

## Population history of the Dniester–Carpathians: evidence from *Alu* markers

Alexander Varzari · Wolfgang Stephan · Vadim Stepanov · Florina Raicu · Radu Cojocaru · Yuri Roschin · Cristiana Glavce · Valentin Dergachev · Maria Spiridonova · Horst D. Schmidt · Elisabeth Weiss

Received: 11 November 2006 / Accepted: 3 January 2007 / Published online: 16 February 2007  
© The Japan Society of Human Genetics and Springer 2007

**Abstract** The area between the Dniester and the eastern Carpathian mountain range is at a geographical crossroads between eastern Europe and the Balkans. Little is known about the genetics of the population of this region. We performed an analysis of 12 binary autosomal markers in samples from six Dniester–Carpathian populations: two Moldavian, one Romanian, one Ukrainian and two Gagauz populations. The results were compared with gene frequency data from culturally and linguistically related populations from Southeast Europe and Central Asia. Small genetic differences were found among southeastern European populations (in particular those of the Dniester–Carpathian region). The observed homogeneity suggests either a very recent common ancestry of all south-

eastern European populations or strong gene flow between them. Despite this low level of differentiation, tree reconstruction and principle component analyses allowed a distinction between Balkan–Carpathian (Macedonians, Romanians, Moldavians, Ukrainians and Gagauzes) and eastern Mediterranean (Turks, Greeks and Albanians) population groups. The genetic affinities among Dniester–Carpathian and southeastern European populations do not reflect their linguistic relationships. The results indicate that the ethnic and genetic differentiations occurred in these regions to a considerable extent independently of each other. In particular, Gagauzes, a Turkic-speaking population, show closer affinities to their geographical neighbors than to other Turkic populations.

---

A. Varzari (✉)  
National Center of Reproductive Health  
and Medical Genetics, Burebista str. 82,  
2062 Kishinev, Moldova  
e-mail: a\_varsahr@yahoo.com

A. Varzari · W. Stephan · E. Weiss  
Biocentre, Ludwigs-Maximilian University Munich,  
Grosshaderner Strasse 2, 82152 Planegg-Martinsried,  
Germany

F. Raicu · C. Glavce  
Anthropological Research Centre “Francisc Rainer”,  
Romanian Academy, B-dul Eroii Sanitari 8, Sector 5,  
Bucharest, Romania

R. Cojocaru  
National Scientific and Practical Centre  
for Preventive Medicine, Ministry of Health,  
Gheorghii Asachi 67, 2028 Kishinev, Moldova

Y. Roschin  
Medical Diagnostical Centre “Modus Vivendi”, Kishinev,  
Moldova, Burebista str. 80,  
2062 Kishinev, Moldova

A. Varzari · V. Dergachev  
Institute of Archaeology and Ethnography,  
Academy of Sciences of Moldova,  
Banulescu-Bodoni str. 35,  
2012 Kishinev, Moldova

V. Stepanov · M. Spiridonova  
Research Institute of Medical Genetics,  
Russian Academy of Medical Sciences,  
Nab. Ushayky 10, 634050 Tomsk, Russia

H. D. Schmidt  
Department of Anthropology, University of Ulm,  
Parkstrasse 11, 89073 Ulm, Germany

**Keywords** *Alu* insertion · DNA polymorphism · Moldavians · Romanians · Ukrainians · Gagauzes · Population structure

## Introduction

The area between the Dniester and the eastern Carpathian mountain range, which we refer to as the Dniester–Carpathian region, represents an important geographical link between eastern Europe and the Balkans. Beginning from the Neolithic epoch, advanced farming traditions and technologies spread from the Balkan Peninsula into eastern Europe through the Dniester–Carpathian territory. Nomadic tribes of the Kurgan cultures moved in the opposite direction, from the Eurasian heartlands towards the west and south-west. The constant interaction of the resident agricultural populations with nomadic tribes was an important factor for the demographic development in this area (for review see Dergachev 1999; Velicanova 1975). At the end of the second millennium BC the Dniester–Carpathian region was inhabited by different Thracian tribes. Later, during the Roman domination in southeastern Europe, they were subjected to Romanization. In the fifth to seventh centuries AD different Slavic clans peopled southeastern Europe and had close contact with the Romanized populations. As a result of these interactions, the old Romanic community north of the Danube appeared. The complicated history of the Dniester–Carpathian region caused a considerable fragmentation of its cultural and linguistic landscape. Nowadays, people in the Dniester–Carpathian region still have different cultural and linguistic habits. Together with the prevalent Romanian-speaking population (Romanians and Moldavians), Slavic peoples (Russians, Ukrainians, and Bulgarians), Gagauzes, Roma (Gypsies) and other nationalities live here.

