

Ivan Nasidze · Tamara Sarkisian · Azer Kerimov
Mark Stoneking

Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome

Received: 9 August 2002 / Accepted: 23 October 2002 / Published online: 14 December 2002

© Springer-Verlag 2002

Abstract A previous analysis of mtDNA variation in the Caucasus found that Indo-European-speaking Armenians and Turkic-speaking Azerbaijanians were more closely related genetically to other Caucasus populations (who speak Caucasian languages) than to other Indo-European or Turkic groups, respectively. Armenian and Azerbaijanian therefore represent language replacements, possibly via elite dominance involving primarily male migrants, in which case genetic relationships of Armenians and Azerbaijanians based on the Y-chromosome should more closely reflect their linguistic relationships. We therefore analyzed 11 bi-allelic Y-chromosome markers in 389 males from eight populations, representing all major linguistic groups in the Caucasus. As with the mtDNA study, based on the Y-chromosome Armenians and Azerbaijanians are more closely-related genetically to their geographic neighbors in the Caucasus than to their linguistic neighbors elsewhere. However, whereas the mtDNA results show that Caucasian groups are more closely related genetically to European than to Near Eastern groups, by contrast the Y-chromosome shows a closer genetic relationship with the Near East than with Europe.

Introduction

The Caucasus, the region between the Caspian and Black Seas, is characterized by extremely high linguistic diver-

sity with four major families (South Caucasian, North Caucasian, Indo-European, and Altaic) spoken by more than 50 ethnic groups. In addition, a major geographic barrier, the Caucasus Mountain range, divides the region into the North and South Caucasus sub-regions. The Caucasus therefore offers the opportunity for studying the impact of linguistic diversity and geographic barriers on the genetic structure of human populations. In particular, the presence of populations who are linguistically-related to populations outside the Caucasus, such as the Armenians (who speak an Indo-European language) and the Azerbaijanians (who speak a Turkic language) raises the question as to whether language affiliation or geographic proximity best explains the genetic relationships of these Caucasian populations.

We have previously addressed this question by studying sequence variation in the first hypervariable segment (HV1) of the mtDNA control region and Alu insertion polymorphisms in Caucasian populations (Nasidze and Stoneking 2001; Nasidze et al. 2001). Both studies have shown that Armenians and Azerbaijanians are more closely related genetically to their geographic neighbors in the Caucasus and not to their linguistic neighbors elsewhere. In addition, both studies have demonstrated that Caucasian populations are genetically intermediate between European and Near Eastern populations, but that they are more closely related to European than to Near Eastern populations.

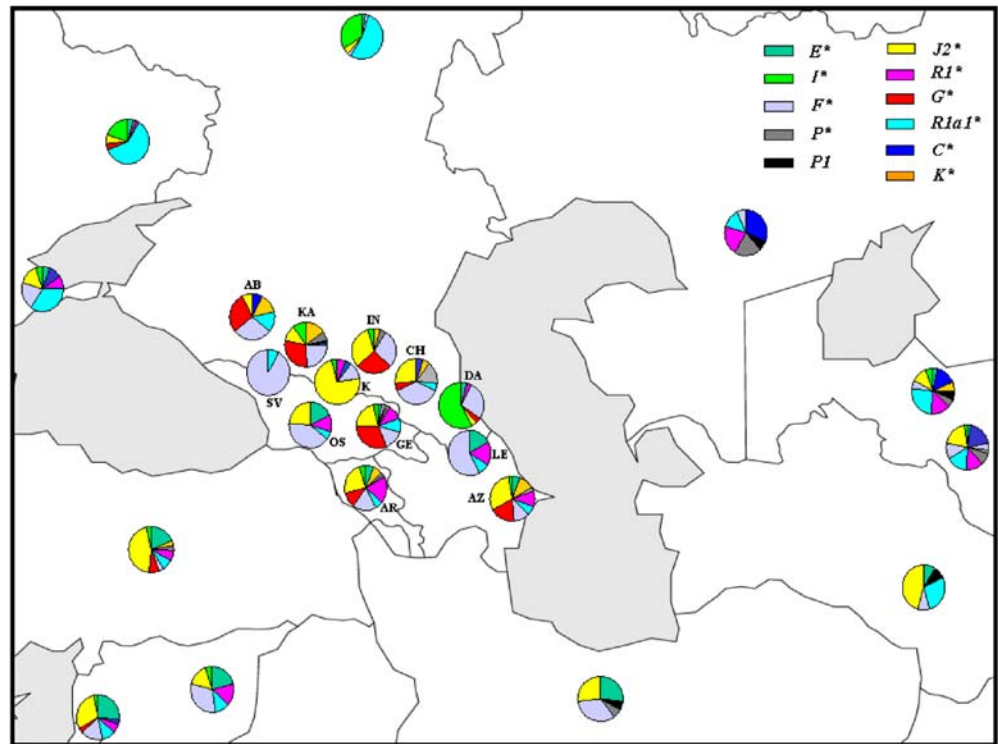
The genetic results thus suggest that both the Armenian and Azerbaijanian languages represent language replacements in the Caucasus. The origins of the Armenian language are obscure, but the Azerbaijanian language was probably introduced in the 11th century AD by central Asian nomads (Johanson 1998). A common mechanism of language replacement is elite dominance (Renfrew 1991), whereby the language of a small invading group is adopted by the larger resident population, either because it is imposed by force or because it is considered socially desirable to speak the language of the invaders. If the invading group is primarily male, then one might expect patterns of Y-chromosome variation to retain some trace of the invaders. To test this hypothesis of language re-

