doi: 10.1046/j.1529-8817.2004.00131.x

Genetic Evidence Concerning the Origins of South and North Ossetians

I. Nasidze[1*], D. Quinque[1], I. Dupanloup[2], S. Rychkov[3], O. Naumova[3], O. Zhukova[3], and M. Stoneking[1]

[1] *Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103, Leipzig, Germany*
[2] *Center of Integrative Genomics University of Lausanne CH-1015 Lausanne Dorigny, Switzerland*
[3] *Vavilov Institute of General Genetics, Gubkin str.3, 119991, Moscow, Russia*

Summary

Ossetians are a unique group in the Caucasus, in that they are the only ethnic group found on both the north and south slopes of the Caucasus, and moreover they speak an Indo-European language in contrast to their Caucasian-speaking neighbours. We analyzed mtDNA HV1 sequences, Y chromosome binary genetic markers, and Y chromosome short tandem repeat (Y-STR) variability in three North Ossetian groups and compared these data to published data for two additional North Ossetian groups and for South Ossetians. The mtDNA data suggest a common origin for North and South Ossetians, whereas the Y-haplogroup data indicate that North Ossetians are more similar to other North Caucasian groups, and South Ossetians are more similar to other South Caucasian groups, than to each other. Also, with respect to mtDNA, Ossetians are significantly more similar to Iranian groups than to Caucasian groups. We suggest that a common origin of Ossetians from Iran, followed by subsequent male-mediated migrations from their Caucasian neighbours, is the most likely explanation for these results. Thus, genetic studies of such complex and multiple migrations as the Ossetians can provide additional insights into the circumstances surrounding such migrations.

Keywords:Ossetians, Y chromosome, mtDNA

Introduction

Ossetians are unique among the ethnic groups of the Caucasus region in that they are the only group to reside on both the north and south slopes of the central part of the Caucasus Mountains range. They are also of interest in that they speak an Iranian language, belonging to the Indo-European language family, whereas their geographic neighbours all speak Caucasian languages.

There are two main hypotheses concerning the origin of Ossetians. According to the first hypothesis, from the 7th century BC to the 1st century AD Ossetia came under Scythian-Sarmatian influence, which was succeeded by the Alani, an Iranian-speaking warrior Sarmatian tribe. The Alani are then believed to be the direct ancestors of Ossetians (Miller, 1992). The second hypothesis describes the origin of Ossetians as descendants of one of the autochthonous groups from the Caucasus. According to this view Ossetians adopted an Iranian language, most likely from the Alani, in the early middle ages or possibly even earlier. It is believed that before this event Ossetians spoke a Caucasian language (Miller, 1992).

A number of population genetic studies have been carried out on Ossetian groups. Classical genetic markers (blood groups, serum proteins and red cell enzymes) were studied in several groups from South and North Ossetia (Inasaridze *et al*. 1990; Salamatina & Nasidze, 1993), and substantial genetic variability was found in these populations. North Ossetians showed greater similarity with North Caucasian groups than with South

---

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

Caucasian groups. Rychkov *et al.* (1996) also studied classical genetic markers in a number of North Ossetian groups. Genetic differentiation between Ossetian groups was the highest among all comparisons of groups from the Caucasus. Surprisingly, North Ossetians showed greater similarity with Russians than with neighbouring populations from the Caucasus. This observation was interpreted as genetic evidence of an ancient northern Alanian origin for Ossetians.

Eleven polymorphic restriction sites in the first hypervariable segment (HVI) of the mtDNA control region were studied in 40 North Ossetians and in additional groups from Siberia (Rychkov *et al.* 1995). A comparison of these groups with each other and with data from Western and Eastern Europe, as well as Central and Eastern Asia, revealed close similarity of all the Eurasian groups with respect to the frequency of each polymorphic restriction site.

Eleven bi-allelic loci and 9 short tandem repeat (STR) loci on the Y-chromosome, as well as mtDNA HV1 sequence variability, were previously studied in two populations from North Ossetia (Nasidze *et al.* 2004). One South Ossetian population, from Georgia, has also been studied for a number of Y chromosome SNPs (Wells *et al.* 2001) and for mtDNA HV1 sequence variability (Kivisild *et al.* 1999). However, the focus of these studies was on a much larger regional scale and did not address specific questions concerning the origin of Ossetians.

To investigate the origins and relationships of North and South Ossetian groups, here we present data on mtDNA and Y chromosome variation in 70 individuals from three additional North Ossetian groups. These new data were combined with the previously-published data from two other North and one South Ossetian group (Kivisild *et al.* 1999; Wells *et al.* 2001; Nasidze *et al.* 2003; Nasidze *et al.* 2004), to address the following questions: (1) are North Ossetians genetically more closely related to South Ossetians or to other North Caucasus groups; (2) are South Ossetians genetically more closely related to North Ossetians or to other South Caucasus groups; and (3) which of the two prevailing theories concerning Ossetian origins (an Iranian origin, or a Caucasian origin followed by language replacement) receives support from the genetic data.