Among this ethnic variety, the Turkic-speaking Gagauzes deserve special attention since they are linguistically isolated among the Indo-European majority. The Gagauzes speak the Oghuz version of the Turkic languages, to which the Turkish, Azerbaijani and Turkmenian languages also belong. However, two important differences exist between the Gagauzes and the above Turkic peoples: (1) a North-Turkic (Tartar or Kypchak) element is present in the Gagauz language besides the main South-Turkic (Oghuz) element, and (2) the Gagauzes are orthodox Christians. Several hypotheses about the ethnogenesis of the Gagauzes have been proposed (for review see Guboglo 1967). The Gagauzes are considered by some researchers as descendants of the nomadic tribes of the Uzi,

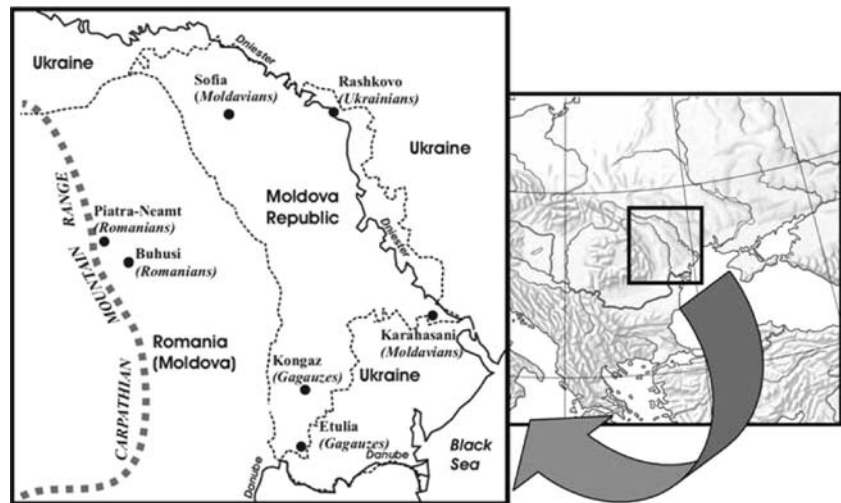
Pechenegs and Polovtsi that are known to have crossed the Danube River and settled in the tenth to thirteenth centuries in the Balkan territories. Other authors consider them as descendants of the Seljuk Turks, who came from the Anatolian Peninsula and established a short-lived state in northeastern Bulgaria in the second half of the thirteenth century AD. According to the third scenario, the Gagauzes are Bulgarians turkicized during the Ottoman occupation of the Balkans in the fifteenth to nineteenth centuries.

In spite of the importance of this region in the population history of Europe, genetic studies of the Dniester–Carpathian population are scarce, and generally restricted to analyses of “classical” markers (Varsahr et al. 2001, 2003, 2006). DNA polymorphisms provide a rich source of information on the genetic structure and evolutionary history of human populations. In an attempt to gain an insight into the genetics of the Dniester–Carpathian region, we studied 12 polymorphic *Alu* markers in six autochthonous Dniester–Carpathian populations. The specific goals were (1) to determine the distribution of *Alu* insertions in the autochthonous Dniester–Carpathian populations, (2) to quantify the degree of genetic differentiation within the Dniester–Carpathian region, and (3) to assess the genetic relationships among the Dniester–Carpathian populations and their relations to linguistically and culturally closely related populations from Europe and Asia.

## Materials and methods

A total of 513 autochthonous unrelated individuals from the Dniester–Carpathian region were analyzed. The six samples comprised Moldavians from village Karahasani ( $n = 123$ ), Moldavians from village Sofia ( $n = 82$ ), Gagauzes from village Kongaz ( $n = 72$ ) and from village Etulia ( $n = 64$ ), Ukrainians from village Rashkovo ( $n = 85$ ) and a sample of Romanians from the Piatra-Neamt and Bacau districts in Romanian Moldova ( $n = 87$ ) (Fig. 1). A specimen of blood (5 ml from each individual by venipuncture in EDTA) was taken after obtaining both the permission of the examined person and a description of his/her ancestral lineage. DNA was extracted from peripheral blood lymphocytes by a salt-based extraction method (Miller et al. 1988) or by using Amersham blood reagents and protocols (Amersham, Little Chalfont, UK). Twelve human-specific autosomal *Alu* polymorphisms were typed in each sample using the primers and PCR conditions described earlier (Batzer et al. 1996; Arcot et al. 1995a, b; Majumder et al. 1999; Comas et al.

**Fig. 1** Locations of the sampled populations in the Dniester–Carpathian region



2000). The PCR products were analyzed in agarose gels. DNA samples were visualized with ethidium bromide under UV light and the results were recorded.

The allele frequencies and heterozygosities were calculated as described by Zhivotovsky (1991). Exact tests for Hardy–Weinberg equilibrium (Guo and Thompson 1992), calculations of gene diversities (Nei 1987), analyses of molecular variance (AMOVA) and genetic distances ( $F_{ST}$ ) (Excoffier et al. 1992) were performed using the software package Arlequin version 2.0 (Schneider et al. 2000) with 10,000 permutations. The values of the genetic differentiation index  $G_{ST}$  were calculated according to Nei (1973). Nei's genetic distances were computed between pairs of populations and were represented in a neighbor-joining tree by means of the PHYLIP 3.5 package (Felsenstein 1993). A total of 1,000 bootstrap replications were performed to assess the strength of the branching structure of the tree. The Kruskal–Wallis' test, Spearman's correlation analysis and principal component (PC) analysis based on the correlation matrix of the *Alu* insertion frequencies were performed using the STATISTICA package (StatSoft 1995).