I. Nasidze (✉) · M. Stoneking
Max Planck Institute for Evolutionary Anthropology,
Inselstrasse 22, 04103 Leipzig, Germany
Tel.: +49-341-9952505, Fax: +49-341-9952555,
e-mail: nasidze@eva.mpg.de

T. Sarkisian
Center of Medical Genetics,
National Academy of Sciences of Republic of Armenia,
5/1 Zakyan Str., 375010 Yerevan, Armenia

A. Kerimov
Scientific-research Institute of Haematology and Transfusiology,
Azerbaijan Republic Ministry of Health,
Gashgay Street 87, Baku, Azerbaijan

Fig. 1 Map of the Caucasus indicating the Y-SNP haplogroup frequencies in European, Near Eastern, Caucasian, and Central Asian populations (*AB* Abazinians, *KA* Kabardinians, *IN* Ingushians, *CH* Chechenians, *DA* Darginians, *SV* Svans, *OS* South Ossetians, *K* Georgians from Kazbegi, *GE* Georgians, *AR* Armenians, *AZ* Azerbaijanians, *LE* Lezginians). *SV*, *K*, and *LE* are from Wells et al. (2001)



placement via elite dominance, we report here an analysis of Y-chromosomal bi-allelic markers in the same set of Caucasian populations previously analyzed for mtDNA HV1 sequence variation (Nasidze and Stoneking 2001).

Materials and methods

Subjects

A total of 389 samples (289 whole blood and 100 cheek cell swabs) from unrelated male individuals were collected in the following eight autochthonous populations (Fig. 1): Georgians (South Caucasian speakers), Armenians (Indo-European speakers), Azerbaijanians (Turkic speakers), and Abazinians, Kabardinians, Ingushians, Chechenians, and Darginians (North Caucasian speakers). DNA from the blood samples had been used previously for analyses of mtDNA HV1 sequence variation and Alu insertion polymorphisms (Nasidze and Stoneking 2001; Nasidze et al. 2001); cheek cell swabs from additional males from Armenia and Azerbaijan were collected in order to increase the sample size for Y-chromosome studies. Informed consent and information about the birthplace of the donor and the donor's parents and grandparents were obtained from all donors. Genomic DNA from blood samples was extracted by using an IsoQuick DNA extraction kit (Orca Research, Bothell, Wash., USA) or a conventional phenol-chloroform method (Maniatis et al. 1982). DNA from cheek cell swabs was extracted by using a conventional salting-out procedure (Miller et al. 1988). Published Y-chromosome single nucleotide polymorphism (Y-SNP) data from the South Caucasus were used from 25 Svans (Georgia), 25 Georgians from Kazbegi (Georgia), 12 Lezgi (Azerbaijan), and 17 South Ossetians (Wells et al. 2001). Additional published data on 21 Azerbaijanians and 47 Armenians (Wells et al. 2001), and 63 Georgians (Semino et al. 2000; Wells et al. 2001), did not differ in any respect from our samples (data not shown) and were not included in the analysis to avoid weighting the results too heavily on these populations. Published Y-SNP data (Semino et al. 2000; Wells et al. 2001) for European, Near Eastern, and Central Asian populations were also included.