Materials and Methods

Samples and DNA Extraction

A total of 70 blood samples from unrelated males, representing three North Ossetian populations (Alagir, N = 24; Zamankul, N = 23; and Zil'ga, N = 23), were collected in the Republic of North Ossetia (Alania), Russian Federation (Figure 1). Informed consent and information about birthplace, parents and grandparents were obtained from all donors. Genomic DNA was extracted from the blood samples using a standard phenol-chloroform method (Maniatis et al. 1982).

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 (Cordaux *et al.* 2003) and H16401 (Vigilant *et al.* 1989) were used to determine sequences for both strands of the PCR products with the DNA Sequencing Kit (Perkin-Elmer), following the protocol recommended by the supplier, and an ABI 3700 automated DNA sequencer. Samples with a C at position 16189, which produces a "C-stretch" between positions 16184-16193 that is difficult to sequence through, were sequenced again in each direction so that each base was determined twice.

Published mtDNA HV1 sequences were also used from two North Ossetian populations (26 Ardon and 30 Digora; Nasidze *et al.* 2004), 201 South Ossetians (Kivisild *et al.* 1999), 23 Abazinians, 23 Chechenians, 44 Cherkessians, 37 Darginians, 35 Ingushians, 51 Kabardinians, 42 Armenians, 41 Azerbaijanians, 57 Georgians (Nasidze & Stoneking, 2001), 16 Balkarians, 32 Avarians, 13 Karachaians, 45 Lezginians, 31 Rutulians, 27 Abkhazians, 79 Iranians from Tehran and 46 from Isfahan, 39 Turks (Nasidze *et al.* 2004), 50 Adyghe (Richards *et al.* 1996), 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 (Corte-Real *et al.* 1996; Richards *et al.* 1996), 102 Russians (Orekhov *et al.* 1999), 18 Slavs (Maliarchuk *et al.* 1995), 45 Israeli Drusi (Macaulay *et al.* 1999), 29 Kurds

Figure 1 A map of the geographic location of Ossetian populations used in this study. 1 – Alagir, 2– Digora, 3 – Ardon, 4 – Zamankul, 5 – Zil'ga, 6– South Ossetians.

(Comas *et al.* 2000), 55 Kazakhs and 94 Kyrgiz (Comas *et al.* 1998).

Y Chromosome bi-Allelic Markers

Ten Y chromosomal SNP markers were genotyped: 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. M9 and RPS4Y were typed by PCR-RFLP assays as described elsewhere (Kayser *et al.* 2000); M89 was typed as described in Ke *et al.* (2001); and the remaining Y-SNP markers were typed using PIRA-PCR assays (Yoshimoto *et al.* 1993) as described in Cordaux *et al.* (2004) and Nasidze *et al.* (2004). The YAP *Alu* insertion was typed as described in Hammer & Horai (1995). The samples were genotyped according to the hierarchical order of the markers (Underhill *et al.* 2000). The Y-SNP haplogroup nomenclature used here is according to the recommendations of the Y Chromosome Consortium (2002).

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.

Y Chromosome STR Loci

Samples belonging to the SNP haplogroup G*(M201) 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 quadruplex PCRs, or alternatively in a single nanoplex PCR, and detected on an ABI PRISM 377 DNA sequencer (Applied Biosystems), as described elsewhere (Kayser *et al.* 1997, 2001). In order to distinguish between the DYS385a and DYS385b loci, 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 mtDNA and 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 HVI sequence data was carried out using the software package NETWORK version 3.1 (Bandelt *et al.* 1999).

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). 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 Variability.

A total of 377 bp of the mtDNA HV1 region, comprising nucleotide positions 16024 to 16400 (Anderson *et al.* 1981), were determined for 70 individuals from the North Ossetian groups of Zil'ga, Zamankul and Alagir. For the purposes of comparing the sequences reported here with published data for other Ossetian populations (Kivisild *et al.* 1999; Nasidze *et al.* 2004) and for groups from the Caucasus, Europe, and West and Central Asia (see the list of populations and corresponding references in the Materials and Methods section), further analyses were restricted to 365 bp (nucleotide positions 16024-16388) of HV1. 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 mutations were found in the Ossetian HV1 sequences (analysis not shown). The sequences will be deposited in the HVRbase database (www.HVRbase.de) at the time of publication and are also available upon request from the corresponding author (I.N.).

