# Population History of the Dniester-Carpathians: Evidence from *Alu* Insertion and Y-Chromosome Polymorphisms

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München


vorgelegt von
**Alexander Varzari**
aus Moldawien
27.07.2006

Erster Gutachter: Prof. Dr. Elisabeth Weiß
Zweiter Gutachter: Prof. Dr. Wolfgang Stephan
Sondergutachter: Prof. Dr. Vadim Stepanov

Tag der mündlichen Prüfung: 12.09.2006

Prüfungskommission:
Prof. Dr. Elisabeth Weiß
Prof. Dr. Wolfgang Stephan
Prof. Dr. Gisela Grupe
Prof. Dr. Thomas Cremer

*In memory of my mother, Varzari Lilia Ivanovna*

TABLE OF CONTENTS

# 1 SUMMARY

The Dniester-Carpathian region has attracted much attention from historians, linguists, and anthropologists, but remains insufficiently studied genetically. We have analyzed a set of autosomal polymorphic loci and Y-chromosome markers in six autochthonous Dniester-Carpathian population groups: 2 Moldavian, 1 Romanian, 1 Ukrainian and 2 Gagauz populations. To gain insight into the population history of the region, the data obtained in this study were compared with corresponding data for other populations of Western Eurasia.

The analysis of 12 *Alu* human-specific polymorphisms in 513 individuals from the Dniester-Carpathian region showed a high degree of homogeneity among Dniester-Carpathian as well as southeastern European populations. The observed homogeneity suggests either a common ancestry of all southeastern European populations or a strong gene flow between them. Nevertheless, tree reconstruction and principle component analyses allow the distinction between Balkan-Carpathian (Macedonians, Romanians, Moldavians, Ukrainians and Gagauzes) and Eastern Mediterranean (Turks, Greeks and Albanians) population groups. These results are consistent with those from classical and other DNA markers and are compatible with archaeological and paleoanthropological data.

Haplotypes constructed from Y-chromosome markers were used to trace the paternal origin of the Dniester-Carpathian populations. A set of 32 binary and 7 STR Y-chromosome polymorphisms was genotyped in 322 Dniester-Carpathian Y-chromosomes. On this basis, 21 stable haplogroups and 171 combination binary marker/STR haplotypes were identified. The haplogroups E3b1, G, J1, J2, I1b, R1a1, and R1b3, most common in the Dniester-Carpathian region, are also common in European and Near Eastern populations. Ukrainians and southeastern Moldavians show a high proportion of eastern European lineages, while Romanians and northern Moldavians demonstrate a high proportion of western Balkan lineages. The Gagauzes harbor a conspicuous proportion of lineages of Near Eastern origin, comparable to that in Balkan populations. In general, the Dniester-Carpathian populations demonstrate the closest affinities to the neighboring southeastern and eastern European populations. The expansion times were estimated for 4 haplogroups (E3b1, I1b, R1a1, and R1b3) from associated STR diversity. The presence in

the studied area of genetic components of different age indicates successive waves of migration from diverse source areas of Western Eurasia.

Neither of the genetic systems used in this study revealed any correspondence between genetic and linguistic patterns in the Dniester-Carpathian region or in Southeastern Europe, a fact which suggests either that the ethnic differentiation in these regions was indeed very recent or that the linguistic and other social barriers were not strong enough to prevent genetic flow between populations. In particular, Gagauzes, a Turkic speaking population, show closer affinities not to other Turkic peoples, but to their geographical neighbors.

# **2** INTRODUCTION

## 2.1 Molecular DNA markers in human populations

### 2.1.1 An overview of DNA markers

People have always been curious about their history. They were deeply interested in issues such as ancestry and the original motherland of mankind, the basis and the dynamics of the morphological diversity, the geographic and the chronological aspects of ethnic differentiation. These questions have always been addressed by experts from various fields, and biologists often played a notable and sometimes a decisive role in deciphering our population histories.

Early the human evolution was studied at the morphological level by means of detailed descriptions and the measurement of various excavation finds of ancient man, as well as comparing and correlating hundreds of populations in various regions of the globe. However, the fossil record is spotty, and the morphological variation often affected by environment. Genetic data offer another way of viewing human evolution.

The pattern of genetic variation in modern human populations depends on our demographic history (including population migrations, bottlenecks and expansions) as well as gene specific factors such as mutation and recombination rates and selection pressure. By examining patterns of genetic polymorphisms we can infer how past demographic events and selection have shaped variation in the genome. Thus, the study of human genetic variation has important implication for evolutionary biology.

Until recently, evolutionary studies were limited by a paucity of useful genetic markers. These were based on the analysis of protein polymorphisms, which are usually referred to as ‘classical polymorphisms’ to distinguish them from those obtained by DNA testing. The large scale population studies of blood group and protein polymorphisms demonstrated that the gene pool is not a simple sum of genes, which are common in the population, but is a dynamic system, which is hierarchally organized and which maintains the memory of past events in the history of populations (Mourant *et al.* 1976; Nei and Roychoudhury 1988; Cavalli-Sforza *et al.* 1994; Walter 1997; Rychkov *et al.* 2000; Altukhov *et al.* 1996).

In the beginning of 1980, after the discovery of DNA polymorphism (Kan and Dozy 1978), a new class of genetic markers appeared due to the progress in gene cloning, and the availability of restriction enzymes. The advantages of analyzing genetic polymorphisms at the DNA level, rather than that of gene products, are manifold. Since the majority of the

genome does not take part in known gene functions (Kass and Batzer 2001), the corresponding non-coding DNA exhibit polymorphisms that outnumber by far the known protein variability (Nei 1987).

DNA polymorphisms were first studied by Southern analysis of DNA digested with restriction enzymes. At present over several hundred restriction enzymes are available. This type of polymorphism is called restriction fragments length polymorphism (RFLP), as alleles differ in the length of the restriction fragments obtained upon digestion. The most common reason of RFLPs is a nucleotide replacement in the recognition site, infrequently a loss or addition of one nucleotide. This type of polymorphism is called SNP (single nucleotide polymorphism). SNPs constitute the great majority of variations in the human genome. According to Tishkoff and Kidd (2004) the human genome contains approximately 4.5 million validated SNPs. At present, due to the improvement and automation of sequencing procedure, and the development of DNA microarrays (Gibson 2002), these markers are extensively studied in the human genome for their association with different complex diseases (Cargill *et al.* 1999; Halushka *et al.* 1999; Tishkoff and Kidd 2004), for understanding various aspects of population differentiation and evolution of humans (Przeworski *et al.* 2000; Jorde and Wooding 2004; Tishkoff and Kidd 2004).

Since the pioneering studies of Bowock *et al*. (1994), special attention has been paid to polymorphisms of repeated sequences. Repetitive sequence elements are distributed over almost the entire genome, and they are subdivided into tandemly arrayed (for example minisatellites and telomere repeats) or interspersed (for example *Alu* repeat) repetitive sequences (Weiner *et al.* 1986; Kass and Batzer 2001; Nikitina and Nazarenko 2004; Grover *et al.* 2005). The attention of researchers is focused on minisatellites consisting of repeated copies (motif) of nine or ten to hundred base pairs each and microsatellites, whose copies are typically two to four, sometimes six nucleotides in length. Microsatellites are also called STRs (short tandem repeats). Minisatellites and microsatellites can be highly variable and thus are excellent tools for genetic individualization. These loci are characterized by rapid evolution. Spontaneous mutation rates of mini- and microsatellite loci are on average several orders of magnitude higher than in the remaining DNA (Weber and Wong 1993), which allows for direct estimation of evolutionary transformation rate in genomic nucleotide sequences (Zhivotovsky *et al.* 2003). Interspersed repeated DNA sequences can be divided into two classes: short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs) The most extensively studied class of SINEs

are *Alu* insertions due their abundance (genomic coverage ~11% in human genome) as well as their association with many biological functions (Batzer and Deininger 2002).

DNA analysis facilitates the study of haplotypes, arrays of alleles at closely linked loci along a chromosome. These regions are short enough to show very little or no recombination and behave as blocks every of which has a single unique genealogical history. Mitochondrial (mtDNA) and Y chromosomal DNA serve as vivid example of such arrayed polymorphisms. The mitochondrial genome offers a large perspective on human evolution (Wallace 1995). Because mtDNA is inherited through the maternal cytoplasm, variation in mtDNA provides a record of the maternal lineages of our species. Whereas Y chromosome DNA (except the recombining pseudoautosomal regions) documents the paternal lineage (Jobling and Tyler-Smith 2003; Lell and Wallace 2000).

Many additional types of polymorphism can be studied at the DNA level. The selection of the genetic markers for a concrete research is determined by the ability of the given marker to solve the tasks and by the technical support. In this chapter we shall dwell on two genetic marker systems, selected for this work, by describing their genetic nature, advantages and limitations for their use to analyze the structure and the evolution of the populations.

### 2.1.2 The mobile genetic element *Alu* in the human genome

*Alu* insertional elements represent the largest family of SINEs in humans. They are named due to the presence of an *Alu*I recognition site in the sequence (Houck *et al.* 1979). The human genome contains about 1,100,000 *Alu* repeats, which account for ~11% of the total nuclear DNA (Lander *et al.* 2001). Like other SINEs, *Alu* repeats are often located in non-coding regions (intergenic spacers, introns) (Batzer *et al.* 1990). *Alu* insertions are of approximately 300 bp in length, dimeric in structure, and composed of two nearly identical monomers joined by a middle A-rich region along with a 3’ oligo(dA)-rich tail and short flanking direct repeats (see Figure 2.1) (Economou *et al*. 1990; Novick *et al.* 1996; Rowold and Herrera 2000). The left monomer contains two promoter elements for RNA polymerase III, blocks A and B, which are about 10 bp each (Jurka and Zuckerkandl 1991).

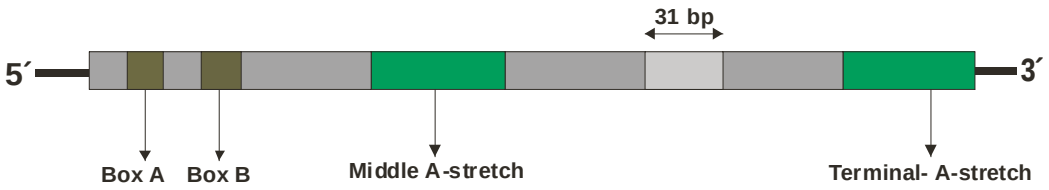


**Figure 2.1** The dimeric structure of the *Alu* element. The two halves are linked by an adenine-rich area. The right monomer includes a 31-base pair insertion, and the left half contains the RNA polymerase III promoter (boxes A and B). The total length of each *Alu* sequence is ~300 bp, depending on the length of the 3' oligo(dA)-rich tail.

Based on sequence homology, *Alu* elements are considered to originate from 7SL RNA (Ullu and Tschudi 1984). The origin of the fossil *Alu* monomer (FAM) can be traced back to the very beginning of the mammalian radiation (~112 *mya*) (Kapitonov and Jurka 1995). The ancestral dimeric *Alu* sequence originated from a head to tail fusion of two distinct forms of the fossil Alu monomer (Quentin 1992), linked by an oligo(dA) tract. The fusion of two monomers occurred after the Rodentia line was branched from Primates approximately 100 mya, but before the primate radiation approximately 65 *mya* (Kapitonov and Jurka 1995). Subsequently, throughout primate evolution the number of mutations has accumulated resulting in a hierarchical subfamily structure, or lineage, of *Alu* repeats (Batzer *et al.* 1996; Kapitonov and Jurka 1996). The youngest subfamilies Ya (also known as HS/PV or human specific/predicted variant) and Yb8 (also known as Sb2) have integrated into the human genome in the past 4-5 million years after the divergence of humans and African apes (Arcot *et al.* 1996; Kapitonov and Jurka 1996; Batzer *et al.* 1996; Roy-Engel *et al.* 2001). It has been estimated that the Ya5/8 and Yb8 subfamilies comprise 500-2000 and 500 members respectively within the human genome (Arcot *et al.* 1996; Batzer *et al.* 1996; Stoneking *et al.* 1997). Approximately 25% of the young Ya5/8 and Yb8 *Alu* elements have retrotransposed so recently that the corresponding loci are polymorphic for the presence/absence of the *Alu* sequence. These insertions have presumably occurred after the arising of the modern humans about 150,000 years ago (Stoneking *et al.* 1997).

*Alu* elements increase in number by retrotransposition – a process that involves reverse transcription of an Alu-derived RNA polymerase III transcript (Novick *et al.* 1996; Batzer and Deininger *et al.* 2002). The mechanisms for the amplification of *Alu* elements require the presence of two enzymes – reverse transcriptase and endonuclease. Since *Alu* elements do not encode these enzymes, they are probably derived from long interspersed elements (LINEs) (Mathias *et al.* 1991). Although *Alu* elements have a functional internal RNA

polymerase III promoter, most *Alu* copies are transcriptional silent. Host sequences upstream of the promoter have been found to be important for *in vivo* expression (Ullu and Weiner 1985; Batzer and Deininger 2002) unless inserted into favorable genomic locations. In addition, due to their CpG content *Alu* elements are especially susceptible to transcriptional silencing by methylation (Batzer *et al.* 1990). Methylation of CpG motifs both nearby and within *Alu* insertions could minimize or eliminate their retrotransposition capability, since transcription factors are unable to bind to methylated promoter elements (Deininger and Batzer 1993; Schmid and Maraia 1992). *Alu* amplification rate is highly variable, with periods of high and low amplification rates. The *Alu* amplification peak was observed around 35 million years ago (Shen *et al.* 1991; Britten RJ 1994). The expansion rate estimated for that time was approximately one new *Alu* insertion in every primate birth. Presently *Alu* elements amplify at a rate 100-200 folds lower (Deininger and Batzer 1999).

Because the abundance of *Alu* repeats in primate genomes and a high degree of sequence similarity among members of this repeat family they might act as nucleation points for unequal homologous recombination (Deininger and Batzer 1999). These recombination events result in the deletion, duplication or translocation of chromosomal segments.

Since the *Alu* repeats affect the composition, organization and expression of the genome, they play a significant role in the occurrence of human genetic diseases. Pathological disorders due to *Alu* insertions can be divided into three classes: disorders caused by retroposition, disorders caused by recombination and disorders caused by exonisation (for review see Deininger and Batzer 1999; Grover *et al.* 2005). *Alu* insertion in primate genome speeds up the rate of gene evolution by generating new proteins that can take up new functions, and by acquiring important regulatory elements. Across all evolutionary time frames *Alu*-mediated recombination led to genetic exchanges and shuffling which, coupled with natural selection, influenced the evolution of the functional genome and thereby contributed to speciation (Batzer and Deininger 2002; Grover *et al.* 2005).

*Alu* repeats are convenient genetic markers. First, the insertion of an *Alu* element at a certain chromosomal site is most probably a unique event in evolutionary history, in other words, the individuals that share *Alu* insertion polymorphisms have inherited the *Alu* elements from a common ancestor, which makes the *Alu* insertion alleles identical by descent. In contrast, other DNA markers like STRs or SNPs are not identical by descent. The same allele may have arisen several times during human evolution. Second, they are

stable polymorphisms - once inserted, the elements are fixed in the genome, as there does not exist any specific mechanism for removing them from the genome. Even when a rare deletion occurs, a significant remnant is left behind, since an exact excision of an insertion is most improbable. And third, the ancestral state of the *Alu* insertion is known to be the absence of the insertion. Polymorphic *Alu* elements are human specific and absent in non-human primates. It is possible to create a hypothetical ancestral population with frequencies of zero for all human specific *Alu* insertions used as DNA markers. The knowledge about the hypothetical ancestral population enables to root phylogenetic trees. The possibility of rooting a tree supplies more information about the origin of human populations. The previous finding that the root of population tree is located near the African Sub-Saharan populations presented evidence for an African origin of modern human populations (Batzer *et al*. 1994; Stoneking *et al.* 1997). Moreover, the populations from Australia and New Guinea are also close to the hypothetical ancestral population, possibly indicating an early expansion of human populations in the tropics (Batzer *et al*. 1994; Stoneking *et al.* 1997).

### 2.1.3 The human Y-chromosome: structure, function and evolution

The Y-chromosome with a length of about 60 Mb is among the smallest in the human genome (Jobling and Tyler-Smith 2003). Two end segments (the pseudoautosomal regions), flanking the Y chromosome, do recombine with respective regions on the X chromosome, and comprise 5% of the chromosome's length. The rest is non-recombining region (NRY), does not undergo sexual recombination and is present only in males (see Figure 2.2). This segment of the Y-chromosome is divided into euchromatic and heterochromatic portions (for review see Skaletsky *et al.* 2003). The heterochromatic sequences consist of massively amplified tandem repeats of low sequence complexity. Nearly all of the euchromatic sequences fall into three classes: X-transposed, X-degenerate and ampliconic. The X-transposed sequences exhibit 99% identity to the X chromosome and are the result of a massive X-to-Y transpositon that occurred 3 - 4 million years ago, after the divergence of the human and chimpanzee lineages. The X-degenerate sequences are relics of ancient autosomes, from which the modern X and Y-chromosomes co-evolved. The ampliconic sequences include large regions (about 35% of the male-specific (MS) Y euchromatin), where sequence pairs show greater than 99.9% identities, which are maintained by frequent gene conversion events (Skaletsky *et al.* 2003).

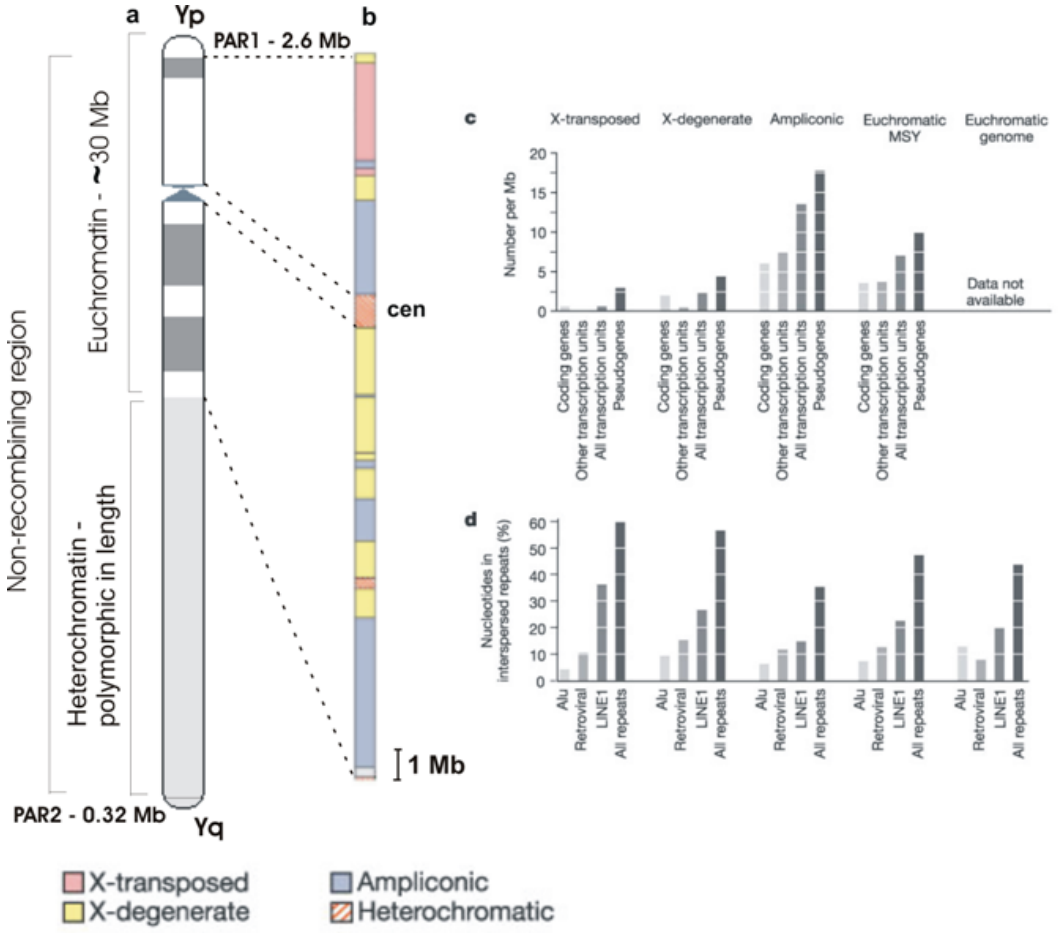


**Figure 2.2** Structure of the Y chromosome. **a**) Cytogenetic features of the chromosome and their approximate locations. Recombination takes place between the Y and X only in the two pseudoautosomal regions (PAR1 and PAR2), and not in the majority of the chromosome which lies between them. **b**) Enlarged view of a 24 Mb portion of the MSY, extending from the proximate boundary of the Yp pseudoautosomal region to the proximal boundary of the large heterochromatic region of Yq. Three classes of euchromatic sequences, as well as heterochromatic sequences are shown. **c**, **d**) Gene, pseudogene and interspersed repeat content of three euchromatic sequence classes. **c**) Densities (numbers per Mb) of coding genes, non-coding transcription units, total transcription units and pseudogenes. **d**) Percentages of nucleotides contained in Alu, retroviral, LINE1 and total interspersed repeats. Redrawn from Skaletsky et al. 2003.

For a long time the Y-chromosome was thought as a sector of inevitable gene decay (Quintana-Murci and Fellous 2001). Now it is understood to be a place of abundant gene conversion (Rozen *et al.* 2003). So far, 156 transcription units, which include 78 protein-coding genes that collectively encode 27 distinct proteins or protein families, have been identified in the human MSY (Jobling and Teylor-Smith 2003; Skaletsky *et al.* 2003). All transcription units are located in euchromatic sequences. The Y-chromosomal genes fall into two functional classes largely on the basis of their expression profile (Skaletsky *et al.* 2003, Lahn and Page 1997). Genes in the first group are expressed in many organs; these housekeeping genes have X homologues that escape X inactivation. The second group, consisting of Y-chromosomal gene families expressed specifically in testes, may account

for infertility among men with Y deletions. Most broadly expressed genes are located in X-degenerative segments, while the testis-specific genes are concentrated predominantly in ampliconic regions (Skaletsky *et al*. 2003). The most prominent feature of the ampliconic region are eight palindromes, at least six of which contain testis genes (Rozen *et al.* 2003). It is speculated that gene conversion helps to preserve the integrity of Y-chromosomal genes, conserving their function across evolutionary time in the absence of crossing-over (Rozen *et al.* 2003; Skaletsky *et al.* 2003).

Investigations have shown that the Y-chromosome has undergone rapid and unconstrained evolution both in sequence content and organization (Archidiacono *et al.* 1998; Skaletsky *et al.* 2003). Many genes on the human Y chromosome have homologues (analogous genes) on the X chromosome. The presence of these X-degenerate sequences reinforces the idea that the Y chromosome developed from an X-like ancestor. According to the reconstruction by Lahn and Page (1999), the first step towards sex determination via DNA occurred roughly 300 million years ago, when one of the autosomes mutated and acquired the *SRY* gene (Sex-determining Region on Y), which is the master switch for male development. The next stage lied in the maintenance of the appeared divergence. The best way of nature for this was to stop recombination. Accordingly, recombination between X and Y was suppressed in a stepwise fashion during evolution, so that discrete portions of chromosomal material suddenly were unable to recombine. Lahn and Page (1999) believe that at least four chromosomal inversion events were responsible for the start-and-stop evolution of the X and Y chromosomes: the first about 300 million years ago and the last 30 million years ago. Such inversions might have been fixed in ancestral populations either by genetic drift or by selection. Each inversion drove the sex chromosomes further apart. Each inverted piece of the chromosomes added to the length of DNA that could no longer align and recombine. On the Y chromosome, this led to degeneration and shrinking, since deleterious mutations were able to build up faster on this non-recombining chromosome. By contrast, the X chromosome retained its genetic integrity and size, since it could continue to recombine with its partner (the other X) in female meiosis.

The Y chromosome also harbors variations of many different kinds. The polymorphisms fall into two main categories:

- Bi-allelic markers: SNPs, short insertion/deletion polymorphisms and *Alu* insertions;
- Multiallelic markers: microsatellites and a minisatellites.

Base substitutions have very low mutation rates about 5 x $10^{-7}$ per site per generation (Hammer 1995). These unique or near unique markers (SNPs and indels) can easily be combined into haplotypes, known as haplogroups. The absence of recombination means that these monophyletic haplogroups can be related by a single phylogeny using the principle of maximum parsimony. Currently over 400 binary polymorphisms, identified by denaturing high performance liquid chromatography (DHPLC) describe the Y-chromosomal phylogenetic tree (Underhill 2003).

Microsatellites, or STR polymorphisms, are also abundant in Y chromosomal genome and can be easily genotyped and scored; they have thus become a useful tool for the elucidation of human population history and for forensic purposes (Buttler 2003; Kayser *et al.* 2004). The number of markers that are suitable to discriminate unrelated males are constantly increasing. In contrast to SNPs, STR loci have substantially higher mutation rates. An average mutation rate of 3 x $10^{-3}$ per locus per generation was estimated by studying Y chromosome in father/sons pairs (Kayser *et al.* 2000), and an effective mutation rate of 6,9 x $10^{-4}$ per generation was defined on the basis of genetic distances (Zhivotovsky *et al.* 2004).

When high-resolution binary lineages are coupled to more rapidly mutating microsatellites the combination of linked polymorphic markers provides a powerful tool for understanding diversity across different time frames (de Kniff 2000; Mountain *et al.* 2002). The combination of slow- and fast-mutating polymorphisms has added values. The typing of STRs within haplogroups allows the investigation of the origin and dispersal of certain haplogroups (Hurles *et al.* 1999; Bosh *et al.* 1999; Mountain et al. 2002).

Due to several special properties, MSY offers an opportunity to reconstruct paternal genealogies. The Y-chromosome is passed down paternal lineages virtually intact except by the gradual accumulation of mutations. This is in contrast to the X chromosome and autosomes, which are continually being reshuffled by recombination. Thus a comparison of Y-chromosomes is a direct comparison of individuals (Jobling and Tyler-Smith 2000; Lell and Wallace 2000). Assuming equal numbers of males and females, the number of Y-chromosomes in the population is one quarter the number of any autosome, hence in the population as a whole, the effective population size of the Y-chromosome is one-quarter of that of a given autosome and one-third of the X chromosomes. In addition one should note that male and female behavior differs with regard to population genetics. The majority of modern societies practice patrilocality (Murdock 1967; Cavalli-Sforza *et al.* 1994),

meaning that wives generally move into their husband's natal domicile. These properties result in strong geographical and social clustering of Y-chromosome variants (Figures 2.3 and 2.4). Gene differentiation parameters $F_{ST}$ and $G_{ST}$ within and between main geographic regions (Africa, Asia and Europe) based on variation of the Y-chromosome are two to three times higher than estimates from autosomal systems and mtDNA (Seielstad *et al.* 1998; Jorde *et al.* 2000).

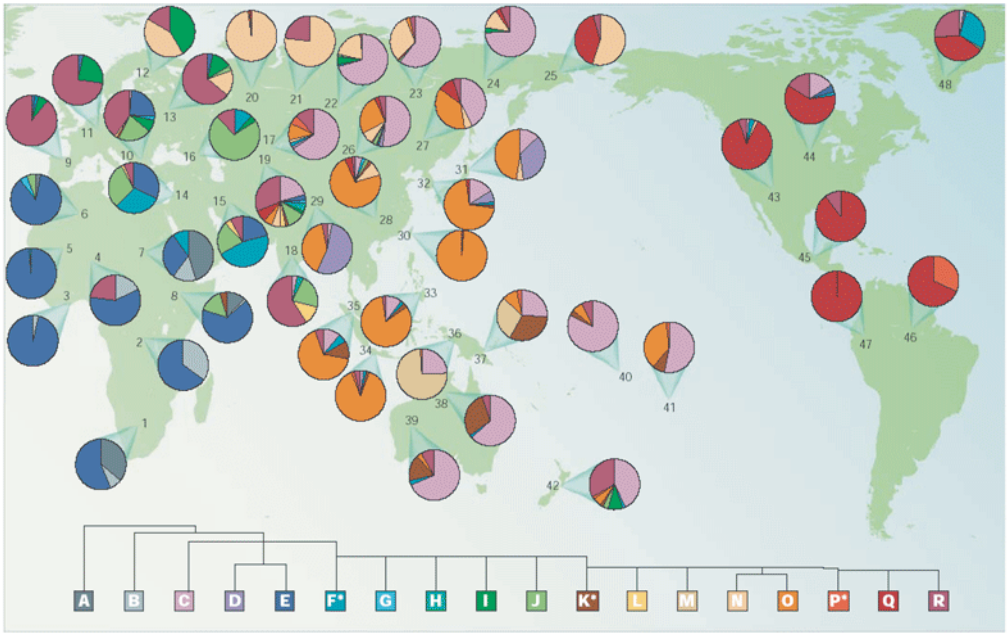


**Figure 2.3** Geographical distribution of the major Y-chromosomal DNA clades (haplogroups) (adopted from Jobling and Tyler-Smith 2003). Each major clade is assigned a color reflecting its position in the phylogeny (below) and its frequency in population samples is shown in the pie charts.

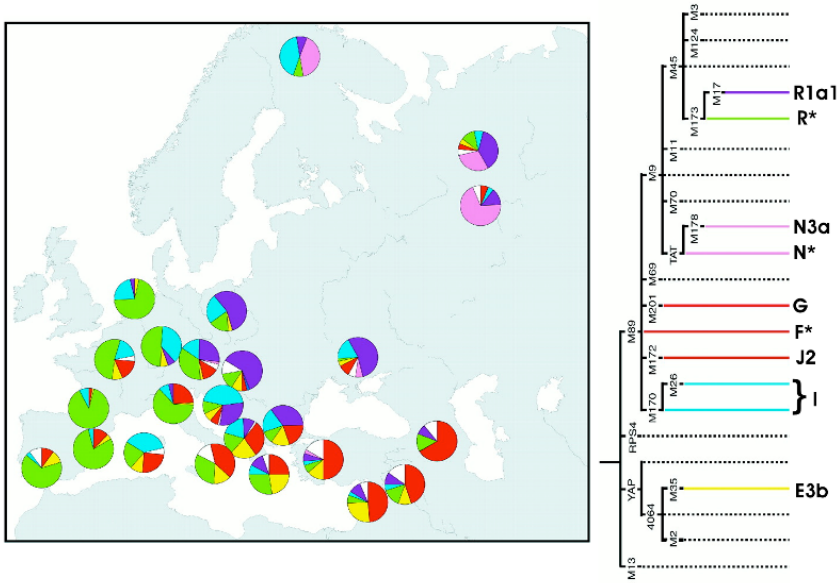


**Figure 2.4** Distribution of major Y-chromosomal haplogroups within Europe. Redrawn from Semino *et al.* 2000. Pie charts show the relative frequencies of different haplogroups, proportional to sector area. The tree right the maps shows the phylogenetic relationships and names of the haplogroups, using YCC nomenclature.

The results of studying Y-chromosomal marker distributions allowed us to reconstruct the origin and the settling of contemporary man as a first approach (Karafet *et al.* 1999; Jin and Su 2000; Underhill *et al.* 2001; Underhill *et al.* 2003). The European sub-continent has been extensively analyzed in respect of the genetic diversity. Nevertheless, the specific features and the formation of regional European genetic pools remain open. This is also the case for the Dniester-Carpathian region, although the history of the various inhabitants has been a subject of considerable interest for historians, linguists and geneticists.

