Hum Genet (2003) 112 : 255–261
DOI 10.1007/s00439-002-0874-4

ORIGINAL INVESTIGATION

Ivan Nasidze · Tamara Sarkisian · Azer Kerimov
Mark Stoneking

Contrasting patterns of mtDNA and Y-chromosome variation in Caucasus populations: a predominantly northern origin for Y-chromosome lineages

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

Abstract A survey of mtDNA variation in the Caucasus has found that Indo-European-speaking Armenians and Turkic-speaking Azerbaijanians have similar mtDNA frequencies (with Caucasian-specific clades) that are different from those of Caucasian-speaking populations, in accordance with linguistic affiliations. However, Armenians and Azerbaijanians differ significantly in their Y-chromosome frequencies, with Armenians showing greater affinity to Caucasian-speaking populations. We typed 11 bi-allelic Y-chromosome markers in 389 males from 8 populations in the Caucasus. As with the mtDNA study, based on Y-chromosome frequencies Armenians and Azerbaijanians are more similar to each other than to Caucasian-speaking populations. However, whereas the mtDNA results suggest that Caucasian populations are more closely related to Near Eastern populations, the Y-chromosome results suggest a closer relationship between Europeans and Caucasians.

Introduction

The Caucasus, the region between the Caspian and Black Seas, is characterized by tremendous linguistic diversity (Nichols 1997). Three major language families are represented in the region: Indo-European, Caucasian and Turkic. Armenians speak an Indo-European language, Azerbaijanians speak a Turkic language, and Georgians speak a Caucasian language. Although the Caucasus is a relatively small region, it has been the site of numerous migrations and invasions, and has been inhabited by many different peoples (Jones 1997). Thus, the Caucasus is an ideal region for studying the genetic consequences of linguistic and cultural diversity.

Previous studies of mtDNA variation in the Caucasus have found that Armenians and Azerbaijanians have similar mtDNA frequencies, with Caucasian-specific clades, that are different from those of Caucasian-speaking populations, in accordance with linguistic affiliations (Comas et al. 1998; Nasidze et al. 2001). However, Armenians and Azerbaijanians differ significantly in their Y-chromosome frequencies, with Armenians showing greater affinity to Caucasian-speaking populations (Nasidze et al. 2001). Here we present a more detailed study of Y-chromosome variation in the Caucasus, with the goal of understanding the origins of the Armenian and Azerbaijanian gene pools and the relationship between linguistic and genetic diversity in the region.

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 Hematology and Transfusiology,
Azerbaijan Republic Ministry of Health,
Gashgay Str. 87, Baku, Azerbaijan

laV m nt y(a l(t m(nanV, w r rt((r an anay s(s f Y-Vr m s ma bl-al l(V mar rs (n t((sam s t f Cu Vas(an u lat(ns r y(u s(y anay z f r mtDNA HV1 s q n V yar(at(n (Nas(z an St n (n9 2001).