Parameters summarizing some characteristics of the mtDNA HV1 sequence variability in Ossetian populations are presented in Table 1. The haplotype diversity varied from 0.931 (Alagir) to 0.977 (Digora), which is

Table 1 MtDNA HV1 sequence variability among Ossetian populations

Population	N	no. of haplotypes	% of shared haplotypes	Haplotype diversit and SD	MPD	Tajima's D	Source
Alagir	24	15	37.5	0.931 +/− 0.034	6.70	-1.76^{*}	present sudy
Zamankul	23	18	21.7	0.941 +/− 0.043	6.68	-2.43^{**}	present sudy
Zil'ga	23	18	21.7	0.976 +/− 0.020	5.24	-1.58^{*}	present sudy
Ardon	26	19	26.9	0.948 +/− 0.034	4.65	-1.66^{*}	Nasidze *et al.*, (2004)
Digora	30	21	30.0	0.977 +/− 0.014	5.46	−1.26	Nasidze *et al.*, (2004)
Total North ossetians	126	61	51.6	0.967 +/− 0.008	6.81	-2.52^{**}	present sudy
South Ossetians	201	65	67.7	0.969 +/− 0.005	6.44	−1.40	Kivisild *et al.*, (1999)

$^{*}P < 0.05$, $^{**}P < 0.01$.

Table 2 Mean pairwise F_{st} values between North and South Ossetian groups, and aucausian, Iranian, West and East European, and West and Central Asian groups. Below diagonal - paiwise Fst values based on Y-SNP haplogroups; above diagonal - pairwise Fst values based on mtDNA HVI sequences

	N.Oss.	S.Oss.	N.Cau.	S.Cau.	Iranians	West Eu.	East Eu.	Centr. Asia	West Asia
N. Oss.		0.037	0.020	0.030	0.024	0.043	0.035	0.035	0.029
S. Oss.	0.330		0.037	0.033	0.015	0.036	0.030	0.073	0.018
N. Cau.	0.231	0.101		0.019	0.021	0.022	0.026	0.023	0.018
S. Cau.	0.298	0.126	0.190		0.016	0.025	0.027	0.021	0.020
Iranians	0.199	0.112	0.105	0.124		0.018	0.016	0.055	0.011
West Eu.	0.387	0.338	0.316	0.209	0.264		0.023	0.078	0.026
East Eu.	0.317	0.292	0.296	0.191	0.116	0.275		0.087	0.027
Centr. Asia	0.303	0.166	0.245	0.225	0.106	0.306	0.149		0.056
West Asia	0.211	0.087	0.135	0.253	0.093	0.243	0.183	0.121	

within the range of haplotype diversity values for other Caucasian groups. The mean number of pairwise nucleotide differences (MPD) varied from 4.65 (Ardon) to 6.70 (Alagir); these values are at, or exceed, the upper limit of the range of MPD values for European populations (3.15-5.03; Comas *et al.* 1997) as well as for the rest of the Caucasus (4.40-5.87; Nasidze *et al.* 2004). Tajima's D values were negative in all Ossetian groups (Table 1), significantly so in all groups except the Digora and the South Ossetian group. The negative D values suggest past population expansions in the Ossetian groups.

For overall pairwise comparisons of North and South Ossetians with neighbouring populations from the Caucasus, as well as groups from West and East Europe and West and Central Asia, we calculated average pairwise F_{st} values, pooling the five North Ossetian groups, as F_{st} values for almost all comparisons between North Ossetian groups are not significantly different from zero (not shown). Pairwise F_{st} comparisons (Table 2) indicate that South and North Ossetians are no closer to each other (average $F_{st} = 0.037$) than to North or South Caucasus groups (average F_{st} values range from 0.020-0.037). These average F_{st} values are not significantly different from one another ($t = 0.976$, $p = 0.332$). Indeed, South and North Ossetians are more distant from each other than are other South Caucasus groups from other North Caucasus groups (average $F_{st} = 0.019$), although the difference between these values is also not significant ($t = 0.817$, $p = 0.460$). North Ossetian populations are more similar to their geographic neighbours from the North Caucasus (average $F_{st} = 0.020$) than South Ossetians are to their neighbours from the South Caucasus (average $F_{st} = 0.033$). The average pairwise F_{st} value between North Ossetians and groups from the South Caucasus (excluding South Ossetians) is lower (0.030) than that between South and North Ossetians (0.037), while the average pairwise F_{st} value for South Ossetians and North Caucasus groups (excluding North Ossetians) is the same (0.037) as that between South and North Ossetians, although these estimates are not significantly different from each other ($t = 0.669$, $p = 0.518$).

Overall, Ossetians are more distant from the other Indo-European-speaking populations from the Caucasus (Armenians; average $F_{st} = 0.030$) than from Caucasian-speaking populations (average $F_{st} = 0.026$), although these values are not significantly different ($t = 1.430$, $p = 0.212$). However, Ossetians are significantly closer to Iranian-speaking populations from Isfahan and Tehran (average $F_{st} = 0.019$) than to Caucasian-speaking populations (average $F_{st} = 0.027$; $t = -2.564$, $p = 0.026$). The same trend holds when we compare haplotype sharing between Ossetian and Iranian populations versus Ossetians and their closest geographic neighbors from the Caucasus. South Ossetians share just 4% of their mtDNA sequences with Georgians, whereas they share 12% and 19% of their mtDNA sequences with Iranian-speaking groups from Isfahan and Tehran respectively. The haplotype sharing between North Ossetians and Iranian groups varies from 13% to 31%. With Ingushians, their closest eastern geographic neighbours, North Ossetians share from 22% to 33% of their mtDNA sequences. With Kabardinians, their closest western geographic neighbours, North Ossetians share 26% to 54% of their mtDNA sequences. This relatively high percentage of shared haplotypes between North