## 2.2 Ethnohistorical background

The Dniester-Carpathian region belongs to the areas, which were inhabited and developed by man from early periods (Chetraru 1973). Its key location at the crossroads of three large subdivisions of the European continent –Eastern, Southeastern and Middle Europe - as well as favorable natural conditions facilitated contacts and interaction of peoples with different cultural and ethnic backgrounds in the course of history. Numerous archeological and historical sources characterize the Carpathian-Dniester region as the contact zone (Dergachev 1990; Dergachev 1999). Despite the available ample set of ethnological, linguistic, archaeological, and anthropological data, an unambiguous opinion on the ethnogeny of the peoples living in the Dniester-Carpathian region is lacking. Let us consider the main issues in the ethnogenesis of the peoples of the Dniester-Carpathian region and the adjacent territories in a chronological order.

Since the book of Childe ‘The Down of European Civilization’ (1968), the contribution of the Neolithic migrants to the reformation of the genetic and cultural landscape of Europe and the Middle East is much discussed. The fact that agriculture arose in the Near East ≈10,000 years before present is not disputed; the argument has arisen over the means of its subsequent dispersal. The demic-diffusion model proposed by Ammerman and Cavalli-Sforza (1984) postulates that extensive migrations of Near Eastern farmers during the Neolithic who brought agricultural techniques to the European continent. In contrast, others have proposed a cultural-diffusion model (Dannell 1983), in which the transfer of agriculture technology occurred without significant population movement.

In the Neolithic and Early Eneolithic the Balkan influences had a major impact on the cultural and historical development of the Carpathian-Dniester region (Figure 2.4). Many surveys showed that virtually all Neolithic cultures of the Dniester-Carpathian region originated from the cultural-historical community of the Balkan-Danubian countries

(Marchevic 1973; Mongait 1973; Comşa 1987; Dergachev *et al.* 1991; Dergachev 1999; Larina 1999).

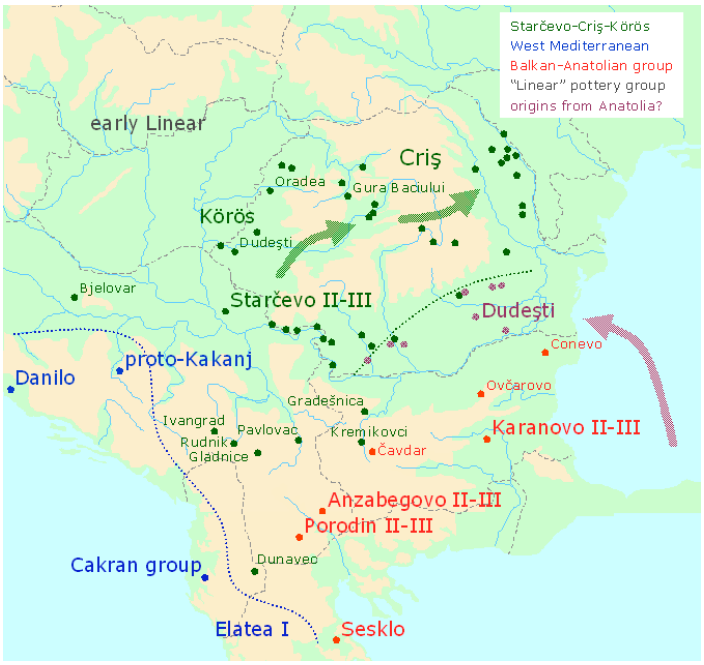


**Figure 2.5** Middle Neolithic period (6,000 – 5,500 BC) in Southeastern Europe. The Starčevo-Körös-Criş culture was the first agricultural community in the Dniester-Carpathian region. This extended across Serbia (Starčevo), East Pannonia (Körös), western Romania, Oltenia and Transylvania (Criş) and in the later phases Moldavia. Adopted from www.eliznik.org.uk/RomaniaHistory/balkans-map/

The Eneolithic on the Moldova territory is characterized by one of the most vivid ancient community of Europe – the Cucuteni-Tripolye culture. The Tripolye cultural community was formed in the Southeastern foothills of the Carpathians at the beginning of the 5th millennium BC on the basis of Neolithic farming cultures of Central Europe and the Balkans (Dergachev and Marchevic 1987; Dergachev 1999). Having spread on the vast territory, stretching from the Southeastern Carpathians to the Dnieper, the Cucuteni-Tripolye culture developed during the 5th-4th millennium BC.

Paleoanthropological data from Neolithic sites of Southeastern and Central Europe support massive migrations from the East Mediterranean area during the Neolithic epoch. The people entering the Balkan Neolithic circle were characterized mainly as narrow faced, of fairly gracial meso-/dolichocran anthropological type, which was classified by the researchers as the Mediterranean one, which differed considerably from the protomorphous European variants of the marginal European pre-Neolithic cultures (Necrasov and Cristescu 1963; Gohman 1966; Potehina 1999; Kruts *et al.* 2003). However, the morphological gracilisation might have occurred as a result of hormonal modeling under the influence of new diets and life styles without a considerable genetic impact. In this

connection the assessment of the inheritance of the Middle Eastern farmers in the gene pool of the contemporary peoples of the Dniester-Carpathian region is of essential interest.


**Figure 2.6** Bronze Age transition (3,500 – 3,000 BC). Beginning from the middle of the Eneolithic (ca. 4,400 BC) and till the end of the Bronze Age (ca. 1,200 BC) the eastern European factor in the history of the Dniester-Carpathian region played the leading role (Dergachev 1999). Adopted from www.eliznik.org.uk/RomaniaHistory/balkans-map/

The influence of the southeastern European factor begins to considerably fade from the middle of the Eneolithic (ca. 4,400 BC) as the steppe East-European factor increases. The Cultural transformation in the Middle Eneolithic – Early Bronze period (4,400 – 2,500 BC) embraced the major part of the European sub-continent. Gimbutas linked the emergence of steppe elements in the Balkan culture to the dissemination of the Indo-European population (Gimbutas 1970). Despite their role in the ethnic history of Europe, the nature of these transformations is subject to hot debates among archeologists, anthropologists and geneticists. The followers of the migratory theory estimate the emergence of the Pit-grave (or Kurgan) traditions in the Balkan and in Central Europe as massive eastern steppe invasions (Gimbutas 1970; Ecsedi 1979; Dergachev 1986; Todorova 1986; Dergachev 1999; Dergachev 2000; Nicolova 2000). Accordingly, the formation of the ancient Pit-grave and later the Pit-grave and the Battle Axis communities was accompanied by the expansion of the cattle-breeding area. The eastern cattle-breeding tribes penetrated deeply into the Carpathian-Danubian area, where they came in direct contact with the local farming population (Figure 2.6). Cavalli-Sforza *et al.* (1994) explain the third principle component of European classical polymorphisms, which accounts for 11% of the total

genetic diversity, by the spread of pastoral nomads during the Eneolithic-Bronze epoch. Also the analysis of Y-chromosomal polymorphism in European populations carried out by Rosser *et al.* (2000) shows a significant cline, stretching from the north of the Black Sea westwards. In contrast, the cultural diffusion model explains the mutual occurrence of elements of the livestock breeding cultures in the environment of the early farming communities of Europe as a process of cultural-historical interactions, based on mutually advantageous exchange and trade (Rassamakhin 1994, Manzura 2000). The leading role of the East-European factor in the history of the Dniester-Carpathian region persists throughout the entire Bronze Age (3,000 – 1,200 BC). However, the livestock breeding tribes of the Kurgan cultures, which penetrated onto this territory from the beginning of the Middle Eneolithic, originated from various regions of a vast territory, stretching from the Dniester in the West to the foothills of the Northern Caucasus and the Southern Ural in the East. They were ethnically and anthropologically heterogeneous (Kruts 1972; Velikanova 1975; Necrasov 1980), pointing to genetic heterogeneity, as well.

The cultural-historical significance of the southeastern factor in the Carpathian-Dniester region was strengthened with the transition to the Early Iron Age (12th-10th centuries BC) (Dergachev 1997; Dergachev 1999). During this period the Carpathian-Dniester region was included into the area of Thracian cultural communities, which were developed here until Late Roman Time. The ethnogeny of the Northern Thracians (for review see Dergachev 1997) is in dispute. Some researchers, following the cultural diffusion model, view them as the immediate heirs of the local population during the Late Bronze Age. Others, followers of the migratory theory, consider the tribes of the Middle Danube as the initial link in the ethnogeny of the Northern Thracians. In this scenario, the local Dniester-Carpathian population was partially assimilated and partially ousted into the Black Sea steppe by the newcomers (Dergachev 1997). In the East the close neighbors of the Thracian were the Cymmerians, who were later ousted by Scythes (Ilynskaya and Terenozhkin 1983). Both the peaceful and militarian ties of the Thracians with their eastern neighbors exerted a great influence on the material culture of the Thracians (Mongait 1974; Melukova and Niculiță 1987).

The ongoing process of the development of the North-Thracian community was stopped at the beginning of the second century AD, when some of the Thracian tribes came under the rule of the Roman Empire. As a consequence of the Roman regime the Romanized population emerged in the Danubian-Carpathian lands (Kolosovskaya 1987; Fedorov

1999). The non-Romanized Thracian population came into contact with numerous migrating tribes of the Dniester-Carpathian region from the north and the east, including the Slavs, the German tribes of the Goths and the Bastarns, the Iranian peoples of the Scythes and the Sarmats. These tribes, which differed in their origin and culture, contributed to the new Cherhyakhov culture, which emerged within an enormous area, stretching from the Dnieper left bank to the Carpathian-Danubian region (Chaplygina 1987; Rickman 1987; Gudkova 1999; Sharov and Bazhan 1999; Sedov 2002; Shschukin 2005). The tribes of the Chernyakhov community attacked constantly the Danubian provinces of the Romans. The internal crisis of the Empire and the increasing pressure of «barbarian» tribes made the Romans leave Dacia (modern day Romania). As soon as the Romans left Dacia, the tribes from the neighboring lands intruded and mixed with the local Romanized population (Fedorov 1999).

From the end of the 5th century AD numerous Slavic tribes of the middle European and East-European plains moved in large numbers to the Danube (Sedov 2002). The eastern path of the Slavs crossed the Dniester-Carpathian lands. In the second half of the 6th century AD they traversed the border of the Byzantine Empire and by the middle of the next century occupied considerable spaces of the Balkan Peninsula right up to the shores of the Adriatic and the Aegean Sea (Sedov 2002). The contribution of the Slavs to the language and the culture of the Romanians and the Moldovians remains a subject for hot disputes among historians, archeologists and politicians. Judging by historical and archeological data, the Slavs constituted the ethnic majority in the Early Middle Ages in the Carpathian basin (Fedorov 1999; Sedov 2002). In that case it appears unclear how the Slavic ethnic community was replaced by the Romanic one. Did the withdrawal of the Slavs from the territory of the Carpathian basin precede the East-Romanic expansion or were they assimilated by the outnumbering Romanic population? Was it the influx of the Romanic population into the North-Danubian lands from the territory of the Balkan peninsula, as the scholars of the migration concept of the origin of the Rumanians and the Moldavians assert, or according to the scenario of the autochthonous development the Rumanians and the Moldavians are the direct successors of the Romanized Thracians, which stayed in the Carpathian basin after the withdrawal of the Roman legions?

In the 13th–14th centuries the Volokhs (the name for old-Romanian communities in the Middle Ages) expanded outside the limits of the internal Carpathian plateau and the Balkans and infiltrated the Eastern foothills of the Carpathians (Zelenchuk 1987; Fedorov

1999). The ethnic development of the Dniester-Carpathian Volokhs proceeded in interaction with the Slavic population that had arrived in the Dniester-Carpathian lands from West Ukraine. A new east-Romanic ethnic community – the Moldavian nationality was formed, which set up its own feudal statehood in 1359 – the Principality of Moldova (Tsaranov *et al.* 1982; Paraska and Sovetov 1987; Fedorov 1999). Before the establishment of the Moldavian protectorate over the territory between the Dniester and the Pruth a considerable part of this territory was a part of the Golden Horde and was inhabited by ethnically diverse peoples (Cumans, Iranians, Slavs, Volokhs) (Polevoy 1987; Russev 1999). It is possible that part of the Golden Horde population remained and was assimilated by east Romanic peoples. The last assumption is supported by craniological surveys (Velikanova 1975). The history of the Moldavian Principality as an independent State was short (Tsaranov et al. 1982). Having reached its bloom under the rule of Prince Steven the Grade (1457-1504), the Moldavian Principality came under the vassalage of the Ottoman Empire by the middle of 16th century after a severe struggle. In 1812, in accordance with the Bucharest Treaty between Russia and the Ottoman Empire, half of the Moldavian Principality, bearing the name of Bessarabia and lying between the Pruth and the Dniester, was transferred to the Russian Empire (Tsaranov *et al.* 1982). From this time the history of the Moldavian people, living on two different Pruth banks, continued independently. In the 19th century the Moldavian people the west of the Pruth river, together with the population of Walachia and later that of Transylvania were integrated into the Romanian nation (Tsaranov *et al.* 1982). The Romanic population of Bessarabia lived in close contact with the Russian and the Ukrainian peoples (Tsaranov *et al.* 1982).

The migration of the Danubians Bulgarians and the Gagauzes into the south of Bessarabia at the end of the 18th to the beginning of the 19th century was an important event in the demographic history of Bessarabia (Radova 1997). Along with the Chuvash, Yakut and Dolgan people of Russia, they are the only ethnic Turkic groups that are predominantly Christian (Eastern Orthodox and some Protestant). The Gagauzes speak the Oghuz branch of the Turkic languages, to which the Turkish, the Azerbaijanian and the Turkmenian languages also belong to. However, the Gagauz language differs from the latter languages by the presence of the Kypchak (Tartar) element (Pokrovskaya 1964; Baskakov 1988). The origin of the Gagauzes remains unclear, and opinions on their ethnogenesis are contradictory (for review see Guboglo 1967; Cimpoies 1997). The Polish turcologist T. Kovalsky concluded from linguistic and cultural-historical data, that three Turkish ethnic

elements took part in the ethnogenesis of the Gagauzes as well as the Deli-Orman Turks: 1) the northern one – the most ancient one, 2) the Seldjuk or the south-Turkic one, referring to the pre-Ottoman epoch in the Balkans and 3) the Turkish-Ottoman one (cited from Pokrovskaya 1964). The presence of some Kypchak «Tartar» linguistic forms in the Gagauz language testifies to the first item. Their usage in the Gagauz language is associated with Turkic tribes (Turk-Bulgarians, the Pechenegs, the Cumans and others), which penetrated into the Balkans from the south-Russian steppes in the 7th– 13th century AD. Part of them settled down on the Balkan Peninsula and mixed with the local population (Guboglo 1967; Cimpoies 1997; Sedov 2002). The participation of the north-Turkic element in the ethnogenesis of the Gagauzes is confirmed by linguistic, and in part by anthropologic and genetic data (Dyachenko 1965; Khit' and Dolinova 1983; Varsahr *et al.* 2001; Varsahr *et al.* 2003). Some researchers interpret the presence of the main south-Turkic (Oghuz) element in the language of the Gagauzes and the Deli-Orman Turks as an inheritance from the Turks-Seldjuks, who were placed in Dobruja in the second half of the 14th century AD by the Byzantine authorities in order to defend the borders of the Empire and to pacify the Bulgarians (for review see Cimpoieş 1997). It is also not ruled out that the Oghuz element was brought to the Balkans by the tribes of the northern nomads, some of which could speak the south-Turkic (Oghuz) dialect. It is thought that not only Turks, but also Bulgarians contributed to the Gagauz ethnic composition (Pokrovskaya 1964).

The contemporary ethnic composition of the indigenous population of the Dniester-Carpathian region is the result of long historical processes. These events are partly fixed in historical chronicles, partly characterized by the archeological and anthropological sources. But bones, stones and chronicles are not the only record of our past. Human DNA, the long, complex molecule that transmits genetic information from one generation to the next, bears the indelible imprint of human history. The contemporary molecular genetic approaches have a sufficient resolution to allow the reconstruction of the genetic connections of the ethnic groups, which are rather close in origin. This thesis is the first attempt to study and explain the molecular-genetic diversity of the ethnically different peoples inhabiting the Dniester-Carpathian region in a single context.

# 3 OBJECTIVE AND TASKS

**The Objective of the thesis was** to investigate the origins and evolution of Dniester-Carpathian populations in the light of the current hypotheses about the history of these populations.

The particular tasks within this general objective were:

1) To characterize the gene pools of the peoples of the Dniester-Carpathian region with molecular marker systems:

a) autosomal *Alu* insertion polymorphisms;

b) compound haplotypes of Y-chromosomes constructed with STR and binary loci localized in the non-recombinant part of the chromosome;

2) To establish the microsatellite diversity within the Y-chromosomal haplogroups, to perform a phylogenetic analysis of the microsatellite haplotypes, and to estimate the time of the origin of the haplogroups most common in the Dniester-Carpathian region;

3) To estimate the level of genetic differentiation among Dniester-Carpathian populations;

4) To estimate the degree of correspondence between the genetic and linguistic variation in the region under study;

5) To analyze the relations between various populations of the Dniester-Carpathian region basing on genetic data and to estimate the genetic position of these populations among western Eurasian peoples.

# 4 MATERIAL AND METHODS

## 4.1 Populations and samples

The objects for this study were DNA probes, extracted from peripheral blood leucocytes. A total 513 blood samples were gathered from unrelated males and females aged 18 years and older in six populations from Dniester-Carpathian region. A specimen of blood was taken from the ulnar vein after obtained both the permission of the examined person and the description of his/her ancestral lineage. The territorial distribution of the surveyed populations is shown in Figure 4.1. The samples of the Moldavians, the Gagauzes and Ukrainians are from the Republic of Moldova. The Moldovans are represented by two rural populations: the northern sample (N=82) is formed from the inhabitants of the Village of Sofia, the Balți district; the southeastern sample of the Moldavians (N=123) is from the Village of Karahasani, the Tighina district. Two Gagauz samples are from villages, which citizens belong to different ethnic subgroups: the population of Kongaz speaks the northern dialect of the Gagauz language (N=72); the inhabitants of Etulia speak the southern dialect (N=64). The sample of the Ukrainians (N=85) was made up of the inhabitants of the Village of Rashkovo, the Kamenka district, Transdniestria. The Romanian sample (N=87) is represented by the inhabitants of the two adjacent east-Romanian towns: Buhuş (the Bacau district) and Piatra-Neamț (the Piatra-Neamț district), which were joined due to their low size.

The medical personnel of the rural ambulatories collected the materials with the participation of the author. The blood samples from the Town of Buhuş were provided by Doctor Ludmila Ştirbu. Doctor Florina Raicu kindly provided the DNA samples from the Town of Piatra Neamț.

DNA was extracted from peripheral blood lymphocytes using salt-based extraction method (Miller *et al.* 1988) or the Amersham genomic DNA extraction reagents and protocols.

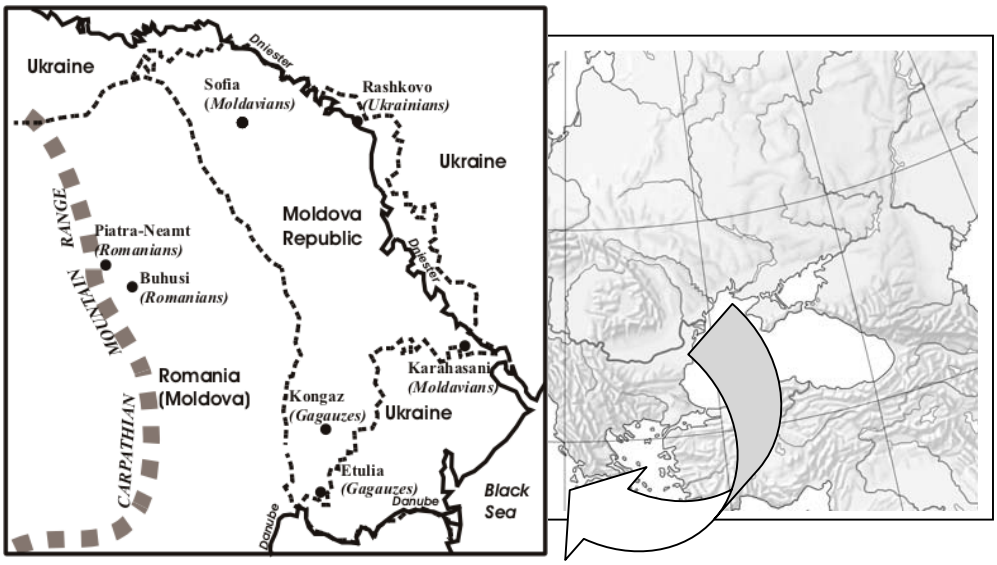


**Figure 4.1** Locations of the studied populations in the Dniester-Carpathian region.

## 4.2 Genotyping

Genetic diversity and population differentiation analyses were conducted using two types of DNA markers: autosomal *Alu* insertion polymorphisms and compound Y-chromosome haplotypes, constructed with the help of STRs and binary loci, localized in the non-recombinant portion of the chromosome.

### 4.2.1 Typing of *Alu* markers

Genotyping was performed by polymerase chain reaction (PCR) in automated Gene Amp PCR System 9600 (Perkin Elmer, USA). PCR amplification was carried out in 20 μl reactions comprising 1.5-3.5 mM $MgCl_2$, 0.2 mM dNTP, 1 units of Taq DNA polymerase (Agrobiogen, Germany), 2 μl 10 x PCR buffer (250 mM Tris-HCl, pH 8.0; 165 mM $((NH_4)_2SO_4)$, 50 ng of target DNA, 10 pmol of each primer. Primers were obtained from (MWG, Germany). The PCR amplification conditions for the ACE, D1, B65, FXIIIB, TPA25, PV92, HS2.43, and HS4.65 were denaturation at 94°C for 1 min, annealing at appropriate temperature (Table 4.1) for 2 min, extension at 72°C for 2 min, for 30 cycles and for the other fore loci (A25, APO, HS3.23, CD4) were as follows: 94°C for 1min, annealing at appropriate temperature (Table 4.1) for 1 min, and 72°C for 1min during 32 cycles. Each sample was subjected to initial denaturation at 94°C for 4 min and to final extension at 72°C for 4 min. 15 μl of PCR product after the addition of 3 μl loading buffer

were electrophoresed in 1.5% agarose 1 x TBE (10 x: 890 mM Tris, 890 mM Borat, 20 mM EDTA) gels. A negative control (all the PCR reagents but not DNA) was carried along with each PCR and the following electrophoresis. DNA bands were visualized by staining with ethidium bromide and photographed under UV light. In order to determine the length of the amplified PCR products, the DNA marker was loaded in each electrophoresis. The presence and absence of an *Alu* insertion in a given locus was designated respectively as *Alu*(+) and *Alu*(-). Individuals were scored as follows: homozygous for the insertion, homozygous for the lack of insertion and heterozygous according the band pattern observed for each locus tested.

**Table 4.1** Autosomal *Alu* markers: Chromosomal location, oligonucleotides for PCR amplification, annealing temperatures and product sizes

| Locus | Ch.l. | Primer sequences (5´-3´) | Annealing Temperature | PCR product sizes (bp) | References |
|---|---|---|---|---|---|
| *A25* | 8 | F: CCACAAATAGGCTCATGTAGAAC<br>R: TATAATATGGCCTGGATTATACC | 63 °C | *Alu* (+): 552<br>*Alu* (–): 268 | Arcot *et al.* 1995a |
| *ACE* | 12 | F: CTGGAGACCACTCCCATCCTTTCT<br>R: GATGTGGCCATCACATTCGTCAGAT | 58 °C | *Alu* (+): 480<br>*Alu* (–): 191 | Batzer *et al.* 1996 |
| *APO* | 11 | F: AAGTGCTAGGCCATTTAGATTAG<br>R: AGTCTTCGATGACAGCGTATACAGA | 56 °C | *Alu* (+): 409<br>*Alu* (–): 97 | Batzer *et al.* 1994;<br>Batzer *et al.* 1996 |
| *B65* | 11 | F: ATATCCTAAAAGGGACACCA<br>R: AAAATTTATGGCATGCGTAT | 52 °C | *Alu* (+): 423/394<br>*Alu* (–): 81 | Arcot *et al.* 1995b |
| *D1* | 3 | F: TGCTGATGCCCAGGGTTAGTAAA<br>R: TTTCTGCTATGCTCTTCCCTCTC | 68 °C | *Alu* (+): 670<br>*Alu* (–): 333 | Arcot *et al.* 1995b |
| *F13B* | 1 | F: TCAACTCCATGAGATTTTCAGAAGT<br>R: CTGGAAAAAATGTATTCAGGTGAGT | 58 °C | *Alu* (+): 700<br>*Alu* (–):410 | Batzer *et al.* 1996 |
| *HS2.43* | 1 | F: ACTCCCCACCAGGTAATGGT<br>R: AGGGCCTTCATCCAGTTTGT | 67 °C | *Alu* (+): 482<br>*Alu* (–): 184 | Arcot *et al.* 1996 |
| *HS3.23* | 7 | F: GGTGAAGTTTCCAACGCTGT<br>R: CCCTCCTCTCCCTTTAGCAG | 60 °C | *Alu* (+): 498<br>*Alu* (–): 200 | Arcot *et al.* 1996 |
| *HS4.65* | 9 | F: TGAAGCCAATGGAAAGAGAG<br>R: ACAGGAGCATCTAACCTTGG | 61 °C | *Alu* (+): 650<br>*Alu* (–): 329 | Arcot *et al.* 1996 |
| *PV92* | 16 | F: AACTGGGAAAATTTGAAGAGAAAGT<br>R: TGAGTTCTCAACTCCTGTGTGTTAG | 54 °C | *Alu* (+): 437<br>*Alu* (–): 122 | Batzer *et al.* 1994;<br>Batzer *et al.* 1996 |
| *TPA25* | 8 | F: GTAAGAGTTCCGTAACAGGACAGCT<br>R: CCCCACCCTAGGAGAACTTCTCTTT | 58 °C | *Alu* (+): 424<br>*Alu* (–): 113 | Batzer *et al.* 1996 |
| *CD4* | 12 | F: AGGCCTTGTAGGGTTGGTCTGATA<br>R: TGCAGCTGCTGAGTGAAAGAACTG | 58 °C | No del*: ~1500<br>Del* : ~1250 | Edwards and Gibbs 1992 |

Note. - Ch.l., Chromosomal location; *CD4* polymorphism is the deletion of 256-bp of a 285-bp *Alu* element at the *CD4* locus; F refers to the forward primer and R refers to the reverse primer for a particular locus.

### 4.2.2 Y-chromosome haplotyping

322 males were examined for 32 binary polymorphisms known to detect variation in West Eurasia (Table 4.2). The samples were examined in a hierarchical way, in agreement with the Y-chromosome phylogeny (YCC 2002). The phylogenetic relationship of the markers analyzed is shown in Figure 4.2. M9 was chosen as the initial marker and surveyed in all samples.

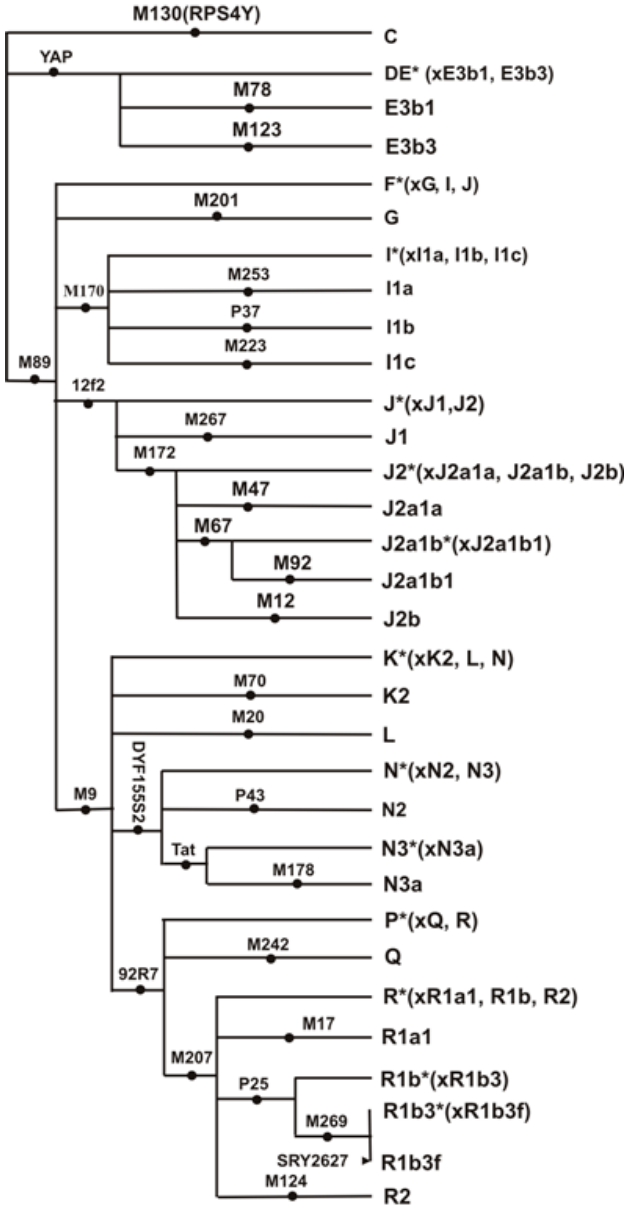


**Figure 4.2** Maximum parsimony phylogeny of the 32 binary markers used in this study. Capital letters indicate haplotypes according to the Y Chromosome Consortium (YCC 2002) with minor modifications (Cinnioğlu *et al*. 2004; Sengupta *et al.* 2006). The M155S2 marker in the tree is a phylogenetical analogue for LLY22g on the YCC tree.