Molecular analysis

Ten SNP markers previously reported to be polymorphic in Europe and the Near East (Semino et al. 2000) were typed in all samples: 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 and Horai 1995) was also typed. For all SNP markers except M130 (RPS4Y), Taqman (Applied Biosystems) assays were designed. Primers and dye-labeled probes were designed by using Primer Express (Version 1 for Macintosh Power PC; PE Biosystems). Primer and probe sequences are given in Table 1. Reaction mixes were prepared as described previously (Morin et al. 1999). M130 was typed by using the polymerase chain reaction/restriction fragment length polymorphism procedure described elsewhere (Kayser et al. 2000), whereas the YAP Alu insertion was typed as described previously (Hammer and Horai 1995). All samples were typed for all markers; no discrepancy was found between the hierarchical order of markers described by Underhill et al. (2000) and our results. The Y-SNP haplogroup nomenclature used here is according to the recent recommendations of the Y Chromosome Consortium (2002). The phylogenetic relationship of the Y-chromosome haplotypes, based on the 11 bi-allelic markers analyzed here, is shown in Fig. 2.

Statistical analysis

Haplogroup diversity and F_{st} values were calculated with Arlequin 2.000 (Schneider et al. 2000), which was also used to carry out Mantel tests for correlations between matrices. Multidimensional scaling (MDS) analysis (Kruskal 1964) of the F_{st} values was carried out with STATISTICA (StatSoft). Programs in PHYLIP3.5c (Felsenstein 1993) were used to construct a neighbor-joining tree based on pairwise F_{st} values.

Table 1 Primer and probe sequences for TaqMan assays with corresponding annealing temperatures (*italic* nucleotide in polymorphic site)

SNP marker	Forward primer	Reverse primer	Probes ^a	Annealing temperature
M9	CCTGAAATACAGAACTGCAAAAGAA	GACATTGAACGTTTGAACATGTCT	CCTAAGATGGTTGAAATGCTCTTTATTTTCTTTAAATTT CCTAAGATGGTTGAAATGCTCTTTATTTTCTTTAAATTT	62°C
M124	AATGCAAAATTCCTGGGCAA	TTCTATAAAGCAAAAGTTGAGGTTG	AACAGGGAAGTGGTTTAAATAATACTGAGTTTGT AACAGGGAAGTGGTTTAAATAATACTGAGTTTGTG	62°C
M45	AACAGTAACTTAGGAGAGGATATCAA	CTATCTCTGGCCTGGACCT	TGAAAAATTATAGATAGGCAAAAAGCTCCITCTG CAGTGAATAATATAGATAAGCAAAAAGCTCCITCT	62°C
M173	TACTGTAACCTCCTAGAAAAITGGAAATAA	CAGTTTCCCAGATCCTGAAAA	CAAGGGCAITTTAGAACAACCTTTGTCAATCTGTTA AGGGCAITTTAGAACCCTTTGTCAATCTGTT	62°C
M17	AAATCAGATTCTGTCTACTCACCAGAG	TCACAAAAAATAGTTTGGCCACTT	TTGCTGGTTGTTACGGGGTTTTTTTTAA TTGCTGGTTGTTACGGGG-TTTTTTAAAGT	60°C
M89	CCTGGATTTCAGCTCTCTTCCT	CCAGCAAAAGGTAGCTGCAAC	TTATGTACAAAAATCTCATCTCTCACITTTGCCCTGA TTATGTACAAAAATCTATCTCTCACITTTGCCCTGA	58°C
M201	GCAATAGTTACTACTTGAGTTACTATATTAGTGCAA	CATCCTATCAGCTTCATCCAAACA	TCCAGTATCAACTGAGGGTTTTCGTAATAGGTACTTA ATCCAGTATCAACTGAGGGTTTTCGTAATAGGTACTTA	62°C
M172	GGATTTTTCATTTTATCCCCC	CCAGGTACAGAGAAAGTTTGGACT	AACCCATTTTGTGCTTAACTTAAAAGGCTTTC ACCCATTTTGTGCTTAACTTAAAAGGCTTTC	60°C
M170	TTTCATATTCTGTGCAATTATACAAATTAATT	CCACACAAAAACAGGTCCTCA	AAAAATCAATTTGCTCAATTTTTCAGTGTGGT AAAAATCAATTTGCTCTTTTTCAGTGTGGG	58°C

