Original Research Article

Searching for the Origin of Gagauzes: Inferences from Y-Chromosome Analysis

ALEXANDER VARZARI,[1,2*] VLADIMIR KHARKOV,[3] WOLFGANG STEPHAN,[1] VALENTIN DERGACHEV,[4] VALERY PUZYREV,[3] ELISABETH H. WEISS,[1] AND VADIM STEPANOV[3]
[1]*Biocentre, Ludwig-Maximilians University Munich, Munich, Germany*
[2]*National Centre of Reproductive Health and Medical Genetics, Kishinev, Moldova*
[3]*Research Institute of Medical Genetics, Russian Academy of Medical Sciences, Tomsk, Russia*
[4]*Department of Archaeology, Institute of Cultural Patrimony, Academy of Sciences of Moldova, Kishinev, Moldova*

ABSTRACT The Gagauzes are a small Turkish-speaking ethnic group living mostly in southern Moldova and northeastern Bulgaria. The origin of the Gagauzes is obscure. They may be descendants of the Turkic nomadic tribes from the Eurasian steppes, as suggested by the "Steppe" hypothesis, or have a complex Anatolian-steppe origin, as postulated by the "Seljuk" or "Anatolian" hypothesis. To distinguish these hypotheses, a sample of 89 Y-chromosomes representing two Gagauz populations from the Republic of Moldova was analyzed for 28 binary and seven STR polymorphisms. In the gene pool of the Gagauzes a total of 15 Y-haplogroups were identified, the most common being I-P37 (20.2%), R-M17 (19.1%), G-M201 (13.5%), R-M269 (12.4%), and E-M78 (11.1%). The present Gagauz populations were compared with other Balkan, Anatolian, and Central Asian populations by means of genetic distances, nonmetric multidimentional scaling and analyses of molecular variance. The analyses showed that Gagauzes belong to the Balkan populations, suggesting that the Gagauz language represents a case of language replacement in southeastern Europe. Interestingly, the detailed study of microsatellite haplotypes revealed some sharing between the Gagauz and Turkish lineages, providing some support of the hypothesis of the "Seljuk origin" of the Gagauzes. The faster evolving microsatellite loci showed that the two Gagauz samples investigated do not represent a homogeneous group. This finding matches the cultural and linguistic heterogeneity of the Gagauzes well, suggesting a crucial role of social factors in shaping the Gagauz Y-chromosome pool and possibly also of effects of genetic drift. Am. J. Hum. Biol. 00:000–000, 2009.

The Gagauzes are a small Turkish-speaking ethnic group living mostly in southern Bessarabia (Moldova Republic, southwestern Ukraine) and southern Dobruja (northeastern Bulgaria, southeastern Romania). The Gagauzes speak the Oghuz version of the Turkic languages, which also includes the Azeri, Turkish, and Turkmeni languages. The Gagauz language is particularly close to the Balkan Turkish dialects spoken in Greece, northeastern Bulgaria, and in the Kumanovo and Bitola areas of Macedonia. The Balkan Turkic languages, including Gagauz, are a typologically interesting case, because they are closely related to Turkish and at the same time contain a North-Turkic (Tartar or Kypchak) element besides the main South-Turkic (Oghuz) element (Pokrovskaya, 1964). The modern Gagauz language has two dialects: central (or "Bulgar") and southern (or maritime) (Pokrovskaya, 1964; Gordon, 2005). It is also important to mention that the Gagauzes are Orthodox Christians, whereas most of the Turkic groups mentioned above are Muslims.

It is historically documented that the Gagauzes migrated to Bessarabia from northeastern Bulgaria (Dobruja) in the beginning of the 19th century fleeing from political and religious oppression by the Ottoman Turks. However, very little is known about their previous history. Several hypotheses about the ethnogenesis of the Gagauzes have been proposed (Pokrovskaya, 1964; Guboglo, 1967). Two of them seem to be most popular among the ethnologists and linguists. One theory considers the Gagauzes as descendants of the Turkic nomadic tribes from the South Russian steppe (Bulgars, Cumans, Pechenegs, or Torks, etc.). According to the other, the Gagauzes descend from the Seljuk Turks that settled in northeastern Bulgaria in the second half of the 13th century, and together with some Turkic tribes from South-Russian steppes they founded a Turkic state there. The military power of the Balkan Turks was used by the Byzantine Empire, because the Turkic hordes ensured its protection against the Slavs (Bulgarians). Having settled in the Balkans the Turkic clans had been converted to Orthodox Christianity before this area was conquered by the Ottoman Turks, i.e., before the 15th century.

The distribution of genetic variation within and among populations has long been used to gain insight into the

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: DFG; Contract grant number: Ste 325/7; Contract grant sponsor: Russian Federation for Basic Research; Contract grant numbers: 03-04-4902, 06-04-48274; Contract grant sponsor: President of the Russian Federation; Contract grant number: MD-840.2003.04; Contract grant sponsors: FCI, DAAD.

*Correspondence to: Alexander Varzari, National Centre of Reproductive Health and Medical Genetics, Burebista str. 82, 2062 Kishinev, Moldova. E-mail: a_varsahr@yahoo.com

Received 24 July 2008; Revision received 23 October 2008; Accepted 11 November 2008

DOI 10.1002/ajhb.20863

Published online in Wiley InterScience (www.interscience.wiley.com).

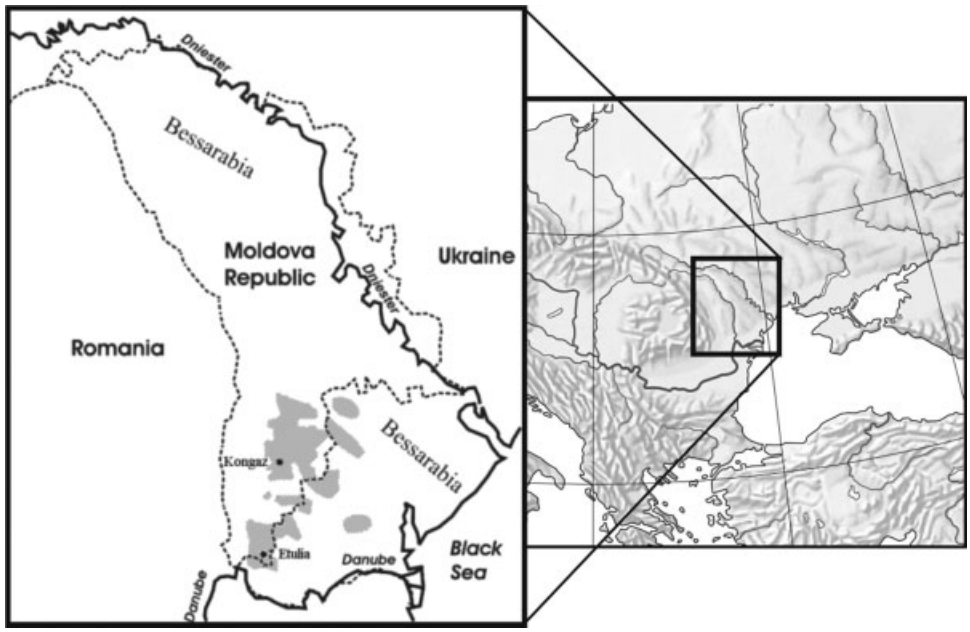

Fig. 1. A map of the geographic location of the Gagauz populations. The Gagauz inhabited areas are shaded.

demographic history of humans. The previous genetic studies in the Dniester-Carpathian region based on classical and autosomal DNA markers showed closer affinities of the Gagauzes to their geographical neighbors from Southeast Europe than to other Turkic populations (Varsahr et al., 2001, 2003; Varzari et al., 2007). A recent analysis of mtDNA and Y-chromosome variation in one Moldavian and one Gagauz population has confirmed similarities between the Gagauzes and their geographical neighbors (Nasidze et al., 2007). The authors of the article suggest that the Gagauzes gained their genetic similarity with Moldavians as a result of extensive gene exchange between them after the Gagauzes migrated from Turkey to Moldavia two hundred years ago and were converted to Orthodox Christianity. However, this conclusion disagrees with the historical and linguistic data; it was based on a small set of binary markers and a relatively small sample size (50 individuals), and did not take into account the degree of substructuring of the Gagauz population.