**Table 4.2** Y chromosomal binary markers: type of polymorphism, detection methods, oligonucleotide primers, annealing temperatures and PCR/RFLP product sizes

| Marker | How detected | Primers used (5´-3´) | Annealing temperature | Allele (product sizes in bp) | Reference |
|---|---|---|---|---|---|
| M1(YAP) | PCR | F: CAGGGGAAGATAAAGAAATA<br>R: ACTGCTAAAAGGGGATGGAT | 51°C | *Alu*(-)→*Alu*(+) | Hammer and Horai 1995 |
| M9 | PCR-RFLP (*Hin*fI) | F: GCAGCATATAAAACTTTCAGG<br>R: AAAACCTAACTTTGCTCAAGC | 58°C | C(182/93/66)→G(248/93) | Underhill *et al.* 2001; Hurles *et al.* 1998 |
| M12 | PCR-RFLP (*Hin*fI) | F: ACTAAAACACCATTAGAAACAAAGG<br>R: CTGAGCAACATAGTGACCCGAAT [a] | 62 °C | G(23/67/219)→T(90/219) | Underhill *et al.* 2001; Kharkov, personal communication |
| M17 | Allele specific PCR | F1: TGTGGTTGCTGGTTGTTACGGGG<br>F2: TGTGGTTGCTGGTTGTTACGGG<br>R: TGAACCTACAAATGTGAAACT | 56°C | F1: no del.(287)→del.(0)<br>F2: no del.(287)→del.(286) | Underhill *et al.* 2001; Kharkov *et al.* 2004 |
| M20 | PCR-RFLP (*Ssp*I) | F: GATTGGGTGTCCTCAGTGCT<br>R: CACACAACAAGGCACCAT | 61°C | A(295/118)→G(413) | Underhill *et al.* 2001; Qamar *et al.* 2002 |
| M46 (Tat) | PCR-RFLP (*Hsp*92II) | F: GACTCTGAGTGTAGACTTGTGA<br>R: GAAGGTGCCGTAAAAGTGTGAA | 60°C | T(85/27)→C(112) | Zerjal *et al*. 1997 |
| M47 | PCR-RFLP (*EcoR*I) | F: AGATCATCCCAAAACAATCATAA<br>R: GAAATCAATCCAATCTGTAAATTTTATGTAGAATT | 61°C | G(35/395)→A(430) | Underhill *et al.* 2001; Kharkov, personal communication |
| M67 | PCR-RFLP (*Ssp*I) | F: CCATATTCTTTATACTTTCTACCTGC<br>R: GTCTTTTCACTTGTTCGTGGACCCCTCAATAT | 60°C | A(379/30)→(T)409 | Underhill *et al.* 2001; Kharkov, personal communication |
| M70 | PCR-RFLP (*Hae*III) | F: ACTATACTTTGGACTCATGTCTCCATGAGG<br>R: TTTGTCTTGCTGAAATATATTTTA | 56°C | A(231)→C(201/30) | Underhill *et al.* 2001; Kharkov, personal communication |
| M78 | PCR-RFLP (*Aci*I) | F: CTTCAGGCATTATTTTTTTTGGT<br>R: ATAGTGTTCCTTCACCTTTCCTT | 54 °C | C(196/105)→T(301) | Underhill *et al.* 2001; Flores *et al.* 2003 |
| M89 | Allele specific PCR | F: AGAAGCAGATTGATGTCCC<br>R1: TCAGGCAAAGTGAGAGATG<br>R2: TCAGGCAAAGTGAGAGATA | 59°C | R1: C(365)→T(0)<br>R2: C(0)→T(365) | Underhill *et al.* 2001; Kharkov *et al.* 2004 |
| M92 | PCR-RFLP (*BstSN*I) | F: TTGAATTTCCCAGAATTTTGC<br>R: TTCAGAAACTGGTTTTGTGTCC | 61°C | T(470)→C(340/130) | Underhill *et al.* 2001; Kharkov, personal communication |

(Contd.)

| Marker | How detected | Primers used (5´-3´) | Annealing temperature | Allele (PCR/PCR-RFLP product sizes in bp) | Reference |
|---|---|---|---|---|---|
| M123 | PCR-RFLP (*BstSN*I) | F: TTGAATTTCCCAGAATTTTGC<br>R: TTCAGAAACTGGTTTTGTGTCC | 61°C | T(470)→C(340/130) | Underhill *et al.* 2001; Kharkov, personal communication |
| M124 | Allele specific PCR | F: TGGTAAACTCTACTTAGTTGCCTTT<br>R1: CACAAACTCAGTATTATTAAACCG<br>R2: CACAAACTCAGTATTATTAAACCA | 63°C | R1: C(269)→T(0)<br>R2: C(0)→T(269) | Underhill *et al.* 2001; Kharkov *et a*l. 2005 |
| M130 (RPS4Y) | PCR-RFLP (*Bsc*4I) | F: TATCTCCTCTTCTATTGCAG<br>R: CCACAAGGGGGAAAAAACAC | 58°C | C(205)→T(159/46) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| M170 | PCR, sequencing | F: TGCTTCACACAAATGCGTTT<br>R: GAGACACAACCCACACTGAAACAAT | 56°C | A→C | Underhill *et al* 2001 |
| M172 | PCR-RFLP (*Hin*fI) | F: TTGAAGTTACTTTTATAATCTAATGCTT<br>R: TAATAATTGAAGACCTTTTGAGT | 56°C | T(220)→G(197/23) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| M178 | PCR-RFLP (*Bsp*19I) | F: TAAGCCTAAAGAGCAGTCAGAG<br>R: AGTTCTCCTGGCACACTAAGGAGCC | 58°C | C(245)→T(218/27) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| M201 | Allele specific PCR | F1: CTAATAATCCAGTATCAACTGAGGG<br>F2: CTAATAATCCAGTATCAACTGAGGT<br>R: GTTCTGAATGAAAGTTCAAACG | 66°C | F1: G(215)→T(0)<br>F2: G(0)→T(215) | Underhill *et al*. 2001; Kharkov *et a*l. 2005 |
| M207 | PCR-RFLP (*Dra*I) | F: AGGAAAAATCAGAAGTATCCCTG<br>R: CAAAATTCACCAAGAATCCTTG | 56°C | A(346/77)→G(423) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| M223 | PCR-RFLP (*Mfe*I) | F: AGTCTGCACATTGATAAATTTACTTACAAT<br>R: CCTTTTTGGATCATGGTTCTT | 54°C | C(172)→T(145/27) | Underhill *et al*. 2001; Kharkov, personal communication |
| M242 | PCR-RFLP (*Bbv*12I) | F: AACTCTTGATAAACCGTGCTGTCT<br>R: TCCAATCTCAATTCATGCCTC | 58°C | C(179/187)→T(366) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| M253 | PCR-RFLP (*Hin*dII) | F: GCAACAATGAGGGTTTTTTTG<br>R: CAGCTCCACCTCTATGCAGTTT | 54°C | C(120/280)→T(400) | Cinnioğlu *et al*. 2004; Kharkov, personal communication |
| **M267** | PCR-RFLP (*BstSN*I) | F: TTATCCTGAGCCGTTGTCCCTG<br>R: CTAGATTGTGTTCTTCCACACAAAATACTGTACGT | 60°C | T(150/33)→G(183) | Cinnioğlu *et al*. 2004; Kharkov, personal communication |

(Contd.)

| Marker | How detected | Primers used (5´-3´) | Annealing temperature | Allele (PCR/PCR-RFLP product sizes in bp) | Reference |
|---|---|---|---|---|---|
| M269 | PCR-RFLP (*Bst*2UI) | F: CTAAAGATCAGAGTATCTCCCTTTG<br>R: ACTATACTTCTTTTGTGTGCCTTC | 58°C | T(427)→C(357/68) | Underhill *et al*. 2001; Kharkov *et al*. 2005 |
| $SRY_{2627}$ | PCR-RFLP (*Bbv*12I) | F: AAACATATAGATGGTTGGACATATGTATA<br>R: CAAAAGTCCTTGAATCAGTGGTTTGG | 56°C | C(918)→T(277/641) | Veitia *et al*. 1998 |
| 92R7 | PCR-RFLP (*Hin*dIII) | F: GACCCGCTGTAGACCTGACT<br>R: GCCTATCTACTTCAGTGATTTCT | 63°C | C(512/197)→T(709) | Mathias *et al*. 1994 |
| 12F2 | PCR | F1: TCTTCTAGAATTTCTTCACAGAATTG<br>R1: CTGACTGATCAAAATGCTTACAGATC<br>F2: CTTGATTTTCTGCTAGAACAAG<br>R2: TGTCGTTACATAAATGGGCAC | 53°C | No del.(820/500)→del.(820) | Rosser *et al*. 2000 |
| P25 | Allele specific PCR | F1: TATCTGCTGCCTGAAACCTGCCTGC<br>F2: TATCTGCTGCCTGAAACCTGCCTGA<br>R: CCAACAATATGTCACAATCTC | 58°C | F1: C(269)→A(0)<br>F2: C(0)→A(269) | Kharkov *et al*. 2005 |
| P37 | PCR-RFLP (*Bst4*cI) | F: CGTCTATGGCCTTGAAGA<br>R: TCCGAAAATGCAGACTTT | 63°C | T(447)→C(136/311) | Kharkov *et al*. 2005 |
| P43 | PCR-RFLP (*Nla*III) | R: GAAGCAATACTCTGAAAAGT<br>F: TTTGGAGGGACATTATTCTC | 58°C | G(519)→A(251/268) | Karafet *et al*. 2002 |

Note. - F refers to the forward primer and R refers to the reverse primer for a particular locus; mismatched bases are underlined.

Tekhnologia Tertsik" thermal cycler (Russia). The reaction mixture for amplification comprised 1.5-2.5 mM $MgCl_2$, 0.2 mM dNTPs, 1.5 units of Taq DNA polymerase (Sibenzyme, Russia), 1.5 µl 10 x PCR buffer (250 mM Tris-HCl, pH 8.0; 165 mM $((NH_4)_2SO_4)$, 10 pmol of each primer, 25 ng genomic DNA in a total reaction of 15 µl. Primers were obtained from "Medigen" and "Sibenzyme". The PCR conditions were initial denaturation at 94°C for 4 min; denaturation at 94°C for 30 s, annealing at appropriate temperature (Table 4.2) for 30 s, extension at 72°C for 45 s, for 37 cycles; final extension at 72°C for 4 min. 23 resulting amplicons (M9, M20, M46 (Tat), M47, M67, M70, M78, M92, M123, M130 (RPS4Y), M172, M178, M207, M223, M242, M253, M267, M269, $SRY_{2627}$, 92R7, P37, P43) were digested with the appropriate enzyme (Table 4.2) according to the manufacturer's instructions (Sibenzyme, Russia and New England BioLabs, UK).

Directly after PCR or enzyme digestion the fragments were analyzed on a 2% or 3% agarose gel. DNA bands were staining with ethidium bromide and detected through UV fluorescence with Bio-Rad Gel Doc EQ (USA) using Analysis Software Version 4.4.

Samples were also typed with 7 microsatellites, of which *DYS392* is trinucleotide; and *DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS393* are tetranucleotide. The information on 7 STRs examined in this study is listed in Table 4.3. Y-STR alleles are named on the basis of the number of repeat units they contain, as established through sequenced reference DNA samples. Allele length for DYS389b was obtained by subtraction of the DYS389I allele length from that of DYS389II.

All loci were PCR-amplified using primers and conditions described elsewhere (de Knijf et al. 1997; Kayzer et al. 1997). All 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* (Table 4.3). Fluorescently labeled primers were obtained from Perkin-Elmer Oligo Factory (Germany). These 7 microsatellites were than organized into one multiplex PCR assay and were analyzed on an ABI Prism 310 sequencer (Perkin-Elmer) using GeneScan500-TAMRA (red) as the internal standard. Data were than analyzed using GeneScan 3.7 Macintosh version. An example of the result from ABI 310 Analyzer using designed Y-STR 7plex is displayed in Figure 4.3.

**Table 4.3** Information on Y-STR markers typed

| Locus | Repetitive DNA sequence | Length range (bp) | Repeat count range | Primer sequences (5´-3´) | Annealing temperature | Reference |
|---|---|---|---|---|---|---|
| *DYS389I* | CTAT | 240-260 | 7-12 | F: **HEX** - CCAACTCTCATCTGTATTATCTATG<br>R: TCTTATCTCCACCCACCAGA | 56°C | Cooper *et al.* 1996 |
| *DYS389b* | CTGT/CTAT | 111-135 | 14-20 | | | |
| *DYS390* | CTGT/CTAT | 202-222 | 21-26 | F: **TET** - TATATTTTACACATTTTTGGGCC<br>R: TGACAGTAAAATGAACACATTGC | 56°C | Kayzer *et al.* 1997; de Knijf *et al.* 1997 |
| *DYS391* | TCTA | 276-288 | 9-12 | F: **TET** - CTATTCATTCAATCATACACCCA<br>R: GATTCTTTGTGGTGGGTCTG | 56°C | Kayzer *et al.* 1997; de Knijf *et al.* 1997 |
| *DYS392* | TAT | 237-261 | 8-16 | F: **FAM** - TCATTAATCTAGCTTTTAAAAACAA<br>R: AGACCCAGTTGATGCAATGT | 56°C | Kayzer *et al.* 1997; de Knijf *et al.* 1997 |
| *DYS393* | AGAT | 116-128 | 12-15 | F: **FAM** - GTGGTCTTCTACTTGTGTCAATAC<br>R: AACTCAAGTCCAAAAAATGAGG | 56°C | Kayzer *et al.* 1997; de Knijf *et al.* 1997 |
| *DYS394* (*DYS19*) | TAGA | 185-205 | 13-18 | F: **HEX** - CTACTGAGTTTCTGTTATAGT<br>R: ATGGCATGTAGTGAGGACA | 51°C | Kayzer *et al.* 1997; de Knijf *et al.* 1997 |

Note. - F refers to the forward primer and R refers to the reverse primer for a particular locus.

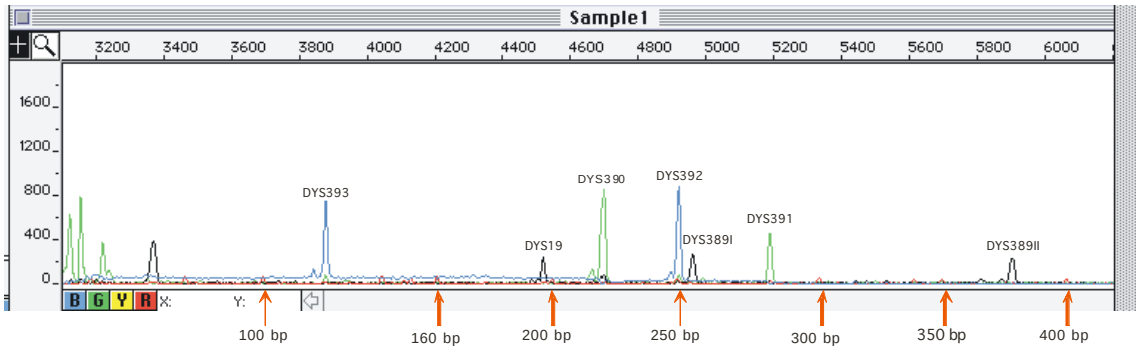


**Figure 4.3** Result from ABI 310 Genetic Analyzer viewed in GeneScan® using a Y-STR 7plex. The PCR products are labeled in three different dye colors with a forth dye (GeneScan 500 TAMRA) used to label an internal-sizing standard.

## 4.3 Statistical analysis

Definitions:

*N* is the sample size (number of individuals or genotypes);

*n* is the number of gene copies in the sample;

*L* is the number of loci;

*k* is the number of alleles or haplotypes;

*H* is heterozygosity;

*V is* variance.

### 4.3.1 Analysis of gene frequencies

Allelic frequencies were calculated using the gene counting method (Li 1976). That is,

$$p_i = n_i / n,$$

where $n_i$ is the number of the *i*-th allele.

Hardy-Weinberg equilibrium was assessed by an exact test. The test was done using a modified version of the Markov-chain random walk algorithm described by Guo and Thomson (1992).

Observed heterozygosity was calculated as

$$H_O = N_O / N,$$

where $N_O$ is the number of heterozygotes.

The theoretical (expected) Hardy-Weinberg heterozygosity of a population for a particular locus was calculated as

$$H_e = 1 - \sum_{i=1}^{k} p_i$$ or, equivalently,

$$H_e = \sum_{i=1}^{k} p_i^2 (1 - p_i).$$

As equivalent to the expected heterozygosity for diploid data, gene diversity and its sampling variance were calculated (Nei 1987) both for autosome and Y-chromosome markers:

$$D = n/(n-1)(1 - \sum_{i=1}^{k} p_i)$$

$$V(D) = 2/(n-1)\{2(n-2)[\sum_{i=1}^{k} p_i^3 - (\sum_{i=1}^{k} p_i^2)] + \sum_{i=1}^{k} p_i^2 - (\sum_{i=1}^{k} p_i^2)^2\}$$

Mean number of differences between all pairs of haplotypes in the sample and its total variance, assuming no recombination between sites and selective neutrality, was obtained as

$$\hat{\pi} = \sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}$$

$$V(\hat{\pi}) = \frac{3n(n+1)\hat{\pi} + 2(n^2+n+3)\hat{\pi}^2}{11(n^2-7n+6)},$$

where $\hat{d}_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes i and j (Tajima 1993).

### 4.3.2 Measures of gene differentiation among populations

The measure of gene differentiation among populations was conducted through the analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992). The primary goal of AMOVA is to assess the amount of variance that can be attributed to different levels of

population organization. The total molecular variance ($\sigma_T^2$) in the case of two hierarchical population structure is the sum of the covariance component due to differences among haplotypes within a population ($\sigma_b^2$) and the covariance component due to differences among the populations ($\sigma_a^2$). Then, the measure of genetic differentiation of populations ($F_{ST}$), or fixation index, is defined by

$$F_{ST} = \sigma_a^2 / \sigma_T^2 .$$

The same framework could be extended to additional hierarchical levels. The genetic structure among population samples was analyzed with (in the case of STR and binary haplotypes) and without (for binary haplotypes only) consideration for molecular differences between individual haplotypes. Confidence intervals for these statistics were constructed using non-parametric permutation approach described in Excoffier *et al*. (1992).

As equivalent to AMOVA, Nei's method (Nei 1987) in the case of *Alu* polymorphisms was also applied. The value of gene differentiation was estimated as the difference between the expected heterozygosities of different levels (total population and subpopulations):

$$D_{ST} = H_T - H_S ,$$

where $H_T$ is the expected heterozygosity of the total population (pooled sample) or the total genetic diversity of the population and $H_S$ is the averaged expected heterozygosity of different samples (subpopulations). The coefficient of genetic differentiations ($D_{ST}$) measures the proportion of population genetic variability accounted for by between-population (between-subpopulation) differences. $G_{ST}$ was calculated according to Nei's formula and expressed in percent:

$$G_{ST} = (D_{ST} / H_T) \times 100\%$$

$G_{ST}$ was estimated for single loci and for all of the loci using heterozygosity values averaged over the loci.

### 4.3.3 Analyses of genetic distances and identity

Genetic distances between pairs of populations in the case of autosomal markers were computed according to the method of Nei (1973; 1987)

$$D = -\ln(I)$$

$$I = (\sum p_i(1-p_i)) \Big/ \sqrt{\sum p_i^2 \sum (1-p_i)^2}\ .$$

Single-locus estimators were combined over all loci as unweighted average of single-locus ratio estimators.

Genetic relationships between the different populations, based on the Y-STR and binary haplotypes, as well as 12 autosomal markers were explored by analysis of molecular variance (AMOVA). The genetic structure among population samples, based on the Y-haplotypes, was analyzed with consideration for the molecular differences between individual haplotypes, in addition to differences in haplotype frequencies, resulting in estimates of $R_{ST}$ (for STR haplotypes) and $\Phi_{ST}$ (for binary haplogroups), an $F_{ST}$ analogues. Significance levels of $R_{ST}$ and $\Phi_{ST}$ values were estimated by use of 10,000 permutations.

In the case of Y-STR haplotypes the probability of identity, *m*, between European population pairs (which reflects the haplotype-sharing index) was estimated, according to the method of Melton *et al.* (1995), as $m = \sum_{i,j}^{k} x_i x_j$ where $x_i$ and $x_j$ are, respectively, the frequencies of a haplotype in populations *i* and *j*, summed over the *k* haplotypes in the two populations.

### 4.3.4 Tree reconstruction and multidimensional scaling analyses

Phylogenetic trees were obtained to display the genetic distances among the samples studied. The trees were constructed using Neighbor-Joining method (Saitou and Nei 1987). This method starts from the genetic distances and is based on sequentially pooling pairs of populations to minimize the sum of branches of the whole tree. In the case of *Alu* markers a total of 1,000 bootstrap replications were performed to assess the strength of the branching structure of the tree.

$F_{ST}$ (for binary haplotypes) and $R_{ST}$ (for Y-STR haplotypes) distances among European population samples were used for multidimensional scaling analysis (MDS). This is an ordination technique for representing the dissimilarity among objects (e.g., populations) in an n-dimensional graph, such that the interpoint distances in the graph space correspond as well as possible to the observed genetic differences between populations. The goodness of

fit between the distances in the graphic configuration and the original genetic distances is measured by a statistic called “stress”, wherein a value of 0 is a perfect fit and a value of 1 is a total mismatch.

### 4.3.5 Barrier analysis

We also used the genetic distance matrix, constructed on the basis of Y-chromosome binary haplotype frequencies, to carry out a barrier analysis. This analysis allows identifying the zones of greatest allele frequency change within a genetic landscape. Monmonier’s maximum difference algorithm was used to find the boundaries (Manni and Heyer 2004). Genetic boundaries were displayed on Delaunay triangulation connections. To assess the robustness of computed barrier, we have obtained 1,000 bootstrap matrices by randomly resampling original data (Y-chromosome haplogroups).

### 4.3.6 Principle component analysis

It is also possible to visualize genetic relationship among populations using the raw data of allele frequencies, rather than genetic distances. Principle component (PC) is an example of this approach, used in this work. The intention is to simplify the multivariate data with a minimum loss of information, that a two-three dimensional graphical representation of the multidimensional data becomes possible. It can be viewed as a rotation of the existing axes to new positions in the space defined by the original variables. In this new rotation, there will be no correlation between the new (imaginary) variables defined by the rotation. The uncorrelated variables are linear combinations of the original variables. The first new variable contains the maximum amount of variation; the second new variable contains the maximum amount of variation unexplained by the first and orthogonal to the first, etc. There can be as many possible dimensions as there are original variables. A plot of the populations on the axes representing the two or three most variable imaginary genes gives a good picture of the biological relationship or distances among the populations.

### 4.3.7 Phylogenetic analysis of STR haplotypes

To investigate the affinities between microsatellite haplotypes within each haplogroups, median-joining networks were constructed. At the population level, phylogenetic networks are more convenient than strictly hierarchical trees to represent relationships among

closely related sequences because the former allow the display of all equally parsimonious hypotheses (i.e., ambiguous relationships) on a single figure. In the median joining algorithm (Bandelt *et al*. 1999) the resulting mass of equally plausible minimum spanning trees are combined within a single network. Then, using the parsimony criterion, inferred intermediate haplotypes (median vectors) are added to network in order to reduce overall tree length. Network construction method takes into account the nature of molecular variability of microsatellite loci. Having analyzed 10844 parent/child meiosis, Brinkman *et al.* (1998) showed that the mutation tempo of the STR loci depends both on the size of the repeated motif and on the average number of repeats in the locus. Therefore, for network calculation, each Y STR locus was weighted according to its estimated mutation rate as given by Kayser *et al.* (2000), so that loci with the highest mutation rates were given the lowest weights (ratio of weights for DYS393:DYS392:DYS19:DYS389I:DYS389II: DYS391:DYS390= 10:10:5:5:2:2:1).

For the ancestral haplotype or the founder haplotype, the following conditions were considered: 1) minimal average distance from the other haplotypes within the sample; 2) frequency in various populations; 3) high frequency of the haplotype in the sample.

### 4.3.8 Age estimates

STR variation data were also used to estimate haplogroup specific expansion times by two methods. Both approaches assume a stepwise mutation model, an average evolutionary STR mutation rate of 6,9 x $10^{-4}$ per STR locus per generation (Zhivotovsky *et al*. 2003), whose value is based upon a generation time of 25 years. One of the methods refers to a median network (Forster *et al*. 2000). In this case relative time estimates calculated by means of $\rho$ , the average distance between founder haplotype and the node of interest, as measured in single repeat differences, and are transformed to absolute time estimates by multiplication with evolutionary mutation. According to the second method (Zhivotovsky *et al.* 2004) the age of STR variation of a subclade was estimated as the average squared difference in the number of repeats between all sampled chromosomes and the founder haplotype, divided by mutation rate (*w).*

For the purposes of estimation of the time since population divergence the $T_D$ estimator was used:

$$T_D = (D_1 - 2V_o)/2w,$$

where $D_1$ is the average squared difference between two alleles sampled from two populations (Goldstein *et al.* 1995; 1996), corrected for bias (Zhivotovsky 2001), $V_0$ is the within-population variance in the number of repeats in the ancestral population prior to its subdivision. The age of divergence, estimated with $T_D$, letting $V_0$=0, gives its upper bound. Time since population divergence was analyzed only in populations with a sample size of at least five individuals.

### 4.3.9 Detecting admixture

We were also interested in the proportions in which different source areas of Y-chromosome are represented in the Dniester-Carpathian paternal gene pools. Admixture proportions were estimated from haplogroup frequencies using the method suggested by Chakraborty (1986) as implemented in the program Admix_2 by Dupanloup and Bertorelle (2001).

### 4.3.10 Mantel test

Mantel's tests were used for assessing the relationships between genetic and geographic distance matrices. This test involves measuring the association between the elements in two matrices by a suitable statistic, and then assessing the significance of this statistic by comparison with the distribution found by randomly reallocating the order of the elements in one of the matrices (Smouse and Long 1992).

### 4.3.11 Software used in the work

Analysis of gene frequencies based on *Alu* frequencies, population differentiation parameters ($D_{ST}$ and $G_{ST}$), age estimates and divergence times of Y-chromosome lineages as described by Zhivotovsky *et al.* (2004) were calculated using Microsoft EXEL. Correlation, multidimensional scaling (MDS) and correspondence analyses were conducted using STATISTICA v.5.5 software (StatSoft Inc. 1995). Genetic distances (as pairwise values of $F_{ST,}$ $\Phi_{ST}$, and $R_{ST}$), genetic diversity parameters (heterozygosities, gene diversity, mean number of differences, average gene diversity), the analysis of molecular variance (AMOVA), and Mantel tests were calculated by use of the ARLEQUIN version

2000 software (Schneider *et al.* 2000). Significance of $F_{ST,}$ $\Phi_{ST}$, and $R_{ST}$ statistics was obtained with 10,000 permutations. The genetic distance matrices in the case of *Alu* insertion polymorphisms were calculated by the GENDIST program in PHYLIP 3.5 (Felsenstein 1993). 1,000 bootstrap replicates were generated with SEQBOOT, and consensus tree was built with CONSENSE as implemented in the PHYLIP 3.5 program package. The program used for constructing the trees was NEIGHBOR in PHYLIP 3.5 (Felsenstein 1993) and for representing the trees – TREEVIEW (Page 1996). The geographical location of putative genetic barriers in the Y-chromosome genetic landscape of Europe was analyzed by means of the Barrier version 2.2 program (Manni and Heyer 2004). Before that 1,000 matrices were generated with the GENDIST and SEQBOOT programs of PHYLIP 3.5 (Felsenstein 1993), which afterwards were involved in the Barrier analysis. Median joining networks and the age of STR variation as described by Forster *et al.* (2000) were calculated by use of the Network 4.111 program (Fluxus Technology Ltd.; www.fluxus-engineering.com). Admixture proportions as well as their standard deviations based on 1,000 bootstrap runs were estimated using computer program ADMIX2_0 (Dupanloup and Bertorelle 2001).

# 5 RESULTS

## 5.1 *Alu* insertion polymorphisms in the Dniester-Carpathian populations

### 5.1.1 Allele frequencies and genetic diversity within populations

The genotype distributions, allele frequencies, concordance of the genotype distributions to Hardy-Weinberg equilibrium, and expected and observed heterozygosities for each locus in the six population samples typed and in the total population sample are shown in Table 5.1. All loci were polymorphic in all populations: no case of allele fixation was found. The observed and expected phenotype frequencies were in sufficient agreement in most populations. Only three out 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.

All the studied groups have similar frequency values of the insertion polymorphisms, which fall in the range of European values (for comparisons see Stoneking *et al.* 1997; Comas *et al.* 2000; Romualdi *et al.* 2002; Comas *et al.* 2004). The only exception was the insertion rate at the *TPA25* locus in the Moldavian sample from the Sofia settlement (0.659), where its value approaches the world maximum (in Madras, India (0.690) and Sri Lanka (0.724) (Antunez-de-Mayolo *et al.* 2002). Besides this, we mark out the little decrease of the *Alu* insertion frequencies at loci *TPA25, B65, D1* and *A25* in the Gagauzes from Etulia, and at the *HS3.23* locus in the Gagauzes from Kongaz. Table 5.2 shows the average gene diversity by locus and population. The loci analyzed in the present samples show significant differences in their gene diversity (Kruskal-Wallis' test, $P<0.0001$) that is a consequence of the observation that, at some loci, both alleles have similar frequencies, whereas in others, one of the alleles is rare, because of random fluctuations. Six out of twelve loci: *ACE, TPA25, FXIIIB, B65, D1, CD4del* exhibited high diversity level (nearly 0.5). For four loci, APOA1, A25, HS2.43, and HS4.65, the diversity level was low (0.06 – 0.17). When we focus on average gene diversity by population, no significant differences between samples are found (Kruskal-Wallis' test, $P=0.9957$), This was expected since similar *Alu* insertion frequencies were found in all the samples analyzed.