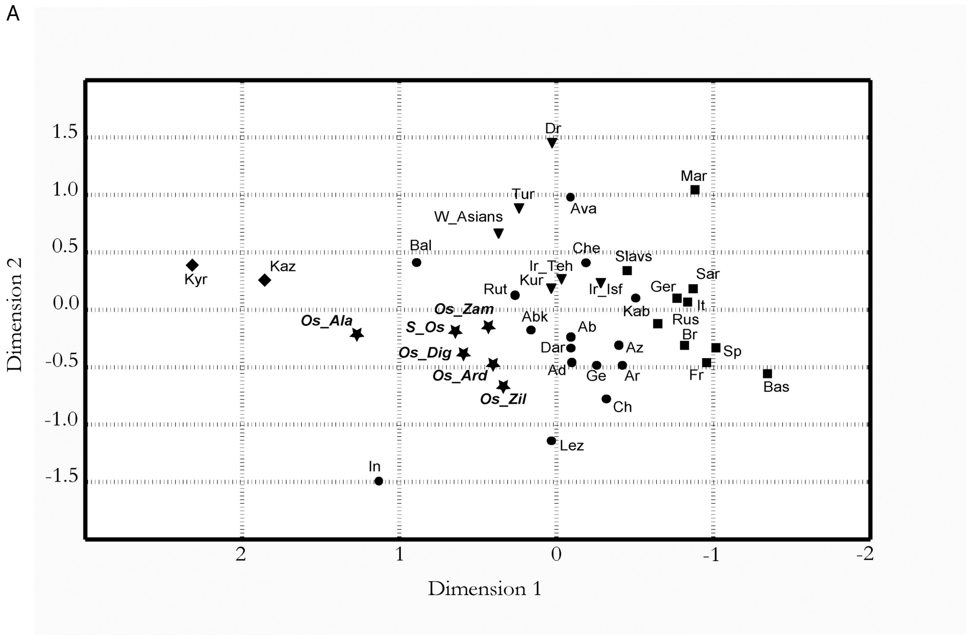

Figure 2 MDS plots based on pairwise F_{st} values, showing relationships among the North and South Ossetians, Caucasian, European, Central and West Asian populations. Ossetians are represented by stars; other Caucasus groups are represented by circles; squares correspond to populations from Europe; Central Asian groups are represented by diamonds; and West Asian populations by triangles. A. Based on mtDNA HVI sequence data. The stress value for the MDS plot is 0.104. B. Based on Y chromosome SNP data. The stress value for the MDS plot is 0.126. The populations are given the following abbreviations: Os_Dig – Ossetians from Digora, Os_Ala – Ossetians from Alagir, Os_Zil – Ossetians from Zil'ga, Os_Zam – Ossetians from Zamankul, Os_Ard – Ossetians from Ardon, S_Os – South Ossetians, Sv – Svans, Rut – Rutulians, Kazb – Kazbegi, Lez_Dag – Lezginians from Dagestan, Lez_Az – Lezginians from Azerbaijan, In – Ingushians, Che – Chechenians, Ch – Cherkessians, Ava – Avarians, Kab – Kabardinians, Ab – Abazinians, Ge – Georgians, Az – Azerbaijanians, Ar – Armenians, Abk – Abkhazians, Bal – Balkarians, Ir_Isf – Persians from Isfahan, Ir_Teh – Persians from Tehran, Leb – Lebanese, Dr – Israeli Drusi, Syr – Syrians, Tur – Turks, Kur – Kurds, Rus - Russians, Ukr – Ukrainians, Mar – Mari, Pol – Polish, Hung – Hungarians, Gr – Greeks, Sar – Sardinians, Cz_Sl – Chez and Slovaks, Sp - Spanish, Fr – French, Ger – Germans, It – Italians, Br – British, And – Andalusians, Dut – Dutch, Cat – Catalans, Bas – Basques, Fr_Bas – French Basques, Sp_Bas – Spanish Basques, Turkm – Turkmans, Kara – Karakalpak, Kyr – Kyrgiz, Kaz – Kazakhs, Ishk – Ishkinasi. The abbreviated names of the populations are also used in the Y-SNP MDS plot.

Ossetians and their closest geographic neighbours can be explained by recent gene exchange among these groups.

An MDS plot (Figure 2A) based on pairwise F_{st} values illustrates the above relationships. In particular, the Ossetian populations cluster together, near groups from the Caucasus and West Asia, and are the closest of any of the Caucasian groups to Central Asia.