In the present study we have further characterized the genetic structure of the Gagauzes, using a comprehensive Y-chromosome analysis of two Gagauz populations from southern Moldavia that speak different dialects, and have compared them with the populations from the Balkans, Anatolia and Central Asia that have had a great historical influence on the Gagauzes. Analysis of Y-chromosome variation is particularly useful in studies of the Gagauz origins because invasions were primarily carried out by males and, therefore, one might expect the Gagauz male pool to retain some trace of the invaders.

MATERIALS AND METHODS

Samples and DNA extraction

A total of 89 unrelated male samples were collected from two Gagauz locations: the Kongaz settlement, $N = 48$, and the Etulia settlement, $N = 41$ (see Fig. 1). Informed consent was obtained from all participants in this study, and information about geographic and ethnic origins of their parents and grandparents was recorded. DNA was extracted from peripheral blood lymphocytes by a salt-based extraction method (Miller et al., 1988) or by using the Amersham genomic DNA extraction reagents and protocols.

Y-chromosome polymorphisms

The Y-chromosomal haplotype composition and structure was examined using two genetic marker systems from the nonrecombining portion of the Y-chromosome: binary markers, mostly represented by SNPs, and multiallelic, highly variable microsatellites (STRs).

Binary markers. Y-chromosome haplogroups were defined by the analysis of 28 binary markers. 23 markers were typed according to previous reports, namely YAP [DYS287] (Hammer and Horai, 1995), 12f2 [DYS11] (Rosser et al., 2000), M17 and M89 (Kharkov et al., 2004), 92R7 (Mathias et al., 1994), Tat [M46] (Zerjal et al., 1997), M9 (Hurles et al., 1998), M70 and M223 (Kharkov et al., 2007), M78 (Flores et al., 2003; Underhill et al., 2001), M123 (Flores et al., 2003), P25 and P37 (Kharkov et al., 2005), M130, M172, M178, M201, M207, M242, and M269 (Underhill et al., 2001; Kharkov et al., 2005), M253 (Chinnioğlu et al., 2004; Kharkov et al., 2007), and SRY-2627 (Hurles et al., 1999). M170 was typed by sequencing from the forward primer (Underhill et al., 2001). In addition, we genotyped five polymorphisms reported previously, namely M12, M47, M67, and M92 (Underhill et al., 2001), and M267 (Cinnioğlu et al., 2004). Primer sequences for each of these five markers were used as previously described, or were designed by introducing a mismatched base to produce a variable restriction site on the amplification products (Table 1). The samples were examined in a hierarchical way, in agreement with the Y-chromosome

TABLE 1. PCR-RFLP protocols developed for five binary markers

Marker	Primers used (5′-3′)	T^a	Size[b]	Digestion	Fragment/s (allele)[c]
M12	F: ACTAAAACACCATTAGAAACAAAGG R: CTGAGCAACATAGTGACCCGAAT[d]	62	309	*Hin*fI	23/67/219 (G) → 90/219 (T)
M47	F: AGATCATCCCAAAACAATCATAA R: GAAATCAATCCAATCTGTAAATTTTATGTAGAATT	61	430	*Eco*RI	35/395 (G) → 430 (A)
M67	F: CCATATTCTTTATACTTTCTACCTGC R: GTCTTTTCACTTGTTCGTGGACCCCTCAATAT	60	409	*Ssp*I	379/30 (A) → 409 (T)
M92	F: TTGAATTTCCCAGAATTTTGC R: TTCAGAAACTGGTTTTGTGTCC	61	470	*Bst*SNI	470 (T) → 340/130 (C)
M267	F: TTATCCTGAGCCGTTGTCCCTG R: CTAGATTGTGTTCTTCCACACAAAATACTGTACGT[d]	60	183	*Bst*SNI	150/33 (T) → 183 (G)

F refers to the forward primer, and R refers to the reverse primer for a particular locus.
[a]PCR annealing temperature in C°.
[b]PCR product size in base pairs.
[c]RFLP fragments in base pairs.
[d]Mismatched primer (mismatched bases are underlined).

haplogroup tree (Karafet et al., 2008). The genealogical relationship of the haplogroups (defined by the markers) is shown in Figure 2. M9 was chosen as the initial marker and surveyed in all samples.

Microsatellite markers. The following seven Y-specific microsatellites were analyzed: *DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392*, and *DYS393*. These markers form the so-called minimal haplotype (de Knijff et al., 1997) and are the most comprehensively studied Y-STR set in different world populations. All loci were PCR-amplified using primers and conditions described elsewhere (de Knijf et al., 1997; Kayser et al., 1997). The forward primers were labeled with TET (green) for *DYS390* and *DYS391*, FAM (blue) for *DYS392* and *DYS393*, and HEX (yellow) for *DYS19, DYS389I,* and *DYS389II*. The amplification products were then pooled together and run on an ABI Prism 310 sequencer (Applied Biosystems) using GeneScan500-TAMRA (red) as the internal standard. Gene Scan Analysis Software v.3.7 (Applied Biosystems) was used to analyze fragment sizes. The alleles were named according to the number of repeat units they contain. The number of repeat units was established through the use of sequenced reference DNA samples as suggested by de Knijf et al., (1997). Allele length for *DYS389b* was obtained by subtraction of the *DYS389I* allele length from that of *DYS389II*.

Statistical analysis

The software package Arlequin 2.000 (Schneider et al., 2000) was used to calculate several population genetic parameters, including diversity of haplogroups and haplotypes, exact tests of population differentiation, as well as pairwise F_{ST} (for haplogroup) and R_{ST} (for haplotype) values. The significance of these statistics was examined with 10,000 permutations. F_{ST} and R_{ST} distances among compared populations were represented in two dimensions with multidimensional scaling (MDS), using the STATISTICA 5.5 software package (StatSoft, Inc 1995). Arlequin was also used to perform analyses of molecular variance (AMOVA) of the SNP and STR data. Network analysis of the STR data was carried out with the software package NETWORK version 4.2.0.1 (http://www.fluxus-technology.com). Networks were calculated by the median-joining method ($\gamma = 0$) after having processed the data with the reduced median method (Bandelt et al., 1999). To score different mutation rates upon the networks construction, each STR locus was weighted in accordance with the values estimated previously (Kayser et al., 2000).