## Results

### Allele frequencies and genetic diversity within populations

The number of chromosomes examined and the allele frequencies of the 12 *Alu* markers are given in Table 1. All loci were polymorphic in all populations; no case of allele fixation was found. The frequencies of insertion polymorphisms were similar for all studied groups, generally falling within the range of European popu-

lations (for comparisons see Stoneking et al. 1997; Comas et al. 2000, 2004; Romualdi et al. 2002). However, the insertion rate at the *TPA25* locus in the Moldavian sample from the Sofia settlement (0.659) lies slightly outside the European range and is close to the maximum found thus far [in Madras, India (0.690) and Sri Lanka (0.724); Antunez-de-Mayolo et al. 2002]. Furthermore, we note the relatively low insertion frequencies at the loci *TPA25*, *B65*, *D1* and *A25* in the Gagauzes from Etulia, and at the *HS3.23* locus in the Gagauzes from Kongaz.

The observed and expected genotype frequencies were in agreement in most populations. Only 3 of 72 tests for Hardy–Weinberg equilibrium showed significant departures from equilibrium (*D1* in Ukrainians, *HS3.23* in Romanians and *HS2.43* in the Gagauz sample from Kongaz). Since none of the deviations are assigned to a particular locus or population, they probably represent random statistical fluctuations.

Six of twelve loci (*ACE*, *TPA25*, *FXIIIIB*, *B65*, *D1*, *CD4del*) exhibited high heterozygosity levels (nearly 0.5) (Table 2). For four loci, *APOA1*, *A25*, *HS2.43*, and *HS4.65*, the diversity level was low (0.06–0.18). Average heterozygosity per population does not differ significantly between the samples when analyzed by the Kruskal–Wallis test ( $P = 0.9957$ ). This was expected since similar *Alu* insertion frequencies were found in all samples analyzed.

### Genetic relationships between populations

To analyze the genetic relationships among the populations, two approaches were followed: tree reconstruction and principle component (PC) analysis. We based these calculations on the published results for 11 *Alu* loci in southeastern European populations

**Table 1** *Alu* insertion frequencies in Dniester–Carpathian populations. The frequency indicated for each bi-allelic marker refers to the presence of the insert, except for *CD4del*. *n* Number of chromosomes. Moldavians: *K* Karahasani, *S* Sofia. Gagauzes: *K* Kongaz, *E* Etulia

| Locus          | Moldavians K |           | Moldavians S |           | Gagauzes K |           | Gagauzes E |           | Ukrainians |           | Romanians |           | Total    |           |
|----------------|--------------|-----------|--------------|-----------|------------|-----------|------------|-----------|------------|-----------|-----------|-----------|----------|-----------|
|                | <i>n</i>     | Frequency | <i>n</i>     | Frequency | <i>n</i>   | Frequency | <i>n</i>   | Frequency | <i>n</i>   | Frequency | <i>n</i>  | Frequency | <i>n</i> | Frequency |
| <i>ACE</i>     | 244          | 0.492     | 164          | 0.476     | 144        | 0.465     | 128        | 0.461     | 166        | 0.404     | 174       | 0.391     | 1,020    | 0.450     |
| <i>PV92</i>    | 242          | 0.236     | 164          | 0.268     | 144        | 0.201     | 128        | 0.258     | 170        | 0.235     | 172       | 0.250     | 1,020    | 0.241     |
| <i>TPA25</i>   | 246          | 0.508     | 164          | 0.659     | 144        | 0.514     | 128        | 0.453     | 170        | 0.518     | 174       | 0.581     | 1,026    | 0.540     |
| <i>FXIIIIB</i> | 246          | 0.468     | 164          | 0.518     | 144        | 0.528     | 128        | 0.555     | 168        | 0.441     | 172       | 0.488     | 1,022    | 0.494     |
| <i>APOA1</i>   | 246          | 0.951     | 164          | 0.982     | 144        | 0.972     | 128        | 0.953     | 168        | 0.964     | 170       | 0.977     | 1,020    | 0.966     |
| <i>B65</i>     | 246          | 0.553     | 164          | 0.598     | 144        | 0.583     | 128        | 0.438     | 168        | 0.530     | 174       | 0.569     | 1,024    | 0.549     |
| <i>D1</i>      | 246          | 0.394     | 164          | 0.366     | 144        | 0.389     | 128        | 0.281     | 170        | 0.412     | 172       | 0.424     | 1,024    | 0.383     |
| <i>A25</i>     | 246          | 0.122     | 164          | 0.109     | 144        | 0.097     | 128        | 0.047     | 168        | 0.077     | 174       | 0.115     | 1,024    | 0.099     |
| <i>HS4.65</i>  | 246          | 0.037     | 164          | 0.018     | 142        | 0.035     | 128        | 0.047     | 168        | 0.012     | 164       | 0.037     | 1,012    | 0.031     |
| <i>HS3.23</i>  | 240          | 0.871     | 164          | 0.872     | 144        | 0.729     | 128        | 0.789     | 170        | 0.877     | 174       | 0.833     | 1,020    | 0.835     |
| <i>HS2.43</i>  | 246          | 0.065     | 164          | 0.061     | 144        | 0.104     | 128        | 0.117     | 170        | 0.117     | 174       | 0.086     | 1,026    | 0.089     |
| <i>CD4del</i>  | 244          | 0.344     | 164          | 0.268     | 144        | 0.368     | 124        | 0.331     | 164        | 0.366     | 174       | 0.362     | 1,014    | 0.340     |

