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

ORIGINAL INVESTIGATION

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 September 2002 / Published online: 14 December 2002
Springer-Verlag 2002

Abstract A number of hypotheses have been put forward concerning the origins of the Caucasus linguistic diversity. Some scholars have suggested that Armenians and Azerbaijanis were originally Caucasian-speaking populations that adopted Indo-European languages (with Caucasian substrates) relatively recently. Armenian and Azerbaijani populations have also been characterized by relatively high frequencies of Y-chromosome haplogroup J, which is thought to have originated in the Near East and spread into the Caucasus with Neolithic agriculture. We have typed 11 Y-chromosome biallelic markers in 389 males from populations representing major Caucasus linguistic groups. As with other Y-chromosome studies, we find relatively high frequencies of haplogroup J in Armenians and Azerbaijanis, but we also find that haplogroup J is present at high frequencies in many Caucasus populations. However, we find that the Armenians and Azerbaijanis have relatively high frequencies of haplogroup R1a1, which is rare in the Caucasus and Near East, but is found at high frequencies in Europe and Central Asia. Haplogroup R1a1 is also found at high frequencies in Caucasus populations that speak Indo-European languages, but is rare in populations that speak Caucasian languages. Thus, the Armenians and Azerbaijanis have Y-chromosome frequencies that are similar to those of Caucasian-speaking populations, but have languages that are Indo-European. Our results suggest that the Armenians and Azerbaijanis have undergone a substantial replacement of their Y-chromosome gene pool, but not their languages, during the past few millennia.

Introduction

The Caucasus, traditionally considered one of the Cradles of Civilization, is characterized by a high degree of linguistic diversity. Three major language families are represented in the region: Indo-European, Caucasian and Altaic. Indo-European languages are spoken by the majority of the populations in the Caucasus, including Armenians, Ossetians, Talysh, Kurds, Tats, Russians, Greeks and Assyrians. Caucasian languages are spoken by the Georgians, Abkhazians, Adygeans, Kabardians, Chechens, Ingushes, Avars, Lezgins, Tabasarans, Rutulians, Tsakhurs, Dargins, Aguls, Lakians, Tsezians, Khvarshians, Hinukhians, Bagvalians, Khinalugians, Udis, Batsbians and Andians. Altaic languages are spoken by the Azerbaijanis, Turkmens, Nogays, Kumyks, Karachays, Balkars and Turks. Linguistic diversity is also found within the Caucasus language families, with the Caucasian languages being the most diverse. Although the Caucasian languages are spoken by relatively few people, they are divided into three distinct groups: Northwest Caucasian, Northeast Caucasian and Kartvelian. Northwest Caucasian languages are spoken by the Abkhazians, Adygeans, Kabardians, Ubykhians and Shapsugians. Northeast Caucasian languages are spoken by the Chechens, Ingushes, Avars, Lezgins, Tabasarans, Rutulians, Tsakhurs, Dargins, Aguls, Lakians, Tsezians, Khvarshians, Hinukhians, Bagvalians, Khinalugians, Udis, Batsbians and Andians. Kartvelian languages are spoken by the Georgians, Svans, Mingrelians and Laz.

The origins of the Caucasus linguistic diversity have been the subject of much speculation. Some scholars have suggested that the Armenians and Azerbaijanis were originally Caucasian-speaking populations that adopted Indo-European languages relatively recently (Diakonoff 1984; Diakonoff and Starostin 1986; Gamkrelidze and Ivanov 1995). Armenian and Azerbaijani populations have also been characterized by relatively high frequencies of Y-chromosome haplogroup J (Nebel et al. 2001), which is thought to have originated in the Near East and spread into the Caucasus with Neolithic agriculture (Semino et al. 2000). Although the Armenians and Azerbaijanis have relatively high frequencies of haplogroup J, they also have relatively high frequencies of haplogroup R1a1, which is rare in the Caucasus and Near East, but is found at high frequencies in Europe and Central Asia (Semino et al. 2000). Haplogroup R1a1 is also found at high frequencies in Caucasus populations that speak Indo-European languages, but is rare in populations that speak Caucasian languages (Nebel et al. 2001). Thus, the Armenians and Azerbaijanis have Y-chromosome frequencies that are similar to those of Caucasian-speaking populations, but have languages that are Indo-European. Our results suggest that the Armenians and Azerbaijanis have undergone a substantial replacement of their Y-chromosome gene pool, but not their languages, during the past few millennia.