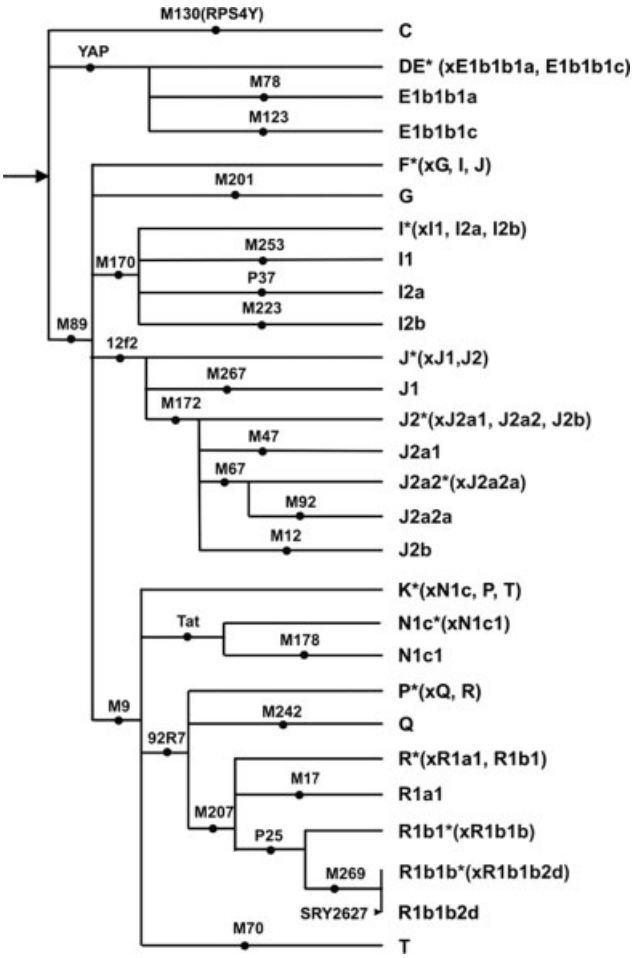

Fig. 2. Maximum parsimony phylogeny of the haplogroups defined by the 28 binary markers used in this study. The names of haplogroups are according to Karafet et al., (2008).

RESULTS

Y-chromosome lineages in the Gagauzes

Haplogroup frequencies in the Gagauz samples and in the pooled sample are reported in Table 2. A total of 23 of

TABLE 2. Haplogroup counts and frequencies, together with Y-chromosome diversities in the Gagauz populations

Haplogroup					
Lineage-based name	Mutation-based name		Kongaz	Etulia	Total
E1b1b1a	E-M78	*N*	6	4	10
		%	12.5	9.8	11.2
E1b1b1c	E-M123	*N*	2	0	2
		%	4.2	0.0	2.2
G	G-M201	*N*	5	7	12
		%	10.4	17.1	13.5
I1	I-M253	*N*	4	0	4
		%	8.3	0.0	4.5
I2a	I-P37	*N*	9	9	18
		%	18.8	22.0	20.2
I2b	I-M223	*N*	2	1	3
		%	4.2	2.4	3.4
J*(xJ1,J2)	J-12f2	*N*	1	0	1
		%	2.1	0.0	1.1
J1	J-M267	*N*	1	0	1
		%	2.1	0.0	1.1
J2*(xJ2a1,J2a2,J2b)	J-M172	*N*	1	2	3
		%	2.1	4.9	3.4
J2a2*(xJ2a2a)	J-M67	*N*	0	1	1
		%	0.0	2.4	1.1
J2b	J-M12	*N*	1	0	1
		%	2.1	0.0	1.1
N1c1	N-M178	*N*	2	0	2
		%	4.2	0.0	2.2
R1a1	R-M17	*N*	6	11	17
		%	12.5	26.8	19.1
R1b1b*(xR1b1b2d)	R-M269	*N*	5	6	11
		%	10.4	14.6	12.4
T	T-M70	*N*	3	0	3
		%	6.3	0.0	3.4
Total		*N*	48	41	89
		%	100.0	100.0	100.0
$H \pm$ SD			0.9131 ± 0.0173	0.8366 ± 0.0262	0.8795 ± 0.0147
No of STR haplotypes			37	25	57
$D \pm$ SD			0.9885 ± 0.0067	0.9646 ± 0.0150	0.9794 ± 0.0008

H, haplogroup diversity; *D*, microsatellite haplotype diversity; SD, standard deviation.

the 28 genotyped binary polymorphisms were informative and defined 15 distinct haplogroups. Two major haplogroups in Gagauz males are haplogroup I-P37 and haplogroup R-M17, comprising 20.2% and 19.1%, respectively, of all Gagauz Y-chromosomes. These were followed by haplogroups G-M201 (13.5%), R-M269 (12.4%), and E-M78 (11.2%). All of the remaining lineages were present at frequencies of less than 5% in the Gagauz paternal gene pool. No lineages representing distant areas (Central/East Asia or Africa) were found in the present study. The haplogroup distributions were similar in the two samples (exact test; $P = 0.1028$) and were in agreement with those reported previously for the Gagauz population (Nasidze et al., 2007) or neighboring populations (Supp Info Table 1). Although Y-haplogoup distribution patterns in two Gagauz populations were not significantly different from each other and from those in other southeastern European populations (exact test; $P > 0.05$), we note a twofold higher frequency of the R-M17 haplogroup in the Etulia sample compared with samples from Kongaz and Comrat (Nasidze et al., 2007), as well as an increased frequency of the G-M201 haplogroup in the two studied samples compared with most Balkan populations, including the Gagauzes from Comrat. The Gagauzes in total are characterized by high haplogroup diversity, comparable with other groups from southeastern Europe (Supporting Information Table 1) that exceed diversity values from other European provinces whose gene pools are dominated by certain haplogroups.

Y-STR polymorphisms were studied to obtain a more detailed view of Y variation in the Gagauz populations. 57 different STR haplotypes were observed among 89 individuals. Thirty-eight haplotypes were found in only one individual, 11 in two individuals, 6 in 3 individuals, and 1 in 5 (ht1) and six (ht47) individuals. Haplotype diversities in the total sample of the Gagauzes (0.979) and in the sample from Kongaz (0.989) were among the values observed in the Balkan ethnic groups (Supporting Information Table 2). The diversity in the Gagauzes from Etulia was lower (0.965); however, this did not differ significantly from other Balkan populations (*T*-test; *t-value* $= -1.429$; df = 13; $P = 0.177$). Both the STR and binary data showed that the Etulia population had less genetic diversity than the Kongaz population. Most haplotypes were found to be population-specific. In all cases but one, the chromosomes sharing a haplotype belonged to the same haplogroup. Hence, 58 compound binary-STR haplotypes were observed (Table A1).

Relationships and population structure

We used genetic distance analysis to compare the present data with those reported for Balkan, Anatolian, and Central Asian populations (Supporting Information Table 1). Pairwise F_{ST} comparisons based on the Y-haplogroup frequencies showed that the Gagauz samples were very similar to each other ($P = 0.26$) and to other Balkan populations (Supporting Information Table 3). They were less

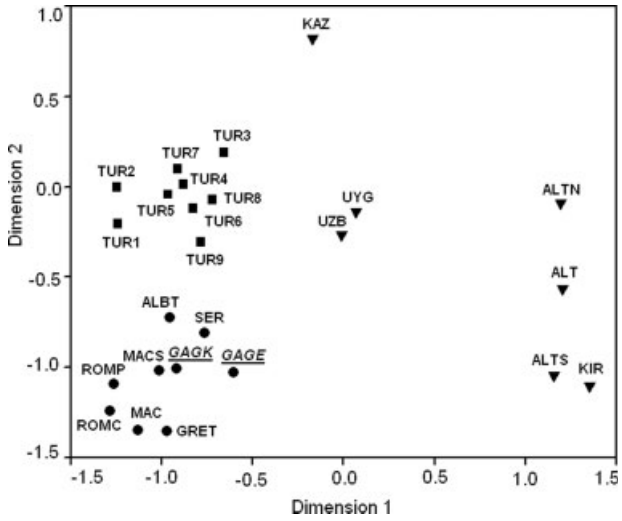

Fig. 3. Plot based on a multidimensional scaling (MDS) analysis of FST values from Y chromosome haplogroup frequencies, showing genetic affinities between the Gagauz and some Balkan, Anatolian, and Central Asian population samples. The stress value for the MDS plot is 0.090. The populations presented are: GAGK = Gagauzes from Kongaz; GAGE = Gagauzes from Etulia (present study); TUR1-TUR9 = Turks (Cinnioğlu et al., 2004); ROMC = Romanians from Constanta; ROMP = Romanians from Ploiesti; GRET = Thracian Greeks; MACS = Macedonians from Scopie, Republic of Macedonia; ALBT = Albanians from Tirana, Albania (Bosch et al., 2006); MAC = Macedonians from Republic of Macedonia; SER = Serbs (Pericić et al., 2005b); ALT = Altais; KAZ = Kazakhs; KIR = Kirghiz; UYG = Uygurs; UZB = Uzbeks (Karafet et al., 2002); ALTN = northern Altais; ALTS = southern Altais (Kharkov et al., 2007). The Balkan groups are indicated by circles; Anatolian groups by squares; and Central Asian groups by triangles.