Y-SNP Haplogroups

Overall, ten Y-SNP haplogroups were found in Ossetians (Table 3), with the number of haplogroups in each individual Ossetian group ranging from four to seven. Haplogroup E* (YAP) was found only in South Ossetians (Wells *et al.* 2001), and haplogroup C* (RPS4Y) was completely absent. The most frequent haplogroup among North Ossetians was G* (M201). Unfortunately, M201 was not typed in South Ossetians by Wells *et al.* (2001); therefore, it is impossible to distinguish between haplogroups G* (M201) and F* (M89) in this group: the latter is the most frequent haplogroup in South Ossetians. In our analyses these individuals were classified as haplogroup F* (M89), although some unknown proportion could be haplogroup G* (M201). We therefore classified all haplogroup G* (M201) North Ossetian individuals as haplogroup F* (M89) in order to be able to

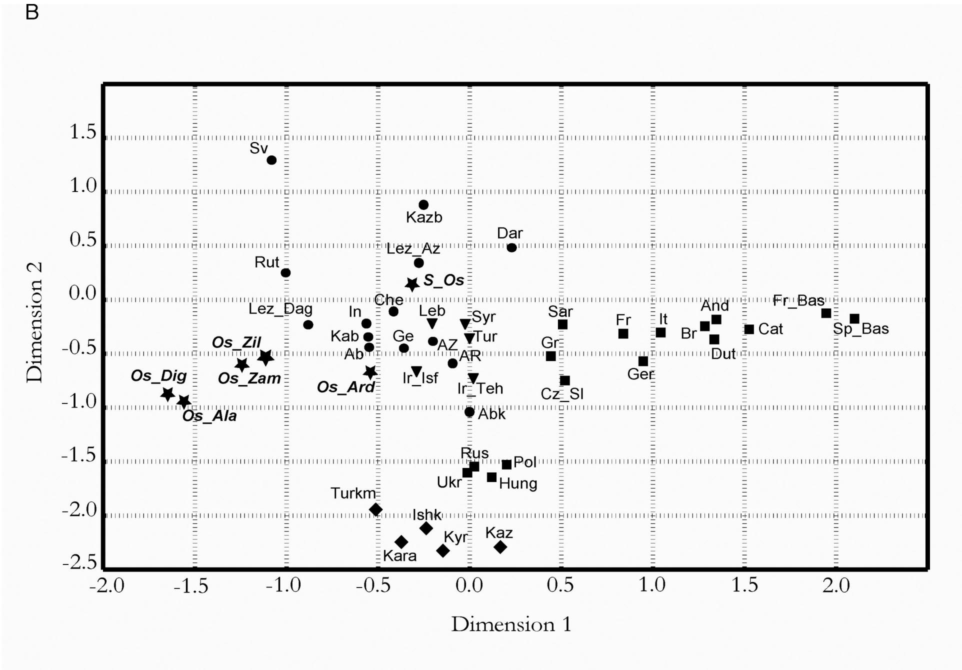

Figure 2 (Continued.)

Table 3 Y chromosome haplogroup frequencies in Ossetian groups

Groups	N	Haplogroups E* YAP	C* RPS4Y	K* M9	P1 M124	P* M45	R1* M173	R1a1* M17	F* M89	G* M201	J2* M172	I* M170	HD	Source
North Ossetia (Russia)														
Digora	**31**	0	0	0	0	0.06	0	0	0.03	0.74	0.03	0.13	**0.440**	Nasidze *et al.*, (2004)
Ardon	**28**	0	0	0.07	0	0.04	0	0.04	0.04	0.21	0.29	0.32	**0.788**	Nasidze *et al.*, (2004)
Zil'ga	**23**	0	0	0.130	0	0	0	0	0.043	0.565	0.261	0	**0.620**	Present study
Zamankul	**23**	0	0	0.217	0	0	0	0	0.087	0.609	0.087	0	**0.590**	Present study
Alagir	**24**	0	0	0.083	0.083	0	0.042	0	0	0.750	0.042	0	**0.440**	Present study
South Ossetia (Georgia)														
S. Ossetians	**17**	0.18	0	0	0	0	0.12	0.06	0.41	-	0.24	0	**0.772**	Wells *et al.*, (2001)

use Y-SNP data from South Ossetians for the MDS and F_{st} analyses. Haplogroup I* (M170) was found in substantial frequencies in groups from Digora and Ardon, whereas this haplogroup is absent from the rest of the groups. Haplogroup J2* (M172) was found in all groups, with frequencies ranging from 0.03 in the Digora group to 0.29 in the Ardon group. Haplogroup K* (M9) was found in all groups except for South Ossetians and the Digora group. The remaining haplogroups were found at low frequencies in one or two groups.

Haplogroup diversity varied substantially, from 0.44 in the populations from Digora and Alagir to 0.79 in Ardon. These values are lower than the diversity values for most other Caucasus groups (Nasidze *et al.* 2004) typed for these markers, except for Svans from Georgia (Wells *et al.* 2001).