**Table 2** Heterozygosities and genetic differentiation indices for individual and for all loci (considered jointly). Moldavians: *K* Karahasani, *S* Sofia; Gagauzes: *K* Kongaz, *E* Etulia

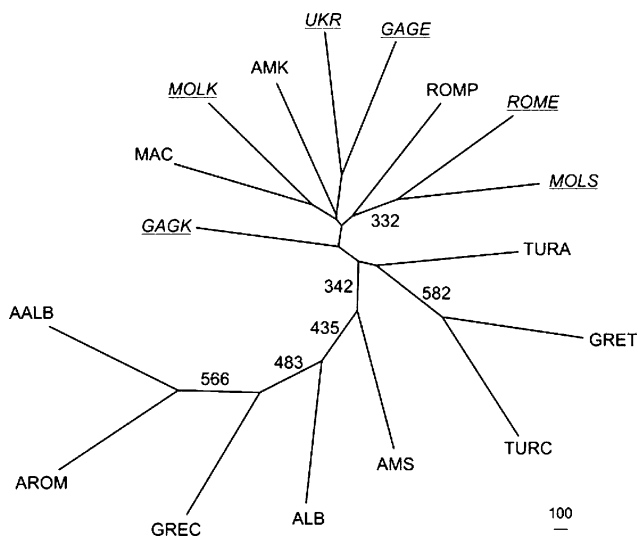
| Locus           | Heterozygosities |              |            |            |            |           |         | Genetic differentiation $F_{ST}$ (%) |
|-----------------|------------------|--------------|------------|------------|------------|-----------|---------|--------------------------------------|
|                 | Moldavians K     | Moldavians S | Gagauzes K | Gagauzes E | Ukrainians | Romanians | Average |                                      |
| <i>ACE</i>      | 0.508            | 0.512        | 0.514      | 0.516      | 0.381      | 0.437     | 0.478   | 0.15 NS <sup>a</sup>                 |
| <i>PV92</i>     | 0.388            | 0.439        | 0.292      | 0.359      | 0.302      | 0.361     | 0.360   | 0                                    |
| <i>TPA25</i>    | 0.504            | 0.439        | 0.528      | 0.531      | 0.488      | 0.448     | 0.488   | 1.37*                                |
| <i>FXIIIIB</i>  | 0.512            | 0.500        | 0.500      | 0.484      | 0.548      | 0.535     | 0.515   | 0.07 NS                              |
| <i>APOA1</i>    | 0.081            | 0.037        | 0.056      | 0.063      | 0.071      | 0.047     | 0.061   | 0                                    |
| <i>B65</i>      | 0.504            | 0.463        | 0.500      | 0.531      | 0.577      | 0.540     | 0.519   | 0.50 NS                              |
| <i>D1</i>       | 0.447            | 0.463        | 0.528      | 0.469      | 0.337      | 0.407     | 0.439   | 0.33 NS                              |
| <i>A25</i>      | 0.228            | 0.195        | 0.167      | 0.094      | 0.129      | 0.230     | 0.181   | 1.74*                                |
| <i>HS4.65</i>   | 0.073            | 0.037        | 0.042      | 0.094      | 0.024      | 0.072     | 0.057   | 0.23 NS                              |
| <i>HS3.23</i>   | 0.225            | 0.232        | 0.403      | 0.297      | 0.221      | 0.241     | 0.262   | 0                                    |
| <i>HS2.43</i>   | 0.114            | 0.122        | 0.125      | 0.172      | 0.165      | 0.163     | 0.141   | 0.21 NS                              |
| <i>CD4del</i>   | 0.508            | 0.390        | 0.514      | 0.403      | 0.410      | 0.471     | 0.455   | 0.02 NS                              |
| <i>All loci</i> | 0.341            | 0.319        | 0.347      | 0.334      | 0.304      | 0.329     | 0.330   | 0.38*                                |

<sup>a</sup> Non-significant

\*  $P < 0.01$

(Stoneking et al. 1997; Romualdi et al. 2002; Comas et al. 2004) and our own data. Figure 2 shows the most probable (consensus) tree from the 1,000 trees generated by resampling. In the consensus tree, the compared populations do not constitute strongly pronounced groups. The low bootstrap support for the tree topology suggests the absence of considerable genetic barriers within southeastern Europe. However, the bootstrap is known to underestimate the true level of statistical support (Sitnikova et al. 1995). From simple visual inspection of the tree, a major distinction between North and South appears. The results of PC analysis confirmed the pattern observed in the consensus tree (Fig. 3). Thus, the first PC, which explains 24% of the variation in allele frequencies, tends to separate the southern and western Mediterranean