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

T. Sarkisian
Center of Medical Genetics,
Andranik Margarian State University of Armenia,
5/1 Zoravar Str., 375010 Yerevan, Armenia

A. Kerimov
Scientific-research Institute of Hematology and Transfusiology,
Azerbaijan Republic Ministry of Health,
Gasanly Str. 87, Baku, Azerbaijan

m nt t mn n , w r rt r n n y s s f y - r m s m - m r s n t s m s t f Cu s n u t n s r u s y n y z f r m t D A HV1 s q n r t n (s z n St n ng 2001).

Materials and methods

S j ts

A t t f 389 s m s (289 w n 100 sw s) fr mu nr t m n u s w r t n t f - w ng g tu t t nu s u t ns (Fg. 1): G r ns (Su t Cu s n s rs), Arm n ns (In -Eur n s rs), Az r- j n ns (ur s rs), n A z n ns, K r n ns, In- s ns, C n ns, n D rg n ns (rt Cu s n s - rs). D A fr m t s m s mu s r u s y fr n y s s f m t D A HV1 s q n r t n n Au ns rt n y m r sms (s z n St n ng 2001; s z t . 2001); sw s fr m t n m s fr m Arm n n Az r j n w r t n r r t n r s t s m s z fr y - r m s m st s. Inf rm ns nt n nf rm t n u t t rt f t n r n t n r's r nts n g r n r- nts w r t n fr m n rs. G n m D A fr m s m s w s tr t u s ng n Is u D A tr t n t (s r , B t , W s ., USA) r n nt n n - r f rm m t (M n nt s t . 1982). D A fr m sw s w s tr t u s ng n nt n s t ng-u t r u r (M r t . 1988). u s y - r m s m s ng n t y m r sm (y-S) t fr m t Su t Cu s s w r u s fr m 25 S ns (G r g), 25 G r g ns fr m K z g (G r g), 12 L z g (Az r j n), n 17 Su t s s t ns (W s t . 2001). A t n u s t n 21 Az r j n ns n 47 Arm n ns (W s t . 2001), n 63 G r g ns (S m n t . 2000; W s t . 2001), n t ff r n m r s, t fr m u r s m s (t n t s w n) n w r n t nu nt n y s s t w g t- ng t r s t s t y n t s u t ns. u s y -S t (S m n t . 2000; W s t . 2001) f r E r n, r E st rn, n C ntr As n u t ns w r s nu .

M u r n y s s

/ n S m r s r u s y r rt t y m r n E r n t r E st (S m n t . 2000) w r y n s m s: S4y (M130), M9, M89, M124, M45, M173, M17, M201, M170, n M172 (U n r t . 2000 n r f r n s t r n); t y A Au ns rt n y m r sm (H mm r n H r 1995) w s s y . F r S m r s t M130 (S4y), / qm n (A- B s st ms) ss s w r s g n . r m r s n y - r s w r s g n u s ng r m r E r ss (V rs n 1 f r M - nt s w r C; E B s st ms). r m r n r s q n s r g n n / 1. t n m s w r r r s s r r - u s y (M r n t . 1999). M130 w s y u s ng t y - m r s n r t n/r str t n fr g m nt n g t y m r sm r u r s r s w r (K y s r t . 2000), w r s t y A Au ns rt n w s y s s r r u s y (H mm r n H r 1995). A s m s w r y fr m r r s; n s- r n y w s fu n tw nt r r r f m r r s - s r y U n r t . (2000) n u r r s ts. / y -S - g r u n m n t r u s r s r n g t r n t r m- m n t ns f t y C r m s m C ns rtu m (2002). / y - g n t r t n s f t y - r m s m y s, s-

St t st n y s s

H g r u rs y n F_{st} u s w r u t w t Ar q n 2.000 (S n r t . 2000), w w s s u s t r y u t M nt t sts f r r r t ns tw n m tr s. M t m n s n s ng (MDS) n y s s (Kr s 1964) f t F_{st} u s w s r- r u t w t S /A /IS /ICA (St tS ft). r g r m s n HW LI 3.5 (F s nst n 1993) w r u s t nstr t n g r-j n ng tr .0256 /w (A M $_{st}$ u s.