Ma a

Sb Vts

A t tal f 389 sam l s (289 w(Z l bl an 100 Z(V((swabs) fr mu nr lat mal (n y(u als w r V((M (n t((f l-l w(n9 b(t u t V(t((nu s u lat(ns (F g. 1): G r ans (Su t((Cu Vas(an s a rs), Arm n ans (In -E r an s a rs), Az r-bal an ans (T r (V s a rs), an Abaz(n ans, Kabar(n ans, In-9 s(ans, C(Z(n ans, an Dar9(n ans (N rt((Cu Vas(an s a-rs). DNA fr m t((bl sam l s (a b nu s r y(u s(y f r anay s s f mtDNA HV1 s q n V yar(at(n an Al ns rt(n V m r (ms (Nas(z an St n (n9 2001; Nas(z t al. 2001); M(V((swabs fr m a t((nal mal s fr m Arm n a an Az rbal anv r V((M (n r r t (n V as t((sam l s z f r Y-Vr m s m su l s. Inf rm V ns nt an (nf rmat(n abu t t((bl rt((l aV f t((n r an t((n r's ar nts an 9ran ar-nts w r btal n fr m all n rs. G n m(V DNA fr m bl sam l s was tra M bj u s(n9 an Is Q (V DNA tra M(n kt (Or a R s arV(, B t((ll, Wash., USA) r a V ny nt(nal Z(n l-V((r f rm m t(((Man anl(s t al. 1982). DNA fr m V(V((swabs was tra M bj u s(n9 a V ny nt(nal sal((u t r V u r (M ll r t al. 1988). B bl(s(Y-Vr m s m s(n9 n V t(V m r (ms (Y-SNP) ata fr m t((Su t((Cu Vas s w r u s fr m 25 S vans (G r(a), 25 G r ans fr m Kazb 9 (G r(a), 12 L z 9 (Az rbal an), an 17 Su t((Oss t ans (W lls t al. 2001). A t((nal u bl(s(ata n 21 Az rbal an ans an 47 Arm n ans (W lls t al. 2001), an 63 G r ans (S m(n t al. 2000; W lls t al. 2001), (n t (ff r(n an y r s, M fr m u r sam l s (ata n t s(wn) an w r n t(n9 (n t((anay s(s t av l w (9t-(n9 t((r s lts t ((ay n t((s u lat(ns. B bl(s(Y-SNP ata (S m(n t al. 2000; W lls t al. 2001) f r E r an, N ar East rn, an C ntral Asian u lat(ns w r al s (n9 .

M l V lar anay s(s

T n SNP mar rs r y(u s(y r rt t b y m r ((V(n E r an t((N ar East (S m(n t al. 2000) w r y (n al sam l s: RPS4Y (M130), M9, M89, M124, M45, M173, M17, M201, M170, an M172 (Un r(V(t al. 2000 an r f r n V s t((r (n); t((YAP Al ns rt(n y m r (ms (Hamm r an H ral 1995) was als y . F r all SNP mar rs V t M130 (RPS4Y), Taqman (A-l(Bl s(st ms) assa s w r s(9n . Prim rs an y lab l r b s w r s(9n bu s(n9 Prim r E r ss (V rs(n 1 f r Ma V(nt sh P w r PC; PE Bl s(st ms). Prim r an r b s q n V s ar 9iv n (n Tab. 1. R al V(n m(s w r r ar as s V(b r-y(u s(y (M (n t al. 1999). M130 was y bu s(n9 t((y-m ras Vma(n r a V(n/r str V(n fra9m nt l n t((y m r (ms r V u r s V(b l s w((r (Ka s r t al. 2000), w((r as t((YAP Al ns rt(n was y as s V(b r y(u s(y (Hamm r an H ral 1995). All sam l s w r y f r all mar rs; n s-V anV was fu n b tw n t((((rar V(al r r f mar rs-s V(b bj Un r(V(t al. (2000) an u r r s lts. T((Y-SNP ha-l 9ru n m n Vat r u s ((r s a V r (n9t t((r V nt r V m-m n at(ns f t((Y C(r m s m C ns rt(m (2002). T((y l-9 n t(V r lat(ns(f t((Y-Vr m s m ha l y s, bs-

Statist(Val anay s(s

Ha l 9ru y rst(y an F_{st} val s w r Val(V lat w t((An q (n 2.000 (S m(n r t al. 2000), w((((was als u s t Varl u t Mant l t sts f r V rr lat(ns b tw n matr(V s. M l t((m ns(nal s al(ns (MDS) anay s(s (Kr s al 1964) f t((F_{st} val s was Var-r(u t w t((STATISTICA (StatS ft). Pr 9rams (n PHYLIP3.5V (F ls nst (n 1993) w r u s t V nstr M an (Z b r-(n(n9 tr .0256Tw (A [$_{st}$ yal s.