^aAncestral probe is shown on the first line and the mutant probe on the second

Results and discussion

Y-SNP haplogroups in the Caucasus

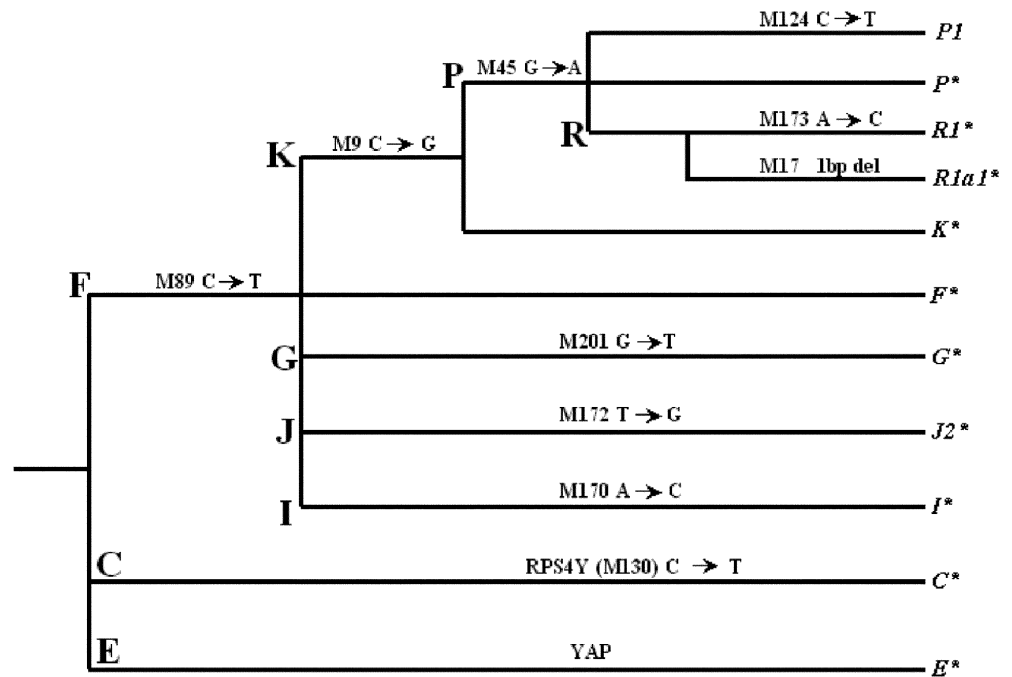
Eleven Y-SNP haplogroups (Fig. 1) were found in the Caucasus (Table 2, Fig. 1). The most frequent haplogroups were F*, G* and J2*; together the frequency of these three haplogroups was 0.53–0.86 in all groups except for the Darginians. The Darginians had a high frequency of haplogroup I* (0.58), which elsewhere was found at a frequency of 0.1 or less. Svans were previously reported (Wells et al. 2001) to have a high frequency (0.92) of haplogroup F*, and Kazbegi to have a high frequency of haplogroup J2* (0.72). No other groups had any single haplogroup at a frequency greater than 0.5, with the exception of the small sample of 12 Lezgi, for which the frequency of haplogroup F* was 0.58. The Darginians, Lezgi, Svans, and Kazbegi correspondingly had the lowest haplogroup diversities (0.153–0.652), whereas for the other groups, the haplogroup diversity was 0.779–0.855.