Pairwise F_{st} values indicate substantial differences between South and North Ossetians (Table 2). The average F_{st} value between these groups was 0.330, significantly higher than between the other North and South Caucasus groups (average F_{st}= 0.190; t = 4.687, p = 0.009); by contrast, the average F_{st} value between South Ossetians and other North Caucasian populations (excluding North Ossetians) was only 0.101, and the average F_{st} between South Ossetians and the other South Caucasus groups was 0.126. North Ossetian populations were almost as distant from South Caucasian groups as from South Ossetians (average F_{st} value, excluding South

Ossetians, was equal to 0.298), while the average F_{st} value between North Ossetians and other North Caucasian populations was significantly lower (average $F_{st} = 0.231$; $t = 3.48$, $p = 0.001$). Despite the fact that North Ossetians showed greater similarity with other North Caucasus populations compared with South Ossetians, the difference in F_{st} values is not significant ($t = -0.734$, $p = 0.503$).

Pairwise F_{st} comparisons between Ossetians and Armenians (the other Indo-European speaking group from the Caucasus) showed that Ossetians are closer to this group (average $F_{st} = 0.144$) than to the Caucasian-speaking groups (average $F_{st} = 0.213$). However, the difference in average F_{st} values is not statistically-significant ($t = 0.559$, $p = 0.583$). Also, Ossetians are about as distant from their direct linguistic relatives, Iranian-speaking groups from Isfahan and Tehran (average $F_{st} = 0.204$), as from non-Indo-European speaking West Asian groups (average $F_{st} = 0.215$; $t = 0.103$, $p = 0.920$).

The MDS analysis (Figure 2B) further illustrates the patterns observed in the pairwise F_{st} comparisons. South Ossetians are distinct from a loose cluster of the North Ossetian groups. The South Ossetians cluster with South and North Caucasian groups and with West Asian groups. Ossetians from Ardon are somewhat separated from the other North Ossetian groups, being closer to the cluster of Caucasian and West Asian groups. The overall high level of Y-chromosome diversity exhibited by Ossetian populations is reflected in the large area they occupy on the MDS plot. We also repeated the MDS analysis without South Ossetians, in order to use information for the haplogroup G* (M201). The result of this analysis (not shown) did not reveal any differences from the analysis without haplogroup G* (M201).

Y Chromosome STRs

Since haplogroup G* (M201) has an unusually high frequency in North Ossetians (average frequency = 0.57) compared with other groups from the Caucasus (average frequency = 0.21), we typed 9 Y-STR loci in individuals with this Y-SNP haplogroup to determine if this elevated frequency indicates a bottleneck effect. We compared the results with the same set of loci on the same Y-SNP background typed in other groups from the Caucasus (Nasidze *et al.* 2003). Haplotype diversity on the background of haplogroup G* (M201) is significantly reduced in Ossetians (0.722 ± 0.071) compared with the rest of the Caucasus (0.992 ± 0.005). A median network of Y-STR haplotypes on the background of haplogroup G* (M201) revealed two clearly separated clusters (Figure 3). One of them almost exclusively contains haplotypes found in the Digora group. The second cluster contains the remaining North Ossetian groups, suggesting either different sources of introduction of haplogroup G* (M201) or isolation and genetic drift in the Digora group.

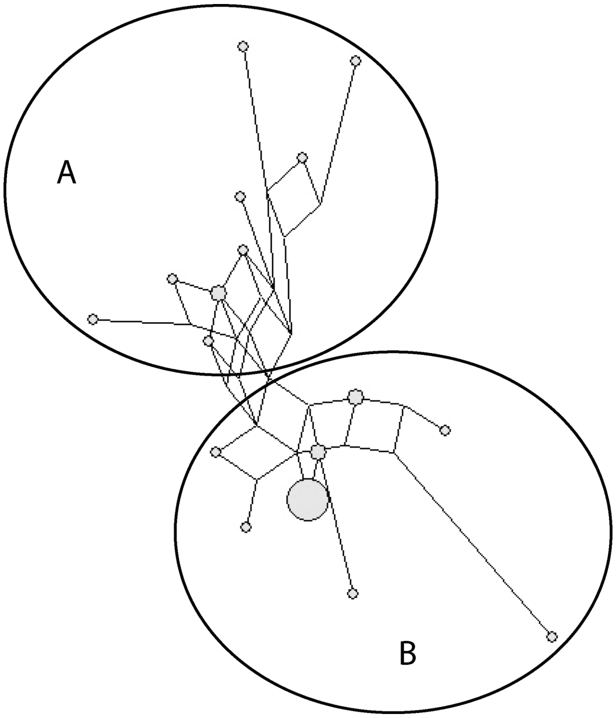

Figure 3 Median network constructed based on Y chromosome STRs on the background of Y-SNP haplogroup *G* (M201).* Cluster A corresponds to Y-STR haplotypes found in the Digora population, while cluster B corresponds to the Y-STR haplotypes found in the rest of Ossetian groups.