Results

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 their frequency in the haplogroups was 0.53, 0.86 in all groups except for the Dargins. The Dargins had a different frequency of haplogroup I* (0.58), which was not as frequent as in the other groups (0.1 or less). Svans were characterized by a different frequency (0.92) of haplogroup F*, and Kazbegis by a different frequency of haplogroup J2* (0.72). None of the groups had any significant haplogroup at a frequency greater than 0.5, with the exception of the small sample of 12 Lezgis, for which the frequency of haplogroup F* was 0.58. The Dargins, Lezgis, Svans, and Kazbegis were characterized by haplogroup diversities (0.153–0.652), while 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 like Lebanese, Turks, Syrians (Semino et al. 2000), and Iranians (Wells et al. 2001), with frequencies of 0.165 and 0.28, respectively, but not in Europe, but not in Europe (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 Southern Caucasus at frequencies similar to the 0.1, while in the Northern Caucasus, it is absent or rare and only Haplogroup R1a1*, which is observed in high frequencies in the southwestern Middle East, and Central Asia, is present at a low frequency in the Caucasus, similar to the Near East. The other Caucasus Y-haplogroups are very rare or low frequency.

The Dargin, Svan, and Kazbegi groups are quite different from the other Caucasus groups. The F_{st} value was highest between Svans and the other Caucasus groups (average F_{st}= 0.332), followed by Kazbegis (average F_{st}= 0.286) and Dargins (average F_{st}= 0.25), while the average pairwise F_{st} value among the remaining Caucasus groups was only 0.047. This high F_{st} value, due to the low frequency and low number of haplogroups (Table 2) in the Dargins, Svans, and Kazbegis, are mostly due to the presence of different haplogroups in small or isolated populations.

The correlation between the geographic and genetic (pairwise F_{st}) distances among the Caucasus populations was not statistically significant (Mantel test: Z = 0.113, P =0.671). Removal of the populations (Svans, Kazbegis, and Dargins) results in a correlation that was higher but still non-significant (Z =0.301, P =0.134). The terminal within the Caucasus Mountains and an

n t V stanV f r tri Su tri an N rtri Ca Vas s a-
rat Vj , b t n tri r was statistiVally signifiVant (N rtri Ca-
Vas s: Z=0.331, *P* =0.133; Su tri Ca Vas s: Z= 0.346, *P* =
0.65). M r v r, tri av ras F$_{st}$ val b tw n Su tri an
N rtri Ca Vas s u lati ns (ni ns u tri rs) was
0.075, simi ar t thiat am ng N rtri Ca Vas s u lati ns
(0.096) an thiat am ng Su tri Ca Vas s u lati ns
(0.040). Tri r f r, tri Ca Vas s Mu ntai ns a ar n t
siav via a t Vabl ini n V n tri s n tiVstur M r f
Ca Vas s u lati ns; inst a , s n tiV rift rati ns in
small, isi at u lati ns s ms t siav mi nat tri
s n tiVstur M r f Ca Vas s u lati ns.

C m ari s n f Ca Vasian, Ei r an, an N ar East rn Y-via l 9ru s

Tri via l 9ru v rsti in tri Ca Vas s (av ras val :
0.797) is a m st as ri as thiat in C ntra Asia (av ras
val : 0.824) an tri N ar East (av ras val : 0.769)
an is sign ifiVantly hi9 r (*t* -t st, *P* =0.024) thian tri via-
l 9ru v rsti in Ei r (av ras val : 0.633). An
MDS l t an n 9b r- ini ns tr bas n F$_{st}$ val s
(Fi9. 3A, B) s l tt Ei r an u lati ns int W st rn an
East rn 9ru s as sias b n bs rv r u sti (S mi n
t a . 2000), wiri C ntra Asian u lati ns fa in9 in b-
tw n tri W st rn an East rn Ei r an 9ru s. Tri
Ca Vas s u lati ns ar int rmi ni at wiri N ar East rn
u lati ns.