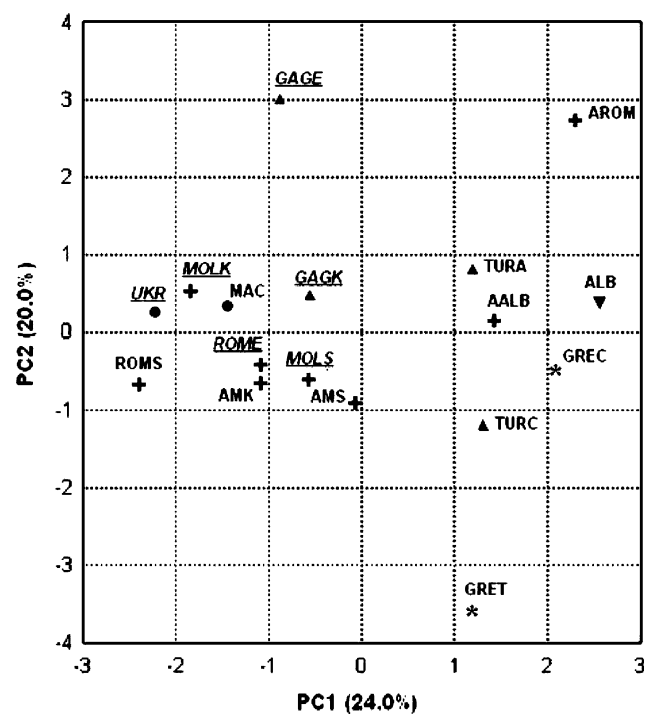
(Turkish Cypriots, Greek Cypriots, Turks, northeastern Greeks, Albanians, Albanian Aromuns) from the northern and Balkan–Carpathian populations (Macedonians, Macedonian Aromuns, Romanians, Moldavians, Ukrainians, Gagauzes). High negative correlations of the first PC with geographical latitude (Spearman’s  $R = -0.6887$ ;  $P = 0.00233$ ) and distance from the Ukrainian settlement (Spearman’s  $R = -0.6078$ ;  $P = 0.00964$ ) provide quantitative evidence for the observations revealed in the tree and PC plot. Along the second principle axis, which explains 20% of the total genetic variance, the Gagauzes from Etulia and the Romanian Aromuns stand apart from the rest of the populations at the positive pole and the north-eastern Greeks at the negative pole. No correspondence between the linguistic affiliation of the



**Fig. 2** Consensus tree depicting the relationships among the southeastern European populations analyzed for 11 *Alu* polymorphisms. Numbers on the branches are bootstrap values based on 1,000 replications. *GAGE* Gagauzes from Etulia, *GAGK* Gagauzes from Kongaz, *MOLK* Moldavians from Karahasani, *MOLS* Moldavians from Sofia, *ROME* Romanians from Piatra-Neamt and Buhusi, *UKR* Ukrainians, *AALB* Albanian Aromuns, *ALB* Albanians, *AMK* Macedonian Aromuns from Krusevo, *AMS* Macedonian Aromuns from Stip, *AROM* Romanian Aromuns, *GRET* Greeks from Thrace, *MAC* Macedonians, *ROMP* Romanians from Ploiesti, *TURA* Turks from Anatolia (Comas et al. 2004), *TURC* Turkish Cypriots, *GREC* Greek Cypriots (Stoneking et al. 1997; Romualdi et al. 2002). The populations examined in the present study are *underlined* and in *italics*

compared populations and genetic differentiation is observed in the plot.

Since we are particularly interested in the origin of the Gagauzes, we assessed the genetic relationship of the Gagauzes and Turkic populations from Central Asia. To determine the genetic relationship of the southeastern European populations and Central Asian populations, we used the information on eight polymorphic *Alu* loci (*ACE*, *TPA25*, *PV92*, *APOA1*, *FXIII B*, *A25*, *B65*, *D1*) previously published for the Uyghurs (Xiao et al. 2002), Uzbeks, Kazakhs, and northern and southern Kyrgyzes (Khitrinskaya et al. 2003). These 8 loci are a subset of the 12 loci we typed (data from the others were not available in the literature). The topology of the consensus tree (Fig. 4) generally reflects the racial classification of populations (Alexeev 1974). The Kyrgyzes and the Kazakhs, which are assigned to the Mongoloid race, cluster together in the tree with considerable distance to the European populations. The bootstrap values observed within the European population cluster are very small and neither geographic nor linguistic relationships are observed between the European samples in the tree, suggesting