Comparison of mtDNA and Y-Chromosome Data

The geographic and linguistic structure of Ossetians, other Caucasus groups, and European, West and Central Asian groups, as assessed by mtDNA and Y chromosome variation, was investigated by the AMOVA procedure (Table 4). As is typically seen in human

Table 4 AMOVA results according to different classifications

	mtDNA					Y-SNP
Classifications	Among groups	Among populations within groups	Within populations	Among groups	Among populations within groups	Within populations
Geography 1	1.98	1.50	96.52	9.52	13.73	76.75
Geography 2	1.83	1.81	96.36	9.45	14.32	76.23
Linguistic 1	1.28	2.33	96.39	7.21	16.36	76.43
Linguistic 2	1.19	2.37	96.44	6.63	16.69	76.68

Geography 1 - South and North Caucasus, Europe, West Asia, Central Asia.
Geography 2 - Caucasus, Europe, West Asia, Central Asia.
Linguistic 1 - Caucasian, Indo-European and Turkic.
Linguistic 2 - South Caucasian, North Caucasuan, Indo-European and Turkic.

populations, the within-populations proportion of the variance was much higher for mtDNA (about 96%) than for the Y chromosome (about 76-77%). For both the mtDNA and the Y-SNP data, the geographic classification of populations gave a slightly better fit to the genetic data (in terms of higher among-group variance and lower among-populations-within-groups variance) than did linguistic classifications (Table 4). Further classifying the Caucasian groups into South and North groups did not significantly improve the fit of either classification to the data (Table 4).

The correlation between pairwise F_{st} values based on mtDNA and Y-SNP data in Ossetians was high but not significant (Mantel test, $Z = 0.531$, $p = 0.084$). The correlation between pairwise F_{st} values based on mtDNA and geographic distances between Ossetians and their geographic neighbours was significant (number of groups = 22; $Z = 0.651$, $p = 0.013$), whereas the correlation between pairwise F_{st} values based on Y-SNP haplogroups and the geographic distances between Ossetians and their geographic neighbours was not significant (number of groups = 20; $Z = 0.659$, $p = 0.149$).

The correlation between the geographic distances and the mtDNA F_{st} values for all Caucasus groups was not statistically significant ($Z = -0.216$, $p = 0.751$). A similar result was obtained for the correlation between geographic distances and the Y-SNP haplogroup F_{st} values (Mantel test, $Z = -0.107$, $p = 0.718$). Moreover, the correlation between the F_{st} distances among pairs of Caucasus groups based on mtDNA and Y-haplogroups was not statistically significant ($Z = 0.063$, $p = 0.297$), suggesting differences in the genetic structure of these groups based on mtDNA and the Y chromosome.

We also carried out a spatial analysis of molecular variance (SAMOVA) to look for regions of large genetic change between Ossetian groups, as well as among Ossetians and other neighbouring populations from the Caucasus, West Asia and East Europe that might define genetic barriers. This analysis follows the same principle as used in the AMOVA analysis to define groups of samples that are best supported by genetic data. As observed previously, a stronger genetic structure is observed for Y chromosome data than for mtDNA (Table 5). As expected, both genetic structures are in strong agreement with the MDS plots shown in Figure 2. SAMOVA applied to mtDNA data for Ossetians alone suggests the genetic isolation of North Ossetian groups from Alagir and Zamankul. The Y-chromosome data underline the genetic isolation of the North Ossetian samples from Ardon and Zamankul.

Discussion

North and South Ossetians are the only ethnic group found on both slopes of the Caucasus Mountains. They speak a language which belongs to the Iranian branch of the Indo-European language family; hence, Ossetians are a linguistic isolate, surrounded by Caucasian-speaking populations. By surveying mtDNA and Y-chromosome variation in Ossetians, we sought answers to several questions concerning the origins and genetic relationships of Ossetians. First, are North and South Ossetians more genetically similar to each other, or to their geographic neighbours (i.e., Caucasian-speaking

Table 5 Fixation indices corresponding to the groups of populations as inferred by SAMOVA
A. Only Ossetian groups were considered in SAMOVA analysis. B. Along with Ossetian groups, other populations from the Caucasus and East Europe were included.

Locus	Group composition	F_{SC}	F_{ST}	F_{CT}
mtDNA	*A. Between Ossetian groups*	-0.011	0.035^{***}	0.045
	1. North Ossetians (Alagir)			
	2. North Ossetians (Zamankul)			
	3. Other samples			
	B. Between Ossetians and other neighboring groups			
	1. Karachaians	0.019^{***}	0.061^{***}	0.043^{***}
	2. North Ossetians (Alagir)			
	3. North Ossetians (Zamankul)			
	4. Other samples			
Y chromosome	*A. Between Ossetian groups*			
	1. South Ossetians	-0.011	0.035^{***}	0.045
	2. North Ossetians (Ardon)			
	3. North Ossetians (Zamankul)			
	4. Other samples			
	B. Between Ossetians and other neighboring groups			
	1. Russians	0.100^{***}	0.270^{***}	0.189^{**}
	2. Georgians (Kazbegi)			
	3. Other samples			

