

Above diagonal - pairwise Fst values calculated based on Y-SNP data are given.
Below diagonal - pairwise Fst values calculated based on mtDNA data are shown.
NC - North Caucasus; SC - South Caucasus; W. Europe - West Europe;
E. Europe - East Europe; W. Asia - West Asia; C. Asia - Central Asia.

populations. More than 27% of the Y chromosome genetic variance occurs between populations, and 79.7% of the pairwise $F_{st}$ values between populations are significantly different from zero ($P < 0.05$). By contrast, there is much lower heterogeneity between populations based on mtDNA; less than 5% of the mtDNA genetic variance occurs between populations, and 56.5% of the pairwise $F_{st}$ values are significantly different from zero ($P < 0.05$).

One possible explanation for much lower $F_{st}$ values for mtDNA than for the Y chromosome might be the difference in mutation rates, as the mtDNA HV1 sequences evolve much more rapidly than the Y-SNPs. However, differences in mutation rate will influence total levels of diversity, but not the apportioning of genetic diversity within *vs*. among populations, as measured by $F_{st}$ values. Empirical evidence for this assertion was demonstrated recently by Excoffier & Hamilton (2003), who show that the components of genetic variance for global human populations due to variability between groups, between populations within groups, and between individual populations, did not differ for several genetic systems with widely varying mutation rates. A more likely explanation for the observed pattern of greater differentiation between Caucasus populations for the Y chromosome than for mtDNA, which is often observed in human populations (Seielstad *et al.* 1998), is a higher female migration rate due to patrilocality (Oota *et al.* 2001).

This genetic heterogeneity does not correlate with either linguistic diversity or geographic barriers in the Caucasus. In particular, Indo-European speaking Armenians and Turkic-speaking Azerbaijanians are genetically most closely related (for both mtDNA and the Y-chromosome) to other Caucasus groups and not to other Indo-European or Turkic-speaking groups (Figures 2A, 2B). Moreover, linguistic classifications of the Caucasus groups

correspond poorly with their genetic structure (Table 4).

Although geographic classifications do provide a better fit to the genetic structure (Table 4), correlations between genetic and geographic distances between groups were statistically non-significant. We also did not detect any influence of the Caucasus Mountains as a significant geographic barrier; the average genetic distances between North and South Caucasus groups for mtDNA ($F_{st} = 0.019$) and Y-haplogroups ($F_{st} = 0.190$) were comparable to the average genetic distances within geographic groups (North Caucasus: mtDNA $F_{st} = 0.024$, Y-haplogroup $F_{st} = 0.185$; South Caucasus: mtDNA $F_{st} = 0.026$, Y-haplogroup $F_{st} = 0.195$). Instead, the spatial analysis of molecular variance (Figure 3 and Table 5) identified a few outlier populations for each genetic system that partially overlapped. Almost all of these outliers are small, isolated populations residing in the highland region of the Caucasus and hence are likely to have undergone genetic drift. In sum, there is no evidence that the Caucasus Mountains have served as a barrier to gene flow.

The primary difference between this study and previous studies lies in the genetic relationships of the Caucasus populations with European and West Asian populations. Previously, mtDNA indicated a closer relationship of the Caucasus with Europe (Nasidze *et al.* 2001), while the Y chromosome indicated a closer relationship with West Asia (Nasidze *et al.* 2003). However, the present study finds that the Caucasus is slightly closer genetically to West Asian than to European populations (Table 6) with respect to mtDNA. The reason for this discrepancy appears to be the inclusion of new data from Iranian (this study) and Kurdish (Comas *et al.* 2000) groups, which associate more closely with the Caucasus groups than do the previously studied West Asian groups (Figure 2A). When the Iranians and Kurds were removed from the analysis, Caucasian groups were much closer to European than to West Asian groups with respect to mtDNA (average $F_{st}$ between Caucasus and West Asia is 0.033, *vs.* 0.021 between Caucasus and Europe). When we included these populations, the average $F_{st}$ value between the Caucasus and West Asia dropped to 0.026. These results emphasize the importance of obtaining data from all relevant neighbouring groups when investigating the genetic relationships of Caucasus groups; we are currently expanding our sampling of Iranian and Kurdish groups to further study their relationships with the Caucasus groups.

With respect to the Y-chromosome relationships of the Caucasus with Europe and West Asia, the present study reinforces previous results (Nasidze *et al.* 2003) in that we find that the Caucasus are more closely related to West Asia than to Europe. In particular, the most common Y chromosome haplogroups in the Caucasus ($F^*$, $G^*$ and $J2^*$) are all probably of West Asian origin (Semino *et al.* 2000), suggesting a strong West Asian paternal influence on the Caucasus region. This interpretation is supported by the MDS plot based on Y-haplogroups (Figure 2b); Caucasus and West Asian groups are intermingled, whereas European populations form distinct and separate clusters. Moreover, these conclusions do not change when the South and North Caucasus are compared separately with Europe and West Asia; there are no significant differences between average pairwise $F_{st}$ values for North and South Caucasus comparisons with Europe and West Asia (t-tests, $t = -0.74$, $P = 0.466$, and $t = -2.86$, $P = 0.06$ for comparing North *vs.* South Caucaus to West Asia and to Europe, respectively). We suggest that the Y-chromosome results may reflect recent "invading" migrations from the Near East that probably mostly involved males. Different Near Eastern groups invaded the Caucasus numerous times during the last two millennia, including the occupation of Georgia by Arab calips after 654 AD, the Seljuc Turks invasion of the South Caucasus in the 11$^{th}$ century, and repeated invasions by Turks and Persians (Muskhelishvili, 1977 and references therein).