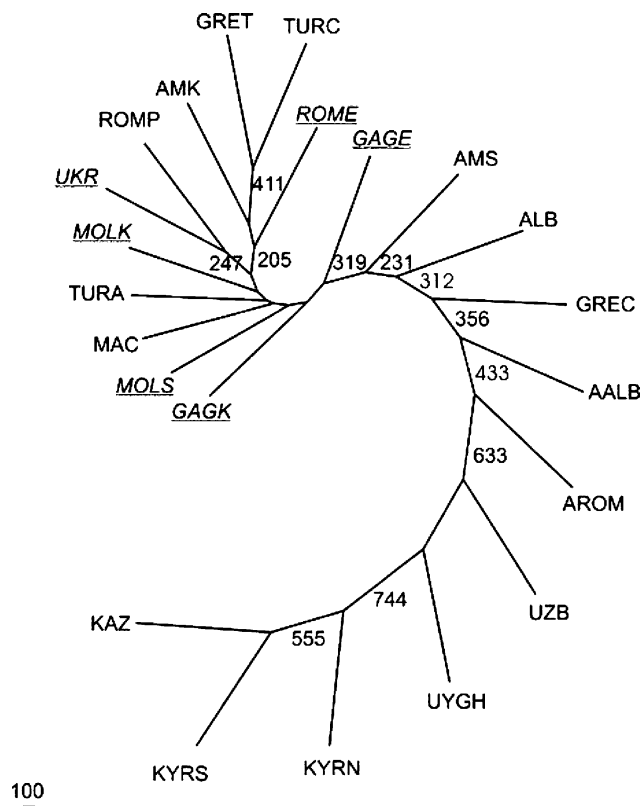


**Fig. 3** Genetic affinities among 17 southeastern European populations based on the first two principle components (PCs) of allele frequencies at 11 *Alu* loci. Population codes as in Fig. 2. The populations examined in the present study are *underlined* and in *italics*. Symbols on the PC plot represent linguistic classification of the samples: Plus signs Italic, filled circles Slavic, filled triangles Turkic, inverted filled triangles Albanian, asterisks Greek

that information based on the eight *Alu* polymorphic loci was insufficient to resolve the relationship between these geographically close populations. The Uzbeks and the Uyghurs, who are considered as a mixed Mongoloid–Caucasoid population, occupy an intermediate position in the tree. The nodes separating Uzbeks and Uyghurs from the Mongoloid and Caucasoid clusters show relatively strong bootstrap support after 1,000 iterations. Both Gagauz samples are grouped together with the European samples.

#### Genetic differentiation between populations

In order to assess interpopulation variability within the Dniester–Carpathian region, an AMOVA was performed (Table 2). The contribution of individual loci to the interpopulation variability of the region was low. For two loci (*TPA25* and *HS3.23*), the values of the differentiation index ( $F_{ST}$ ) were significantly different from zero. The  $F_{ST}$  for all loci means that only 0.38% of the total variance in allele frequencies at these loci is due to differences between the populations, while the rest is due to differences within the populations. Although this



**Fig. 4** Consensus tree of southeastern European and central Asian populations analyzed for eight *Alu* polymorphisms. Numbers on the branches are bootstrap values based on 1,000 replications. Codes for the southeastern European populations are as in Figs. 2 and 3; those for the other populations are as follows: *KAZ* Kazakhs, *KYRN* northern Kyrgyzes, *KYRS* southern Kyrgyzes, *UZB* Uzbeks (Khitrinskaya et al. 2003), *UYGH* Uyghurs (Xiao et al. 2002). The populations examined in the present study are underlined and in *italics*

value implies a very low level of population subdivision, our  $F_{ST}$  analysis suggests significant population differentiation within the Dniester–Carpathian region.

We have also analyzed genetic differentiation within Southeast Europe using the same set of populations

and loci as in the phylogenetic analyses. Since only allele frequency data was available from the literature, we had to resort to Nei’s (1973)  $G_{ST}$  approach. Within Southeast Europe, the fraction of the genetic variance attributable to differences among populations ( $G_{ST}$ ) was 1.61%, indicating that the level of genetic differentiation within Southeast Europe was two times lower than the level within the whole of Europe (see Kutuev et al. 2006).

When Dniester–Carpathian populations were divided into three groups defined by language (Romanian, Gagauz and Ukrainian), no significant difference was observed between population groups (Table 3). When we extended the analysis of genetic differentiation to Southeast Europe, the component of genetic variance due to differences among linguistic groups (defined as in Fig. 3) was very low (0.50%)—even lower than the component due to differences among populations within groups (1.11%). These findings suggest that language does not explain the genetic affinities among the Dniester–Carpathian and south-eastern European populations.

**Discussion**

**Pattern of *Alu* insertion variation in southeastern Europe**

Previous analyses based on classical polymorphic markers (Cavalli-Sforza et al. 1994), autosomal DNA polymorphisms (Nasidze et al. 2001; Jorde and Wooding 2004; Tishkoff and Kidd 2004) and mitochondrial DNA (mtDNA) (Calafell et al. 1996) have revealed that Europe is a genetically homogeneous continent. This conclusion is supported by two lines of evidence. First, by small differentiation indexes: the  $F_{ST}$  value for Europe is 2–7 times lower than in other Continents and geographical areas. Second, by small genetic distances:

**Table 3** Components of genetic variance (%) at three levels of population subdivision; populations were pooled according to their affiliation to linguistic groups

| Source of variation             | Southeast Europe <sup>a</sup> | Dniester–Carpathian region <sup>b</sup> |                         |
|---------------------------------|-------------------------------|---|-------------------------|
|                                 | Nei (1987)                    | Nei (1987)                              | Excoffier et al. (1992) |
| Among groups                    | 0.50                          | 0.35                                    | 0.20 NS <sup>c</sup>    |
| Among populations within groups | 1.11                          | 0.49                                    | 0.24 NS                 |
| Within populations              | 98.39                         | 99.16                                   | 99.56                   |

<sup>a</sup> Calculated on the basis of the data on 11 insertion frequencies from the present study, Stoneking et al. (1997), Romualdi et al. (2002), and Comas et al. (2004). The 17 populations were pooled into five groups as defined in Fig. 3