similar to Turkish samples, and most distant to Central Asian groups. All pairwise differences between the Gagauz and the Turkic samples, including those from Anatolia, were statistically significant ($P < 0.05$). The MDS analysis based on the F_{ST} distance matrix summarizes these patterns (see Fig. 3). The samples from Anatolia and the Balkans fall into two contiguous clusters. The positions of the populations within these clusters correspond well with their assignments to specific regional groups. The populations from Central Asia exhibit the most considerable interpopulation variability, showing significant distances to Anatolian and Balkan groups ($P < 0.05$). Both Gagauz samples clearly cluster with the Balkan samples, thus showing a general similarity with geographically close populations.

Fast mutating markers may be more suitable to study genetic differentiation between populations that are rather closely related genetically. We therefore also used STR haplotype frequencies and molecular differences between haplotypes for phylogenetic reconstructions within Balkans and Anatolia. Phylogenetic analysis was performed by pooling the data of the present study with those of Zaharova et al., (2001), Robino et al., (2002), Barbarii et al., (2003), Cinnioğlu et al., (2004), Robino et al., (2004), Bosch et al., (2006), Lauc et al., (2005), Pericic et al., (2005a), Spiroski et al., (2005) (Supporting Information Table 2). Eighteen of twenty-four compared samples were same as in the previous analysis based on the binary polymorphisms. This enables us to compare the results of the two analyses. Results of MDS based on R_{ST} genetic distances (Supporting Information Table 4) are shown in Figure 4. As in the case of the binary markers, the compared populations are grouping according to major geographic regions. Both Gagauz samples have close affinity to the Balkan ethnic groups; however, they exhibit substantial dissimilarities if compared with each other ($P = 0.04$). In terms of genetic distances, the Gagauzes from Etulia show the highest affinity to the northern Greeks, Serbs and Romanians from Constanta and Ploiesti, and the lowest to the Turkish groups, whereas the Gagauzes from Kongaz show close affinity with the majority of the Balkan populations, including the Bulgarian Turks, as well as with the three Turkish groups from Anatolia. Remarkably, the affinity of the Gagauzes from Kongaz to the Turks is not higher than affinity of the latter to some non-Turkic ethnic groups from the Balkans.

The pairwise F_{ST} and R_{ST} comparisons show that Gagauzes are similar to surrounding populations and distant to the Turkic ones. However, F_{ST} and R_{ST} analyses are known to be influenced by multiple-testing problems. To avoid these problems, AMOVA analyses were performed (Table 3). Within Anatolia the genetic variance attributable to differences among populations was not significantly different from zero for both data sets ($P > 0.05$), suggesting that Anatolian populations are highly homogeneous. In the Balkan region a high genetic homogeneity was revealed only for Y-haplogroups ($P > 0.05$), whereas for Y-STR haplotypes a significant heterogeneity was found ($P < 0.001$). Likewise, the analysis showed no significant differences in haplogroup and significant differences in haplotype compositions between the Gagauz populations. The highest level of population differentiation was observed in Central Asia, with 7.7% of the total Y-haplogroup variation being attributable to differences among populations. Previous genetic analyses based on Y-chromosome and mtDNA data also revealed substantial genetic diversity among Central Asian populations. Such findings seem to be strongly determined by the historical past of Central Asia, which in turn is largely influenced by its geographical location at the crossroads between major Eurasian subdivisions. The AMOVA for the Y chromosome showed significant differences in haplogroup and Y-STR haplotype composition ($P < 0.001$) between major geographic regions. No significant differences were found between Gagauz and non-Gagauz populations in the Balkans when considering both sets of markers ($P > 0.05$). In contrast, we observed striking genetic differences between Gagauz and Turkic-speaking groups from Central Asia and Anatolia ($P < 0.05$). Thus, this set of analyses, in agreement with phylogenetic analyses, shows that the Gagauz Y-pools belong to the Balkan pools of Y-chromosomes.

The R-M17 chromosomes could penetrate into the gene pool of the Gagauzes from Central Asia, where in some Turkic populations they are present in a very high frequency (Karafet et al., 2002; Kharkov et al., 2007; Wells et al., 2001; Zerjal et al., 2002). To explore the genetic similarities of the R-M17 Gagauz chromosomes with those from Central Asia and the Balkans, a median network based on Y-chromosome STR haplotypes on the background of M17 was generated (see Fig. 5). In the median network, the Balkan and Asian haplotypes tend to cluster according to geography and most of the Gagauz haplotypes cluster with the Balkan haplotypes. In particular, we could not find any Y-haplotypes typical for Central

Asia (that are absent on the Balkans) in the Gagauz gene pool. Pairwise R_{ST} comparisons for Y-STR haplotypes within haplogroup R-M17 further indicate that the Gagauz R-M17 chromosomes are closely related to the Balkan R-M17 chromosomes (0.0207; $P > 0.05$) than to those from Central Asia (0.3522; $P < 0.001$).

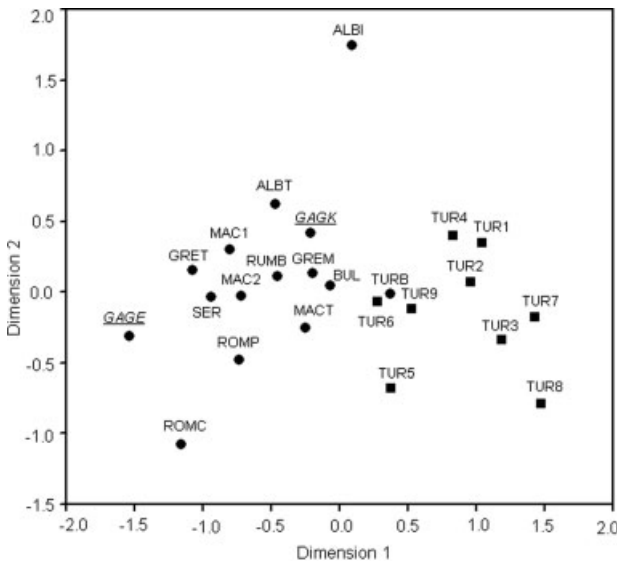

Fig. 4. Plot from multidimensional scaling (MDS) analysis of RST values from Y chromosome STR haplotype frequencies, showing genetic affinities among Balkan and Anatolian populations. The stress value for the MDS plot is 0.065. The populations presented are: GAGK = Gagauzes from Kongaz; GAGE = Gagauzes from Etulia (present study); TUR1-TUR9 = Turks (Cinnioğlu et al., 2004); BUL = Bulgarians; TURB = Bulgarian Turks (Zaharova et al., 2001); MAC1 = Macedonians from Republic of Macedonia (Pericić et al., 2005a); MAC2 = Macedonians from Republic of Macedonia (Spiroski et al., 2005); SER = Serbs (Lauc et al., 2005); ALBT = Albanians from Tirana; GRET = Thracian Greeks; MACT = Macedonians from Tirana, Republic of Macedonia; ROMC = Romanians from Constanta; ROMP = Romanians from Ploiesti (Bosch et al., 2006); GREM = Macedonian Greeks (Robino et al., 2004); ALBI = Albanians (Robino et al., 2002); ROMB = Romanians from Bucharest (Barbarii et al., 2003). The populations investigated in the present study are in italic and underlined. The Balkan groups are indicated by circles and Anatolian groups by squares.