**Table 5.1** Distribution of genotypes and frequencies in the populations studied and tests for Hardy-Weinberg equilibrium

| Population | N | Genotype | | | *Alu* frequency | $H_o$ | $H_e$ | *P* |
|---|---|---|---|---|---|---|---|---|
| | | +/+ | +/- | -/- | | | | |
| | | | | | *ACE* | | | |
| Moldavians K | 122 | 29 | 62 | 31 | 0.4918 | 0.5082 | 0.5019 | 1.0000 |
| Moldavians S | 82 | 18 | 42 | 22 | 0.4756 | 0.5122 | 0.5076 | 1.0000 |
| Gagauzes K | 72 | 15 | 37 | 20 | 0.4653 | 0.5139 | 0.5085 | 1.0000 |
| Gagauzes E | 64 | 13 | 33 | 18 | 0.4609 | 0.5156 | 0.5092 | 1.0000 |
| Ukrainians | 83 | 18 | 31 | 34 | 0.4036 | 0.3810 | 0.4918 | 0.0654 |
| Romanians | 87 | 15 | 38 | 34 | 0.3908 | 0.4368 | 0.4834 | 0.4991 |
| Total | 510 | 108 | 243 | 159 | 0.4500 | 0.4775 | 0.4966 | 0.4274 |
| | | | | | *PV92* | | | |
| Moldavians K | 121 | 5 | 47 | 69 | 0.2355 | 0.3884 | 0.3679 | 0.4604 |
| Moldavians S | 82 | 4 | 36 | 42 | 0.2683 | 0.4390 | 0.4039 | 0.3914 |
| Gagauzes K | 72 | 4 | 21 | 47 | 0.2014 | 0.2917 | 0.3350 | 0.4701 |
| Gagauzes E | 64 | 5 | 23 | 36 | 0.2578 | 0.3594 | 0.3973 | 0.7414 |
| Ukrainians | 85 | 7 | 26 | 52 | 0.2353 | 0.3023 | 0.3617 | 0.2230 |
| Romanians | 86 | 6 | 31 | 49 | 0.2500 | 0.3605 | 0.3859 | 0.7780 |
| Total | 510 | 31 | 184 | 295 | 0.2412 | 0.3601 | 0.3674 | 0.7125 |
| | | | | | *TPA25* | | | |
| Moldavians K | 123 | 30 | 65 | 28 | 0.5081 | 0.5041 | 0.5017 | 1.0000 |
| Moldavians S | 82 | 36 | 36 | 10 | 0.6585 | 0.4390 | 0.4525 | 0.8078 |
| Gagauzes K | 72 | 18 | 38 | 16 | 0.5139 | 0.5278 | 0.5031 | 0.8086 |
| Gagauzes E | 64 | 12 | 34 | 18 | 0.4531 | 0.5313 | 0.5080 | 0.6238 |
| Ukrainians | 85 | 23 | 42 | 20 | 0.5176 | 0.4884 | 0.5079 | 0.8287 |
| Romanians | 87 | 31 | 39 | 17 | 0.5805 | 0.4483 | 0.4965 | 0.5118 |
| Total | 513 | 150 | 254 | 109 | 0.5400 | 0.4883 | 0.4979 | 0.7270 |
| | | | | | *FXIIIB* | | | |
| Moldavians K | 123 | 26 | 63 | 34 | 0.4675 | 0.5122 | 0.4999 | 0.8573 |
| Moldavians S | 82 | 22 | 41 | 19 | 0.5183 | 0.5000 | 0.5024 | 1.0000 |
| Gagauzes K | 72 | 20 | 36 | 16 | 0.5278 | 0.5000 | 0.5019 | 1.0000 |
| Gagauzes E | 64 | 20 | 31 | 13 | 0.5547 | 0.4844 | 0.4979 | 1.0000 |
| Ukrainians | 84 | 14 | 46 | 24 | 0.4405 | 0.5476 | 0.4959 | 0.3839 |
| Romanians | 86 | 19 | 46 | 21 | 0.4884 | 0.5349 | 0.5083 | 0.6611 |
| Total | 511 | 121 | 263 | 127 | 0.4941 | 0.5147 | 0.5004 | 0.5315 |
| | | | | | *APOA1* | | | |
| Moldavians K | 123 | 112 | 10 | 1 | 0.9512 | 0.0813 | 0.1009 | 0.2462 |
| Moldavians S | 82 | 79 | 3 | 0 | 0.9817 | 0.0366 | 0.0481 | 1.0000 |
| Gagauzes K | 72 | 68 | 4 | 0 | 0.9722 | 0.0556 | 0.0679 | 1.0000 |
| Gagauzes E | 64 | 59 | 4 | 1 | 0.9531 | 0.0625 | 0.1050 | 0.1141 |
| Ukrainians | 84 | 78 | 6 | 0 | 0.9643 | 0.0706 | 0.0799 | 1.0000 |
| Romanians | 85 | 81 | 4 | 0 | 0.9765 | 0.0471 | 0.0577 | 1.0000 |
| Total | 510 | 477 | 31 | 2 | 0.9657 | 0.0607 | 0.0681 | 0.1081 |
| | | | | | *B65* | | | |
| Moldavians K | 123 | 37 | 62 | 24 | 0.5528 | 0.5041 | 0.4964 | 1.0000 |
| Moldavians S | 82 | 30 | 38 | 14 | 0.5976 | 0.4634 | 0.4912 | 0.8166 |
| Gagauzes K | 72 | 24 | 36 | 12 | 0.5833 | 0.5000 | 0.4976 | 1.0000 |
| Gagauzes E | 64 | 11 | 34 | 19 | 0.4375 | 0.5313 | 0.5048 | 0.6171 |
| Ukrainians | 84 | 20 | 49 | 15 | 0.5298 | 0.5765 | 0.5005 | 0.1959 |
| Romanians | 87 | 26 | 47 | 14 | 0.5690 | 0.5402 | 0.4933 | 0.3936 |
| Total | 512 | 148 | 266 | 98 | 0.5488 | 0.5185 | 0.4964 | 0.3137 |

**Table 5.1** (Contd.)

| Population | N | Genotype | | | Alu frequency | $H_o$ | $H_e$ | $P$ |
|---|---|---|---|---|---|---|---|---|
| | | +/+ | +/- | -/- | | | | |
| | | | | *D1* | | | | |
| Moldavians K | 123 | 21 | 55 | 47 | 0.3943 | 0.4472 | 0.4845 | 0.4566 |
| Moldavians S | 82 | 11 | 38 | 33 | 0.3659 | 0.4634 | 0.4669 | 1.0000 |
| Gagauzes K | 72 | 9 | 38 | 25 | 0.3889 | 0.5278 | 0.4924 | 0.4547 |
| Gagauzes E | 64 | 3 | 30 | 31 | 0.2813 | 0.4688 | 0.4075 | 0.3520 |
| Ukrainians | 85 | 21 | 28 | 36 | 0.4118 | 0.3372 | 0.4944 | 0.0075 |
| Romanians | 86 | 19 | 35 | 32 | 0.4244 | 0.4070 | 0.4876 | 0.1748 |
| Total | 512 | 84 | 224 | 204 | 0.3828 | 0.4386 | 0.4722 | 0.1254 |
| | | | | *A25* | | | | |
| Moldavians K | 123 | 1 | 28 | 94 | 0.1220 | 0.2276 | 0.2231 | 1.0000 |
| Moldavians S | 82 | 1 | 16 | 65 | 0.1098 | 0.1951 | 0.1966 | 1.0000 |
| Gagauzes K | 72 | 1 | 12 | 59 | 0.0972 | 0.1667 | 0.1893 | 0.5012 |
| Gagauzes E | 64 | 0 | 6 | 58 | 0.0469 | 0.0938 | 0.1050 | 1.0000 |
| Ukrainians | 84 | 1 | 11 | 72 | 0.0774 | 0.1294 | 0.1529 | 0.3923 |
| Romanians | 87 | 0 | 20 | 67 | 0.1149 | 0.2299 | 0.2148 | 0.5908 |
| Total | 512 | 4 | 93 | 415 | 0.0986 | 0.1813 | 0.1794 | 0.8103 |
| | | | | *HS4.65* | | | | |
| Moldavians K | 123 | 0 | 9 | 114 | 0.0366 | 0.0732 | 0.0786 | 1.0000 |
| Moldavians S | 82 | 0 | 3 | 79 | 0.0183 | 0.0366 | 0.0481 | 1.0000 |
| Gagauzes K | 71 | 1 | 3 | 67 | 0.0352 | 0.0423 | 0.0820 | 0.0693 |
| Gagauzes E | 64 | 0 | 6 | 58 | 0.0469 | 0.0938 | 0.1050 | 1.0000 |
| Ukrainians | 84 | 0 | 2 | 82 | 0.0119 | 0.0238 | 0.0354 | 1.0000 |
| Romanians | 82 | 0 | 6 | 76 | 0.0366 | 0.0723 | 0.0817 | 1.0000 |
| Total | 506 | 1 | 29 | 476 | 0.0306 | 0.0572 | 0.0613 | 0.3773 |
| | | | | *HS3.23* | | | | |
| Moldavians K | 120 | 91 | 27 | 2 | 0.8708 | 0.2250 | 0.2332 | 1.0000 |
| Moldavians S | 82 | 62 | 19 | 1 | 0.8720 | 0.2317 | 0.2368 | 1.0000 |
| Gagauzes K | 72 | 38 | 29 | 5 | 0.7292 | 0.4028 | 0.3977 | 1.0000 |
| Gagauzes E | 64 | 41 | 19 | 4 | 0.7891 | 0.2969 | 0.3510 | 0.4537 |
| Ukrainians | 85 | 65 | 19 | 1 | 0.8765 | 0.2209 | 0.2258 | 1.0000 |
| Romanians | 87 | 62 | 21 | 4 | 0.8333 | 0.2414 | 0.3114 | 0.0020 |
| Total | 510 | 359 | 134 | 17 | 0.8353 | 0.2622 | 0.2808 | 0.0009 |
| | | | | *HS2.43* | | | | |
| Moldavians K | 123 | 1 | 14 | 108 | 0.0650 | 0.1138 | 0.1297 | 0.4059 |
| Moldavians S | 82 | 0 | 10 | 72 | 0.0610 | 0.1220 | 0.1267 | 1.0000 |
| Gagauzes K | 72 | 3 | 9 | 60 | 0.1042 | 0.1250 | 0.2004 | 0.0230 |
| Gagauzes E | 64 | 2 | 11 | 51 | 0.1172 | 0.1719 | 0.2223 | 0.1883 |
| Ukrainians | 85 | 3 | 14 | 68 | 0.1176 | 0.1647 | 0.2192 | 0.0815 |
| Romanians | 87 | 0 | 15 | 72 | 0.0862 | 0.1628 | 0.1611 | 1.0000 |
| Total | 513 | 9 | 73 | 431 | 0.0887 | 0.1406 | 0.1623 | 0.0097 |
| | | | | *CD4del* | | | | |
| Moldavians K | 122 | 11 | 62 | 49 | 0.3443 | 0.5082 | 0.4615 | 0.2352 |
| Moldavians S | 82 | 6 | 32 | 44 | 0.2683 | 0.3902 | 0.4039 | 1.0000 |
| Gagauzes K | 72 | 8 | 37 | 27 | 0.3681 | 0.5139 | 0.4684 | 0.4487 |
| Gagauzes E | 62 | 8 | 25 | 29 | 0.3306 | 0.4032 | 0.4515 | 0.5666 |
| Ukrainians | 82 | 13 | 34 | 35 | 0.3659 | 0.4096 | 0.4828 | 0.2494 |
| Romanians | 87 | 11 | 41 | 35 | 0.3621 | 0.4713 | 0.4761 | 1.0000 |
| Total | 507 | 57 | 231 | 219 | 0.3402 | 0.4547 | 0.4509 | 0.8434 |

Note. - *N*, number of individuals analyzed; $H_o$, observed heterozygosity; $H_e$, expected heterozygosity; *P*, *P*-value of the test for goodness-of-fit to Hardy-Weinberg expectations. The frequency indicated for each bi-allelic marker is that of the present of the insert for insertion-deletion markers except *CD4del*; presence of the deletion for *CD4del*. Moldavians: K=Karahasani, S=Sofia; Gagauzes: K=Kongaz, E=Etulia.

**Table 5.2** Average gene diversity by locus and by population

| Locus | | Population | |
|---|---|---|---|
| *ACE* | 0.4948±0.0104 | Moldavians Karahasani | 0.3320±0.1844 |
| *PV92* | 0.3676±0.0251 | Moldavians Sofia | 0.3172±0.1776 |
| *TPA25* | 0.4916±0.0197 | Gagauzes Kongaz | 0.3458±0.1916 |
| *FXIIIB* | 0.5001±0.0027 | Gagauzes Etulia | 0.3350±0.1865 |
| *APOA1* | 0.0649±0.0234 | Ukrainians | 0.3216±0.1797 |
| *B65* | 0.4934±0.0060 | Romanians | 0.3354±0.1863 |
| *D1* | 0.4685±0.0310 | | |
| *A25* | 0.1711±0.0470 | | |
| *HS4.65* | 0.0600±0.0249 | | |
| *HS3.23* | 0.2802±0.0732 | | |
| *HS2.43* | 0.1668±0.0417 | | |
| *CD4dl* | 0.4491±0.0279 | | |

### 5.1.2 Genetic differentiation

In order to determine the interpopulation variability, genetic differentiation indices, $G_{ST}$ (Nei 1987) and $F_{ST}$ (Excoffier 1992), were generated for each *Alu* insertion and for all loci considered jointly for the Dniester-Carpathian population (Table 5.3). The contribution of individual loci to the interpopulation variability of the region under study was in the range of low values. For *TPA25* and *HS3.23* loci the $F_{ST}$ values were statistically significantly different from 0. The $G_{ST}$ and $F_{ST}$ values for all loci were 0.0084 and 0.0038, respectively, which could mean that only 0.84/0.38 percents of the total variance in allele frequencies at these loci were due to differences between the populations, where the rest was due to differences within the populations. Although, these values imply a very low level of population genetic subdivision in the Dniester-Carpathian region, the $F_{ST}$ value was significantly different from zero ($P$=0.0098). We have also computed the genetic differentiation $G_{ST}$ on the basis of 11 *Alu* insertion frequencies among six our and eleven southeastern European populations published previously (Stoneking *et al.* 1997, Romualdi *et al.* 2002; Comas *et al.* 2004). Within Southeast Europe, the fraction of the genetic variance attributable to differences among populations was 1.61%.

When the hierarchical approach was taken, populations were pooled together according to linguistic group (Table 5.4). Within the Dniester-Carpathian region, the genetic variance attributable to differences among groups was not significantly different from zero. If we extended the analysis of gene differentiation ($G_{ST}$) to Southeast Europe, the fraction of the genetic variance attributable to differences among populations within the limits of one group prevailed above the component attributable to differences among groups. Moreover, the latter component was very low. These findings suggested that linguistic classification

does not appear to explain the genetic relationships among Dniester-Carpathian, as well as southeastern European populations.

**Table 5.3** Genetic differentiation analyses for 12 individual loci and for all loci considered jointly in the Dniester-Carpathian region

| Locus | Nei (1987) | | | | Excoffier *et al.* (1992) | |
|---|---|---|---|---|---|---|
| | $H_T$ | $H_S$ | $D_{ST}$ | $G_{ST}$ (%) | Among populations (%) | Within populations (%) |
| *ACE* | 0.4950 | 0.4919 | 0.0031 | 0.61 | 0.15 ns | 99.85 |
| *PV92* | 0.3660 | 0.3652 | 0.0008 | 0.22 | -0.33 ns | 100.33 |
| *TPA25* | 0.4968 | 0.4890 | 0.0078 | 1.57 | 1.37* | 98.63 |
| *FXIIIB* | 0.4999 | 0.4972 | 0.0027 | 0.55 | 0.07 ns | 99.93 |
| *APOA1* | 0.0663 | 0.0660 | 0.0003 | 0.41 | -0.09 ns | 100.09 |
| *B65* | 0.4949 | 0.4900 | 0.0049 | 0.99 | 0.50 ns | 99.50 |
| *D1* | 0.4726 | 0.4690 | 0.0036 | 0.76 | 0.33 ns | 99.67 |
| *HS3.23* | 0.2740 | 0.2686 | 0.0054 | 1.97 | 1.74* | 98.26 |
| *A25* | 0.1777 | 0.1765 | 0.0012 | 0.68 | 0.23 ns | 99.77 |
| *HS4.65* | 0.0594 | 0.0592 | 0.0002 | 0.45 | -0.05 ns | 100.05 |
| *HS2.43* | 0.1617 | 0.1607 | 0.0010 | 0.66 | 0.21 ns | 99.79 |
| *CD4del* | 0.4487 | 0.4464 | 0.0023 | 0.51 | 0.02 ns | 99.98 |
| All loci | 0.3344 | 0.3316 | 0.0028 | 0.84 | 0.38* | 99.62 |

Note. - $H_T$, total genetic variability; $H_S$, variability explained by inter-individual differences within populations; $D_{ST}$, interpopulation differences; $G_{ST}$, the coefficient of genetic differentiation.

ns: non-significant; *$P$<0.01.

**Table 5.4** Components of genetic variance (%) at three levels of population subdivision; populations were pooled according to their affiliation to linguistic group (Turkic, Romanic, Albanian, Greek or Slavic)

| Source of variation | Southeast Europe | Dniester-Carpathian region | |
|---|---|---|---|
| | Nei (1987) | Nei (1987) | Excoffier *et al.* (1992) |
| Among groups | 0.50 | 0.35 | 0.20 ns |
| Among populations within groups | 1.11 | 0.49 | 0.24 ns |
| Within populations | 98.39 | 99.16 | 99.56 |

Note. - ns: non-significant.

### 5.1.3 Genetic relationships between populations

To examine genetic relationships between the populations analyzed, we used phylogenetic analysis. $F_{ST}$ genetic distances between the six local Dniester-Carpathian populations are given in Table 5.5. Six out of fifteen pairwise comparisons gave the results at a significant level ($P$<0.05). Population comparisons revealed that the Gagauzes from Etulia is the most distant population. It is significantly different from four out of five compared populations ($P$<0.05). As a consequence of this the Gagauz sample from Etulia occupies the most removed position in the neighbor joining (NJ) tree, constructed on the basis of $F_{ST}$

distances (Figure 5.1). No significant distances were revealed between linguistically related populations.

**Table 5.5** Matrix of genetic distances ($F_{ST}$) between the Dniester-Carpathian populations constructed on the basis of 12 autosomal polymorphisms analyzed

| Population | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gagauzes E | - | | | | | |
| 2. Gagauzes K | 0.0043 ns | - | | | | |
| 3. Moldavians K | 0.0070* | 0.0012 ns | - | | | |
| 4. Moldavians S | 0.0175** | 0.0084* | 0.0039 ns | - | | |
| 5. Romanians M | 0.0101* | 0.0002 ns | -0.0006 ns | 0.0017 ns | - | |
| 6. Ukrainians R | 0.0081* | 0.0030 ns | -0.0013 ns | 0.0077* | -0.0026 ns | - |

Note. - ns: non-significant; *$P$<0.05; **$P$<0.01.

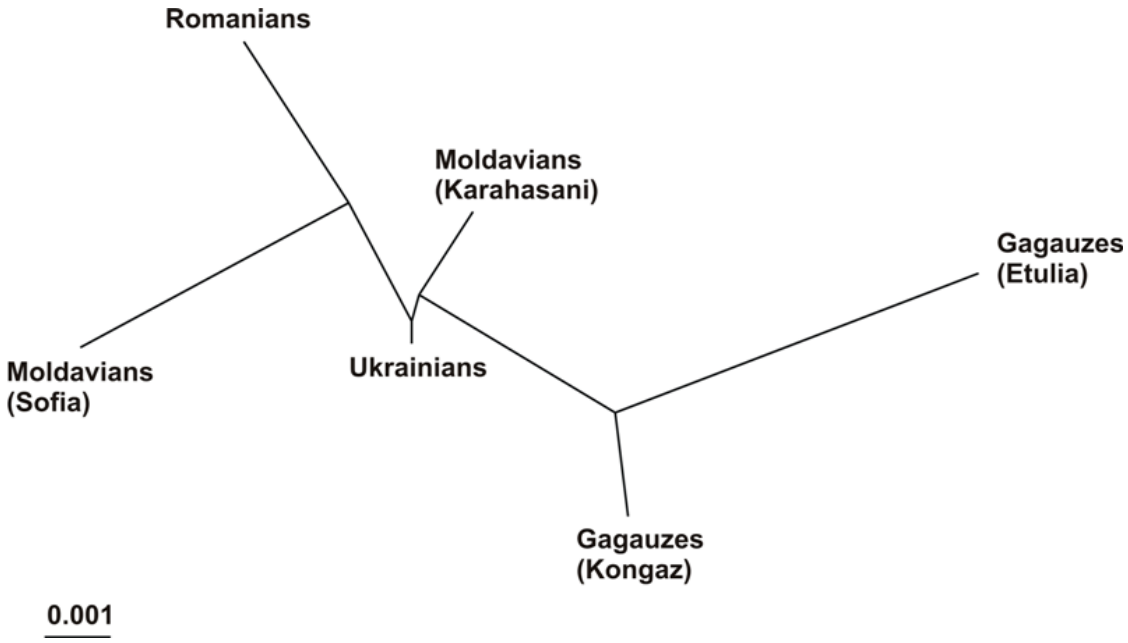


**Figure 5.1** NJ tree based on $F_{ST}$ for pairwise comparisons between the six population samples considered.

To determine the genetic relationships of the Dniester-Carpathian populations with populations of Southeast Europe, we have used the data on 11 Alu insertion loci presented in Comas *et al*. (2004), Stoneking *et al.* (1997) and Romualdi *et al.* (2002) that are common with our study (the list of populations is presented in Table 5.6). In order to visualize the relationships among the populations, two approaches were followed: tree reconstruction and principle component (PC) analysis.

Based on the data on allelic frequencies, the matrix of genetic distances between the 17 populations was constructed using Nei's method (1987) (Appendix 1). To obtain the most probable tree configuration (i.e., the consensus tree), the bootstrap method was used

(Felsenstein 1985). In the consensus tree the compared populations do not constitute strongly pronounced groups (Figure 5.2). The low bootstrap supports (<58.2%) point to the absence of strong phylogenetic links between the neighboring populations in the tree, suggesting the absence of considerable genetic barriers within the southeastern European genetic landscape. However, the bootstrap is known to underestimate the true level of statistical support (Sitnikova *et al.* 1995). It is evident that the topology of the tree in general reflects the geographical proximity of the populations to the south or to the north of the region. Moreover, an additional comparison of the remote groups in the population tree, for example, the Moldavians from Sofia and the eastern Romanians with the Turkish Cypriots and the Greeks, shows strong bootstrap support (91.3%) in the branch linking them (not shown), suggesting at the same time certain distinction between geographically distant populations.

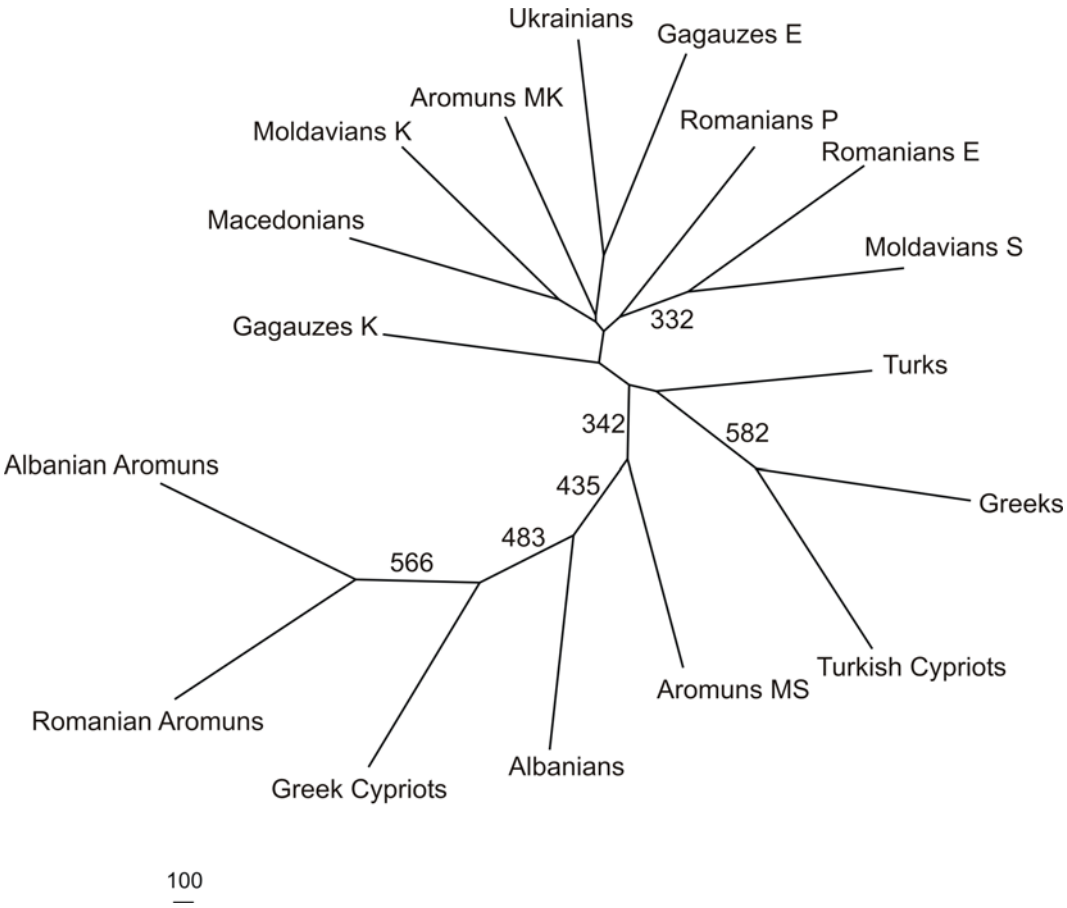


**Figure 5.2** Consensus tree depicting the relationships among the southeastern European populations analyzed for 11 *Alu* polymorphisms. Numbers on the branches are bootstrap values based on 1,000 replications. Moldavians: K=Karahasani; S=Sofia; Gagauzes: K=Kongaz; E=Etulia; Romanians: P=Ploiesti; E=Piatra Nemti and Buhus.

The result from the PC analysis confirms the pattern observed in the consensus tree. The principle component scores generated for each population are presented in Table 5.6. These scores were used to generate the two-dimensional graph of Figure 5.3. As is seen, the first principle component, which explains 24% of the variation in allele frequencies, tends to separate the western Mediterranean (Turkish Cypriots, Greek Cypriots, Turks, northeastern Greeks, Albanians, Albanian Aromuns) from the Balkan-Carpathian populations (Macedonians, Macedonian Aromuns, Romanians, Moldavians, Ukrainians), being the first characterized (absolute correlations greater than 0.63) by high frequencies of *B65* and *FXIIIB Alu* insertions. 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 in the positive pole and the northeastern Greeks in the negative pole.

**Table 5.6** Geographical parameters for the southeastern European populations considered in present study and the corresponding scores of the first two principle components based on the allele frequencies of 11 polymorphic *Alu* repeats (*ACE, TPA25, PV92, APO, FXIIIB, D1, A25, B65, HS2.43, HS3.23, HS4.65*)

| | Location | | Scores | |
|---|---|---|---|---|
| | Latitude | Longitude | PC1 | PC2 |
| Karahasani (Moldavians) | 46°28’N | 29° 48’E | -1.83535 | 0.52894 |
| Sofia (Moldavians) | 47°56’N | 27°52’E | -0.58196 | -0.57637 |
| Etulia (Gagauzes) | 45°31’N | 28°27’E | -0.86375 | 3.00488 |
| Kongaz (Gagauzes) | 46°06’N | 28°35’E | -0.54804 | 0.45505 |
| Rashkovo (Ukrainians) | 47°57’N | 28°50’E | -2.22068 | 0.22354 |
| Piatra-Neamt (Romanians) | 46°58’N | 26°26’E | -1.02891 | -0.45436 |
| Ploiesti (Romanians) | 44°55’N | 26°02’E | -2.40453 | -0.66472 |
| Skopie (Macedonians) | 41°59’N | 21°28’E | -1.43125 | 0.29943 |
| Tirana (Albanians) | 41°19’N | 19°49’E | 1.43242 | 0.18245 |
| Kogalniceanu (Aromuns) | 44°21’N | 28°26’E | 2.32085 | 2.73939 |
| Krusevo, Macedonia (Aromuns) | 41°22’N | 21°15’E | -1.06494 | -0.62908 |
| Stip, Macedonia (Aromuns) | 41°44’N | 22°11’E | -0.02889 | -0.93793 |
| Andon Poci, Albania (Aromuns) | 40°25’N | 20°37’E | 2.53801 | 0.38017 |
| Istanbul (Turks) | 41°00’N | 28°57’E | 1.21251 | 0.78902 |
| Komotini (Greeks) | 41°07’N | 25°25’E | 1.18236 | -3.62290 |
| Nicosia, Cyprus (Greeks) | 35°11’N | 33°22’E | 1.99301 | -0.49993 |
| Nicosia, Cyprus (Turks) | 35°11’N | 33°22’E | 1.32915 | -1.21757 |

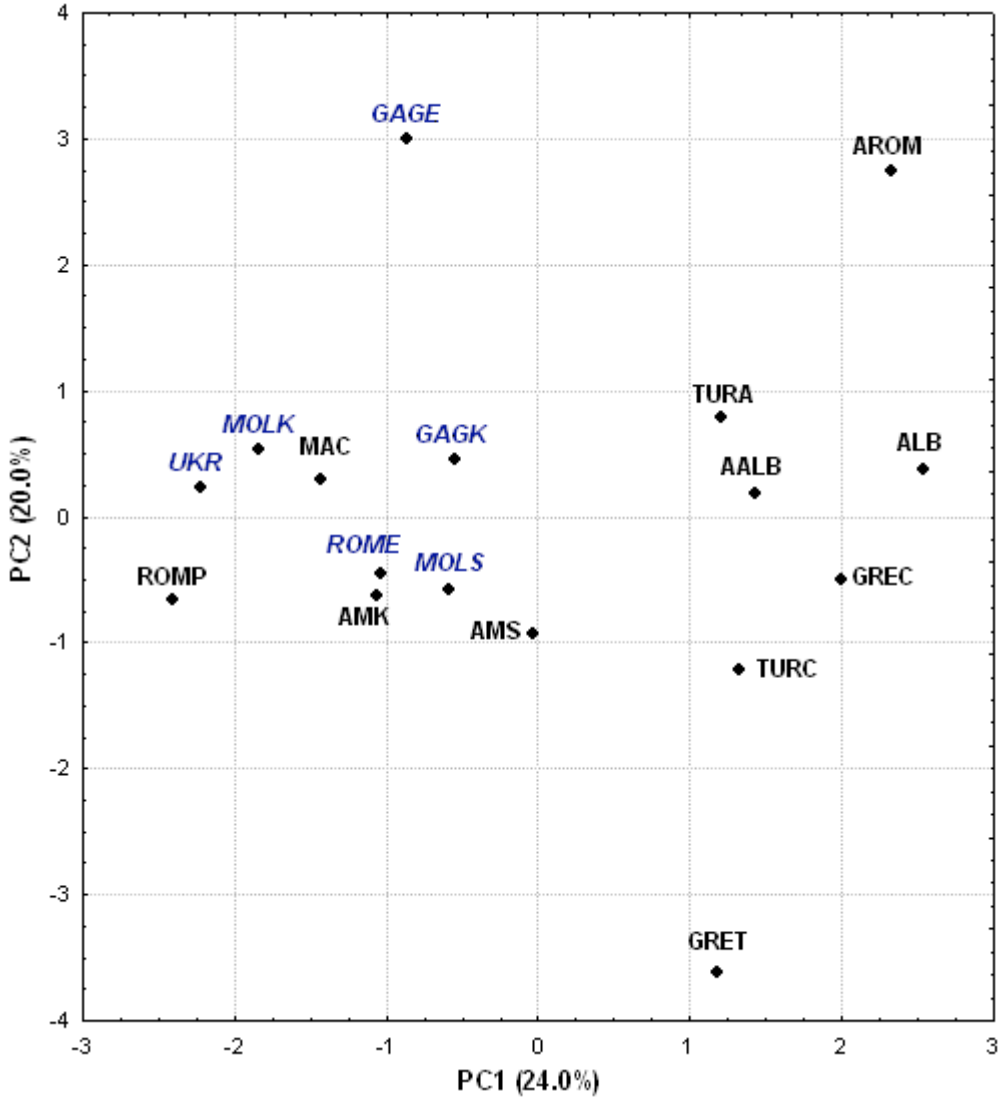


**Table 5.3** Genetic affinities among 17 southeastern European populations based on first two principle components of allele frequencies at 11 *Alu* loci. GAGE=Gagauzes from Etulia; GAGK=Gagauzes Kongaz MOLK=Moldavians from Karahasani; MOLS=Moldavians from Sofia; ROME=Romanians from Piatra-Neamti and Buhusi; UKR=Ukrainians (present study); 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 investigated in the present study populations are marked in blue.

To examine the geographical differentiation of the extracted components more quantitatively we have correlated their scores with geography. The latitude and longitude values assigned to the populations are given in Table 5.6 together with the PC scores. The first principle component is significantly correlated with latitude showing higher values towards South (see Table 5.7). Despite the distinct pattern in geographic distribution the correlation between the geographical and genetic distances, calculated on the basis of Mantel test ($r$=0.165), was not significant ($P$=0.120).

**Table 5.7** Percentage of total variance explained by the two principal components (PCs) and correlation of the PCs scores with geography

| Principle Component | Variance explained (%) | Spearman's correlation coefficient | | |
|---|---|---|---|---|
| | | Latitude | Longitude | Distance from the Ukrainian settlement Rashkovo |
| PC1 | 24.0 | -0.6887* | -0.0245 | 0.6078* |
| PC2 | 20.0 | -0.2598 | 0.1985 | -0.3211 |

Note. - *$P$<0.01 Spearman; $n$=17.