Two of the three common Caucasus haplogroups (F* and J2*) are also common in Near Eastern populations Lebanese, Turks, Syrians (Semino et al. 2000), and Iranians (Wells et al. 2001), with average frequencies of 0.165 and 0.28, respectively, but present in lower frequencies in Europe (average frequencies 0.021 and 0.074, respectively). The third common Caucasus haplogroup, G*, is rare in Europe (frequency=0.061) and, in the Near East, has been reported only in the Turkish and Lebanese groups (Semino et al. 2000). Haplogroup R1*, which is common in Western and Central Europe, is observed in the South Caucasus at frequencies higher than 0.1, whereas in the North Caucasus, it is absent or nearly so. Haplogroup R1a1*, which is observed in high frequencies in the south-west Mediterranean region, Eastern Europe, and Central Asia, is present at a low frequency in the Caucasus, similar to the Near East. The other Caucasus Y-haplogroups occur at low frequency.

The Darginian, Svan, and Kazbegi groups appear to be outliers compared with the other Caucasus groups. The F_{st} value was highest between Svans and other Caucasus groups (average F_{st} = 0.332), followed by the Kazbegi (average F_{st} = 0.286) and Darginians (average F_{st} = 0.25), whereas the average pairwise F_{st} value among the remaining Caucasus groups was only 0.047. These high F_{st} values, coupled with the lower haplogroup diversity and reduced number of haplogroups (Table 2) in the Darginians, Svans, and Kazbegi, are most likely the result of genetic drift operating in small isolated populations.

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 = 0.671). Removal of the outliers (Svans, Kazbegi, and Darginians) resulted in a correlation that was higher but still non-significant (Z = 0.301, P = 0.134). To determine whether the Caucasus Mountains have an influence on the genetic structure of Caucasian populations, we examined the correlation between geographic and ge-

Fig. 2 Phylogenetic relationship of Y-chromosome haplogroups, based on 10 Y-SNPs and the YAP marker



netic distance for the South and North Caucasus separately, but neither was statistically significant (North Caucasus: $Z=0.331$, $P=0.133$; South Caucasus: $Z=-0.346$, $P=0.65$). Moreover, the average F_{st} value between South and North Caucasus populations (excluding outliers) was 0.075, similar to that among North Caucasus populations (0.096) and that among South Caucasus populations (0.040). Therefore, the Caucasus Mountains appear not to have had a detectable influence on the genetic structure of Caucasus populations; instead, genetic drift operating in small isolated populations seems to have dominated the genetic structure of Caucasus populations.

Comparison of Caucasian, European, and Near Eastern Y-haplogroups

The haplogroup diversity in the Caucasus (average value: 0.797) is almost as high as that in Central Asia (average value: 0.824) and the Near East (average value: 0.769) and is significantly higher (t -test, $P=0.024$) than the haplogroup diversity in Europe (average value: 0.633). An MDS plot and neighbor-joining tree based on F_{st} values (Fig. 3A, B) split European populations into Western and Eastern groups, as has been observed previously (Semino et al. 2000), with Central Asian populations falling in between the Western and Eastern European groups. The Caucasus populations are intermingled with Near Eastern populations.

These patterns have further been confirmed by the pairwise F_{st} comparisons; the mean pairwise F_{st} value for the Caucasus vs Europe is 0.254, whereas the mean F_{st} value for the Caucasus vs the Near East is 0.079, which is significantly lower (t -test based on average F_{st} values jackknifed over populations, $P < 0.001$). On average, South

Caucasian populations are more similar to both Near Eastern and European populations (average $F_{st}=0.038$ and 0.222, respectively) than are North Caucasian populations (average $F_{st}=0.097$ and 0.303); however, both the South Caucasus and the North Caucasus are more similar to the Near East than to Europe with respect to Y-SNP haplogroups.

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* (Fig. 2), was not analyzed (Table 2). In the above analyses, these individuals were classified as haplogroup F*, although some unknown proportion could be haplogroup G*. To determine whether this inability to distinguish between haplogroups F* and G* for some groups influenced the results of the MDS and F_{st} analyses, we classified all haplogroup G* individuals 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.