$^{***}P < 0.001$, $^{**}P < 0.005$.

populations in the North and South Caucasus, respectively)? The results are somewhat different for mtDNA vs. the Y-chromosome. North and South Ossetians do cluster somewhat in the MDS plot based on mtDNA (Fig. 2A), which may indicate a common origin. However, for the Y-chromosome, North Ossetians are more similar to other North Caucasian populations, and South Ossetians to other South Caucasian populations, than to each other. The SAMOVA analysis also identifies a boundary between South Ossetians and other groups for the Y chromosome, but not for mtDNA. Thus, there is no indication in the Y-chromosome of a particularly close genetic relationship between N. Ossetians and S. Ossetians. If they did have a common origin in the past, it has apparently become obscured by subsequent gene flow with their geographic neighbours on the same sides of the Caucasus Mountains.

Putting together the archaeological and genetic data, and assuming a common origin of South and North Ossetians (which is supported by the mtDNA data), a plausible scenario is that "alteration" of the initial Ossetian Y-chromosome gene pool took place in North Ossetians via other North Caucasus groups. This assumption is enforced by the fact that the genetic distances between North Ossetians and South Caucasus groups are similar to those between North Ossetians and South Ossetians, but the genetic distances between North Ossetians and other North Caucasus groups are much smaller. Moreover, there are differences in genetic structures based on Y chromosome and mtDNA, as the correlation between F_{st} distances among pairs of Caucasus groups based on mtDNA and Y-haplogroups was not statistically significant. The different patterns observed between South and North Ossetians for the Y chromosome may also have been reinforced by the traditional patrilocal social structure of this population, leading to a higher degree of differentiation for the Y chromosome than for mtDNA.

The Ossetians speak an Iranian language; is this because they are directly descended from the Alani (an Iranian-speaking group), or is it rather that genetically the Ossetians resemble their geographic neighbours in the Caucasus, and hence replaced their ancestral Caucasian language with an Iranian language, after contact with the Alani (or another group)? Average pairwise F_{st} values are smaller between Ossetians and Iranians than between Ossetians and Caucasians for both mtDNA and the Y chromosome, significantly so for mtDNA, which suggests an Iranian origin of Ossetians. Subsequent and largely male-mediated migrations between Ossetians and neighbouring groups in the North and South Caucasus, respectively, would explain the greater similarity between Ossetians and Caucasians for the Y-chromosome, as discussed previously.

In conclusion, the genetic results are supported by the archaeological record, in that they reflect a common Iranian origin of South and North Ossetians, as well as a genetic footprint of ancient migrations in the North Caucasus that mostly involved male individuals. Thus, genetic studies of such complex and multiple migrations as the Ossetians can provide additional insights into the circumstances surrounding such migrations.

Acknowledgments

We are grateful to the original donors of samples. This research was supported by funding from the Max Planck Society, Germany.

References

Anderson, S., Bankier, A. T., Barrell, B. G., 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.

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.

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

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 CentralAsian 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 Anthrop* **112**, 5–16.

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.

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.

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**, 331–350.

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.

Hammer M. & Horai S. (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet*. **56**, 951–962.

Inasaridze, Z., Nasidze, I., Shengelii, L. & Shneider, Iu.V (1990) Genetics of the native inhabitants of the Caucasus: distribution of various immunological and biochemical markers in North Osetiia and Checheno-Ingushety. *Genetika* **9**, 1648–1659.

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., 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., 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., Well,s, 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**, 304–314.

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 mt DNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet*. **64**, 232–249.

Maliarchuk, B. A., Derenko, M. V. & Solovenchuk, L. L. (1995): Types of regulatory regions in mitochondrial DNA in eastern Slavs. *Genetika* **31**, 846–851.

Miller, V. F. (1992) Ossetian essays. Vladikavkaz, SOIGI.

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

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

Nasidze, I., Schadlich, H. & Stoneking, M. (2003) Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Sci Int* **137**, 85–93.

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*. (in press)

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.

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.

Rychkov, S. Yu., Naumova, O. Iu., Falunin, D. A., Zhukova, O. V., Rychkov, Iu.G. (1995) Gene pool of residents of northeastern Eurasia in light of data on polymorphism of mitochondrial DNA. I. New data on polymorphism of restriction sites of the D-loop of mtDNA in aboriginal populations of the Caucasus and Siberia. *Genetika* **1**, 118–127.

Rychkov, Yu.G., Balanovskaya, E. V., Zhukova, O. V., Nurbaev, S. D. & Shneider, Yu.V. (1996) Gene pool, gene geography and sickness rate in a population (the case of North Ossetians). *Advances in modern genetics* **20**, 133–199.

Salamatina, N. & Nasidze, I. (1993) Genetic polymorphisms in a rural Osetian community. *Gene Geography* **3**, 251–255.

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.

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., Shanmugalakshimi, S., Balakrishnan, K. & *et al.* (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 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: 14 April 2004
Accepted: 28 June 2004