Since the Gagauz language belongs to the Turkic linguistic group, it is of particular interest to assess the genetic relationship of the Gagauzes with Turkic populations from Central Asia. To determine the genetic relationship of the Dniester-Carpathian populations with Central Asian populations we have used the information on eight *Alu* polymorphisms (*ACE, TAT25, PV92, APOA1, F13B, A25, CD4, B65, D1*) previously published in Uyghurs (Xiao *et al.* 2002), Uzbeks, Kazakhs and Kyrgyzes (Khitrinskaya *et al.* 2003). The topology of the consensus tree (Figure 5.4) in general reflects the racial classification of populations. The Kyrgyzes and the Kazakhs, which are assigned to the Mongoloid race, cluster together in the tree with considerable distance to the European populations (Appendix 2). The bootstrap values observed within the European population cluster were very small and neither geographic nor linguistic relationships were observed between the European samples in the tree, pointing that information based on 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 strong bootstrap supports after 1,000 iterations. Both Gagauz samples are grouped together with the European samples.

The plot of the first two components (Figure 5.5), which accounts 62.7% of the total genetic variance (46.7 and 16.0 percents respectively), confirms the trend observed in the consensus tree. Along the first axis the Asian groups are clearly distinguished from the European ones, within which the Gagauz sample from Etulia exhibits a slow approach to the Asian cluster. The insertions at *APO* and *B65* loci, which show their maximum frequency in the European samples (absolute correlations greater than 0.63) and the insertions at *ACE, PV92, FXIIIB,* and *D1,* frequent in the populations from Central Asia,

are the main determinants (with absolute correlations greater than 0.69) of the first principle component.

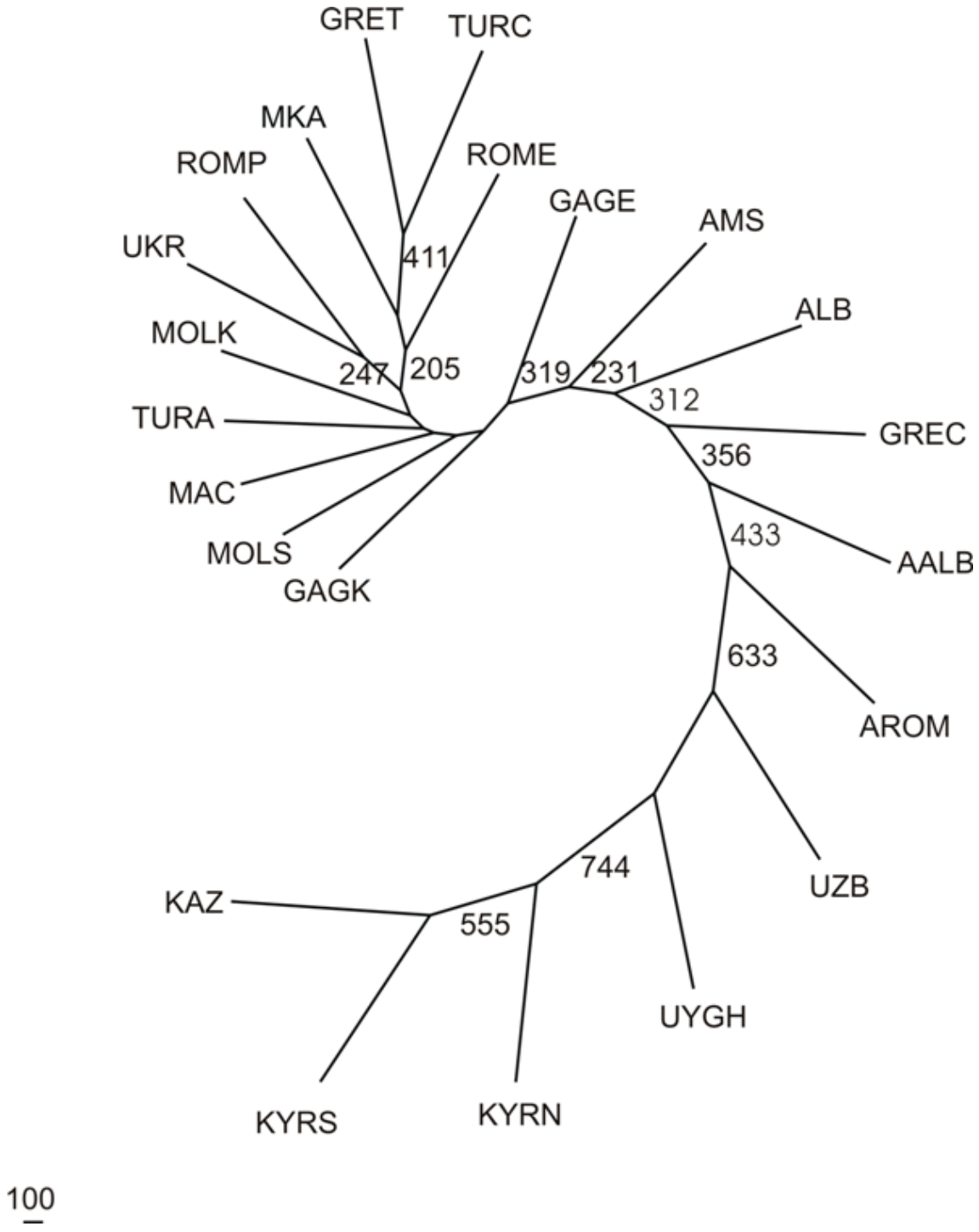


**Figure 5.4** Consensus tree of southeastern European and central Asian populations analyzed for 8 *Alu* polymorphisms. Numbers on the branches are bootstrap values based on 1,000 replications. Codes for the southeastern European populations are as in Figure 5.3; KAZ=Kazakhs; KYRN=northern Kyrgyzes; KYRS= southern Kyrgyzes; UZB=Uzbeks (Khitrinskaya *et al.* 2003); UYGH=Uyghurs (Xiao *et al.* 2002).

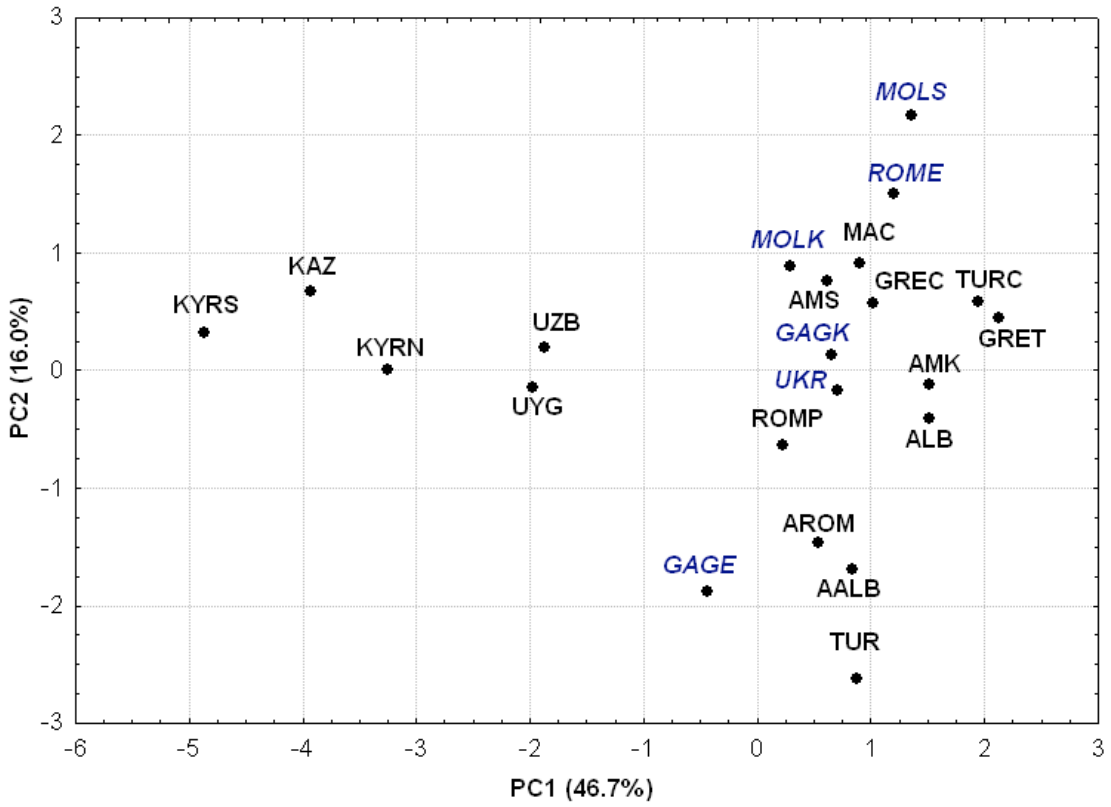


**Figure 5.5** Genetic affinities among southeastern European and Central Asian populations based on first two principle components of allele frequencies at 8 *Alu* loci. Population codes are as in Figures 5.3 and 5.4. The investigated in the present study populations are marked in blue.

## 5.2 Y-chromosome variation: binary-lineage diversity

### 5.2.1 Haplogroup distribution

A total of 28 out of 32 binary polymorphisms genotyped were informative and defined 21 distinct haplogroups. The frequency distribution of Y-haplogroups in all the samples as well as in the joint population sample from the Dniester-Carpathian region is listed in Table 5.8. Within the Dniester-Carpathian region we found a significant heterogeneity of haplogroup frequencies ($\chi^2$=70.1554, *d.f.*=35, *P*<0.001). The exact test of population differentiation implied that eight out of fifteen pairwise comparisons of haplogroup distribution gave the results at a significant level (*P*<0.05). Haplogroups G-M201 and R1a1-M17 out of the predominant lineages are the main contributors to the observed in the region heterogeneity.

Two most frequent in the region haplogroups R1a1-M17 and I1b-P37 comprise together 50.6 percents of all Y-chromosome lineages. The R1a1-M17 is unevenly distributed among the Dniester-Carpathian samples ($\chi^2$=11.33, *df*=5, *P*<0.05). Its frequency in the Ukrainians from Trans-Dniestria and in the Moldavians from Karahasani falls to the lower edge of the eastern European population range (Semino *et al.* 2000; Wells *et al.* 2001; Kharkov *et al.* 2004; Kayser *et al.* 2005; Kharkov *et al.* 2005). In the rest of the samples the frequency of R1a1-M17 is lower and corresponds to the values observed in Southeast Europe (Semino *et al.* 2000; Bosch *et al.* 2006; Marjanovic *et al.* 2005; Pericic *et al.* 2005). With the highest frequency haplogroup I1b-P37 was revealed in the Romanian sample. This peculiarity aligns them with western Balkan populations (Rootsi *et al.* 2004; Marjanovic *et al.* 2005; Pericic *et al.* 2005). Haplogroup I1b-P37 preserves substantial frequency in other samples from the Dniester-Carpathian region. The major western European diagnostic lineage R1b-P25 is the third most prevailing one (15.2%) with the majority belonging to R1b3-M269. With the similar frequency it occurs in all populations studied. Its frequency in the Dniester-Carpathian region well corresponds with the values from Southeast Europe (Semino *et al.* 2000; Bosch *et al.* 2006; Marjanovic *et al.* 2005; Pericic *et al.* 2005) and is stably higher than the values from Eastern Europe (Semino *et al.* 2000; Wells *et al.* 2001; Kharkov *et al.* 2004; Kharkov *et al.* 2005).

**Table 5.8** Y-chromosome haplogroup frequencies (%) and haplogroup diversity in six Dniester-Carpathian populations studied

| Haplogroup | Moldavians (Karahasani) | Moldavians (Sofia) | Romanians (Buhusi, Piatra-Neamt) | Ukrainians (Rashkovo) | Gagauzes (Kongaz) | Gagauzes (Etulia) | Total |
|---|---|---|---|---|---|---|---|
| Sample size | 72 | 54 | 54 | 53 | 48 | 41 | 322 |
| E3b1-M78 | 8.3 | 13.0 | 7.4 | 0.0 | 12.5 | 9.8 | 8.4 |
| E3b3-M123 | 4.2 | 0.0 | 0.0 | 0.0 | 4.2 | 0.0 | 1.6 |
| G-M201 | 0.0 | 1.9 | 5.6 | 0.0 | 10.4 | 17.1 | 5.0 |
| I*-M170 | 1.4 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.6 |
| I1a-M253 | 2.8 | 7.4 | 3.7 | 3.8 | 8.3 | 0.0 | 4.3 |
| I1b-P37 | 16.7 | 25.9 | 40.7 | 20.8 | 18.8 | 22.0 | 23.9 |
| I1c-M223 | 4.2 | 1.9 | 1.9 | 0.0 | 4.2 | 2.4 | 2.5 |
| J*-12f2 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 | 0.0 | 0.3 |
| J1-M267 | 5.6 | 1.9 | 0.0 | 1.9 | 2.1 | 0.0 | 2.2 |
| J2*-M172 | 2.8 | 3.7 | 1.9 | 1.9 | 2.1 | 4.9 | 2.8 |
| J2a1a-M47 | 0.0 | 0.0 | 0.0 | 1.9 | 0.0 | 0.0 | 0.3 |
| J2a1b*-M67 | 1.4 | 0.0 | 0.0 | 0.0 | 0.0 | 2.4 | 0.6 |
| J2a1b1-M92 | 0.0 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.3 |
| J2b-M12 | 0.0 | 0.0 | 1.9 | 3.8 | 2.1 | 0.0 | 1.2 |
| K2-M70 | 0.0 | 1.9 | 0.0 | 0.0 | 6.3 | 0.0 | 1.2 |
| N2-P43 | 1.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 |
| N3a-M178 | 0.0 | 3.7 | 0.0 | 5.7 | 4.2 | 0.0 | 2.2 |
| Q-M242 | 0.0 | 1.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 |
| R1a1-M17 | 34.7 | 20.4 | 20.4 | 41.5 | 12.5 | 26.8 | 26.7 |
| R1b*-P25 | 0.0 | 0.0 | 0.0 | 5.7 | 0.0 | 0.0 | 0.9 |
| R1b3-M269 | 16.7 | 16.7 | 13.0 | 13.2 | 10.4 | 14.6 | 14.3 |
| Haplogroup diversity | 0.820±0.030 | 0.853±0.024 | 0.779±0.043 | 0.771±0.044 | 0.837±0.026 | 0.913±0.017 | 0.839±0.011 |

The Near Eastern haplogroups amount to 22.7% of all the variety of Y-chromosome in our Dniester-Carpathian sample. Haplogroup DE-YAP occurred in five Dniester-Carpathian populations with a frequency typical for North Balkan and Carpathian populations (Semino *et al.* 2000; Stefan *et al.* 2001; Bosch *et al.* 2006; Pericic *et al.* 2005). It was not found in our Ukrainian sample. Two sub-clades of DE-YAP, E3b1-M78 and E3b3-M123, account for all the DE-YAP variation observed in the region. In all the surveyed populations haplogroup E3b1-M78 occurs more often than E3b3-M123 that, as it is known, is a feature of the European pools (Cruciani *et al.* 2004; Semino *et al.* 2004). On the contrary in the Anatolian populations E3b1-M78 and E3b3-M123 occur approximately at similar frequencies (Cinnioğlu *et al.* 2004). Haplogroup G-M201 is unevenly distributed among the populations ($\chi^2$=22.26, *df*=5, *P*<0.001). High frequency of the G-M201 lineages in the Gagauzes draws them in one circle with the Anatolian, the Transcaucasian and the southern Balkan populations (Semino *et al.* 2000; Di Giacomo *et al.* 2003; Nasidze *et al.* 2003; Cinnioğlu *et al.* 2004). Haplogroup G-M201 was not revealed in the Ukrainians and in the Moldavians from Karahasani. On the whole haplogroup J-12f2 occurs at frequency of 7.7% and is uniformly distributed among the samples. Despite its relatively low average frequency in the Dniester-Carpathian region, almost all J-12f2 sub-clades, which were described earlier in the European samples (Di Giacomo *et al.* 2004), were found in the Dniester-Carpathian population sample. Haplogroup J2-M172 was proved the most common variant both in the European populations and in the majority of the Dniester-Carpathian samples (Di Giacomo *et al.* 2004; Semino *et al.* 2004), with the majority of lineages belonging to J2*-M172. Though it should be noted that its sister haplogroup J1-M267 was revealed in five out of six Dniester-Carpathian samples and it occurred more frequently than in the neighboring populations (Di Giacomo *et al.* 2004; Bosch *et al.* 2006; Pericic *et al.* 2005).

Other lineages observed in the Dniester-Carpathian samples, namely N3a-M178, N2-P43, K2-M70, Q-M242, I1a-M253, I1c-M223 and I*-M170, which have different origins and distribution patterns (Seielstad *et al.* 2003; Rootsi *et al.* 2004; Tambets *et al.* 2004), were found at low frequencies, less than 5%, in the Dniester-Carpathian paternal gene pool.

The distinctive haplogroup diversity in the eastern Transcarpathians is reflected in their intrapopulation diversity value (Table 5.8). Indeed, Ukrainians and Romanians are the most homogeneous groups (gene diversity coefficients are 0.771 and 0.779 respectively); the Gagauzes from Kongaz have the highest haplogroup diversity (0.913); the haplogroup

diversity in other samples has intermediate values (0.820–0.853). Relative low diversity in the Ukrainians and the Romanians is due to the predominance in their gene pool the R1a1-M17 and the I1b-P37 lineages, respectively.

### 5.2.2 Analysis of Molecular Variance (AMOVA)

AMOVA was performed to test genetic homogeneity among populations as well as their linguistic aggregates. As reported in Table 5.9, a low but significant level of genetic differentiation is observed among Dniester-Carpathian populations, both when molecular information is used ($\Phi_{ST}$=2.37%, $P$=0.001) and when it is not used ($F_{ST}$=1.83%, $P$=0.003). It was essentially lower than the $F_{ST}$ of 11.2% for 42 western Eurasian populations used for comparison (see Appendix 3 for the list of samples). When Dniester-Carpathian populations were divided into 3 groups defined by language (Romanian, Gagauz and Ukrainian), the genetic variance attributed to differences among groups was non-significant ($F_{ST}$=0.66%, $P$=0.248; $\Phi_{ST}$=1.23%, $P$=0.271), whereas differences among populations within groups were significant at the 0.05 level ($F_{ST}$=1.36%, $P$=0.031; $\Phi_{ST}$=1.49%, $P$=0.043), suggesting that linguistic affiliation is not a good predictor of the genetic structure in the eastern Transcarpathians.

**Table 5.9** Analysis of Molecular Variance (AMOVA) among Dniester-Carpathian populations

| Source of variation | Values for binary markers | | Values for microsatellites |
|---|---|---|---|
| | $F_{ST}$ | $\Phi_{ST}$(%) | $R_{ST}$(%) |
| No grouping | | | |
| Among populations | 1.83** | 2.37** | 2.05** |
| Within populations | 98.17 | 97.63 | 97.95 |
| Linguistic grouping strategy | | | |
| Among populations within groups | 1.36** | 1.49* | 1.33* |
| Among groups | 0.66 ns | 1.23 ns | 1.01 ns |
| Within populations | 97.98** | 97.27** | 97.66** |

Note. - ns: non-significant; *$P$<0.05; **$P$<0.01.

### 5.2.3 Population affinities

In order to asses the relationship between the populations analyzed, $F_{ST}$ and $\Phi_{ST}$ pairwise genetic distances were calculated (Table 5.10) and depicted in NJ trees (Figures 4.6A and 4.6B). The results of phylogenetic analysis were markedly similar irrespective to whether the molecular information was used or not. In both cases population comparisons revealed

that the Gagauz sample from Etulia was not significantly different from the other tested populations ($P$>0.05). The most considerable differences were revealed between the Gagauzes from Kongaz and the Ukrainians from Transdniestria ($P$<0.002).

**Table 5.10** Analysis of genetic differentiation among Dniester-Carpathian populations: pairwise $F_{ST}$-values below the diagonal and pairwise $\Phi_{ST}$-values above the diagonal

| Population | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Moldavians (Karahasani) | - | 0.0060 ns | 0.0388* | 0.0037 ns | -0.0017 ns | 0.0440** |
| 2. Moldavians (Sofia) | 0.0078 ns | - | -0.0022 ns | 0.0381 | -0.0067 ns | 0.0008 ns |
| 3. Romanians | 0.0382** | 0.0017 ns | - | 0.0662** | 0.0125 ns | 0.0173 ns |
| 4. Ukrainians | 0.0013 ns | 0.0260* | 0.0448** | - | 0.0230 ns | 0.0966*** |
| 5. Gagauzes (Etulia) | 0.0084 ns | 0.0016 ns | 0.0141 ns | 0.0232 ns | - | 0.0110 ns |
| 6. Gagauzes (Kongaz) | 0.0272* | -0.0036 ns | 0.0225* | 0.0523** | 0.0035 ns | - |

Note. - ns: non-significant; *$P$<0.05; **$P$<0.01; ***$P$<0.001.

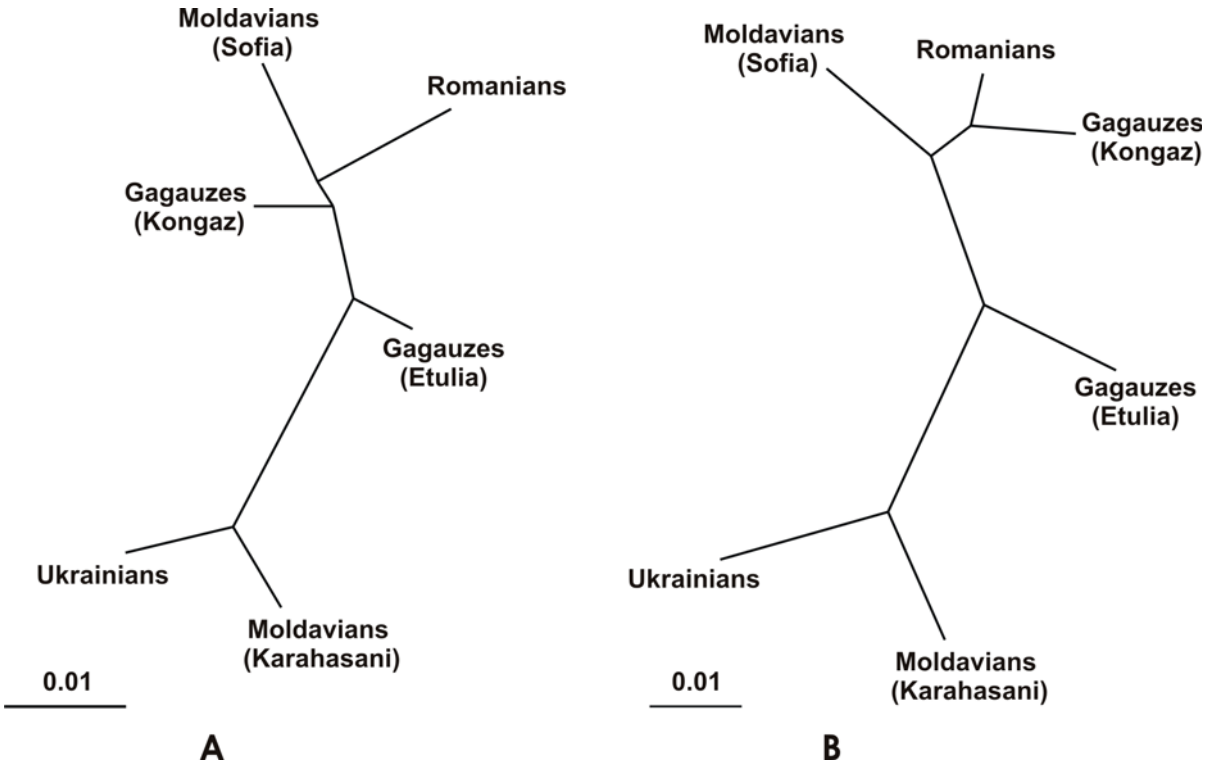


**Figure 5.6** Neighbor-joining (NJ) trees based on pairwise $F_{ST}$ **(A)** and $\Phi_{ST}$-values **(B).**

In order to place Transcarpathian Y-chromosome haplotype diversity within the western Eurasian framework, we compared our samples with 36 western Eurasian samples from the literature (Semino *et al.* 2000; Cinnioğlu *et al.* 2004; Kharkov *et al*. 2005; Bosch et al. 2006; Marjanovic *et al.* 2005; Pericic *et al.* 2005). Results of MDS based on $F_{ST}$ genetic

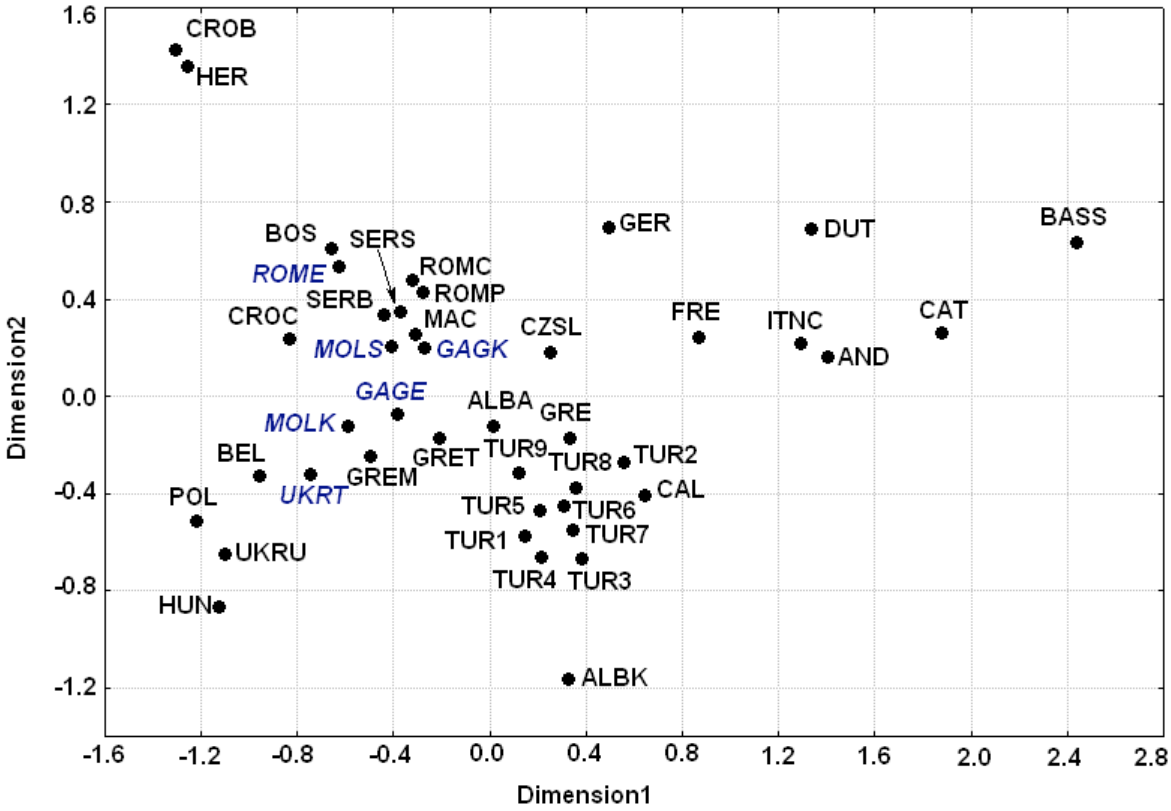


**Figure 5.7** Plot from multidimensional scaling (MDS) analysis of a $F_{ST}$ values from Y chromosome haplogroup frequencies, showing genetic affinities among European and Anatolian populations. The populations presented are: GAGK=Gagauzes from Kongaz; GAGE=Gagauzes from Etulia; MOLK=Moldavians from Karahasani; MOLS=Moldavians from Sofia; ROME=Romanians from Piatra Neamt and Buhus; UKRT=Ukrainians from Trans-Dniestria (present study); AND=Andalusians; BASS=Spanish Basque; CAL=Calabrians; CAT=Catalans; CZSL=Czech and Slovakians; DUT=Dutch; FRE=French; GER=Germans; GRE=Greeks; HUN=Hungarians; ITCN=Central-Northern Italians; GREM=Macedonian Greeks; POL=Poles; UKRU=Ukrainians from Ukraine (Semino *et al.* 2000); TUR1-TUR9=Turks (Cinnioğlu *et al.* 2004); ROMC=Romanians from Constanta; ROMP=Romanians from Ploiesti; GRET=Thracian Greeks (Bosch *et al.* 2006); BEL=Byelorussians (Kharkov *et al.* 2005); CROB=Bosnian Croats, HER=Herzegovinians, SERB=Bosnian Serbs (Marjanovic *et al.* 2005); ALBK=Albanians from Kosovo; SERS=Serbs from Serbia (Pericic *et al.* 2005); CROC=Croats of Croatia (pooled data from Bosch *et al.* 2000 and Pericic *et al.* 2005); ALBA=Albanians from Albania (pooled data from Semino *et al.* 2000 and Bosch *et al.* 2006); BOS=Bosnians (pooled data from Marjanovic *et al.* 2005 and Pericic *et al.* 2005); MAC=Macedonians (pooled data from Bosch *et al.* 2006 and Pericic *et al.* 2005). The investigated in the present study populations are marked in blue.

distances (Appendix 3) are shown in Figure 5.7. A good fit between the two-dimensional graph and the original distance matrix was obtained, demonstrated by the low stress value obtained (0.110). The two Bosnian samples (Croatians and Herzegovinians) and the Albanians from Kosovo demonstrated statistically significant differences from all the other studied populations ($P$<0.05); hence, we regard them as genetic outliers. The remaining populations form the continuous net of genetic relationships, within which four distinctive groups are traced. These population groups can be designated as the Anatolian/southern

Balkan, the western Balkan, the western European and the eastern European clusters. The genetic position of the Dniester-Carpathian populations is ambiguous and is determined by their genetic affinity either to the Balkan or to the eastern European population groups. The Ukrainians from Transdniestria occupy their place among eastern European populations; the eastern Romanians found themselves among western Balkan populations. The Moldavians from Karahasani are genetically closer to eastern European populations; the Moldavians from Sofia and the Gagauzes from Kongaz demonstrate closer affinity to the Balkan cluster. Two Gagauz samples are closer than the other Dniester-Carpathian samples to the Anatolian cluster, though all the pairwise differences between the Gagauzes and the Turks remain statistically significant ($P<0.05$). Three Romanian (one our sample and two from the paper by Bosch *et al.* 2006) and one Moldavian (from Sofia) samples revealed no significant differences ($P>0.05$), whereas the Moldavians from Karahasani show close affinity only with the Moldavians from Sofia ($P>0.05$).

### 5.2.4 Barrier analysis

In order to detect the zones of the sharpest genetic change within European landscape and to see how the Dniester-Carpathian Y-chromosome pools are sorted with these zones, a genetic-barrier analysis on the basis of Monmonier's algorithm was further performed using the same data set as in the comparison analysis. We computed the first three barriers, which were afterwards projected on the European Map (Figure 5.8). The extracted edges with bootstrap supports more than 54% split the European map into three large genetic zones, i.e. Eastern, Western and Southeastern Europe. In its turn, four divisions, namely Anatolia, the South of the Balkan Peninsula, the Dinaric Alps, the Balkan-Carpathian region and central Europe with the sample of the Czechs and Slovaks are traced within the limits of southeastern European zone. In addition to these large population areas, the Albanians from Kosovo, or Kosovars, demonstrate their significant isolation from the neighboring populations with strong bootstrap supports (62–98%). It is interesting to note that the barrier, which separates Eastern Europe from the rest of the continent, passes through the territory of the Dniester-Carpathian region and in the European genetic space the Ukrainians, the southeastern Moldavians and the Gagauzes from Etulia find themselves to the east of the barrier, whereas the Romanians, the northern Moldavians and the Gagauzes from Kongaz clustered together with southeastern European populations.