Results and discussion

Y-S gru s nt Cu ss s

E n Y-S gru s (Fig. 1) r fu n nt Cu ss s (2, Fig. 1). m st frq nt gru s r F*, G* n J2*; t g t r t frq ny f t s tr gru s s 0.53 0.86 n gru s tf r t D rgn ns. D rgn ns g frq ny f gru I* (0.58), s r s fu n t frq ny f 0.1 r ss. S ns r r us r rt (W st . 2001) t g frq ny (0.92) f gru F*, n Kz g t g frq ny f gru J2* (0.72). tr gru s ng gru t frq ny r t rt n 0.5, t t n f t sm s m f 12 Lz, f r t frq ny f gru F* s 0.58. D rgn ns, Lz, S ns, n Kz g rr s n ng t st gru rs t s (0.153 0.652), r s f r t tr gru s, t gru rs t s 0.779 0.855.

f t tr mm n Cu ss gru s (F* n J2*) r s mm n n r E strn u t ns L ns , ur s, Sr ns (S m n t . 2000), n Irn-ns (W st . 2001), t r g frq n s f 0.165 n 0.28, r s t y, ut r s nt n r frq n s n Er (r g frq n s 0.021 n 0.074, r s-t y). t r mm n Cu ss gru , G*, s r r n Er (frq ny=0.061) n , nt r E st, s n r rt ny nt ur s n L ns gru s (S m n t . 2000). H gru 1*, s mm n n W strn n C ntr Er , s sr n t Su t Cu ss t frq n s g rt n 0.1, r s nt rt Cu ss, ts s nt r n ry s . H-gru 1 1*, s sr n g frq n s nt su t-st M trrn n rg n, E strn Er , n C ntr As , s r s nt t frq ny nt Cu-ss, sm rt t rE st. t r Cu ss y- -gru s ur t frq ny.

D rgn n, S n, n Kz g gru s rt ut rs m r t t r Cu ss gru s. F$_{st}$ u s g st tw n S ns n t r Cu ss gru s (rg F$_{st}$= 0.332), f y t Kz g (rg F$_{st}$= 0.286) n D rgn ns (rg F$_{st}$= 0.25), r s t rg rs F$_{st}$ u m ng t r m n ng Cu-ss gru s s ny 0.047. s g F$_{st}$ us, u-t t r gru rs y n r u n m r f gru s (2) nt D rgn ns, S ns, n Kz g, r m st y t r s t f g nt r ft-r t ng n sm s t u t ns.

rr t n tw n t g r n g nt (rs F$_{st}$) st n ss r t ng rs f Cu ss -u t ns s nt st t st y sgn f nt (M nt t st: Z = 0.113, *P* =0.671). m f t ut rs (S ns, Kz-g, n D rgn ns) rs t n rr t n t t s g r u t st n n-sgn f nt (*Z* =0.301, *P* =0.134). t rmn t r t Cu ss Mu nt ns n nfu-

n t st n f r t Su t n rt Cu s s s -
r t y, u t n t r w s st t st y s g n f nt (rt Cu -
s s: Z=0.331, *P* =0.133; Su t Cu s s: Z= 0.346, *P* =
0.65). M r r, t r g F_{st} u tw n Su t n
rt Cu s s u t ns (ng u t rs) w s
0.075, s m r t t t m ng rt Cu s s u t ns
(0.096) n t t m ng Su t Cu s s u t ns
(0.040). ' r f r, t Cu s s Mu nt ns r n t t
t nfu n nt g nt str ut r f
Cu s s u t ns; nst , g nt r ft r t ng n
sm s t u t ns s ms t m nt t
g nt str ut r f Cu s s u t ns.