In order to exclude any possible bias in our conclusions caused by the small size of the samples from some of the groups, we repeated all analyses after excluding all groups with sample size less than 25 (see Table 2). All conclusions remained the same (data not shown).

Genetic relationship between Basques and Caucasians

A common origin of Caucasian and Basque speakers has been proposed, based on (controversial) assertions of the similarity in their languages (Gamkrelidze and Ivanov 1995; Ruhlen 1991). A genetic relationship between Basques and Caucasians is not supported by the Y-SNP data. The

Table 2 Y chromosomal haplogroup frequencies and haplogroup diversities in the Caucasus. Haplogroup designations are according to the nomenclature proposed by the Y Chromosome consortium (2002)

Population	Number	Haplogroups										Haplo-group diversity	Source	
		E*	C*	K*	P1	P*	R1*	R1a1*	F*	G*	J2*			I*
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	Present study
Azerbaijanians	72	0.06	0	0.11	0.03	0	0.11	0.07	0.11	0.18	0.31	0.03	0.84	Present study
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	Present study
Kabardinians	59	0	0	0.15	0	0.07	0.02	0.02	0.24	0.29	0.12	0.10	0.82	Present study
Ingushians	22	0	0	0.05	0	0.05	0	0	0.27	0.27	0.32	0.05	0.78	Present study
Chechenians	19	0	0.05	0.1	0.158	0	0	0.05	0.37	0.05	0.26	0	0.80	Present study
Darginians	26	0.04	0	0	0	0	0.038	0	0.27	0.04	0.04	0.58	0.61	Present study
Abazianians	14	0	0.07	0.14	0	0	0	0.14	0.29	0.29	0.07	0	0.85	Present study
Georgians	63	0	0	0.05	0.06	0	0.14	0.08	0.03	0.30	0.33	0	0.81	Semino et al. 2000
Svans ^a	25	0	0	0	0	0	0	0.08	0.92	-	0	0	0.15	Wells et al. 2001
Kazbegi ^a	25	0	0	0	0	0	0.08	0.04	0.12	-	0.72	0.04	0.48	Wells et al. 2001
Ossetians ^a	17	0.18	0	0	0	0	0.12	0.06	0.41	-	0.24	0	0.77	Wells et al. 2001
Lezgi ^a	12	0.17	0	0	0	0	0.17	0.08	0.58	-	0	0	0.65	Wells et al. 2001

^aNot typed for M201 (haplogroup G*)

average pairwise F_{st} value between Basques and Caucasian groups is much higher ($F_{st} = 0.563$) than that between Basques and Indo-European groups ($F_{st} = 0.311$). The Caucasus groups and Basques are not clustered together in either the neighbor-joining tree or the MDS plot (Fig. 3A, B). These results are in agreement with previous studies based on “classical” markers and mtDNA HV1 sequence data (Bertorelle et al. 1995; Nasidze and Stoneking 2001).

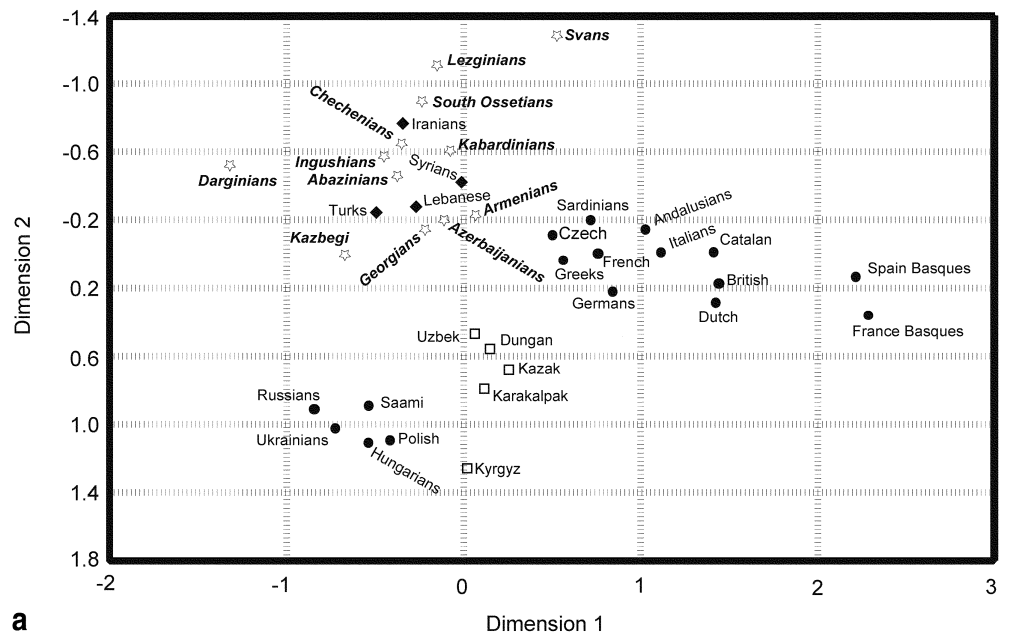
Language replacements and genetic relationships in the Caucasus