Moreover, it is notable that the thickness of the eastern barrier is gradually decreasing from the Baltic to the Black Sea, reaching its minimal values (33–40%) on the edges, dividing Etulia from the neighboring Low Danube populations. This finding suggests that across the history the geographic boundary, dividing Southeast Europe from Eastern Europe was more transparent for the reciprocal flows than the boundary between Eastern and Western Europe.


**Figure 5.8** Barrier analysis based on Y chromosome haplogroup frequencies. Reinolds genetic distances were used to compute the first three barriers. The positions of the first three barriers computed are indicated as red lines on Delaunay connections (thin black lines) between sample localities. The thickness of each edge of a barrier is proportional to the number of times it was included in one of the 1,000 computed barriers and numbers along the edges indicate (in per cent) the fraction of 1,000 bootstrap replicates. Edges with bootstrap support less than 20% were not visualized. The population code as in Figure 11.

### 5.2.5 Admixture analysis

The lineages that belong to a certain haplogroup can penetrate into a population from several (rather than one) other populations, in different periods or simultaneously. In this connection, for assessment of the parental contributions into the contemporary pools of Y-chromosome it is more appropriate not just to compute the shares of the diagnostic haplogroups in the hybrid gene pools, but also to take into account the proportions of the respective haplogroups in the source populations. In our choice of parental populations we proceeded from the archaeological data indicating that the cultural landscape of the studied region was subjected to western Mediterranean, Balkan-Carpathian, eastern European, and

western European influences (Dergachev 1999). Five local samples (three Romanian, one Moldavian and one Gagauz) as well as the pooled Dniester-Carpathian sample demonstrate the prevalence of the western Balkan heritage in their pool of Y-chromosomes (Table 5.11). The influences of Eastern Europe proved to have more priority in the genetic history of the Ukrainians from Transdniestria and the Moldavians from Karahasani. The Anatolian influences played a comparatively large role in the population history of the Gagauzes and the southern Romanians.

**Table 5.11** Relative admixture contributions of Western Balkans, Eastern Europe, Anatolia and Western Europe to the Dniester-Carpathian and southern Romanian populations inferred from Y-chromosome haplogroup frequencies

| Population | Source areas | | | |
|---|---|---|---|---|
| | Western Balkans[d] | Eastern Europe[c] | Anatolia[e] | Western Europe[f] |
| Gagauzes K[a] | 0.6544±0.1804 | -0.0965±0.1372 | 0.4708±0.1427 | -0.0287±0.0871 |
| Gagauzes Ea | 0.3168±0.1907 | 0.3091±0.1883 | 0.3000±0.1432 | 0.0741±0.1019 |
| Ukrainiansa[a] | 0.1134±0.1954 | 0.7377±0.1980 | -0.0577±0.1087 | 0.2066±0.1031 |
| Moldavians K[a] | 0.2727±0.1538 | 0.5068±0.1537 | 0.0858±0.0974 | 0.1347±0.0801 |
| Moldavians S[a] | 0.6133±0.1774 | 0.1493±0.1562 | 0.0941±0.1200 | 0.1434±0.0935 |
| Romanians E[a] | 0.9114±0.1913 | 0.0307±0.1596 | -0.0622±0.1112 | 0.1201±0.0825 |
| DCR[a] | 0.4491±0.0842 | 0.3175±0.0756 | 0.1224±0.0540 | 0.1110±0.0405 |
| Romanians P[b] | 0.9181±0.2023 | -0.3653±0.1088 | 0.4910±0.1776 | -0.0438±0.0960 |
| Romanians C[b] | 0.8957±0.2354 | -0.2432±0.1440 | 0.2077±0.1784 | 0.1398±0.1200 |

Note. - Moldavians: K=Karahasani, S=Sofia; Gagauzes: K=Kongaz, E=Etulia; Romanians: E=eastern, C=Constanta, P=Ploiesti; DCR=Dniester-Carpathian region.
[a] These data.
[b] from Bosch *et al.* (2006).
[c] Averaged frequencies from Ukrainian, Polish (Semino *et al.* 2000) and Byelorussian (Kharkov *et al.* 2005) samples.
[d] Averaged frequencies from Croatian (Semino *et al.* 2000, Pericic *et al.* 2005, Marjanovic *et al.* 2005), Bosnian (Pericic *et al.* 2005, Marjanovic *et al.* 2005), Serbian (Marjanovic *et al.* 2005, Pericic *et al.* 2005) and Herzegovinian (Marjanovic *et al.* 2005) samples.
[e] The poled data of Turks from Cinnioğlu *et al.* (2004).
[f] Averaged frequencies from Basque, Andalusia, Catalan, French, North-Central Italian, Dutch and German samples (Semino *et al.* 2000).

The main western Eurasian genetic components are represented unevenly in the Y-chromosomal pools of the Romanian-speaking populations. The Moldavians from Karahasani demonstrate considerable eastern European proportion. All the remaining eastern Romanic samples are characterized by a prevalence of the western Balkan component over the eastern European one. The Romanians from Constanta, and even more so the Romanians from Ploiesti, differ from the eastern Romanians by a notable Anatolian component in their paternal gene pools.

## 5.3 Y-chromosome variation: STR-haplotype diversity

### 5.3.1 STR haplotypes distribution and genetic diversity within populations

Y-STR polymorphisms were studied to obtain a more detailed view of Y-chromosome variation. Among 310 Dniester-Carpathian males examined a total of 157 haplotypes were revealed. The most common haplotype in our study 13-10-17-24-10-11-13 (13/310: 4.2%) occurs in 307 out of 19637 Europeans from 135 different regions sampled in the Y-STR haplotype reference database (YHRD) (http://www.ystr.org/europe). The next four most frequent haplotypes were found 12, 11, 9 and 8 times (with frequencies of 3.9%, 3.5%, 2.9% and 2.6%, respectively) in the Dnieater-Carpathian region and 224, 102, 62 and 154 times out of 19637 Europeans (1.1%, 0.5%, 0.3% and 0.8%, respectively).

A total of 100 single unique haplotypes (63.29%) were observed. These occur in just a single individual in a single population (according to Kayser *et al.* 2001) (Table 12). Of these haplotypes 38 were not found in 3719 men from Anatolia, Southeast, Central and Eastern Europe (see Appendix 4 for the list of populations). A total of 27 haplotypes (17.09%) are shared by individuals within a single population (according to Kayser *et al.* 2001 these haplotypes are designated “multiple unique”). When we consider the Dniester-Carpathian samples within the scale of population of Anatolia, Southeastern, Central and Eastern Europe, the number of the “multiple unique” haplotypes becomes equal to six. "Single unique" and “multiple unique” haplotypes are present only within one of the samples (these haplotypes are designated "total unique" according to Kayser *et al.* 2001). The remaining 30 haplotypes (19.62%), if we consider the diversity within the Dniester-Carpathian region, and 113 haplotypes in the space of European chromosomes occurred in many male individuals in several populations, in other words they are shared among populations (i.e., not unique). Investigation of haplotype sharing (or identity) within populations (multiple-unique haplotypes) and of population-specific haplotypes (single- and multiple-unique haplotypes) allows some insight into population structure and history. High percentage of unique haplotypes points to an isolated population and high percentage of multiple unique haplotypes requires a strong founder in the population history. High percentage of haplotypes shared among populations suggests a common recent ancestry of the populations and/or extensive gene flow among them. A greater share of multiple unique haplotypes in the rural populations of Moldavia, than in the urban sample of the

Romanians is in good accordance with the size of these populations. For all population samples surveyed except the Romanians, the proportion of population specific haplotypes was either higher than or equal to that of shared by populations. All pairwise population comparisons had shared haplotypes (Table 5.13). The Romanians share the most considerable amount of the haplotypes with the Moldavians from Karahasani (N=12); the greatest probability of identity is observed between the Moldavians from Sofia and the Romanians (0.0125) and between the Moldavians from Karahasani and the Gagauzes from Etulia. On the European scale, the Gagauzes and the Romanians show the most considerable identity index with the Balkan populations, particularly with the Albanians, whereas the Ukrainians from Transdniestria with the Eastern and the Western Slavs, as well as with the Balts and the Slovenians (Appendix 5). The Moldavians show dualism on this parameter and reveal simultaneously the proximity with east European and Balkan populations. At the same time it should be noted that the Moldavians from Karahasani share more chromosomes with east European populations than the Moldavians from Sofia, which share more chromosomes with Balkan populations.

The marked genetic variation of Y-STR haplotypes in the populations under study is mirrored in the haplotype-diversity values. These ranged from 0.9636 in the sample from Etulia to 0.9898 in the sample from Kongaz (Table 5.12). The values of the variability coefficients in various European populations, taken from the literature, are given in Appendix 4 for comparison. These values are lower than those described in studies in which more microsatellite loci were used (Ploski *et al.* 2002). Nevertheless, the haplotype diversity index exceeds 0.97 in 37 out of 39 cases, indicating a high level of genetic diversity within European populations. The values of the haplotype diversity indices in all surveyed samples are within the limits of the European scale and correspond to the average European values in five cases. Only the sample from Etulia shows a lower haplotype diversity index. The lowest haplotype diversity 0.9636 in the Etulia might be due to relatively small sample size (N=39).

**Table 5.12** Y-STR haplotype-sharing statistics in the Dniester-Carpathian populations

| Parameter | | Gagauzes (Kongaz) | Gagauzes (Etulia) | Moldavians (Karahasani) | Moldavians (Sofia) | Ukrainians | Romanians |
|---|---|---|---|---|---|---|---|
| No. of individuals | | 47 | 39 | 72 | 50 | 51 | 51 |
| No. of haplotypes | | 37 | 24 | 44 | 36 | 38 | 34 |
| Discrimination (%) | | 78.72 | 61.54 | 61.11 | 72.00 | 74.51 | 66.67 |
| Haplotype class (within the Dniester-Carpathian region): | | | | | | | |
| Single unique: | No. | 19 | 7 | 18 | 22 | 20 | 14 |
| | Proportion | 0.5135 | 0.2917 | 0.4091 | 0.6111 | 0.5263 | 0.4118 |
| Multiple unique: | No. | 5 | 5 | 7 | 3 | 5 | 2 |
| | Proportion | 0.1351 | 0.2083 | 0.1591 | 0.0833 | 0.1316 | 0.0588 |
| Total unique: | No. | 24 | 12 | 25 | 25 | 25 | 16 |
| | Proportion | 0.6486 | 0.5000 | 0.5682 | 0.6944 | 0.6579 | 0.4706 |
| Nonunique: | No. | 13 | 12 | 19 | 11 | 13 | 18 |
| | Proportion | 0.3514 | 0.5000 | 0.4318 | 0.3056 | 0.3421 | 0.5294 |
| Ratio (unique/nonunique): | | 1.8462 | 1.0000 | 1.3158 | 2.2727 | 1.9231 | 0.8889 |
| Haplotype class (the populations are considered in European context)*: | | | | | | | |
| Single unique: | No. | 8 | 2 | 7 | 11 | 6 | 4 |
| | Proportion | 0.2162 | 0.0833 | 0.1591 | 0.3056 | 0.1579 | 0.1176 |
| Multiple unique: | No. | 0 | 2 | 2 | 0 | 2 | 0 |
| | Proportion | 0.0000 | 0.0833 | 0.0455 | 0.0000 | 0.0526 | 0.0000 |
| Total unique: | No. | 8 | 4 | 9 | 11 | 8 | 4 |
| | Proportion | 0.2162 | 0.1667 | 0.2045 | 0.3056 | 0.2105 | 0.1176 |
| Nonunique: | No. | 29 | 20 | 35 | 25 | 30 | 30 |
| | Proportion | 0.7838 | 0.8333 | 0.7955 | 0.6944 | 0.7895 | 0.8824 |
| Ratio (unique/nonunique) | | 0.2759 | 0.0909 | 0.1892 | 0.4400 | 0.1875 | 0.1333 |
| Haplotype diversity | | 0.9898±0.0065 | 0.9636±0.0163 | 0.9804±0.0067 | 0.9837±0.0077 | 0.9875±0.0064 | 0.9796±0.0082 |

Note. - *See Appendix 4 for the list of populations.

**Table 5.13** Number of shared haplotypes (below the diagonal) and probability of identity (above the diagonal) for all 15 possible population pairs

| Population | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gagauzes (Kongaz) | | 0.0065 | 0.0065 | 0.0094 | 0.0100 | 0.0042 |
| 2. Gagauzes (Etulia) | 5 | | 0.0118 | 0.0067 | 0.0101 | 0.0020 |
| 3. Moldavians (Karahasani) | 7 | 9 | | 0.0081 | 0.0109 | 0.0093 |
| 4. Moldavians (Sofia) | 6 | 3 | 5 | | 0.0125 | 0.0078 |
| 5. Romanians | 8 | 7 | 12 | 7 | | 0.0046 |
| 6. Ukrainians | 4 | 3 | 6 | 6 | 5 | |

### 5.3.2 Analysis of Molecular Variance (AMOVA)

Analyses of the Molecular Variance (AMOVA) were performed to establish the apportionment of the genetic variance found in the present sample set (Table 5.9, page 53). The AMOVA results were broadly similar to those obtained with binary markers. The fraction of the genetic variance resulting from differences between populations ($F_{ST}$) was 2.05 (a value significantly different from zero, $P$=0.003), whereas the rest found within populations. When populations were grouped according to their linguistic affiliation not significantly different from zero variation ($F_{CT}$=1.01%; $P$=0.223) was found among groups, whereas significant differences at 5% level were found within groups ($F_{SC}$=1.33%; $P$=0.046), suggesting that linguistic affiliation has no genetic consistence.

### 5.3.3 Genetic relationships between populations

Binary marker ascertainment bias can lead to quite different conclusions about the same populations (Karafet *et al.* 2001), but this should not occur when unbiased markers are used that are variable in all populations. We therefore used microsatelite haplotype frequencies and the molecular differences between haplotypes to compute population genetic distances in the form of values of $R_{ST}$. The matrix of pairwise $R_{ST}$ values is shown in Table 5.14. Among populations under study $R_{ST}$ values were statistically significant at 5% level in 5 out of the 15 comparisons. The sample of the Gagauzes from Kongaz manifests the greatest differences with any other Dniester-Carpathian sample (P>0.05). The relative distances (as measured by $R_{ST}$ values) among the populations studied are displayed graphically in Figure 5.9. The most remote position occupies the sample from Kongaz. No correlation between genetic affinities and ethnical affiliations is observed in the Dniester-Carpathian tree.

**Table 5.14** Matrix of genetic distances ($R_{ST}$) among Dniester-Carpathian populations based on microsatellite haplotypes

| Population | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gagauzes (Kongaz) | - | | | | | |
| 2. Gagauzes (Etulia) | 0.0396* | - | | | | |
| 3.Moldavians (Karahasani) | 0.0483** | -0.0154 ns | - | | | |
| 4. Moldavians (Sofia) | 0.0175 ns | -0.0052 ns | -0.0028 ns | - | | |
| 5. Romanians | 0.0956** | 0.0062 ns | 0.0051 ns | 0.0103 ns | - | |
| 6. Ukrainians | 0.0459** | -0.0031 ns | 0.0148 ns | 0.0077 ns | 0.0419* | - |

Note. - ns: non-significant; *$P$<0.05; **$P$<0.01.

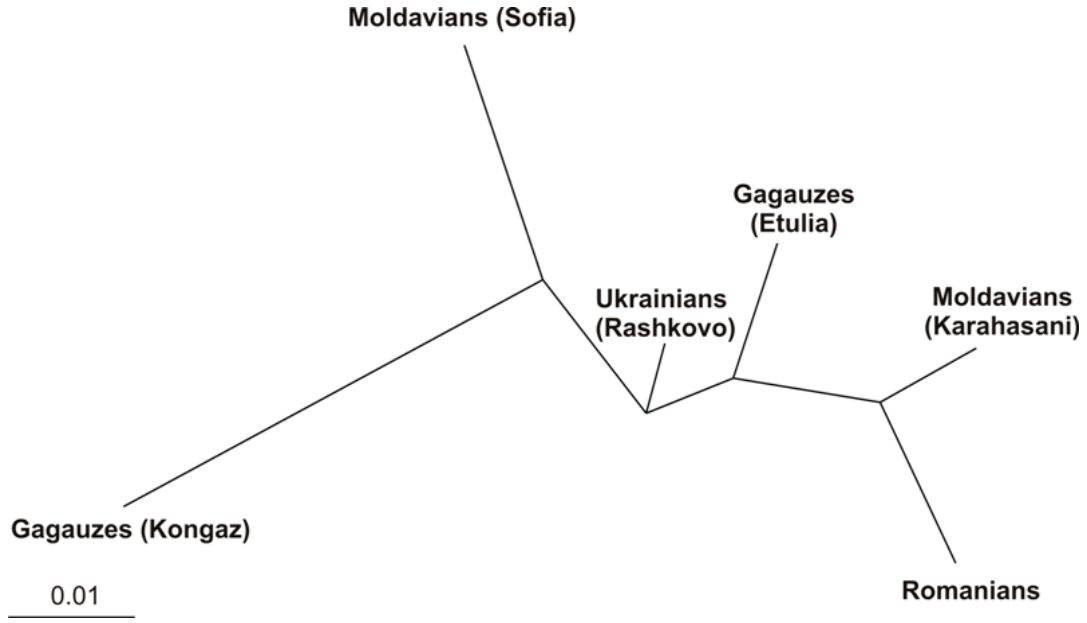


**Figure 5.9** Neighbor-Joining tree based on pairwise $R_{ST}$ values from Y-STR haplotypes of six Dniester-Carpathian populations.

Phylogenetic analyses was also performed by pooling the data of the present study with those of Zaharova *et al.* 2001; Ploski *et al.* 2002; Barac *et al.* 2003; Cinnioğlu *et al.* 2004; Bosch *et al.* 2006; Klaric *et al.* 2005; Pericic *et al.* 2005; Roewer *et al.* 2005. The matrix of pairwise $R_{ST}$ values is presented in Appendix 5. For most of the pairwise population comparisons, the interpopulation differences were significant. Eastern and western Slavic populations (Poles, Russians, Byelorussians, and Ukrainians) demonstrate the closest affinities with each other. This fact points out to a common ancestry for Slavic paternal gene pools. The results of the haplotype sharing and $R_{ST}$ analyses were not always correlated (Appendix 5). For example, the Albanians show considerable genetic distances with all the populations; however, 14 pair of comparisons between the Albanians and the

European populations show a high level of identity. This is explained on the basis of the number of mutational differences between nonidentical haplotypes.

In order to represent the genetic distances between samples, an MDS analysis was performed. Figure 5.10 shows the results of MDS based on $R_{ST}$ genetic distances. A good fit between the two dimensional plot and the source data (paiwise values of $R_{ST}$) was obtained, as demonstrated by the low stress value obtained (0.065). The plot of the first two dimensions of the MDS clustered most of the populations analyzed. The only exceptions are the Albanian and Estonian populations, which lie on the opposite ends of the second dimension and show statistically significant differences with all the samples ($P$<0.05). As expected from the $R_{ST}$ values matrix, the East Slavic populations formed a loose cluster, located on the lower left side of Figure 5.10. The sample from Moscow and the Ukrainian sample from Transdniestria are located at some distance from it. The western European populations occupy an opposite (right) side of the MD plot. On the whole, the first (horizontal) dimension shows a significant correlation ($r$= -0.662; $P$<0.0001) with geographical longitude, while the second (vertical) dimension demonstrates a significant correlation ($r$= -0.419; $P$=0.0079) with latitude. The populations from Southeast Europe demonstrate the most considerable interpopulation variability. Two Moldavian samples, the Gagauzes from Etulia and the Ukrainians from Transdniestria occupy an intermediate place in the MD space between the eastern European and the Balkan-Carpathian populations. The eastern Romanians show the most considerable proximity with the Bosnians and the Croatians. This observation agrees well with the results of the analysis of Y-chromosome binary polymorphisms. The Gagauzes from Kongaz show the most considerable affinity with southern and central Balkan populations - namely, with the Bulgarians, Bulgarian Turks, Macedonians, Northern Greeks, Serbs, as well as with the Hungarian sample from Budapest, the Romanian sample from Ploiesti and the Moldavian sample from Sofia. Though the analysis of the genetic distances did not reveal any considerable differences between the Gagauzes from Kongaz and the Balkan Turks, the differences between the Gagauzes and the Anatolian Turks are statistically significant. It is of worth to note that the two Turkish samples (the Anatolian and the Balkan ones) do not show any significant differences.

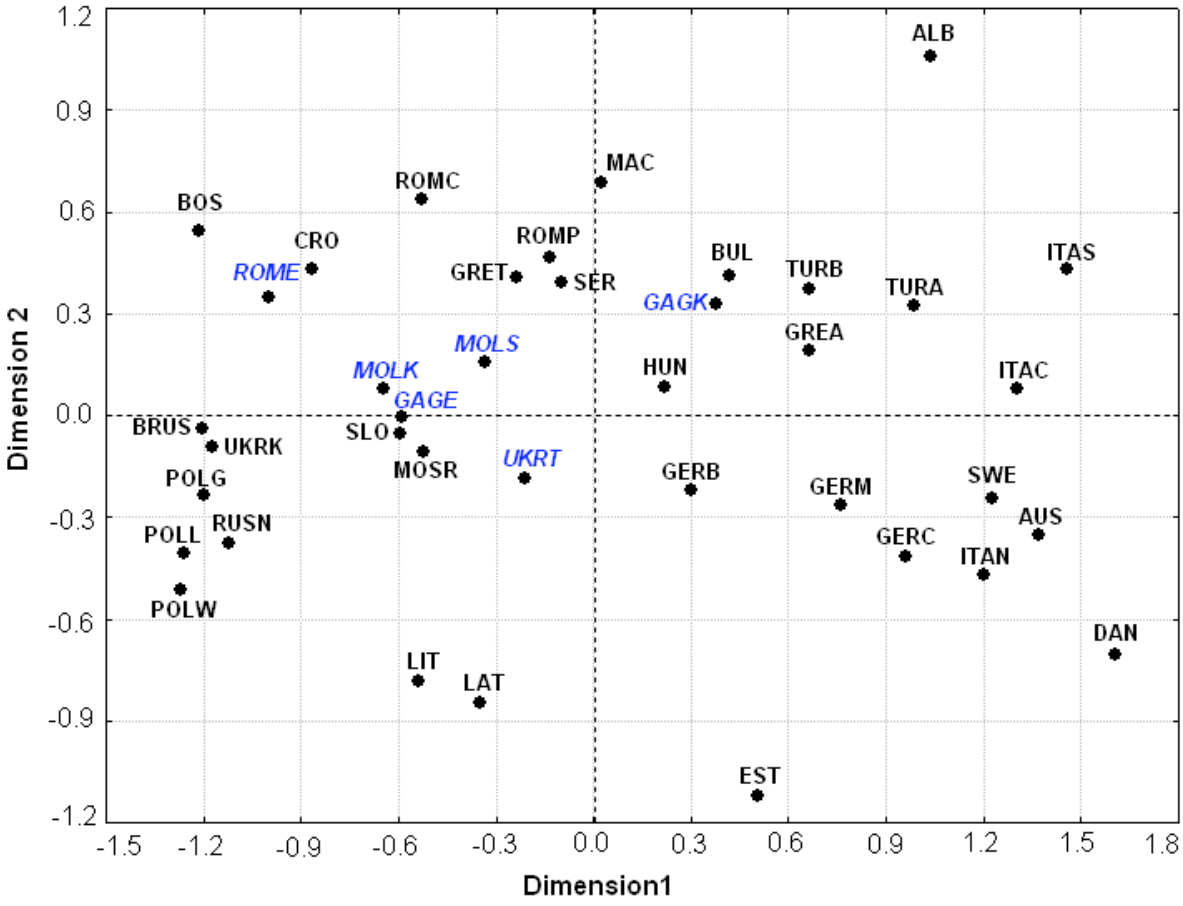


**Figure 5.10** Plot from multidimensional scaling (MDS) analysis of a $R_{ST}$ values from Y chromosome STR haplotype frequencies, showing genetic affinities among European and Anatolian populations. The populations presented are: GAGK=Gagauzes from Kongaz; GAGE=Gagauzes from Etulia; MOLK=Moldavians from Karahasani; MOLS=Moldavians from Sofia; ROME=Romanians from Piatra Neamt and Buhus; UKRT=Ukrainians from Trans-Dniestria (present study); MOSR= Moscow [Russia]; POLG=Gdansk [Poland]; POLW=Wrozlaw [Poland]; POLL=Lublin [Poland]; LIT=Vilnus [Lithuania]; LAT=Riga [Latvia]; EST=Tartu [Estonia]; HUN=Budapest [Hungary] (Ploski *et al.* 2002); RUSN=Novgorod [Russia]; BRUS=Byelorussia; UKRK=Kiev [Ukraine]; ALB=Albania; SLO=Ljubljana [Slovenia]; GERB=Berlin [Germany]; GERC=Cologne [Germany]; GERM=Munich [Germany]; AUS=Vien [Austria]; DAN=Denmark; SWE=Sweden; GREA=Athens [Greece]; ITAS=Sicily [Italy]; ITAC=Lazio [Italy]; ITAN=Lombardy [Italy] (Roewer *et al.* 2005); CRO=Croatia (Barac *et al.* 2003); TUR=Turkey (Cinnioğlu *et al.* 2004); BUL=Bulgaria [Bulgarians]; TURB=Bulgaria [Turks] (Zaharova *et al.* 2001); MAC=Macedonia (Pericic et al. 2005); BOS=Bosnia (Klaric *et al.* 2005); SER=Serbia (Barac Lauc *et al.* 2005); GRET=Thrace [Greece]; ROMC=Constanta [Romania]; ROMP=Ploiesti [Romania] (Bosch *et al.* 2006). The investigated in the present study populations are marked in blue.

### 5.3.4 Microsatellite diversity within haplogroups

Combining the binary markers with the microsatellite loci, we have calculated the diversity of the microsatellite haplotypes within the haplogroups based on the binary markers. A total of 171 combination binary marker/STR haplotypes were produced (Appendix 6). We obtained quantitative estimates of the microsatellite diversity within the haplogroups with the help of AMOVA. STR diversity parameters within haplogroups were calculated only

for haplogoups represented in our material by more than five Y-chromosomes. The phylogenetic trees, based on the algorithm of the median networks, were built with the help of the Network program to present the visual pattern of the existing diversity, as well as to reveal the detailed phylogenetic relations between the microsatellite haplotypes within the haplogroups (Bandelt *et al.* 1995; 1999). The median networks were constructed for the haplogroups that were the most common in the Dniester-Carpathian region - namely, R1a1-M17, I1b-P37, R1b3-M269 and E3b1-M78. The estimations were calculated both from our own data and the available literature data for the western Eurasian populations (Table 5.15) in order to make the phylogenetic analysis, which would reflect the existing pattern of the mutual evolution relations among the haplotypes. Some parameters of the median networks for the pooled Dniester-Carpathian and European population samples respectively are given in Tables 5.16 and 5.17.

**Table 5.15** Source data for the western Eurasian samples used in the STR diversity analyses

| Haplogroup | Samples |
|---|---|
| R1a1-M17 | Moldavians (2), Romanians (1), Ukrainians (1), Gagauzes (2) (this study); Islanders (1) (Helgason *et al.* 2000); Croats (1) (Barac *et al.* 2003); Turks (1) (Chinnioğlu *et al.* 2004); Ukrainians (1), Russians (1), Byelorussians (1) (Kharkov 2005) |
| I1b-P37 | Moldavians (2), Romanians (1), Ukrainians (1), Gagauzes (2) (this study); Turks (1) (Chinnioğlu *et al.* 2004); Ukrainians (1); Russians (1); Byelorussians (1) (Kharkov 2005) |
| R1b3-M269 | Moldavians (2), Romanians (1), Ukrainians (1), Gagauzes (2) (this study); western Europeans (1), Turks (1) (Chinnioğlu *et al.* 2004); Ukrainians (1), Russians (1), Byelorussians (1) (Kharkov 2005) |
| E3b1-M78 | Moldavians (2), Romanians (1), Ukrainians (1), Gagauzes (1) (this study); Turks (1) (Chinnioğlu *et al.* 2004); Albanians (1), Aromuns (3), Greeks (1), Macedonians (1), Romanians (2) (Bosch *et al.* 2006) |

Note. - Number of samples used for calculation of the time since of population divergence (see page 75) is given in brackets.

**Table 5.16** STR diversity parameters of the R1a1-M17, I1b-P37, R1b3-M269 and E3b1-M78 haplogroups in the European samples* considered jointly

| Haplogroup | $n$ | $k$ | $d$ | $\hat{D} \pm SD$ | $\hat{H} \pm SD$ | $\hat{\pi} \pm SD$ | FH |
|---|---|---|---|---|---|---|---|
| R1a1-M17 | 451 | 139 | 209 | 0.9730±0.0029 | 0.3940±0.2313 | 2.76±1.46 | 16-10-17-25-11-11-13 |
| I1b-P37 | 142 | 54 | 65 | 0.9474±0.0099 | 0.3071±0.1900 | 2.15±1.20 | 16-10-18-24-11-11-13 |
| R1b3-M269 | 198 | 103 | 153 | 0.9785±0.0047 | 0.4447±0.2565 | 3.11±1.62 | 14-10-16-24-11-13-13 |
| E3b1-M78 | 128 | 58 | 97 | 0.8519±0.0294 | 0.2343±0.1542 | 1.64±0.98 | 13-10-17-24-10-11-13 |

Note. - *See Table 5.15 for the list of samples. Column headings: $n$, number of chromosomes; $k$, number of haplotypes; $d$, number of mutations; $\hat{D}$, haplotype diversity; $\hat{H}$, averaged over seven STR loci gene diversity; $\hat{\pi}$, mean number of pairwise differences; FH, founder (ancestral) haplotype: DYS19-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393.

**Table 5.17** STR diversity parameters of the R1a1-M17, I1b-P37, R1b3-M269 and E3b1-M78 haplogroups in the Dniester-Carpathian samples considered jointly

| Haplogroup | $n$ | $k$ | $d$ | $\hat{D} \pm SD$ | $\hat{H} \pm SD$ | $\hat{\pi} \pm SD$ | FH |
|---|---|---|---|---|---|---|---|
| R1a1-M17 | 84 | 39 | 55 | 0.9616±0.0093 | 0.3688±0.2213 | 2.58±1.40 | 16-10-17-25-10-11-13 |
| I1b-P37 | 76 | 31 | 39 | 0.9467±0.0108 | 0.3084±0.1918 | 2.16±1.21 | 16-10-18-24-11-11-13 |
| R1b3-M269 | 45 | 30 | 53 | 0.9697±0.0152 | 0.4541±0.2657 | 3.18±1.68 | 14-10-16-24-11-13-13 |
| E3b1-M78 | 24 | 10 | 15 | 0.7500±0.0916 | 0.2329±0.1588 | 1.63±0.10 | 13-10-17-24-10-11-13 |

Note. - Column headings as in Table 5.16.