Tri s att rns via u rtri r b n Vnfirm by tri ar-
ys F$_{st}$ V m ari s ns: tri m an ari ry s F$_{st}$ val f r tri
Ca Vas s vs Ei r s 0.254, wi r as tri m an F$_{st}$ val
f r tri Ca Vas s vs tri N ar East s 0.079, wi ri s sig-
nifiVantly l w r (*t* -t st bas n av ras F$_{st}$ val s aV-
nif y r u lati ns, *P* <0.001). On av ras , Su tri
Ca Vas an u lati ns ar m r simi ar t b tri N ar East-
rn an Ei r an u lati ns (av ras F$_{st}$= 0.038 an
0.222, r s Vi y) thian ar N rtri Ca Vas an u lati ns
(av ras F$_{st}$= 0.097 an 0.303); ri w v r, b tri tri Su tri
Ca Vas s an tri N rtri Ca Vas s ar m r simi ar t tri
N ar East thian t Ei r wiri rs, M t Y-SNP via l-
9ru s.

Tri MDS an F$_{st}$ anaiy s s (ni s m 9ru s fr m
W ls t a . (2001) in wi i tri M201 mar r, wi i tri s-
tini (sri s via l 9ru G* fr m via l 9ru F* (Fi9. 2),
was n t anaiyz (Tabl 2). In tri ab v anaiy s s, tri s
(n Vu ais w r assifi as via l 9ru F*, a thu 9h
s mu n n vn r rti n Vu l b via l 9ru G*. T
t rmi n wi tri r tris nabi ty t ist ini sri b tw n
via l 9ru s F* an G* f r s m 9ru s ini nV tri
r sul ts f tri MDS an F$_{st}$ anaiy s s, v assifi all
via l 9ru G* in Vu ais as via l 9ru F* an r at
tri anaiy s s. Tri r sul ts (n t sri vn), v r ss nti ai y
i nti Vai; tri s, tri nabi ty t ist ini sri b tw n via-
l 9ru s F* an G* in s m 9ru s s n t ini nV
u r V ni sri s.

In r r t ni an y ssib bias in u r V ni-
sri s Vas by tri small siz f tri sam l s fr m s m
f tri 9ru s, w r at all anaiy s s aft r ni ni all
9ru s wiri sam l siz l ss tri n 25 (s Tabl 2). All
V ni sri s r main tri sam (ata n t sri vn).