The presence, in the Caucasus, of groups whose geographic neighbors are not their linguistic neighbors allows us to address the question as to which better explains the genetic relationships of these groups: geography or language? In particular, Azerbaijanians speak a Turkic language but are surrounded by non-Turkic speakers, and Armenians speak an Indo-European language but are surrounded by non-Indo-European speakers. Previous mtDNA analyses have shown that both Azerbaijanians and Armenians are more closely related genetically to other Caucasus groups than to their respective linguistic neighbors (Nasidze and Stoneking 2001), which indicates that the Azerbaijanian and Armenian languages were introduced via language replacements. We have hypothesized that these language replacements have occurred via elite dominance (Renfrew 1991) in which a resident population adopts the language of a small group of migrants either by force or because speaking the language confers social prestige. If such were the case, and the migrants were primarily male, then Y-DNA analyses may more closely reflect the linguistic relationships of Azerbaijanians and Armenians than have mtDNA analyses.

Indeed, the Y-SNP haplogroups give similar results. The Turkic-speaking Azerbaijanians are genetically closer to their geographic neighbors in the Caucasus (average $F_{st}=0.047$) than to other Turkic-speaking groups (average $F_{st}=0.105$; t -test, $P < 0.001$). Similarly, Indo-European-speaking Armenians are genetically closer to other Caucasus groups (average $F_{st} = 0.053$) than to other Indo-European-speaking groups (average $F_{st} = 0.146$; t -test, $P < 0.001$). This conclusion is also supported by the MDS plot and neighbor-joining tree (Fig. 3A, B), in which Armenians and Azerbaijanians both cluster with other Caucasian populations.

Thus, in agreement with previous studies of mtDNA diversity and Alu insertion polymorphisms (Nasidze and Stoneking 2001; Nasidze et al. 2001), Y-chromosome haplogroups indicate that Indo-European-speaking Armenians and Turkic-speaking Azerbaijanians are genetically more closely related to their geographic neighbors in the Caucasus than to their linguistic neighbors elsewhere. A similar observation has been reported recently in a study of Y-chromosomal markers in Armenian populations (Weale et al. 2001). Thus, all of the genetic evidence agrees that the Armenian and Azerbaijanian languages reflect language replacements, which occurred without any detectable genetic contribution of the original Indo-European and

Fig. 3A, B The phylogenetic relationship of Y-chromosome haplogroups, based on 10 Y-SNPs and the YAP marker. **A** MDS plot based on pairwise F_{st} values, showing relationships among the Caucasus, European, Near Eastern, and Central Asian populations (*open stars* with boldface italic population names Caucasus groups, *closed circles* European populations, *open squares* Central Asian groups, *closed diamonds* Near Eastern populations). The stress value for the MDS plot is 0.134. **B** Neighbor-joining tree based on pairwise F_{st} values for the same populations (*boldface* Caucasus groups)



a



b

Turkic groups, respectively. This may still reflect an “elite dominance” scenario, as presumably the original Indo-European/Turkic migrant groups were very small and/or did not mix extensively with the resident groups. In any event, the migrant groups had a negligible genetic impact

on the resident groups. Currently, there are approximately 8 million Azerbaijanians and 3.5 million Armenians, attesting to the remarkable rapidity of these language shifts; such large-scale shifts would have to be accomplished by cultural rather than merely biological means.