Haplogroup R1a1-M17 is the most common lineage in the Dniester-Carpathian region (27.1%). The samples from Kongaz, Sofia, Rashkovo and Karahasani are characterized by a high level of STR diversity within the R1a1-M17 haplogroup (Table 5.18). The sample from eastern Romania is inferior in this respect. The Gagauzes from Etulia show the least diversity level with an unexpressed founder.

**Table 5.18** STR diversity parameters of Hg R1a1-M17 in the Dniester-Carpathian populations

| Population | $n$ | $k$ | $\hat{D} \pm SD$ | $\hat{H} \pm SD$ | $\hat{\pi} \pm SD$ |
|---|---|---|---|---|---|
| Gagauzes (Kongaz) | 6 | 6 | 1.0000±0.0962 | 0.2381±0.1865 | 1.67±1.13 |
| Gagauzes (Etulia) | 11 | 4 | 0.6727±0.1232 | 0.2338±0.1681 | 1.64±1.04 |
| Moldavians (Karahasani) | 25 | 15 | 0.9300±0.0357 | 0.3376±0.2121 | 2.36±1.33 |
| Moldavians (Sofia) | 11 | 9 | 0.9778±0.0540 | 0.4127±0.2675 | 2.89±1.66 |
| Ukrainians | 21 | 13 | 0.9486±0.0231 | 0.3891±0.2390 | 2.72±1.50 |
| Romanians | 10 | 6 | 0.8444±0.1029 | 0.36833±0.2435 | 2.58±1.51 |
| Total | 84 | 39 | 0.9616±0.0093 | 0.3688±0.2213 | 2.58±1.40 |

Note. - Column headings as in Table 5.16.

The European median network of haplogroup R1a1-M17 has a complicated configuration with many reticulations and accumulations of certain haplotypes (Figure 5.11). In the European network the ancestral (or founder) haplotype is associated with East Slavic populations and, at the same time, the Eastern Slavs show the highest haplotype diversity, a fact that testifies to the origin of haplogroup R1a1-M17 within the limits of Eastern Europe (see also Kharkov 2005). The hypothetical ancestral haplotype in the Dniester-Carpathian region is represented in four out of six surveyed populations (Figure 5.12). It does not contain the chromosomes from the Romanian sample and the sample of the Gagauzes from Etulia. It is noteworthy that the ancestral haplotype in the surveyed sample differs from its European analogue by one repeat at the STR391 locus. The latter is present only in three Dniester-Carpathian samples: those of the Ukrainians, the Moldavians from

Karahasani and of the Gagauzes from Etulia. The hypothetical ancestral haplotype in the sample from the Dniester-Carpathian region is also most common haplotype in Anatolia and in the Balkans.

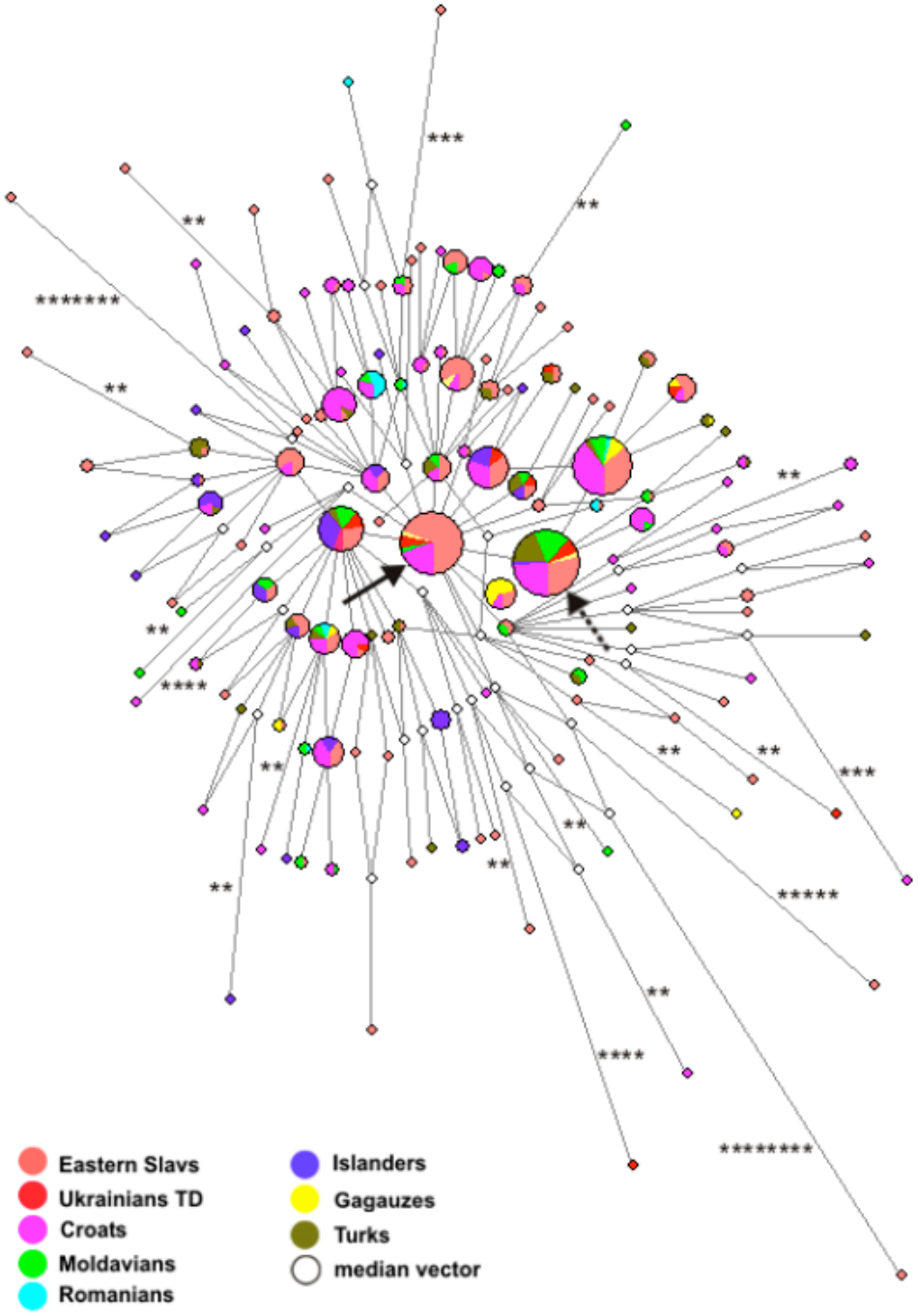


**Figure 5.11** Median-joining network constructed for the western Eurasian population sample, representing microsatellite variation within haplogroup R1a1-M17. Circles represent haplotypes, with areas proportional to the number of individuals they contain. Color indicates population of origin. Branch lengths are proportional to the number of mutational steps and the asterisks along the branches represent two and more mutational changes. The solid arrow points out the founder haplotype in the western Eurasian population sample and the dashed arrow indicates the ancestral haplotype in the Dniester-Carpathian sample.

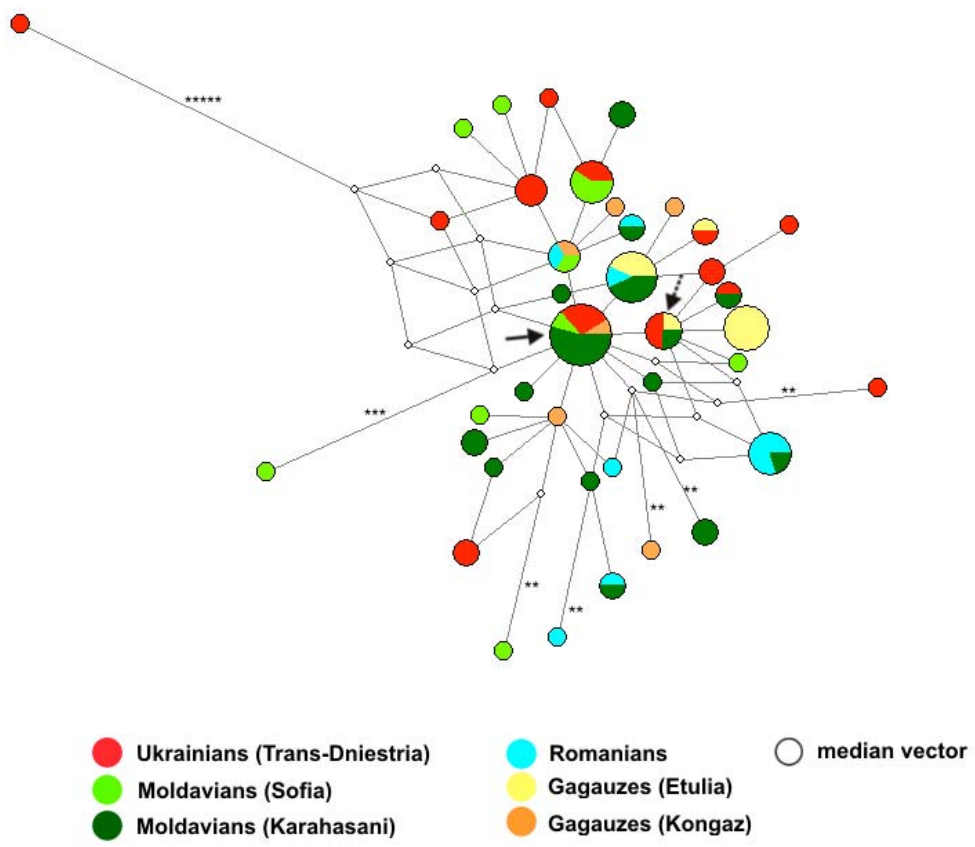


**Figure 5.12** Median-joining network constructed for the Dniester-Carpathian population sample, representing microsatellite variation within haplogroup R1a1-M17. The solid arrow points out the founder haplotype in the Dniester-Carpathian population sample and the dashed arrow indicates the ancestral haplotype in the European sample. Other designations are as in Figure 5.11.

The second most common haplogroup in the region under study is the I1b-P37 lineage. Within Europe it occurs with the highest frequency in the Croatian and the Bosnian samples (Marjanovic *et al.* 2005; Pericic *et al.* 2005), which are also characterized by a high microsatellite diversity within I1b-P37 (Rootsi *et al.* 2004; Pericic *et al.* 2005). Previously, Rootsi *et al.* (2004) revealed a high STR diversity of haplogroup I1b-P37 in the Gagauz sample from Moldavia and an extremely low diversity in the Moldavian sample from central Moldavia. In contrast to the results of Rootsi *et al.* (2004), our Moldavian samples show a considerably higher level of microsatellite diversity (Table 5.19). The extremely high diversity (0.9818) was revealed in the Ukrainian sample from Transdniesria. The Moldavians from Sofia and the Gagauzes from Kongaz are characterized by a relatively lower diversity. Remarkably, the European median network adopts a fairly compact star-like structure (Figure 5.13). The central haplotype from the Dniester-Carpathian sample is an ancestral haplotype in the sample of the European I1b-

P37 chromosomes. However, this haplotype is not predominant in the region. It is encircled by the other haplotypes, which have comparable frequencies (Figure 5.14). The supposed ancestral haplotype contains chromosomes from five surveyed samples. It does not contain the chromosomes from Etulia.

**Table 5.19** STR diversity parameters of Hg I1b-P37 in the Dniester-Carpathian populations

| Population | $n$ | $k$ | $\hat{D} \pm SD$ | $\hat{H} \pm SD$ | $\hat{\pi} \pm SD$ |
|---|---|---|---|---|---|
| Gagauzes (Kongaz) | 9 | 5 | 0.8611±0.0872 | 0.2738±0.1945 | 1.92±1.20 |
| Gagauzes (Etulia) | 9 | 7 | 0.9444±0.0702 | 0.2262±0.1678 | 1.58±1.04 |
| Moldavians (Karahasani) | 12 | 9 | 0.9545±0.0467 | 0.2273±0.1632 | 1.59±1.01 |
| Moldavians (Sofia) | 13 | 7 | 0.8718±0.0670 | 0.2381±0.1678 | 1.67±1.05 |
| Ukrainians | 11 | 10 | 0.9818±0.0463 | 0.4649±0.2924 | 3.26±1.81 |
| Romanians | 22 | 13 | 0.9481±0.0251 | 0.2721±0.1798 | 1.91±1.13 |
| Total | 76 | 31 | 0.9467±0.0108 | 0.3084±0.1918 | 2.16±1.21 |

Note. – Column headings as in Table 15

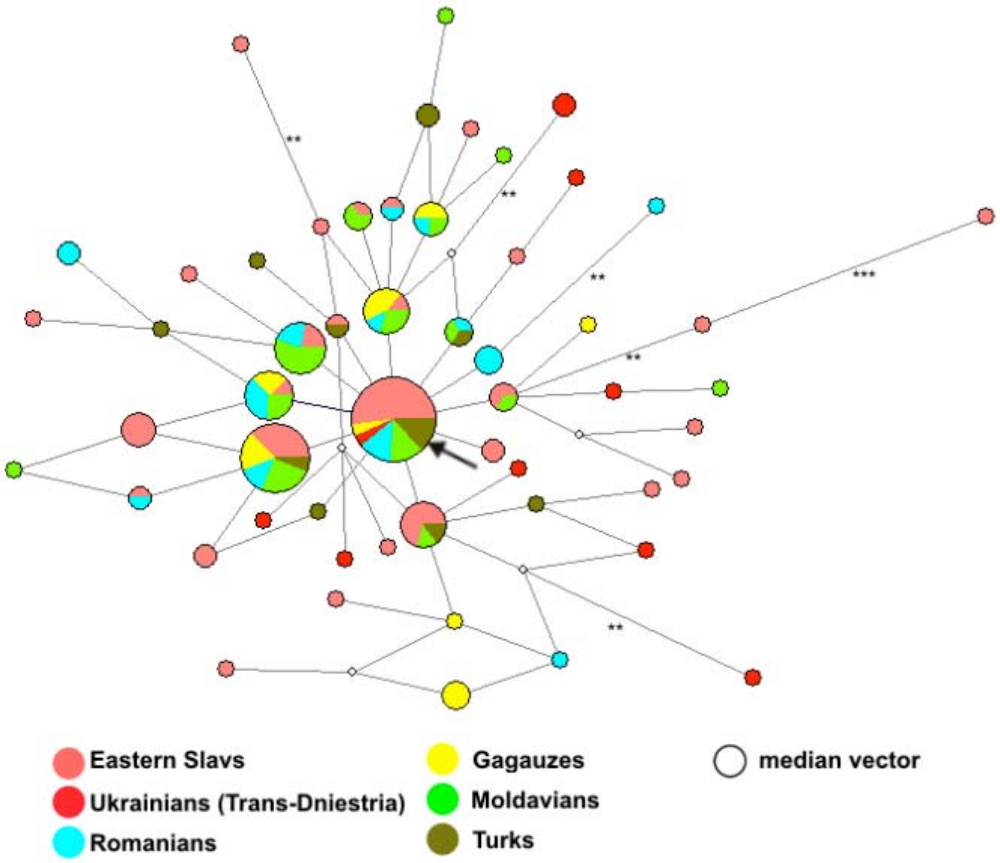


**Figure 5.13** Microsatellite network of the I1b-P37 haplogroup constructed for the western Eurasian population sample. Designations as in Figure 5.11.

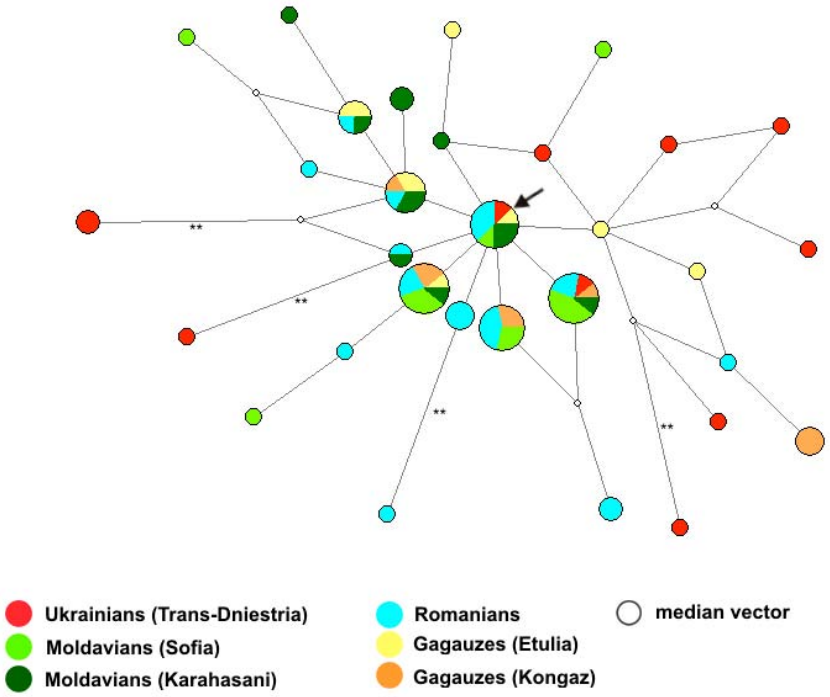


**Figure 5.14** Microsatellite network of the I1b-P37 haplogroup constructed for the Dniester-Carpathian population sample. Designations as in Figure 5.11.

Three samples from the Dniester-Carpathian region (the Ukrainian sample and both Moldavian samples) are characterized by a high STR diversity level within haplogroup R1b3-M269 (Table 5.20). The lowest diversity was noted in the Romanian sample, indicating a strong founder or sample effect. The network of haplogroup R1b3-M269, formed on the basis of the joint European and Anatolian data, is characterized by the presence of two expressed branching centers, which can be conditionally designated as “European” and “Anatolian” (Figure 5.15). They differ by one mutational step at the DYS393 locus. The majority of the chromosomes from the Dniester-Carpathian populations belong to the “European” cluster, and the founder-haplotype in the Dniester-Carpathian sample is a constituent portion of the western Eurasian ancestral haplotype. In the Dniester-Carpathian region the founder haplotyte is represented by the chromosomes from three samples: the Romanian, the Gagauz from Kongaz and the Moldavian from Karahasani (Figure 5.16). More than half of the Romanian R1b3-M269 chromosomes were classified to this haplotype. The majority of the samples have no distinct center and their haplotypes are unevenly scattered around the central haplotype in the region.

**Table 5.20** STR diversity parameters of Hg R1b3-M269 in the Dniester-Carpathian populations

| Population | $n$ | $k$ | $\hat{D} \pm SD$ | $\hat{H} \pm SD$ | $\hat{\pi} \pm SD$ |
|---|---|---|---|---|---|
| Gagauzes (Kongaz) | 5 | 4 | 0.9000±0.1610 | 0.4286±0.3130 | 3.00±1.87 |
| Gagauzes (Etulia) | 6 | 4 | 0.8667±0.1291 | 0.2857±0.2151 | 2.00±1.30 |
| Moldavians (Karahasani) | 12 | 11 | 0.9848±0.0403 | 0.4935±0.3049 | 3.46±1.90 |
| Moldavians (Sofia) | 8 | 7 | 0.9643±0.0772 | 0.4592±0.3014 | 3.21±1.85 |
| Ukrainians | 7 | 6 | 0.9524±0.0955 | 0.4150±0.2830 | 2.91±1.73 |
| Romanians | 7 | 3 | 0.6667±0.1598 | 0.1497±0.1284 | 1.05±0.78 |
| Total | 45 | 30 | 0.9697±0.0152 | 0.4541±0.2657 | 3.18±1.68 |

Note. - Column headings as in Table 5.16

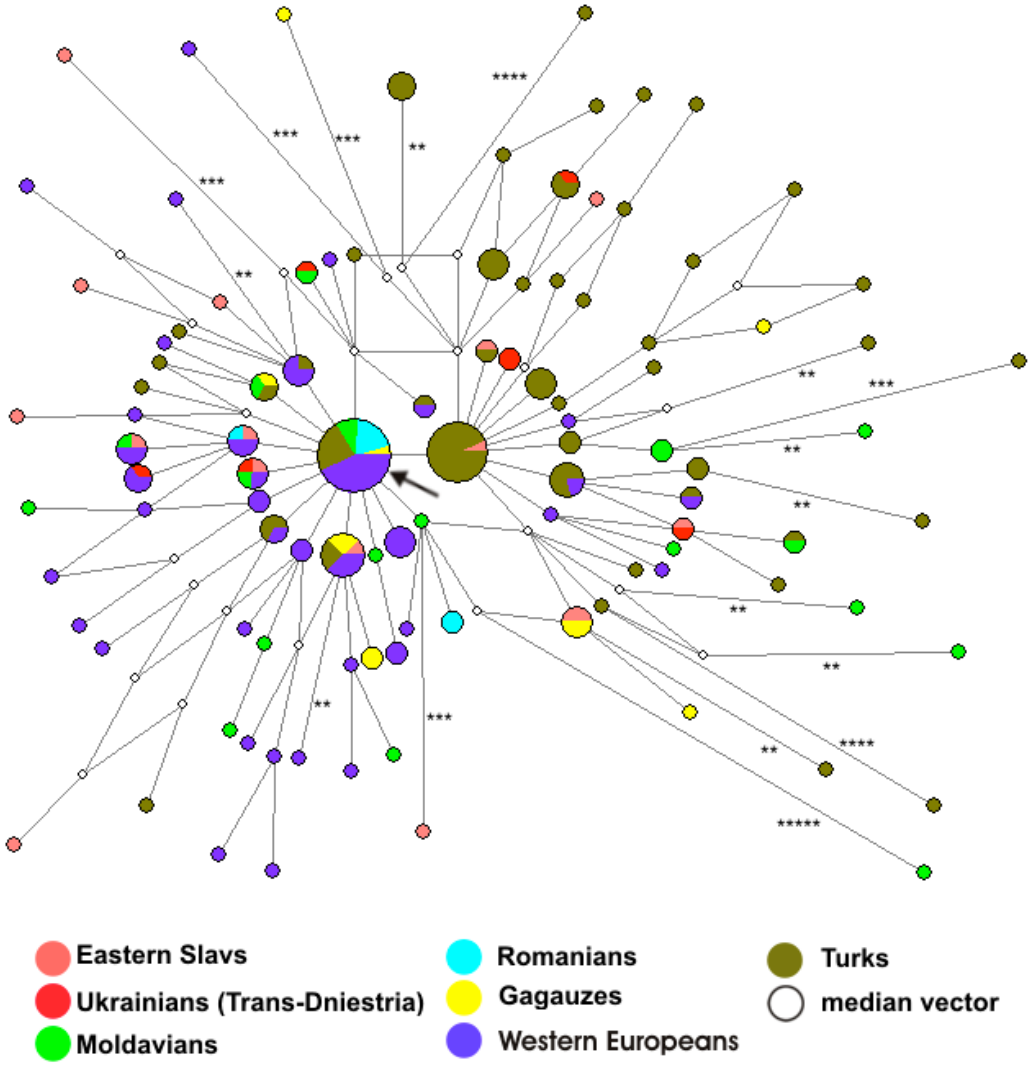


**Figure 5.15** Microsatellite network of the R1b3-M269 haplogroup constructed for the western Eurasian population samples. Designations as in Figure 5.11.

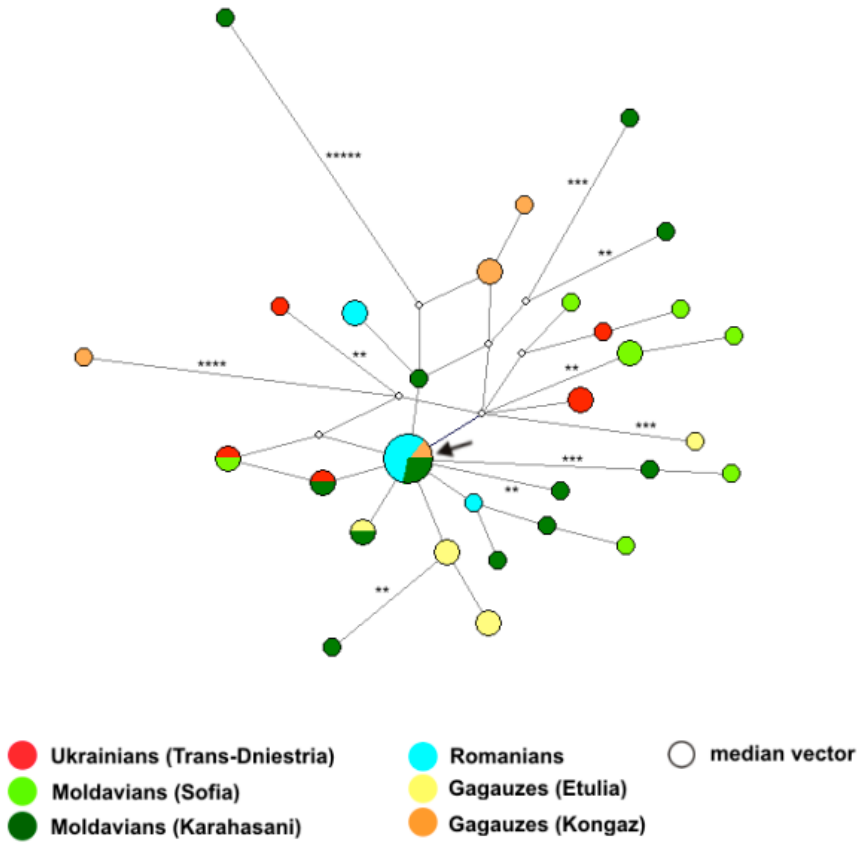


**Figure 5.16** Microsatellite network of the R1b3-M269 haplogroup constructed for the Dniester-Carpathian population samples. See Figure 5.11 for designations.

The E3b1-M78 chromosomes display a star-like network with a marked founder haplotype shared among all compared populations (Figures 5.17 and 5.18). The halplogroup E3b1-M78 is characterized by low microsatellite diversity indices (Table 5.16, 5.17) with a founder-haplotype containing 37 and 50 percents of the E3b1-M78 chromosomes in the European and Dniester-Carpathian population samples respectively. It is notable that this haplotype is the major STR-defined lineage in the pool of the Dniester-Carpathian Y-chromosomes.

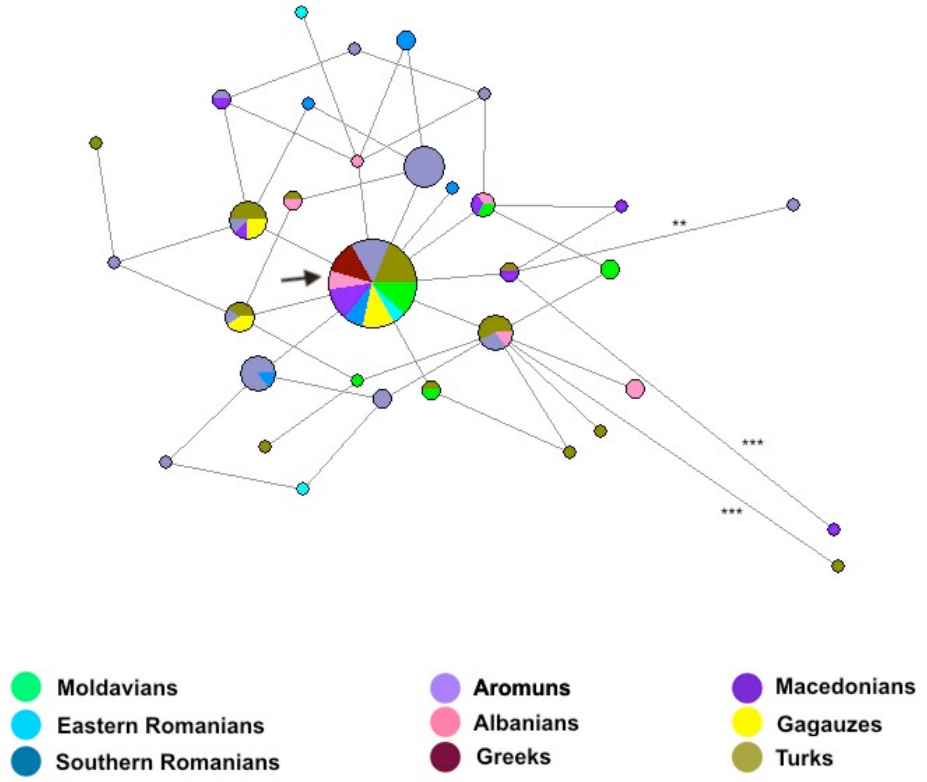


**Figure 5.17** Microsatellite network of the E3b1-M78 haplogroup constructed for the western Eurasian population samples. See Figure 5.11 for designations.

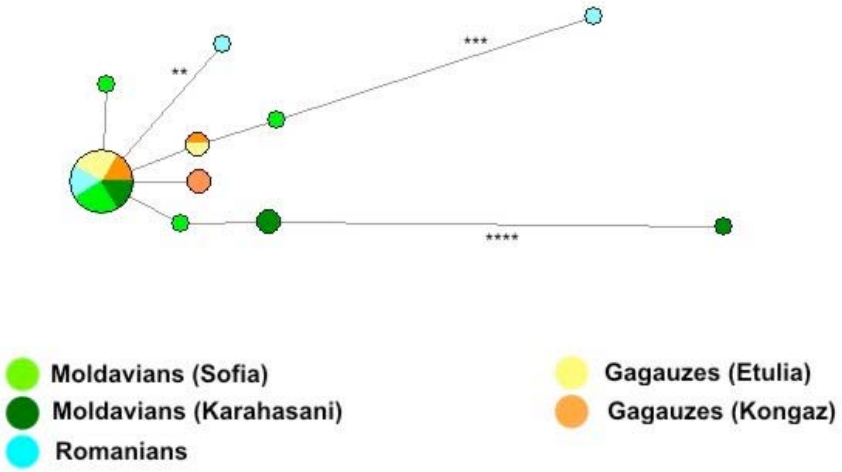


**Figure 5.18** Microsatellite network of the E3b1-M78 haplogroup constructed for the Dniester-Carpathian population samples. See Figure 5.11 for designations.

### 5.3.5 Age estimates of the predominant in the Dniester-Carpathian region haplogroups

The observed inter-molecular diversity of the lineages was the result of the haplotype evolution during a long historical period and is a function of the age of the lineages, the mutation rate of the microsatellites and the demographic history of populations. Since the mutation tempo of the microsatellite loci at the Y-chromosome and the diversity observed in the contemporary populations allow making their estimation, one can calculate the origin time of the lineages or the parameters of the demographic history of the population (the founder size) by estimating their diversity. The time estimates were determined for the four haplogroups which occurred in the Dniester-Carpathian region with the highest frequency. The tempo necessary for the mutation of the microsatellite loci was taken from Zhivotovsky *et al*. (2004) and is equal to $6.9 \times 10^{-4}$ per locus for 25 years. The age of STR variation was determined on the basis of our own and the pooled European data, with the help of two methods described in the papers of Zhivotovsky *et al.* (2004) and Forster *et al.* (1996; 2000). The age of the population expansions was estimated using the method of Zhivotovsky (2001; 2004) on the basis of the European pooled data only. It should be noted that the absolute estimations of the age parameters (in the terms of years) should be interpreted very carefully. As for any model, the methods that were used do not take into consideration a number of specific parameters (the demographic and the molecular ones) and that is why they are not without certain drawbacks (Stepanov 2002). Nevertheless, these methods are very useful for understanding the entire demographic picture. The time estimates of the microsatellite diversity are presented in Table 5.21. The age of STR variation within various haplogroups does not exceed the age of the Last Glacial Maximum (LGM, 14-20 *kya*). Haplogroups R1b3-M269 and R1a1-M17 are the “eldest” ones when compared to the “middle” I1b-P37 and the “youngest” E3b1-M78 lineages. The expansion time of the R1b3-M269 and R1a3-M17 lineages corresponds in geological terms to the interval between the LMG and the Younger Dryas (14-12.5 *kya*). From the viewpoint of archeology it was the period of the large expansion of the Upper Paleolithic culture of Madlen and the beginning of the Mesolithic epoch in Western Europe. The expansion of the I1b-P37 lineages took place obviously not earlier than in the Younger Dryas (12.5–11 *kya*) and no later than in the Neolithic (9–7 *kya*). The E3b1-M78 expansion period corresponds to the Late Neolithic – the Early Bronze Age.