Of the five predominant Y-haplogroups present in the Gagauzes, haplogroups R-M269 and G-M201 are widespread in Anatolia (Cinnioğlu et al., 2004). A detailed microsatellite analysis of these haplogroups in Gagauz, Anatolian, and Balkan populations is presented in Figure 6. For haplogroup G-M201 two Gagauz haplotypes (ht9 and ht13) were found to be shared with Turkish haplotypes, but no haplotype sharing was found between the Gagauzes and the Balkans, implying that at least the two shared with the Turks' G-M201 lineages penetrated into the Gagauzes from Anatolia. In the R-M269 network of haplotypes, of four haplotypes shared by the Gagauzes with other populations one Gagauz Y-STR haplotype (ht51) groups with an Anatolian haplotype, one (ht49) clusters with a Balkan haplotype, and the remaining two haplotypes (ht50 and ht54) could be of either Balkan or Anatolian origin. Besides the three haplotypes mentioned, one belonging to R-M269 (ht51) and two to G-M201 (ht9 and ht13), we did not succeed in finding other haplotypes specific to Anatolian Turks in the Gagauzes.

DISCUSSION

Two different scenarios have been postulated in order to explain the origin of the Gagauzes. Each scenario suggests a different structure of the extant Gagauz gene pool, being either closer to the Central Asian or Anatolian one. The evidence will now be considered in the light of the Y-data.

The Gagauzes may be descendants of the Turkic nomadic tribes from the Eurasian heartlands. This hypothesis would imply genetic similarity between Gagauzes and Turkic-speaking groups from Central Asia. A distinguishing feature of the population of Central Asia is its high genetic heterogeneity (Karafet et al., 2002; Zerjal et al., 2002). Haplogroups Q-M242, C-M130, O-M175 and R-M17, however, are present in every population in Central Asia. The first three of the haplogroups are specific to the Asian region, but very scarce in Europe. The Gagauzes

TABLE 3. AMOVA results

	Grouping	Among groups	Among populations	Within populations
Y-HG (F_{ST}[a])	Gagauzes		0.57 ns	99.43
	Balkans		0.58 ns	99.42
	Anatolia (Turks)		0.49 ns	99.51
	Central Asia		7.70***	92.30
	Balkans, Anatolia	6.88***	0.49*	92.63
	Balkans, Central Asia	11.18***	3.40***	85.42
	Anatolia, Central Asia	9.00***	3.15***	87.85
	Gagauzes, Balkans	0.75 ns	0.31 ns	98.94
	Gagauzes, Anatolia	5.56*	0.47 ns	93.96
	Gagauzes, Central Asia	6.84*	6.35***	86.82
Y-HT (R_{ST}[a])	Gagauzes		3.28*	96.72
	Balkans		1.87***	98.13
	Anatolia (Turks)		0.65 ns	99.35
	Balkans, Anatolia	4.44***	1.37***	94.19
	Gagauzes, Balkans	0.28 ns	1.82***	97.90
	Gagauzes, Anatolia	5.94*	0.83*	93.23

HG, haplogoups; HT, haplotypes; ns, not significant.[a]Distance method applied.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

differ greatly from Central Asian populations with respect to Y-haplogroup frequencies. Indeed, none of 89 Gagauz male chromosomes investigated belongs to the Asian cluster, i.e., to the haplogroups Q-M242, C-M130, and O-M175. Although the haplogroup R-M17 is widely present in the gene pool of Gagauzes, we could not find among the Gagauz R-M17 chromosomes those specific to Central Asian populations. On the contrary, the Gagauz R-M17 chromosomes demonstrate a much higher affinity and identity with R-M17 chromosomes from the Balkans than with the ones from Central Asia, suggesting the plausible European origin of the R-M17 chromosomes in the Gagauz paternal gene pool. Some significant differences between Y-haplogroup frequencies in Gagauzes and in Central Asian populations are mirrored in significant genetic distances between them. Thus, our Y data seems to reject the hypothesis that the Gagauzes are biological descendants of the Turkic nomadic tribes from the Eurasian steppe.

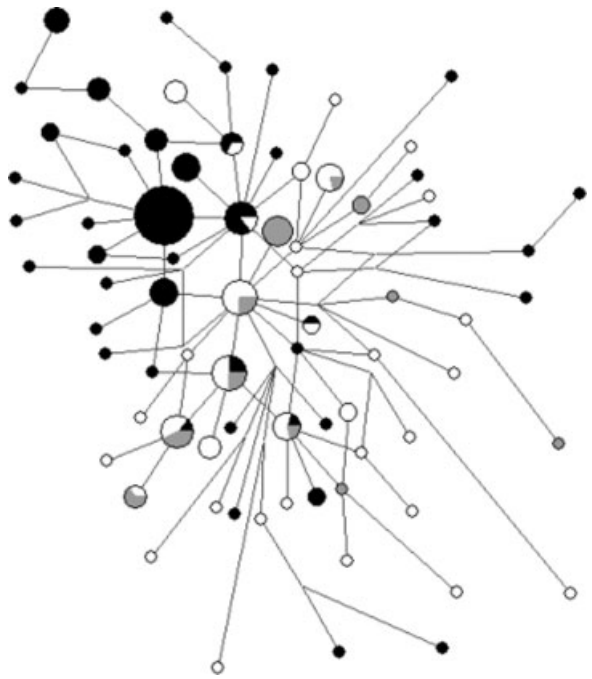

Fig. 5. Median-joining networks showing phylogenetic relationships of the Gagauz, Balkan, and Central Asian Y-haplotypes within haplogroup R-M17. *Gray* Gagauzes (R1a1-M17 chromosomes from present study pooled with those from Nasidze et al., 2007); *white* Balkan (Macedonians, Serbs, Albanians, Greeks from Bosch et al., 2006 and Pericić et al., 2005b); *black* Central Asia (Altais, Kazakhs, Kirghiz, Uygurs, Uzbeks from Zerjal et al., 2002 and Kharkov et al., 2007). The size of each circle is proportional to the haplotype frequency.