However, the Y-haplogroups contrast sharply with the previous mtDNA and Alu insertion polymorphism studies (Nasidze and Stoneking 2001; Nasidze et al. 2001) in terms of the overall relationships of Caucasus populations. For the Y-chromosome, the Caucasus populations are more closely related to Near Eastern populations than to European populations. Evidence for this is provided by the higher diversity values and by the pairwise F_{st} values and MDS plot. In particular, the Caucasus appears to be a “break zone” in the Y-chromosome genetic landscape of Eurasia; haplogroups such as R1a1* are at a high frequency in East Europe but at low frequency in the Caucasus and the Near East, whereas haplogroups such as F* and J2* are common in the Caucasus and the Near East, but rare in Europe (Table 2, Fig. 1). By contrast, both mtDNA and Alu insertion polymorphisms group Caucasian populations with Europe rather than the Near East.

How can these contrasting views on the relationship of Caucasus populations be reconciled? A possible explanation is that the Y-chromosome results reflect repeated “invading” migrations from the Near East, migrations that probably involved mostly males. Various Near Eastern groups invaded the Caucasus numerous times during the last two millennia including the occupation of Georgia by Arab caliphs after 654 AD (Muskhelishvili 1977), the Seljuc Turks invasion of the South Caucasus in the 11th century (Muskhelishvili 1977), and repeated invasions by Turks and Persians (Muskhelishvili 1977 and references therein). Indeed, strong Y-chromosomal gene flow from the Near East to the Caucasus has been clearly demonstrated for the interpolated maps of admixture proportions in a recent study of Central Asian populations (Zerjal et al. 2002). Although these male-mediated invasions from the Near East might possibly explain why the Caucasian Y-haplogroups are of Near Eastern origin, they do not explain why mtDNA and Alu insertion polymorphisms group the Caucasus with Europe. A possible explanation for this latter result would be a common ancestry of Caucasian and European populations. This hypothesized common ancestry could date back to pre-Neolithic times, as suggested by Renfrew (1992) who considers Caucasian languages to reflect human dispersal before 15,000 years ago. On the other hand, 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 and Merpert 1982; Muskhelishvili 1977). Analyses of additional autosomal markers are needed to verify this hypothesized common ancestry of Caucasian and European populations.

Acknowledgements We thank Manfred Kayser for useful discussion, Peter Underhill for control DNA samples, and Hiltrud Schaedlich for technical assistance. This research was supported by funding from the Max Planck Society.

References

Bertorelle G, Bertranpetit J, Calafell F, Nasidze I, Barbujani G (1995) Do Basque-speaking and Caucasian-speaking populations share non-Indo-European ancestors. *Eur J Hum Genet* 3:256–263

Felsenstein J (1993) PHYLIP (phylogeny inference package) v. 3.5c. Department of Genetics, University of Washington, Seattle

Gamkrelidze T, Ivanov V (1995) Indo-European and the Indo-Europeans. De Gruyter, Berlin

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

Johanson L (1998) The history of Turkic. In: Johanson L, Csato E (eds) *The Turkic languages*. Routledge, London, pp 81–83

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhover W, Stoneking M (2000) Melanesian origin of Polynesian Y chromosome. *Curr Biol* 10:1237–1246

Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27

Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Masson VM, Merpert NJ (1982) Archaeology of USSR: Neolith of USSR. Nauka, Moscow, pp 93–164

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215

Morin PA, Saiz R, Monjavez A (1999) High-throughput single nucleotide polymorphism genotyping by fluorescent 5' exonuclease assay. *BioTechniques* 3:538–552

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

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

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M (2001) Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* 9:267–272

Renfrew C (1991) Before Babel: speculations on the origins of linguistic diversity. *Cambridge Archaeol J* 1:13–23

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

Ruhlen M (1991) *A guide to the world's languages*. Stanford University Press, Stanford

Schneider S, Roessli D, Excoffier L (2000) Arlequin v. 2000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, et al (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, et al (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361

Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyann N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum Genet* 109:659–674

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, 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 Res* 2:339–348

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am J Hum Genet* 71:466–482