C m r s n f Cu s n, E r n, n r E st rn y - g ru s

' g ru rst y nt Cu s s (r g u :
0.797) s m st s g s t t n C ntr As (r g
u : 0.824) n t r E st (r g u : 0.769)
n ss g nf nty g r (*t* -t st, *P* =0.024) t nt -
g ru rst y n E r (r g u : 0.633). An
MDS t n ng r-j n ng tr s n F_{st} u s
(F g. 3A, B) s t E r n u t ns nt W st rn n
E st rn g ru s, s s n s r r u sy (S m n
t . 2000), w t C ntr As n u t ns f ng n -
tw n t W st rn n E st rn E r n g ru s. '
Cu s s u t ns r nt rm ng w t r E st rn
u t ns.

' s tt rns f rt r n nf rm y t r-
w s F_{st} m r s ns; t m n r w s F_{st} u f r t
Cu s s E r s 0.254, w r st m n F_{st} u
f r t Cu s s st r E st s 0.079, w ss g -
nf nty w r (*t* -t st s n r g F_{st} u s j -
nf r u t ns, *P* <0.001). r g , Su t
Cu s n u t ns r m r s m r t t r E st-
rn n E r n u t ns (r g F_{st}= 0.038 n
0.222, r s t y) t n r rt Cu s n u t ns
(r g F_{st}= 0.097 n 0.303); w r, t t Su t
Cu s s n t rt Cu s s r m r s m r t t
r E st t nt E r w t r s, t t y-S -
g ru s.

' MDS n F_{st} n y s s nu s m g ru s fr m
W s t . (2001) n w t M201 m r r, w s-
t ng s s g ru G* fr m g ru F* (F g. 2),
w s n t n y z (' 2). In t n y s s, t s
n u s w r ss f s g ru F*, t u g
s m u n n w n r rt n u g ru G*. '
t rm n w t r t s n t y t st ng s tw n
g ru s F* n G* f r s m g ru s nfu n t
r s ts f t MDS n F_{st} n y s s, w ss f
g ru G* n u s s g ru F* n r t
t n y s s. ' r s ts (n t s w n) w r ss nt y
nt ; tu s, t n t y t st ng s tw n -
g ru s F* n G* n s m g ru s s n t nfu n
u r n u s ns.

In r rt u n y ss s nur nu -
s ns u s y t sm sz f t s m s fr m s m
f t g ru s, w r t n y s s ft r u ng
g ru s w t s m sz ss t n 25 (s ' 2). A
n u s ns r m n t s m (t n t s w n).