G n tiV rlati nshi b tw n Basq s an Ca Vasians

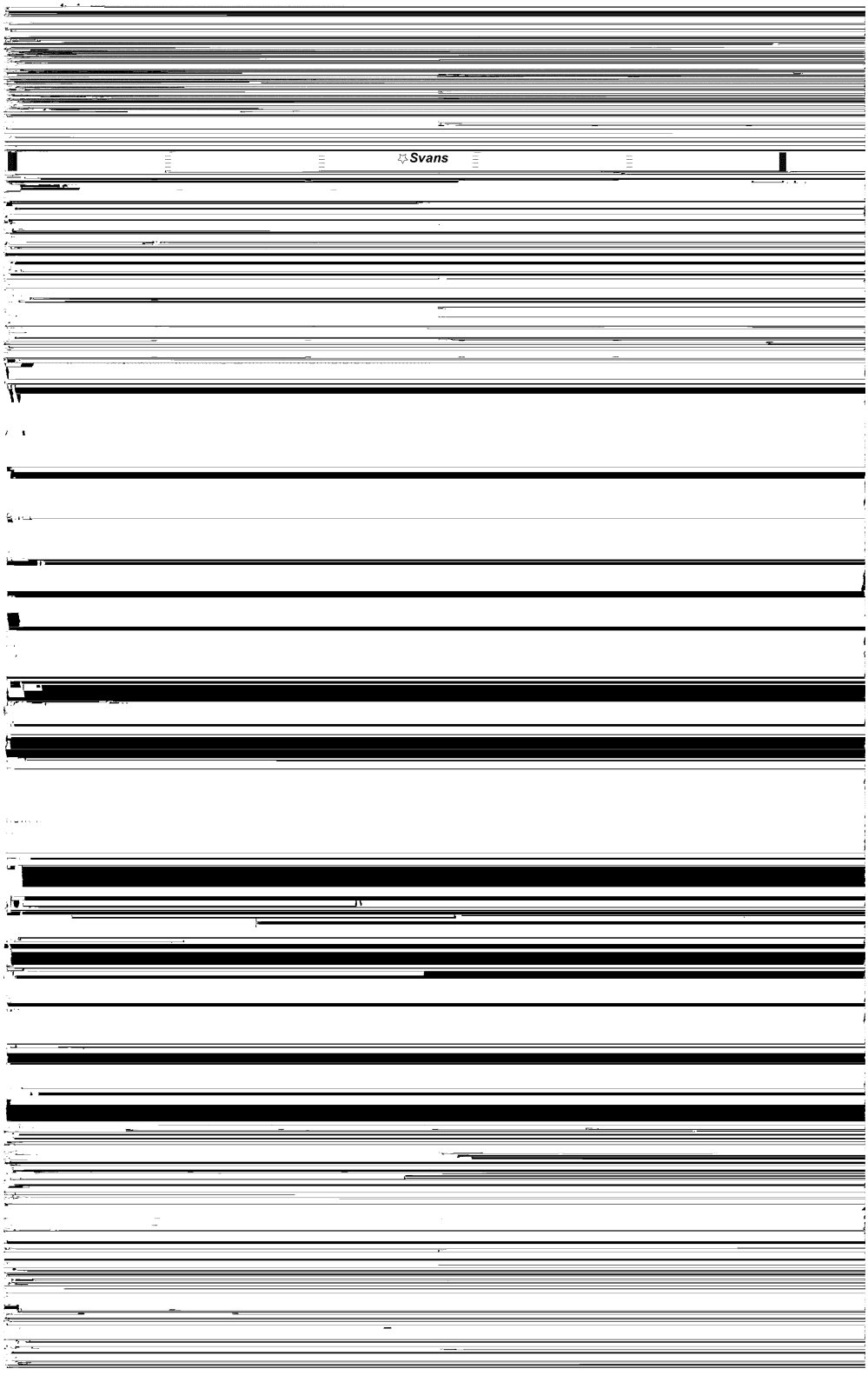
A V mm n ri in f Ca Vasian an Basq s a rs via

av ra a rv s F st ya b tw n Basq s an Ca as an
 ru s s m n r (F_{st} = 0.563) t an t at b tw n
Basq s an In -E r an ru s (F_{st}= 0.311). T Ca -
 as s ru s an Basq s ar n t st r t t r n -
 r t n b r- n ns tr r t MDS t (Fig. 3A, B).
T s r su ts ar n a r m nt w t r y u s st s bas
n ass a mar rs an mtDNA HV1 s q n ata
(B rt r t a . 1995; Nas z an St n n 2001).

Lan a r a m nts an n t r at ns s n t Ca as s

T r s n , n t Ca as s, f ru s w t s -
 ra n n b rs ar n t t r n st n b rs a ws
u s t a r ss t u st n as t n b tt r a ns t
 n t r at ns s f t s ru s: ra r an-

Fig. 3A, B The geographic distribution of Y-chromosome diversity based on 10 Y-SNPs and the YAP marker. **A** MDS plot based on pairwise Fst values, showing relationships among Caucasus, Eastern European, Near Eastern, and Central Asian populations (*open stars* with black font Caucasus populations, *closed circles* Eastern European populations, *open squares* Central Asian populations, *closed diamonds* Near Eastern populations). The stress value of the MDS plot is 0.134. **B** Neighbor-joining tree based on pairwise Fst values (*boldface* Caucasus populations)



The Caucasus, as we have seen. This may still reflect ancient migrations, as it is probably the original Indo-European and Turkic-speaking populations that dominate the region today. However, the migrations were small and/or not accompanied by substantial gene flow. In any event, the migrations did not drastically change the overall genetic landscape of the Caucasus. Currently, there are approximately 8 million Azerbaijanians and 3.5 million Armenians, attesting to the fact that these ancient shifts were followed by substantial population growth.

H w y r, th Y-na l Sru s y ntrast shar w vith th