According to the hypothesis of an Anatolian origin, the Gagauzes are traced to the Seljuk-Turks who migrated to Dobruja from Anatolia in the end of 13th century, and afterwards mixed with Turkic nomads from the Eurasian steppe. This scenario would imply a close genetic relationship between Gagauzes and Anatolian Turks. The haplogroup frequencies in the Gagauzes were also significantly different from those in Anatolian/Turkish populations, though to a lower degree than in Central Asian populations. The Anatolian populations have a high frequency of the Middle Eastern haplogroup J-12f2, whereas European haplogroups I-M170 and R-M17 are present here in much lower frequencies. The Gagauzes, on the contrary, have a low frequency of haplogroup J-12f2 and high or moderately high frequencies of I-M170 and R-M17. The frequencies of these haplogroups in the Gagauzes are very close to those in the Balkans. The Gagauzes also represent the Balkans with respect to the E-M78 to E-M123 ratios; haplogroup E-M78 occurs here much more often than E-M123 (Cruciani et al., 2004; Semino et al., 2004), whereas in Anatolia E-M78 and E-M123 occur at approximately equal frequencies (Cinnioğlu et al., 2004). Visual inspection revealed that the only Y-chromosome lineage that had frequencies in the Gagauzes closer to those in Turks than in the Balkans was G-M201. These frequencies were 0.171, 0.104 (our data) and 0.041 (Nasidze et al., 2007) in the Gagauz populations, 0–0.129 (average 0.055) in the rest of the Balkans and 0.039–0.200 (average 0.112) in Anatolia. This situation could indicate paternal gene flow mediated by the Turks, as suggested by the Seljuk hypothesis. Or, alternatively, genetic drift could be responsible for the increased

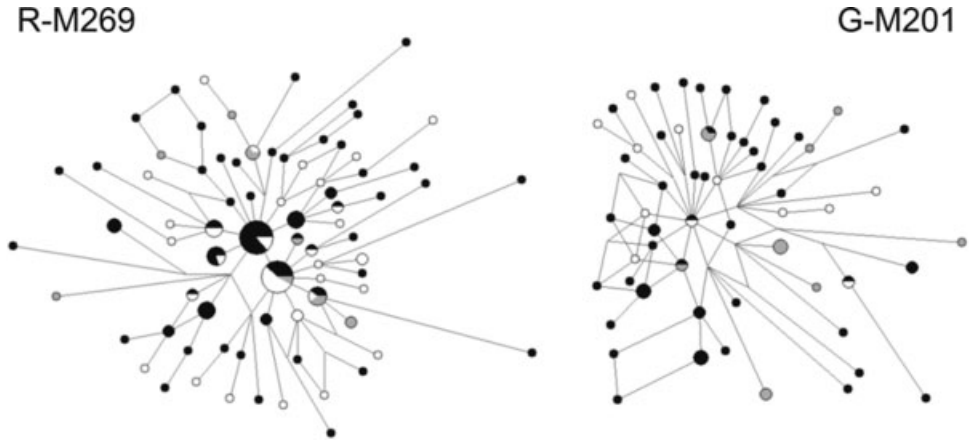

Fig. 6. Median-joining networks showing phylogenetic relationships of the Gagauz, Balkan, and Anatolian Y-haplotypes within haplogroups R-M269 and G-M201. *Gray* Gagauzes (present study); *white* Balkan (Macedonians, Serbs, Albanians, Greeks from Bosch et al., 2006 and Pericić et al., 2005b); *black* Anatolia (Turks from Cinnioğlu et al., 2004). The size of each circle is proportional to the haplotype frequency.

G-M201 frequencies in two Gagauz samples. Analyses of diversity and median networks have demonstrated the plausibility of both assumptions. Indeed, the Gagauzes from Etulia with the highest G-M201 frequency are characterized by a relatively low level of STR haplotype diversity within G-M201 ($D = 0.810$), indicating some effect of genetic drift. At the same time, some sharing between the Kongaz and Turkish G-M201 haplotypes in the absence of any sharing between the Gagauz and Balkan G-M201 haplotypes suggests a direct contribution of the Turks to the Gagauz paternal gene pool and, hence, lends some support to the theory of the Seljuk origin of the Gagauzes.

Although some sharing between Gagauz and Turkish Y-haplotypes implies direct gene flow from Anatolia to the Gagauzes, its impact on the structure of the extant Gagauz gene pool was rather small. This conclusion is supported by three lines of evidence: (1) the Gagauzes represent the Balkans with respect to the Y-haplogroup frequencies; (2) genetic distance analyses based on stable and fast polymorphisms indicate a closer relationship of the Gagauzes to Balkan populations than to any Turkic group, and (3) in the MDS plots the Gagauz samples were not intermediate between the Balkan and Turkic samples, but occupied positions among the Balkan ones. These results are in agreement with previous investigations based on "classical" and DNA markers (Nasidze et al., 2007; Varsahr et al., 2001, 2003; Varzari et al., 2007). Altogether the genetic data indicate that the Gagauz language represents a case of language replacement in southeastern Europe. How has this replacement happened?

In our previous investigation of autosomal DNA markers in the Dniester-Carpathian region (Varzari et al., 2007), we suggested that in the case of the Gagauzes replacement could have occurred via the "elite dominance" model, which means that the original Turkic immigrant groups could be very small such that their genetic effect on the resident groups was negligible (Renfrew, 1987). This hypothesis is supported by numerous historical sources (Guboglo, 1967; Shabashov, 2002). Throughout the Middle Ages the Balkan peninsula was constantly subjected to Turkic invasions and conquests both from the southern Russian steppe and Anatolia. These tribes formed military (for example, that of the Avars, the Pechenegs, and the Cumans) and political (for example, that of the Bulgars and the Seljuks) unions, which also included the local Slavic and Romance populations besides the Turkic newcomers.

Another point of view was offered by Nasidze et al., (2007). The authors consider the Gagauzes as "Orthodox Turks". After their resettlement to southern Moldavia from Turkey 150 years ago (as it is asserted by Nasidze et al., 2007) they were intensively exchanging genes with the Moldavians. As a result they became genetically closer to the Moldavians than to the Turks. Though the general idea of gene exchange between populations as a mechanism for erasing the genetic differences between them is undoubtedly correct, lines of historical, ethnological, and linguistic data provide evidence against this theory. First, the Gagauzes resettled to southern Moldova not from Turkey, but from the Balkan Peninsula, where they formed an independent ethnic group probably before the Ottoman occupation of the Balkans (Guboglo, 1967; Shabashov, 2002). Secondly, the Gagauz language contains a North-Turkic (Tartar or Kypchak) element besides the main South-Turkic (Oghuz) element, which probably entered by the northern route from the Eurasian steppes (Pokrovskaya, 1964). Thirdly, before Bessarabia (Moldavia) joined the Soviet Union in 1940, marriages between the Gagauzes and other nationalities were extremely rare because of the Gagauzes' strong patriarchal way of life, which forbade inter-ethnical marriages (Zelenciuk and Guboglo, 1979; Curoglo and Marunevici, 1983; Kvilinkova, 2007). The number of marriages between different nationalities, however, increased considerably because of social and spiritual transformations among the Gagauzes in the Soviet period (Curoglo and Marunevici, 1983; Varzar' et al., 2003; Zelenciuk and Guboglo, 1979). It should also be noted that we collected DNA samples for our research in ethnically homogeneous localities where Gagauzes constituted more than 95%, and we collected these samples from adult individuals whose ancestors were of the same (i.e. Gagauz) nationality back to the third generation. Altogether, the aforementioned arguments suggest that the genetic affinity between the Gagauzes and the Moldavians is explained by their common "Balkan" ancestry rather than by direct intermarriages.