<sup>b</sup> Calculated on the basis of the data on 12 markers from the present study. Populations were pooled into three groups: Gagauz, Romanian and Ukrainian

<sup>c</sup> Non-significant

in a neighbor-joining tree of the world populations, European populations cluster in a small compact group, while other populations are connected to each other by much longer branches. The Dniester–Carpathian autosomal pool also follows this pattern. Our analysis of 12 autosomal DNA polymorphisms in the Dniester–Carpathian region has shown that the allele frequencies in these populations are similar to each other and to the frequencies observed in other European populations, despite considerable linguistic differences. The genetic homogeneity among southeastern European populations suggests either a recent common ancestry of all southeastern European populations or strong gene flow between populations, which eliminated any initial differences. Taking into account that the region has had a relatively high population density since the Neolithic period and that this region represents a crossroads of routes connecting the cultural centers of Middle East with different European areas, both explanations are plausible.

Despite the apparent low level of genetic differentiation of the southeastern European gene pool and the lack of correlation between linguistic and genetic geographic patterns, we demonstrated some interesting aspects of population structure. The first PC, which explains 24% of the total genetic diversity, is correlated with geographical latitude. The observed pattern of genetic differentiation within southeastern Europe is not surprising. Our results are consistent with those from classical and DNA markers (Cavalli-Sforza et al. 1994; Malaspina et al. 2001) and are also compatible with archaeological and paleoanthropological data. Since the Neolithic (7,500 BC) the eastern Mediterranean area has been a field of constant presence of agricultural communities. These arose from the common Neolithic ‘package’ originating in the Near East (Renfrew 1987). The demographical process in the northern part of Southeast Europe was different from that in the southern part. The Balkan–Mediterranean farming traditions developed here during the Neolithic—Early Eneolithic period (6,500–4,000 BC). Beginning from the Late Eneolithic the nomadic tribes of Kurgan cultures penetrated into the Carpathian basin and the Balkans from the Pontic steppes. These cultures developed on an East European Mesolithic basis (Dergachev 1999). The considerable differences in a set of morphological characters between farming tribes from Southeast Europe and the Mesolithic and nomadic tribes from East Europe (Velicanova 1975) suggest different structures of their respective gene pools. The genetic differences between northern and southern populations of Southeast Europe observed in our work seem to be due to the unequal proportion of

the European (‘Mesolithic’) and Near-Eastern (‘Neolithic’) components in their gene pools. The second PC, which does not appear to exhibit any geographical pattern, has no obvious interpretation.

#### *Alu* insertion polymorphisms and the origins of the Gagauzes

Are the Gagauzes descendants of the Turkic nomadic tribes from the South Russian steppe (Uzi, or Pechenegs, or Polovtsi, etc.) or descendants of the Anatolian Turks (Seljuks and/or Ottomans)? In the first case, the Gagauzes should be genetically more similar to some Turkic populations (from the Eurasian Heartlands), while in the second case they should resemble populations from Anatolia. Our previous analysis of classical polymorphisms in the Dniester–Carpathian region demonstrated that Gagauzes group genetically with their geographic neighbors, rather than with any Turkic populations (Varsahr et al. 2001, 2003). The present analysis, based on autosomal DNA markers, is consistent with our previous results. The Gagauz samples showed considerable genetic distances to Central Asian populations. Furthermore, the genetic position of the Gagauzes in the tree was not intermediate between southeastern European and Central Asian populations. Therefore, our data reject the hypothesis that the Gagauzes are direct biological descendants of the Turkic nomads from the South Russian steppes.

According to the other scenario, the Gagauzes are descendants of the Seljuk Turks who migrated to the Balkans from Anatolia in the second half of the thirteenth century. Although the Gagauzes show closer relationships with the Dniester–Carpathian populations than with the Turks from Anatolia and Cyprus, it should be noted that the differences between the populations mentioned above are not sufficiently large to rule out the hypothesis of a Seljuk origin of the Gagauzes. A problem with this scenario, however, is that it does not explain the presence of the Kypchak (Tartar) element in the Gagauz language, which could only have entered by the northern route from the Eurasian steppes.

The lack of correlation between the linguistic and genetic differentiation in Southeast Europe (in particular in the Dniester–Carpathian region) suggests that ethnic and genetic differentiation occurred here relatively independently of each other. The genetic landscape of Southeast Europe had presumably been formed long before the linguistic/ethnic landscape we now observe was shaped. One other possibility is that the cultural barriers were not strong enough to prevent gene flow between populations. The Turkic language

of the Gagauzes could be a case of language replacement. Replacement could occur via the ‘elite dominance’ model, which means that the original Turkic migrant groups could have been very small such that their genetic effect on the resident groups was negligible (Renfrew 1987). However, the elite dominance scenario is more suitable for larger populations, such as those of Anatolia or Azerbaijan, which consist of 70 million and 8 million people, respectively (Nasidze et al. 2001; Cinnioglu et al. 2004). The Gagauzes are much less numerous (~200,000). Thus, it is still possible that they are a remnant of a once larger Turk-speaking Orthodox group in southeastern Europe.