To explain why mtDNA variation (but not the Y-chromosome) places the Caucasus in an intermediate position between Europe and West Asia, we suggest that this reflects a common ancestry of Caucasus and European populations. This common ancestry could date back to pre-Neolithic times, as suggested by Renfrew (Renfrew, 1992) who considered Caucasian languages to reflect human dispersal over 15,000 years ago. Or, it could reflect a route for Neolithic farmers from the Near East to Europe via the Caucasus. There are several securely-dated Neolithic sites in the Caucasus that

 

are 6,000-7,000 years old (Masson & Merpert, 1982; Muskhelishvili, 1977), which pre-date or coincide with the appearance of agriculture in Europe. Regardless of whether the close relationship between European and Caucasian groups reflects pre-Neolithic or Neolithic events, more recent, primarily male-mediated migrations from West Asia to the Caucasus would have reduced the signature of a common Europe-Caucasus genetic ancestry for the Y-chromosome, but not for mtDNA (Nasidze *et al*. 2003).

In conclusion, the major aspect of the Caucasus population structure seems to be high overall levels of genetic differentiation, much higher for the Y chromosome than for mtDNA. The genetic structure of Caucasian groups is more accurately represented by geographic than linguistic classifications of populations. Overall, it appears that isolation and small population sizes, especially in the highland areas, have led to genetic drift and enhanced genetic differentiation. We also find evidence of different demographic histories for the Y chromosome *vs*. mtDNA, with the Y chromosome indicating a predominantly West Asian influence, whereas mtDNA variation in the Caucasus seems to reflect a more complex interaction of European and West Asian influences.

## Acknowledgements

We thank M. Kayser and B. Pakendorf for useful discussions and suggestions. This research was supported by funding from the Max Planck Society.

## References

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.

Aris-Brosou, S. & Excoffier, L. (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol* **13**, 494–504.

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.

Barbujani, G., Nasidze, I. S. & Whitehead, G. N. (1994a) Genetic diversity in the Caucasus. *Hum Biol* **66**, 639–668.

Barbujani, G., Oden, N. L. & Sokal, R. R. (1989) Detecting regions of abrupt change in maps of biological variables. *Systematic Zoology*, **38**, 376–389.

Barbujani, G. & Sokal, R. R. (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA*. **87**, 1816–1819.

Barbujani, G., Whitehead, G. N., Bertorelle, G. & Nasidze, I. S. (1994b) Testing hypotheses on processes of genetic and linguistic change in the Caucasus. *Hum Biol* **66**, 843–864.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P. & Comas, D. (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* **59**, 63–81.

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., Perez-Lezaun, A., Bosch, E. & Bertranpetit, J. (1997) Mitochondrial DNA variation and the origin of the Europeans. *Hum Genet* **99**, 443–449.

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.

DiRienzo, 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., Schneider, S. & Exoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* **11**, 2571–2581.

Excoffier, L. & Hamilton, G. (2003) Comment on "Genetic structure of human populations". *Science* **300**, 1877b.

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.

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**, 232–49.

Maliarchuk, B. A., Derenko, M. V. & Solovenchuk, L. L. (1995) Types of regulatory regions in mitochondrial DNA in eastern Slavs. *Genetika* **31**, 846–51.

Maniatis, T., Fritsh, E. F. & Sambrook, J. (1982) Molecular clonning. A laboratory manual. New York: Cold Spring Laboratory.

Masson, V. M. & Merpert, N. J. (1982) Archaeology of USSR: Neolith of USSR. Moscow: Nauka Press, pp. 93–164.

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.

Muskhelishvili, D. (1977) The main problems of Georgian historical geography. (Tbilisi: Metsniereba Press).

Nasidze, I., & Stoneking, M. (2001) Mitochondrial DNA variation and language replacements in the Caucasus. *Proc R Soc Lond B*, **268**, 1197–1206.

Nasidze, I., Risch, G. M., Robichaux, M., Sherry, S. T., Batzer, M. A. & Stoneking, M. (2001) Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* **9**, 267–272.

Nasidze, I., Sarkisian, T., Kerimov, A., Stoneking, M. (2003) Testing hypotheses of language replacement in the Caucasus: Evidence from the Y-chromosome. *Hum Genet* **112**, 255–261.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., Stoneking, M. (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat Genet* **29**, 20–21.

Orekhov, V., Poltoraus, A., Zhivotovsky, L. A., Spitsyn, V., Ivanov, P. & Yankovsky, N. (1999) Mitochondrial DNA sequence diversity in Russians. *FEBS letters* **445**, 197–201.

Piercy, R., Sullivan, K. M., Benson, N. & Gill, P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* **106**, 85–90.

Pons, O. & Petit, R. J. (1995) Estimation, variance and optimal sampling of gene diversity. *Theor Appl Genet* **90**, 462–470.

Renfrew, C. (1992) Archaeology, genetics, and linguistic diversity. *Man* **27**, 445–478.

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., Sellitto, D., Cruciani F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A. & Bandelt, H. J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.

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.

Seielstad, M., Minch, E. & Cavalli-Sforza, L. (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* **20**, 278–280.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiæ, M., Mika, A., Mika, B., Primorac, D., A. Santachiara-Benerecetti, S., L. Cavalli-Sforza, L. & Underhill, P. A. (2000) The genetic legasy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*, **290**, 1155–1159.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

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., Shanmugalakshimi, S., Balakrishnan, K., Read, M.,

 

Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A. & Bodmer, W. F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA*, **98**, 10244–10249.

Y chromosome Consortium (2002) A nomenclature system for the tree of Human Y- chromosomal binary haplogroups. *Genome Res* **2**, 339–348.

*Received:* 16 June 2003
*Accepted:* 9 January 2004