The faster evolving microsatellite loci showed that Gagauzes do not represent a homogeneous group. This finding does not contradict the analysis of stable polymorphisms, for which inter-population differences in allele frequencies (although insignificant) have also been found. Molecular differences within shared haplogroups appear to make the main contribution to the observed differentiation of the Gagauzes. The observed genetic heterogeneity correlates well with the cultural and linguistic heterogeneity among the Gagauzes. The Gagauzes from Kongaz speak a central (or "Bulgar") dialect, whereas the Gagauzes from Etulia speak a southern (or maritime) dialect. As ethnologists and linguists maintain, the ethnic differentiation of the Gagauzes had happened on the Balkan Peninsula long before their migration to Bessarabia in the beginning of 19th century (Kvilinkova, 2007; Pokrovskaya, 1964). The "Bulgar" Gagauzes were in the domain of the Bulgarian Orthodox Church, and thus subjected to a strong cultural influence by Bulgarians. The finding that the Kongaz Gagauzes are very close genetically to Bulgarians may be explained, in part, by a culturally enforced mixing between the Bulgarians and "Bulgar" Gagauzes. Alternatively, the Turkic language could be imposed on a group of Bulgarians through the elite-dominance process. The maritime Gagauzes were in the domain of the Greek Orthodox Church and, thus, socially isolated from the Bulgarians because of hostile relations between the two Orthodox Churches on the Balkan Peninsula. The significant distance between the Gagauzes from Etulia, on the one hand, and the Gagauzes from Kongaz and the Bulgarians, on the other hand, implies a limited gene flow between these groups. Alternatively, the differences between the Gagauz groups (either cultural or genetic) may have existed prior to the penetration of the Turkic language into the Balkans. According to this hypothesis, the Turkic language could have been imposed on culturally and genetically diverse groups in the Balkans. Moreover, the genetic heterogeneity of the Gagauzes could have been reinforced by possible fragmentations of their gene pool throughout history and particularly during their migration from the Balkans to Bessarabia in the beginning of the 19th century, possibly facilitated by the effects of genetic drift. The reduction in both haplogroup

and haplotype diversity values in Etulia Ggagauzes agrees well with the action of drift.

In conclusion, our Y-chromosome analysis indicates a strong similarity between Gagauzes and Balkan populations. This finding could support the suggestion previously advanced on the basis of autosomal DNA markers, and the historical information that the Turkic language was imposed on the Balkans according to the elite-dominance model. According to this hypothesis, the Turkic newcomers were small in number such that their genes have been diluted by those of the autochthonous inhabitants. Interestingly, using microsatellite markers, we also discovered some traces of recent Anatolian lineages in the Gagauz paternal gene pool. This discovery matches the hypothesis of a Seljuk (Anatolian) origin of the Gagauz language, which, however, does not rule out a penetration of some Turkic linguistic elements from Eurasian steppes. Furthermore, we demonstrated that at the Balkan scale the Gagauzes are not a genetically homogeneous group. The observed genetic heterogeneity correlates well with the cultural and linguistic diversity among the Gagauzes and was presumably determined by the culturally and/or genetically heterogeneous environment on the Balkans. Genetic drift caused by cultural isolation and migration of Gagauzes from the Balkans to Bessarabia could also have facilitated the genetic differentiation among the Gagauz populations.

ACKNOWLEDGMENTS

The authors are grateful to all donors for providing blood samples and to the people who contributed to their collection. They also thank Boris Joffe and Elisaveta Kvilinkova for helpful advice and comments.

LITERATURE CITED

Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Barbarii LE, Rolf B, Dermengiu D. 2003. Y-chromosomal STR haplotypes in a Romanian population sample. Int J Legal Med 117:312–315.

Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil H.-G, Huckenbeck W, Efremovska L, Mikerezi I, Xirotiris N, Grasa C, Schmidt H, Comas D. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70:459–487.

Cinnioğlu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y chromosome haplotype strata in Anatolia. Hum Genet 114:127–148.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y-chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022.

Curoglo SS, Marunevici MV. 1983. The social transformations in the everyday life and culture of the Gagauz population of the MSSR. Chisinau (Moldova): Stiinta (in Russian).

de Knijff P, Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L. 1997. Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Legal Med 110:134–149.

Flores C, Maca-Meyer N, Perez JA, Gonzalez AM, Larruga JM, Cabrera VM. 2003. A predominant European ancestry of paternal lineages from Canary Islanders. Ann Hum Genet 67:138–152.

Gordon RJ. 2005. Ethnologue: languages of the World, 15th ed. Dallas, TX: SIL International. Available at:www.ethnologue.com.

Guboglo MN. 1967. The ethnic affiliation of the Gagauzes. Sov Etnograf 3:160–167 (in Russian).

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

Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling M, Sykes BC. 1998. European Y chromosomal lineages in Polynesia: a contrast to the population structure revealed by mitochondrial DNA. Am J Hum Genet 63:1793–1806.

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Perez-Lezaun A, Bosch E, Shlumukova M, Shlumukova M, Cambon-Thomsen A, McElreavey K, Lopez De Munain A, Rohl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA. 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 67:1055–1061.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838.

Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF. 2002. High levels of Y chromosome differentiation among Native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum Biol 74:761–789.

Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L. 1997. Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A. 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588.

Kharkov VN, Stepanov VA, Borinskaia SA, Kozhekbaeva ZhM, Gusar VA, Grechanina EIa, Puzyrev VP, Khusnutdinova EK, Yankovskii NK. 2004. Gene pool structure of eastern Ukrainians as inferred from the Y-chromosomal haplogroups. Russ J Genet 40:326–331.

Kharkov VN, Stepanov VA, Medvedeva OF, Spiridonova MG, Voevoda MI, Tadinova VN, Puzyrev VP. 2007. Gene pool differences between northern and southern Altaians inferred from the data on Y-chromosomal haplogroups. Russ J Genet 43:675–687.

Kharkov NV, Stepanov VA, Puzyrev VP, Feschenko SP, Borinskaya SA, Yankovsky NK. 2005. Frequency of Y chromosomal bi-allelic haplogroups in Belorussians. Russ J Genet 41:938–931.

Kvilinkova EN. 2007. The traditional spiritual culture of the Gagauzes: ethno-regional features. Chisinau (Moldova): Business-Elita (in Russian).

Lauc LB, Perićić M, Klarić IM, Sijacki A, Popović D, Janićijević B, Rudan P. 2005. Y chromosome STR polymorphisms in a Serbian population sample. Forensic Sci Int 150:97–101.

Mathias N, Bayes M, Tyler-Smith C. 1994. Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123.

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

Nasidze I, Quinque D, Udina I, Kunizheva S, Stoneking M. 2007. The Gagauz, a linguistic enclave, are not a genetic isolate. Ann Hum Genet 71:379–389.

Perićić M, Klarić IM, Lauc LB, Janićijević B, Dordević D, Efremovska L, Rudan P. 2005a. Population genetics of 8 Y-chromosome STR loci in Macedonians and Macedonian Romani (Gypsy). Forensic Sci Int 154:257–261.

Perićić M, Lauc LB, Klarić IM, Rootsi S, Janicijević B, Rudan I, Terzić R, Colak I, Kvesić A, Popović D, Sijacki A, Behluli I, Dordević D, Efremovska L, Bajeć DD, Stefanović BD, Villems R, Rudan P. 2005b. High-resolution phylogenetic analysis of southeastern Europe (SEE) traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964–1975.

Pokrovskaya LM. 1964. Gagauz grammar: phonetics and morphology. Moscow (Russia): Nauka (in Russian).

Renfrew C. 1987. Archaeology and language: the puzzle of Indo-European origins. London, UK: Jonathan Cape.

Robino C, Gino S, Ricci U, Grignani P, Previdere C, Torre C. 2002. Y-chromosomal STR haplotypes in an Albanian population sample. Forensic Sci Int 129:128–130.