## Conclusion

Our study of *Alu* polymorphisms indicates low levels of population differentiation in the Dniester–Carpathian region and in Southeast Europe. Although the interpopulation differentiation within Southeast Europe is small, tree reconstruction and PC analysis allowed a distinction between southern and northern populations. These observations are in agreement with classical markers and are also compatible with archaeological and paleoanthropological data. The genetic affinities among Dniester–Carpathian and southeastern European populations do not reflect linguistic relationships. Overall, these results indicate that ethnic and genetic differentiation occurred in these regions to a considerable extent independently of each other. Thus, based on our dataset of 12 *Alu* markers, we have accomplished the goals we set out to achieve in this project. Nonetheless, it would be desirable to collect more data to verify the conclusions obtained in this study. In particular, other marker systems with a pronounced ethnic specificity (e.g. Y-chromosome and mtDNA markers) may be useful, and more populations may need to be sampled.

**Acknowledgments** We are grateful to all donors for providing blood samples and to the people who contributed to their collection. We also thank B. Joffe, B. Nürnberger and the anonymous reviewers for helpful comments and suggestions. E.W. was supported by FCI and DFG, W. S. by the Volkswagen-Foundation and DFG, and A.V. by the DAAD. V.S. and M.S. were supported by the Russian Federation for basic research (grant 03-04-4902) and by a grant from the President of the Russian Federation (grant MD-88.2003.04).

## References

- Alexeev VP (1974) Geography of human races (in Russian). Nauka, Moscow
- Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha PS, Hammer M, Yunis JJ, Yunis EJ, Damodaran Ch, Martinez de Pancorbo M, Caeiro HL, Puzyrev VP, Herrera RJ (2002) Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 23:3346–3356
- Arcot SS, Fontius JJ, Deininger PL, Batzer MA (1995a) Identification and analysis of a ‘young’ polymorphic *Alu* element. *Biochim Biophys Acta* 1263:99–102
- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995b) *Alu* repeats: a source for the genesis of primate microsatellites. *Genomics* 29:136–144
- Batzer MA, Arcot SS, Phinney JM, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M (1996) Genetic variation of recent *Alu* insertions in human populations. *J Mol Evol* 42:22–29
- Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60:35–49
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Cinnioglu 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
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A (2000) *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312–319
- Comas D, Schmid H, Braeuer S, Flaiz C, Busquets A, Calafell F, Bertranpetit J, Scheil H-G, Huckenbeck W, Efmovska L, Schmidt HD (2004) *Alu* insertion polymorphisms in the Balkans and the origins of the Aromuns. *Ann Hum Genet* 68:120–127
- Dergachev VA (1999) The particularities of the cultural development of the region between the Carpathians and Dniester (in Russian). *Stratum plus* 2:169–221
- Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Felsenstein J (1993) PHYLIP, version 3.5. University of Washington Press, Seattle
- Guboglo MN (1967) The ethnic affiliation of the Gagauzes (in Russian). *Sov Etnograf* 3:160–167
- Guo S, Thompson E (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Jorde LB, Wooding SP (2004) Genetic variation, classification and ‘race’. *Nat Genet* 36:28–33
- Khitrinskaya IYu, Stepanov VA, Puzyrev VP, Spiridonova MG, Voevoda MI (2003) Genetic differentiation of the population of Central Asia inferred from autosomal markers. *Russ J Genet* 39:1175–1183
- Kutuev I, Khusainova R, Karunas A, Yunusbayev B, Fedorova S, Lebedev Y, Hunsmann G, Khusnutdinova E (2006) From East to West: patterns of genetic diversity of populations living in four Eurasian regions. *Hum Hered* 61:1–9
- Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK (1999) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet* 7:435–446



- Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, Plata E, Kozlov AI, Barbujani G, Vernesi C, Papola F, Ciavarella G, Kovatchev D, Kerimova MG, Anagnou N, Gavrilu L, Veneziano L, Akar N, Loutradis A, Michalodimitrakis EN, Terrenato L, Novelletto A (2001) A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. *Ann Hum Genet* 65:339–349
- 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, 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
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Renfrew C (1987) *Archaeology and language: the puzzle of Indo-European origins*. Cape, London
- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12:602–612
- Schneider S, Roessli D, Excoffier L (2000) *Arlequin Ver 2000* A software for population genetics data analysis. Genetics and biometry laboratory. University of Geneva, Switzerland
- Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. *Mol Biol Evol* 12:319–333
- StatSoft (1995) *STATISTICA for Windows* [Computer program manual]. Tulsa, OK
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7:1061–1071
- Tishkoff SA, Kidd KK (2004) Implications of biogeography of human populations for ‘race’ and medicine. *Nat Genet* 36:21–27
- 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
- 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, Scheil H-G, Schmidt HD (2006) Blood group and serum protein polymorphisms in a population group of Moldavians. *Anthropol Anz* 64:51–58
- Velicanova MS (1975) *Paleoanthropology of the Dniester–Pruth interfluvial* (in Russian). Nauka, Moscow
- Xiao FX, Yang JF, Cassiman JJ, Decorte R (2002) Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China. *Hum Biol* 74:555–568
- Zhivotovsky LA (1991) *Population Biometrics* (in Russian). Nauka, Moscow