G n t r t ns tw n B sq s n Cu s ns

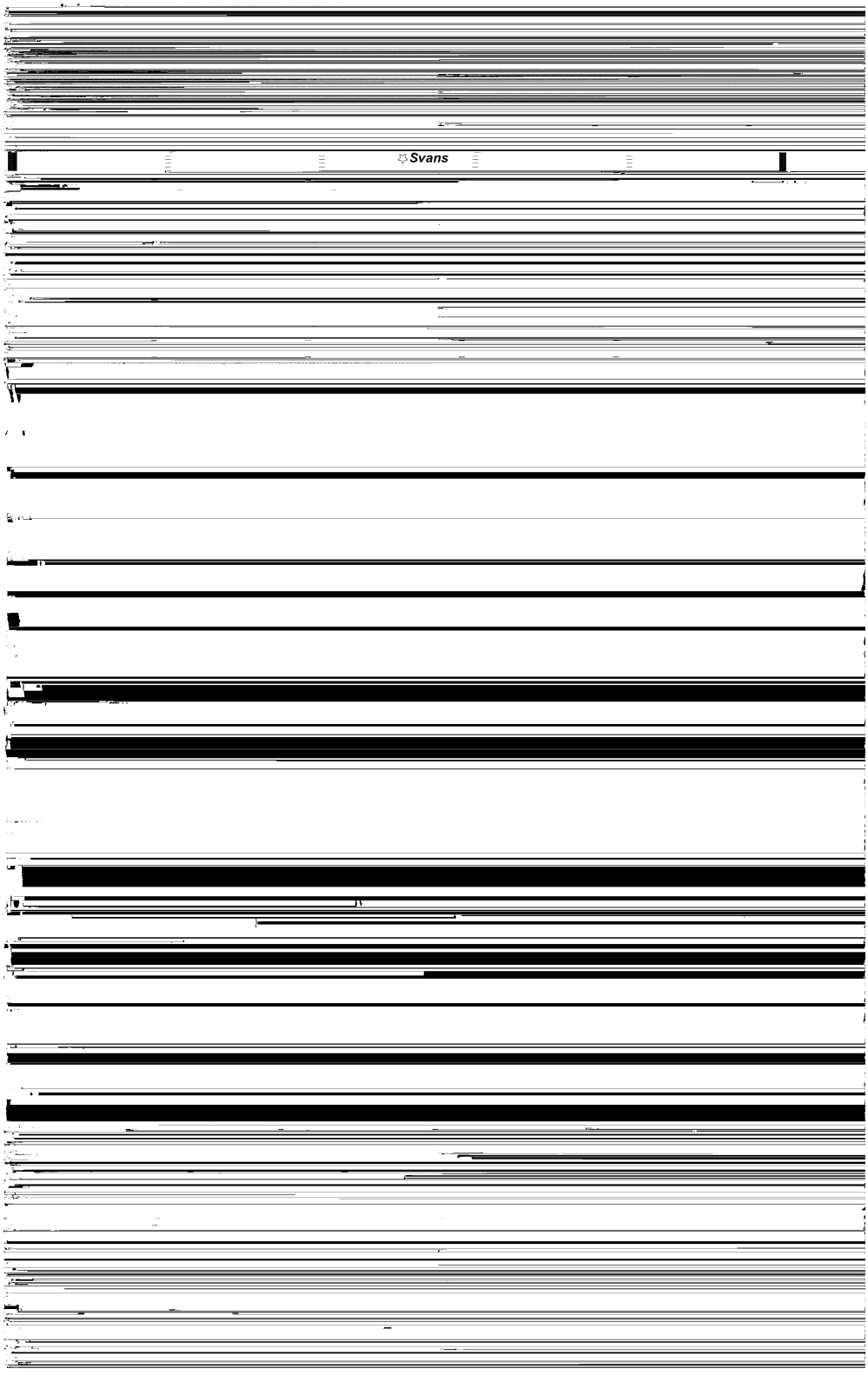
A mm n r g n f Cu s n n B sq s rs s

rg r rvs F_{st} u tv n B sq s n Cu s n gru s s m g r (F_{st} = 0.563) t n t t tv n B sq s n In -Eur n gru s (F_{st}= 0.311). ' Cu - s s gru s n B sq s r n t u st r t g t r n - t r t n g r-j n ng tr r t MDS t (Fg 3A, B). ' s r s ts r n gr m nt v t r u s st s s n ss m r rs n mtD A HV1 s q n t (B rt r t . 1995; s z n St n ng 2001).

L ng g r m nts n g n t r t ns s nt Cu s s

' r s n , n t Cu s s, f gru s v s g - gr n g rs r n t t r ng st n g rs vs u s t r sst q st n st v tt r ns t g n t r t ns s f t s gru s: g gr y r n-

Fig. 3A, B ' y g n t r t ns f y- r m s m g ru s, s n 10 y-S s n t y A m r r. **A** MDS t s n rw s F_{st} u s, s w ng r t n- s s m ng t Cu s s, E- r n, r E st rn, n C n- tr As n u t ns (*open stars* w t f t u- t n n m s Cu s s g ru s, *closed circles* E r n u- t ns, *open squares* C ntr As n g ru s, *closed diamonds* r E st rn u t ns). ' str ss u f r t MDS t s 0.134. **B** g r-j n ng tr s n rw s F_{st} - u s f r t s m u t ns (*boldface* Cu s s g ru s)



u r g ru s, r s t y. ' s m y st r f t n t nt r s nt g ru s. C rr nty, t r r m t y m n n s n r , s r s m y t r g n In -E- 8m n Az r j n ns n 3.5m n Arm n ns, t- r n/u r m g r nt g ru s w r y sm n / r t st ng t t r m r r y f t s ng g s fts; n t m t ns y w t t r s nt g ru s. In n s n g -s s fts wu t m s y nt, t m g r nt g ru s n g g n t m t u u r r t r t nm r y g m ns.