Robino C, Varacalli S, Gino S, Chatzikyriakidou A, Kouvatsi A, Triantaphyllidis C, Di Gaetano C, Crobu F, Matullo G, Piazza A, Torre C. 2004. Y-chromosomal STR haplotypes in a population sample from continental Greece, and the islands of Crete and Chios. Forensic Sci Int 145:61–64.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L,

Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HBSM, de Knijff P, Decorte R, Dubrova YuE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kučinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar A-V, Mitchell RJ, Nafa Kh, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PhC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling M. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin Ver. 2000: A software for population genetics data analysis. Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y chromosome haplogrops E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean Area. Am J Hum Genet 74:1023–1034.

Shabashov AB. 2002. The Gagauzes: kinship terms and ethnic origin. Odessa (Ukraine): Astroprint (in Russian).

Spiroski M, Arsov T, Krüger C, Willuweit S, Roewer L. 2005. Y-chromosomal STR haplotypes in Macedonian population samples. Forensic Sci Int 148:69–73.

StatSoft, Inc. 1995. STATISTICA for Windows [Computer program manual]. Tulsa, OK.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62.

Varsahr AM, Dubova NA, Kutuyev IA. 2003. Serological researches in the south of Moldavia in connection with the problem of the ethnogeny of the Gagauzes, the Moldavians and the Bulgarians. Anthropol Anz 61:395–411.

Varsahr AM, Spitsyn VA, Bychkovskaya LS, Kravchuk OI. 2001. To the research of the gene pool of the Gagauz population of Moldavia. Anthropol Anz 59:11–17.

Varzar' AM, Spitsyn VA, Sheremet'eva VA. 2003. Genetic-demographic study of the Gagauz population of Moldova. Russ J Genet 39:1258–1267.

Varzari A, Stephan W, Stepanov V, Raicu F, Cojocaru R, Roschin Yu, Glavce Ch, Dergachev V, Spiridonova M, Schmidt HD, Weiss E. 2007. Population history of the Dniester-Carpathians: evidence from Alu markers. J Hum Genet 52:308–316.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249.

Zaharova B, Andonova S, Gilissen A, Cassiman JJ, Decorte R, Kremensky I. 2001. Y-chromosomal STR haplotypes in three major population groups in Bulgaria. Forensic Sci Int 124:182–186.

Zelenciuk VS, Guboglo MN. 1979. The national and international aspects in the Soviet way of life. Stiinta, Kishinev, MD (in Russian).

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C. 1997. Genetic relationship of Asians and Northern Europeans, revealed by Y chromosomal DNA analysis. Am J Hum Genet 60:1174–1183.

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.

APPENDIX

TABLE A1. Y-STR haplotypes by haplogroups in the Gagauzes

		Allele status at									
Haplotype	Haplogroup	DYS19	DYS389I	DYS389b	DYS390	DYS391	DYS392	DYS393	Kongaz	Etulia	Total
ht1	**E-M78**	13	13	17	24	10	11	13	**2**	**3**	**5**
ht2	**E-M78**	13	13	17	25	10	11	13	**1**	**1**	**2**
ht3	**E-M78**	13	13	18	24	10	11	13	**3**		**3**
ht4	**E-M123**	13	12	18	24	10	11	14	**1**		**1**
ht5	**E-M123**	13	12	18	24	11	11	14	**1**		**1**
ht6	**G-M201**	14	12	16	23	10	11	15		**2**	**2**
ht7	**G-M201**	14	12	16	24	10	12	13	**1**		**1**
ht8	**G-M201**	14	12	17	23	10	12	14		**1**	**1**
ht9	**G-M201**	15	12	17	21	10	11	14	**1**		**1**
ht10	**G-M201**	15	12	17	23	10	12	14		**3**	**3**
ht11	**G-M201**	16	12	16	21	10	11	13		**1**	**1**
ht12	**G-M201**	16	12	16	22	10	11	13	**1**		**1**
ht13	**G-M201**	16	12	17	22	10	10	14	**2**		**2**
ht14	**I-M253**	13	12	17	23	10	11	13	**1**		**1**
ht15	**I-M253**	14	12	15	22	10	11	14	**1**		**1**
ht16	**I-M253**	14	12	16	23	10	11	13	**2**		**2**
ht17	**I-P37**	14	13	17	24	10	11	13	**3**		**3**
ht18	**I-P37**	15	13	17	24	11	11	13		**1**	**1**
ht19	**I-P37**	15	13	18	24	11	11	13		**1**	**1**
ht20	**I-P37**	16	13	17	24	11	11	13	**1**		**1**
ht21	**I-P37**	16	13	18	24	10	11	13	**2**		**2**
ht22	**I-P37**	16	13	18	24	11	11	13		**1**	**1**
ht23	**I-P37**	16	13	18	24	11	11	15		**1**	**1**
ht24	**I-P37**	16	13	19	24	11	11	13	**2**	**1**	**3**
ht25	**I-P37**	17	13	18	24	11	11	13	**1**	**2**	**3**
ht26	**I-P37**	17	13	19	24	11	11	13		**2**	**2**
ht27	**I-M223**	15	12	16	23	10	12	14	**1**		**1**
ht28	**I-M223**	15	13	16	23	10	12	15	**1**		**1**
ht29	**I-M223**	16	13	17	23	10	12	13		**1**	**1**
ht30	**J-12f2***	15	13	16	23	9	11	12	**1**		**1**
ht31	**J-M267**	15	13	17	23	10	11	12	**1**		**1**
ht32	**J-M172***	14	13	17	23	10	11	12		**1**	**1**
ht33	**J-M172***	15	13	16	23	9	11	12		**1**	**1**
ht34	**J-M172***	16	13	16	24	9	11	14	**1**		**1**

(Continued)

TABLE A1. (Continued)

Haplotype	Haplogroup	Allele status at							Kongaz	Etulia	Total
		DYS19	DYS389I	DYS389b	DYS390	DYS391	DYS392	DYS393			
ht35	**J-M67***	14	13	14	22	10	11	12		**1**	**1**
ht36	**J-M12**	15	12	16	24	10	11	12	**1**		**1**
ht37	**N-M178**	14	14	16	23	10	14	15	**1**		**1**
ht38	**N-M178**	14	14	16	23	11	14	14	**1**		**1**
ht39	**R-M17**	15	13	17	25	10	11	13	**1**		**1**
ht40	**R-M17**	15	13	18	25	10	11	13	**1**		**1**
ht41	**R-M17**	16	10	16	25	10	11	13	**1**		**1**
ht42	**R-M17**	16	13	15	25	10	11	13	**1**		**1**
ht43	**R-M17**	16	13	16	24	10	11	13		**1**	**1**
ht44	**R-M17**	16	13	16	25	10	11	13		**3**	**3**
ht45	**R-M17**	16	13	17	25	10	11	13	**1**		**1**
ht46	**R-M17**	16	13	17	25	11	11	13		**1**	**1**
ht47	**R-M17**	16	13	17	26	11	11	13		**6**	**6**
ht48	**R-M17**	17	13	17	25	10	11	13	**1**		**1**
ht49	**R-M269***	14	13	16	24	11	11	12	**2**		**2**
ht50	**R-M269***	14	13	16	24	11	13	13	**1**		**1**
ht51	**R-M269***	14	13	16	24	12	13	13		**1**	**1**
ht52	**R-M269***	14	13	17	24	11	11	12	**1**		**1**
ht53	**R-M269***	14	14	15	25	10	14	12	**1**		**1**
ht54	**R-M269***	14	14	16	24	11	13	13		**2**	**2**
ht55	**R-M269***	14	14	16	25	10	13	12		**1**	**1**
ht56	**R-M269***	14	15	16	24	11	13	13		**2**	**2**
ht57	**T-M70**	13	14	16	23	10	13	13	**2**		**2**
ht58	**T-M70**	14	15	17	23	10	15	14	**1**		**1**