**Table 5.21** Age estimates of haplogroups R1a1-M17, I1b-P37, R1b3-M269 and E3b1-M78

| Haplogroup | Age (*ky*) of STR variation | | | | Time since population divergence (*ky*) according to Zhivotovsky *et al.* 2001, 2004 (only for European data) | |
|---|---|---|---|---|---|---|
| | According to Zhivotovsky *et al.* 2004 | | According to Forster *et al.* 2000 | | Upper bound | Lower bound |
| | Present study | European data* | Present study | European data* | | |
| R1a1-M17 | 15.40±3.26 | 12.48±2.88 | 10.04±2.35 | 10.94±2.42 | 11.45±2.68 | 1.44±0.40 |
| I1b-P37 | 13.64±4.25 | 9.43±3.48 | 8.37±2.07 | 8.02±1.77 | 9.14±3.35 | 0.90±0.40 |
| R1b3-M269 | 17.60±1.69 | 13.46±1.51 | 12.08±2.17 | 11.94±2.63 | 12.85±1.52 | 2.98±0.68 |
| E3b1-M78 | 5.39±1.54 | 5.80±1.46 | 5.39±1.66 | 4.95±1.35 | 5.58±1.64 | 1.98±0.75 |

Note. - * See Table 5.15 for the list of samples.

# 6 DISCUSSION

## 6.1 *Alu* insertion polymorphisms in the Dniester-Carpathian populations

### 6.1.1 Variation pattern of *Alu* insertions in Southeastern Europe

Classical polymorphic markers (i.e., blood groups, protein electromorphs, and HLA antigens) had revealed that Europe is a genetically homogeneous continent, with only a few outliers (Saami, Sardinians, Icelanders, and Basques) (Cavalli-Sforza *et al.* 1994). Resent studies of autosomal DNA polymorphisms confirmed a high degree of homogeneity among European populations. 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 (Nasidze *et al.* 2001). Second, by small genetic distances: 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 with much longer branches (Jorde and Wooding 2004; Tishkoff and Kidd 2004). The Dniester-Carpathian autosomal pools also follow this rule. Our analysis of 12 autosomal DNA polymorphisms in the Dniester-Carpathian region has shown that the allele frequencies in these populations are strikingly similar to each other, as well as to the frequencies observed in other European populations, despite notable linguistic differences.

The genetic homogeneity among southeastern European populations suggests either a common ancestry of all southeastern European populations or a strong gene flow between populations that eliminated any initial differences. Taking into account that the region had a relatively high population density since the Neolithic, and that this region has been a crossroad of the routes connecting cultural centers of Middle East with different European areas, both explanations are plausible.

The low level of genetic differentiation of the southeastern European gene pool and the lack coordination between linguistic and genetic spatial patterns, make the further analysis of the population structure in this region very difficult. Nevertheless, our results demonstrate a certain degree of differentiation. The first principle component that explains 24% of the total genetic diversity is considerably correlated with the geographical latitude. The other components find no reasonable interpretation. This fact corresponds to only a low correlation between geographical and genetic distances and absence of robust clusters in the consensus tree.

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; Chikhi *et al*. 1998; Malaspina *et al*. 2000) and are 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 those in the eastern Mediterranian area. 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 were penetrating into the Carpathian basin and into the Balkans from the Pontic steppes. These cultures developed on the 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 (Velikanova 1975; Kruts *et al.* 2003) imply different structure of their 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') to the Near-Eastern ('Neolithic') components in their gene pools.

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

Several hypotheses about the origin of the Gagauzes (who speak a Turkic language) have been proposed (Guboglo 1967). The opposing points of view consider them either descendents of the Turkic nomadic tribes from South Russian steppe (Kumans, or Pechenegs, or Torks, etc.) or descendents of the Anatolian Turks (Seljuks and/or Ottomans). Since each of these scenarios implies a distinct genetic structure of the Gagauz populations, they can be tested by the means of population genetic analysis. Importantly, both hypotheses imply that Gagauzes should be genetically more similar to some Turkic populations (from Eurasian Heartlands for the first of hypothesis or from Anatolia for the second one). Our previous analysis of classical polymorphisms in the Dniester-Carpathian region demonstrated however that Gagauzes grouped genetically with their geographic neighbors, rather than with any Turkic populations (Varsahr *et al.* 2001; Varsahr *et al.* 2003). The present analysis, based on the autosomal DNA markers, is consistent with the results shown by classical genetic markers. The Gagauz samples differed from Central

Asian populations in the PC analysis and also revealed considerable genetic distances from them. Moreover, the genetic position of the Gagauzes in the tree was not an intermediate one between southeastern European and Central Asian populations. Therefore our data testify against the hypothesis that the Gagauzes are direct biological descendents of the Turkic nomads from South Russian steppes.

According to another scenario the Gagauzes are descendants of the Seljuk Turks who migrated to the Balkans from Anatolia in the second half of the 13th century. Nevertheless, Gagauzes showed tighter relationships with the Dniester-Carpathian populations, than with the Turks from Anatolia and Cyprus. It should be noted however that the differences between the populations mentioned above are not significant enough to completely exclude the hypothesis of the Seljuk origin of the Gagauzes basing on the used marker system. A drawback of this scenario is that it does not explain the presence of the Kypchak (Tartar) element in the Gagauz language, which could have penetrated in it only by the northern way 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 proceeded here relatively independently from each other. The genetic landscape of Southeast Europe had presumably been formed long before the linguistic/ethnic landscape we now observe was shaped (one more option is that the cultural barriers were not strong enough to prevent genetic flow between populations). A Turkic language of the Gagauzes could be a case of language replacement. Replacement could occur via the “elite dominance” model – in this case the original Turkic migrant groups could be very small which would explain their negligible genetic effect on the resident groups (Renfrew 1987). However, elite dominance scenario is more suitable for more numerous populations, as e.g., those of Anatolia or Azerbaijan (70 and 8 million, respectively). The Gagauzes are much less numerous (~200,000). It is still possible that they are a remnant of a sometime larger Turk-speaking Orthodox group in Southeastern Europe.

In conclusion, our study of *Alu* polymorphisms indicates low level of population differentiation in the Dniester-Carpathian region as well as in Southeast Europe. Although the interpopulation diversity within Southeast Europe is small our population tree and PC plot allow the distinction between South and North. These observations are in agreement with classical and STR markers showing small and clinal geographical variation within the

European subcontinent. The genetic affinities among Dniester-Carpathian and southeastern European populations do not reflect linguistic relationships; overall, these results indicate that the ethnic and genetic differentiations proceeded in these regions to a considerable extent independently from each other.

## 6.2 Y-chromosomal DNA variation in the Dniester-Carpathian region

### 6.2.1 On the origin of Y-chromosome diversity in the eastern Trans-Carpathians

Analyses of molecular variance and population relationships showed that Dniester-Carpathian populations do not constitute a homogenous group with close affinity to a specific western Eurasian cluster. In case of Y-chromosome, the proportion of the genetic variation that is due to interpopulation differences is 5-6 times higher, than in case of aoutosomal loci. This fact is not surprising. A similar situation was found after comparison of proportions both in the world-wide level and within continents. This fact is obviously due to smaller effective size of Y-chromosome in comparison to autosomes. Moreover, autosomal loci are inherited both paternally and maternally. Therefore a lower demographic mobility of males in comparison to females could probably facilitate a higher geographical differentiation of the Y-chromosomal pool in comparison to the autosomal one.

The contribution of various source areas of Western Eurasia to the paternal gene pool of the Dniester-Carpathian region corresponds well to the role of these areas in the cultural development of the region (Dergachev 1999). From the results of gene frequency and admixture analyses we infer that the flows from the Western Balkans and Eastern Europe played a major role in the formation of the structure of the Dniester-Carpathian paternal gene pool. Obviously, it is explained by the geographical proximity of the Dniester-Carpathian region to the Balkan and North Pontic cultural centers and, as a consequence, the region came every now and again under the influence of one of them or was subject to the bilateral influence in the course of various historical periods.

Migrations from the western Balkans were the main source of the I1b-P37 haplogroup in the male pool of the Dniester-Carpathian region. Pericic *et al.* (2005) asserted that the genetic expansion of the I1b-P37 lineages probably took place during the Mesolithic, not earlier than the Younger Dryas to Holocene transition (~11,000 *ya*) and not later than the early Neolithic (~8,500 *ya*). The results of our estimates with the use of seven STR loci places the beginning of the STR variation within haplogroup I1b-P37 (9.43±3.48 by Zhivotovsky *et al.* 2004) somewhere between the Pleistocene and the Holocene and thus support the finding of Pericic *et al.* (2005). The first wave of the spatial expansion of haplogroup I1b-P37 took place during this time. The further expansion of the I1b-P37 lineages from the Middle-Danubian Lowland may be associated at least with two events in

the history of Europe. It is known that the early farming communities of central and eastern Europe originated in the Middle Danube area in the Middle Neolith (8,000-7,500 *ya*) (Mongait 1973; Whittle 1996). The expansion of the haplogroup I1b-P37 lineages from the Middle Danube Lowland in this period is explained by the "wave of advance" model initially suggested by Ammerman and Cavalli-Sforza (1984). According to this scenario, the endemic Balkan lineages should have penetrated into European gene pools together with the Near Eastern haplogroups. The second cataclysm, probably associated with population growth and demographic migrations, is connected with the spread from the Middle Danube Lowland of the advanced Early Hallstatt technologies and traditions in the period of the transition to the Iron Age (1,300–1,100 years BC) (Dergachev 1997).

The R1a1-M17 haplogroup represents the East European stratum in the paternal gene pool of the discussed populations. It shows a clinal frequency distribution across Europe with a frequency peak in Eastern Europe. This spatial pattern has been associated with various ancient population movements from North Black Sea Littoral towards West and Southwest, namely with (1) population expansion from eastern European (Dnieper-Donetsk) transglacial refugium at the end of the Upper Palaeolithic/Mesolithic time (15,000 – 12,500 BC), (2) with the migration of the peoples of Kurgan cultures at the Middle Eneolithic-Bronze epochs (~ 4,400 – 1,200 BC), and (3) with the Great Slavic colonization of Southeast Europe at the beginning of the Middle Ages (5th-7th centuries AD) (Rosser *et al.* 2000; Semino *et al.* 2000; Pericic *et al.* 2005). Our estimates of the age of STR variation within the R1a1-M17 haplogroup support all these scenarios. The geographic expansion of the haplogroup R1a1-M17 lineages began in the post-glacial period (in the Late-Upper Paleolithic/Mesolithic) and continued for several millenniums till the Middle Ages. In Southeast Europe one finds the clearly central haplotype, different from the ancestral European haplotype. Therefore it seems very likely that one of the above mentioned scenarios indeed describes a key-episode in the genetic history of Southeast Europe. A better resolution of these migration patterns requires more extensive sampling of European populations.

Another marker indicating influences from the East is haplogroup N3a, defined by M178. High frequencies of the N3a-M178 haplogroup and the world maximum of its microsattelite diversity were registered in some Finno-Ugric populations, which suggest its origination in Northeastern Europe in the late Upper Paleolithic (Rootsi *et al.* 2000; Tambets *et al.* 2004). The increased frequency of the N3a-M178 lineages in a number of

eastern Slavic populations could be explained by the assimilation of the Finnish tribes during the migration of the Slavs eastwards from Central Europe. This haplogroup is rare or absent in southeastern European populations. The presence of the N3a-M178 lineages in Dniester-Carpathian populations can be explained as the consequence of either the massive Slavic migration during the early medieval period or the earlier migrations of peoples directly from the Volga-Ural-Siberian area. The latter migration is reliably documented in archeological records and may be illustrated by the wide diffusion of the Seiminsko-Turbinsk antiquities in the northwestern Black Sea Littoral in the Late Bronze Age (Hansel 1982; Dergachev and Bochkarev 2002).

The third important genetic stratum in the male pool of the Eastern Trans-Carpathians is represented by lineages defined by mutations at 12f2, YAP and M201 loci. The expansion of the Near Eastern lineages is traditionally associated with the settling of the Near Eastern agriculturists during the Neolithic (Semino *et al.* 2000). The abundant archeological and paleoanthropological Neolithic material from the Danubian-Carpathian region testifies clearly to a strong Near Eastern genetic influx to the Danube-Carpathian area during the Neolithic and the Early Eneolithic periods (Dergachev 1999; Kruts *et al*. 2003). Moreover, recent studies of the diversity of Y-chromosome lineages in western Eurasia showed that the penetration of Middle East lineages into Europe took place also in post-Neolithic time (Cruciani *et al.* 2004; Di-Giacomo *et al.* 2004; Semino *et al.* 2004). Our estimates of the age of the E3b1-M78 haplogroup in Southeast Europe and in the Dniester-Carpathian region conform to these findings.

The E3b1-M78 and J-12f2 haplogroups show in Europe a clear clinal reduction of frequency from Near East towards Europe. The G-M201 haplogroup deviates from this pattern. It has maximal frequency in the North Caucasian region (Semino *et al.* 2000; Nasidze *et al.* 2003; Nasidze *et al.* 2004) and relatively high frequencies in Turkey and southern Italy, (Semino *et al.* 2000; Di Giacomo *et al.* 2003; Cinnioğlu *et al.* 2004) while in the Middle East countries it occurs with a low average frequency (Semino *et al.* 2000; Hammer *et al.* 2000; Al-Zahery *et al.* 2003). Insufficient understanding of the phylogeography of haplogroup G-M201 does not allow us to establish its origin. Interestingly, a higher frequency of G-M201 is observed in the Lower Danube area not only in the Gagauzes, but also in the Romanians (Bosch *et al.* 2006). This fact may be tentatively interpreted as an evidence of close ancient connections of populations from the Lower Danube zone with the North Caucasian and/or the Anatolian populations. The

further studies of the distribution of the G-M201 lineages in western Eurasia are necessary to understand their origin and the ways of their diffusion.

The extensive archaeological studies in the Dniester-Carpathian region demonstrated an only moderate influence of Western Europe on the culture of the region (for review see Dergachev 1999). Our study of Y-chromosome diversity in the Dniester-Carpathian populations conforms to these findings: the western European stratum is the least important one in the Dniester-Carpathian paternal gene pool. There is no evidence of massive migrations of western European tribes into the eastern Transcarpathians for the period from the Mesolithic to Roman times. In this period the penetration of the western European lineages into the Danubian-Carpathian region might have had a diffusive nature. The migrations of the Germanic tribes of the Bastarns and the Goths were the first massive invasions from Western Europe (Sedov 2002; Shschukin 2005). It looks possible that the gene pools of the contemporary peoples of the eastern Transcarpathians owe a considerable proportion of the western European lineages to the Bastarns and the Goths. To test this assumption, larger samples of the major western European lineages R1b3-M269 and I1a-M253 from the area are required.

The frequencies and the STR diversities of the R1b-P25 haplogroup are known to have uncoordinated spatial distributions in Europe (Pericic *et al.* 2005; Cinnioğlu *et al.* 2004). While the haplogroup R1b-P25 frequency shows a decline from western towards eastern and southeastern Europe, the spatial distribution of STR variance within R1b-P25 shows a different pattern, a one with multiple peaks in Europe and Asia Minor. Importantly, a major R1b3-M269 cluster uniting lineages from Asia Minor exceeds in the STR diversity level even the Iberian cluster (Cinnioğlu *et al.* 2004). This fact hinders identification of the origin center of the R1b-P25 haplogroup and the ways of its diffusion. Pericic *et al.* (2005) suggested a possible concurrent dissemination of the R1b-P25 lineages from Asia Minor and Iberian Peninsula during re-peopling of Europe in the Late Paleolithic and Holocene.

### 6.2.2 Origin and population history of the Romanians, the Moldavians and the Gagauzes: evidence from the Y-chromosome

The migrations, associated with the expansion of the major western Eurasian lineages, took place in the remote historical periods. These migrations, which involved the vast areas of the Europe, were the main reasons for the currently observed clines of the genetic frequencies, which crossed over the sub-continent. Sometimes the so-called genetic

boundaries, the geographic areas associated with a considerable genetic change, single out within continues European landscape. One of such boundaries crosses the Dniester-Carpathian region, as for the first time it was demonstrated by Stefan *et al.* (2001) and confirmed by the present survey. It seems, that the reasons, which conditioned the differentiation of the Dniester-Carpathian populations, lie in the recent demographic events of local significance, otherwise the differences between the local populations will disappear in the course of time in the absence of any geographic and linguistic (in the case of the eastern Romanic population) barriers. Moreover, the similar reasoning is supported by historical facts. It is known that the lands to the south and to the east of the Carpathians were poorly populated in the 11th – 13th centuries AD due to devastating raids by the Turkic nomads from the North Pontic steppes (Fedorov 1999). From the 13th century the old-Romanic population (Volokhs), the direct ancestors of the contemporary Moldavians and the Romanians, penetrated there from the adjacent territories of Southeast Europe. Simultaneously or a little later the Slavs settled down predominantly in the Dniester valley.

A high share of the Anatolian/southern Balkan stratum in the male pool of the southern Romanians and as a consequence their close genetic affinity with the autochthonous Balkan populations testify to a significant gene flow from the southern/central Balkans and thus support the migration concept of the origin of the Romanians (for review see Fedorov 1999). A considerable prevalence of the western Balkan component over the Anatolian one and a moderate share of the eastern European component in the pool of the eastern Romanians and the northern Moldavians may be attributable to the peopling of the eastern Transcarpathians from Transylvania and in this way is more consistent with the theory of the autochthonous (within the Carpathian Basin) development of the Romanians and the Moldavians. As we see, no theory (the migration one or that of the autochthonous development) explains completely the observed variability of the Y-chromosome in the gene pool of the Romanians and the Moldavians, but it does not confront with the observed variability either. The results of the study of the Y-chromosome polymorphism testify to the mixed origin of the male pool of the East Romanic population. It seems that probably the East Romanic expansion came from two distinct areas in the Medieval Ages. At the same time the Balkan Volokhs (the old-Romanian community) preferred to settle down on the lands, which were in close vicinity of the Balkans to the South of the Carpathians, whereas the Carpathian Volokhs settled down in the eastern Transcarpathians. The gene pools of at least some Moldavian groups, except the Balkan-Carpathian components, also

included a considerable eastern European component that seems to be attributable to the involvement of settlers of Slavic origin. The presence of multiple Slavic elements in the spoken language and folklore of the Moldavians supports the interpretation that Slavs left significant imprint on the genesis of the present-day paternal pool of the Moldavians.

Among the peoples of the Dniester-Carpathian region the Gagauzes are characterized by the highest proportion of the Near Eastern lineages. This fact agrees with the historically documented information on the migration of the Gagauzes to the Southern Bessarabia from the territory of the Balkan Peninsula at the end of 18th – the beginning of 19th centuries. Despite a comparatively big share of the Anatolian/southern Balkan stratum in the Gagauz paternal gene pool, the proportion of the Near Eastern to European lineages in the Gagauz populations is considerably lower than that of the Turks. Moreover, the distribution of the Near Eastern lineages in the Gagauz and in the Anatolian populations also differs. The Gagauzes represent a European population in respect of the E3b1-M78 to E3b3-M123 and J2-M172 to J1-M267 ratios. This finding testifies to the emergence of the Near Eastern lineages in the Gagauz paternal gene pool, probably, long before the penetration of the Seljuk Turks and the Osman Turks into the Balkans. We come also to this conclusion analyzing STR haplotype sharing among southeastern European and Anatolian populations. The Gagauzes share considerably more haplotypes with the Balkan populations than with the Turks from Anatolia. The analysis of the genetic distances confirms this reasoning. In population comparisons the Gagauzes are more closely related genetically to the neighboring southeastern European groups than to linguistic-related Anatolian populations. All pairwise comparisons between the Gagauz and the Turkish samples were statistically significant ($P<0.01$). A relatively high value of probability of identity and the insignificant genetic distances between the Bulgarian Turks and the Gagauzes presumably suggests their common Balkan ancestry, because both the Bulgarian Turks and particularly the Gagauzes demonstrate close affinities with the Bulgarians and the Macedonians. More considerable distinctions in the distribution of Y chromosome components appeared between the Gagauzes and the Turkic peoples from central Eurasia (Wells *et al.* 2001; Zerjal *et al.* 2002). Thus, none of 89 Gagauz male chromosomes investigated belong to the Asian cluster. In our views on the observed inconsistency between the linguistic and genetic affiliation of the Gagauzes we keep to the viewpoint of T. Kowalski (1933) and P. Mutafchiev (1947) on the stratification of various Turkic waves,

arriving from the South Russian steppes and the Anatolian peninsula (cited from Pokrovskaya 1964). Each new wave included the preceding one within its sphere of influence and alongside with it absorbed a certain share of the non-Turkic (Slavic or Romanic) population. Besides, in virtue of the lack of social barriers between the indigenous and the Turkic-Orthodox populations of the Balkan Peninsula the ongoing intensive reciprocal gene flow was accompanied by the gradual dissolution of the Asian genetic component in the pool of Balkan genes. In this case we deal with the multi-step process of language replacement in accordance with the elite dominance model (Renfrew 1987). We have already proposed in the *Alu* section of the discussion that in the Middle Ages the size of Turkic-speaking Orthodox population in the Balkans could be larger than in present time. In this context the reduction of the population could have facilitated the loss of the Asian diagnostic lineages due to random fluctuations.

In conclusion, our analysis of Y-chromosome polymorphism revealed high level of variability within Dniester-Carpathian male pool and relatively high level of population differentiation for such a small area. The presence of different genetic components of different age in the Dniester-Carpathian region indicates successive waves of migration from diverse source areas of Western Eurasia and thereby highlights the region between the Carpathians and Dniester as a zone of rich contact and interaction of various western Eurasian genetic ‘provinces’. The impacts from the western Balkans and Eastern Europe had priority among them. The heterogeneity of the eastern Romanic gene pool is, probably, the consequence of the recent historical events, connected with the peopling of the lands to the north from the Danube from various areas of Southeastern Europe and the unequal participation of the Slavs in the ethnogenesis of the Romanian sub-ethnic communities. The study has demonstrated that the Y-chromosomal pool of the Gagauzes exhibits a high degree of genetic affinity to geographically neighboring populations, suggesting that the Turkic element in their language was, probably, introduced vie elite dominance. In perspective an analysis of the mtDNA should be of particular interest to yield insights into the evolutionary processes experienced by the female part of population in the context of comparison with the evolutionary processes experienced by the male part.

# 7 APPENDIX

**Appendix 1** Matrix of genetic distances (Nei's method) among southeastern European populations, based on 11 *Alu* markers (*ACE, TPA25, PV92, APO, FXIIIB, D1, A25, B65, HS2.43, HS3.23, HS4.65)*

| Population | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0000 | | | | | | | | |
| 2 | 0.0040 | 0.0000 | | | | | | | |
| 3 | 0.0072 | 0.0123 | 0.0000 | | | | | | |
| 4 | 0.0039 | 0.0064 | 0.0064 | 0.0000 | | | | | |
| 5 | 0.0020 | 0.0057 | 0.0076 | 0.0052 | 0.0000 | | | | |
| 6 | 0.0027 | 0.0028 | 0.0097 | 0.0036 | 0.0018 | 0.0000 | | | |
| 7 | 0.0031 | 0.0056 | 0.0108 | 0.0048 | 0.0031 | 0.0037 | 0.0000 | | |
| 8 | 0.0012 | 0.0036 | 0.0075 | 0.0034 | 0.0032 | 0.0034 | 0.0043 | 0.0000 | |
| 9 | 0.0072 | 0.0050 | 0.0102 | 0.0054 | 0.0105 | 0.0075 | 0.0100 | 0.0039 | 0.0000 |
| 10 | 0.0132 | 0.0104 | 0.0169 | 0.0097 | 0.0166 | 0.0121 | 0.0177 | 0.0110 | 0.0058 |
| 11 | 0.0046 | 0.0077 | 0.0093 | 0.0044 | 0.0035 | 0.0033 | 0.0059 | 0.0037 | 0.0069 |
| 12 | 0.0045 | 0.0051 | 0.0120 | 0.0042 | 0.0085 | 0.0054 | 0.0079 | 0.0040 | 0.0035 |
| 13 | 0.0124 | 0.0104 | 0.0106 | 0.0084 | 0.0133 | 0.0096 | 0.0164 | 0.0104 | 0.0046 |
| 14 | 0.0080 | 0.0117 | 0.0113 | 0.0062 | 0.0098 | 0.0098 | 0.0090 | 0.0058 | 0.0056 |
| 15 | 0.0148 | 0.0145 | 0.0200 | 0.0083 | 0.0126 | 0.0080 | 0.0117 | 0.0138 | 0.0128 |
| 16 | 0.0115 | 0.0099 | 0.0113 | 0.0063 | 0.0145 | 0.0095 | 0.0139 | 0.0088 | 0.0034 |
| 17 | 0.0102 | 0.0091 | 0.0169 | 0.0095 | 0.0087 | 0.0058 | 0.0081 | 0.0105 | 0.0123 |

(Contd.)

| Population | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | |
| 2 | | | | | | | | |
| 3 | | | | | | | | |
| 4 | | | | | | | | |
| 5 | | | | | | | | |
| 6 | | | | | | | | |
| 7 | | | | | | | | |
| 8 | | | | | | | | |
| 9 | | | | | | | | |
| 10 | 0.0000 | | | | | | | |
| 11 | 0.0154 | 0.0000 | | | | | | |
| 12 | 0.0067 | 0.0054 | 0.0000 | | | | | |
| 13 | 0.0044 | 0.0103 | 0.0068 | 0.0000 | | | | |
| 14 | 0.0096 | 0.0099 | 0.0091 | 0.0085 | 0.0000 | | | |
| 15 | 0.0189 | 0.0081 | 0.0121 | 0.0114 | 0.0113 | 0.0000 | | |
| 16 | 0.0081 | 0.0098 | 0.0066 | 0.0033 | 0.0065 | 0.0077 | 0.0000 | |
| 17 | 0.0178 | 0.0117 | 0.0144 | 0.0133 | 0.0080 | 0.0062 | 0.0099 | 0.0000 |

Note. - Populations: 1=Moldavians (Karahasani); 2=Moldavians (Sofia); 3=Gagauzes (Etulia); 4=Gagauzes (Kongaz); 5=Ukrainians (Rashkovo); 6=Romanians (Eastern Romania); 7=Romanians (Ploiesti); 8=Macedonians; 9=Albanians; 10=Aromuns (Romania); 11=Aromuns (Macedonia, Krusevo); 12=Aromuns (Macedonia, Stip); 13=Aromuns (Albania); 14=Turks; 15=Greeks (Northeastern Greece); 16=Greek Cypriots; 17=Turkish Cypriots.

**Appendix 2** Matrix of genetic distances (Nei's method) among southeastern European and Central Asian populations, based on 8 Alu markers (*ACE, TPA25, PV92, APO, FXIIIB, A25, B65, D1*)

| Population | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | | | | | | | | | | |
| 2 | 0.0064 | - | | | | | | | | | |
| 3 | 0.0087 | 0.0171 | - | | | | | | | | |
| 4 | 0.0014 | 0.0054 | 0.0086 | - | | | | | | | |
| 5 | 0.0023 | 0.0085 | 0.0095 | 0.0033 | - | | | | | | |
| 6 | 0.0036 | 0.0039 | 0.0137 | 0.0032 | 0.0020 | - | | | | | |
| 7 | 0.0038 | 0.0079 | 0.0150 | 0.0052 | 0.0031 | 0.0052 | - | | | | |
| 8 | 0.0015 | 0.0054 | 0.0096 | 0.0011 | 0.0047 | 0.0048 | 0.0057 | - | | | |
| 9 | 0.0103 | 0.0067 | 0.0136 | 0.0056 | 0.0140 | 0.0109 | 0.0150 | 0.0049 | - | | |
| 10 | 0.0184 | 0.0134 | 0.0245 | 0.0126 | 0.0231 | 0.0175 | 0.0254 | 0.0151 | 0.0076 | - | |
| 11 | 0.0068 | 0.0117 | 0.0127 | 0.0035 | 0.0046 | 0.0048 | 0.0084 | 0.0054 | 0.0100 | 0.0222 | - |
| 12 | 0.0059 | 0.0067 | 0.0149 | 0.0022 | 0.0100 | 0.0069 | 0.0114 | 0.0043 | 0.0050 | 0.0077 | 0.0071 |
| 13 | 0.0172 | 0.0142 | 0.0144 | 0.0110 | 0.0174 | 0.0135 | 0.0246 | 0.0139 | 0.0067 | 0.0052 | 0.0145 |
| 14 | 0.0098 | 0.0152 | 0.0157 | 0.0083 | 0.0116 | 0.0136 | 0.0131 | 0.0065 | 0.0078 | 0.0130 | 0.0134 |
| 15 | 0.0136 | 0.0133 | 0.0243 | 0.0097 | 0.0083 | 0.0055 | 0.0133 | 0.0117 | 0.0144 | 0.0218 | 0.0050 |
| 16 | 0.0119 | 0.0093 | 0.0132 | 0.0073 | 0.0140 | 0.0103 | 0.0185 | 0.0070 | 0.0025 | 0.0083 | 0.0102 |
| 17 | 0.0116 | 0.0098 | 0.0228 | 0.0126 | 0.0074 | 0.0062 | 0.0107 | 0.0115 | 0.0170 | 0.0241 | 0.0147 |
| 18 | 0.0517 | 0.0468 | 0.0314 | 0.0510 | 0.0594 | 0.0564 | 0.0687 | 0.0526 | 0.0439 | 0.0388 | 0.0697 |
| 19 | 0.0781 | 0.0824 | 0.0856 | 0.0765 | 0.0852 | 0.0792 | 0.0905 | 0.0912 | 0.0955 | 0.0615 | 0.0970 |
| 20 | 0.0587 | 0.0580 | 0.0623 | 0.0580 | 0.0688 | 0.0625 | 0.0748 | 0.0667 | 0.0650 | 0.0354 | 0.0828 |
| 21 | 0.0645 | 0.0657 | 0.0577 | 0.0655 | 0.0698 | 0.0659 | 0.0786 | 0.0770 | 0.0803 | 0.0587 | 0.0853 |
| 22 | 0.0170 | 0.0190 | 0.0252 | 0.0200 | 0.0232 | 0.0219 | 0.0241 | 0.0226 | 0.0281 | 0.0173 | 0.0382 |

(Contd.)

| Population | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | | | | | | | | | | | |
| 4 | | | | | | | | | | | |
| 5 | | | | | | | | | | | |
| 6 | | | | | | | | | | | |
| 7 | | | | | | | | | | | |
| 8 | | | | | | | | | | | |
| 9 | | | | | | | | | | | |
| 10 | | | | | | | | | | | |
| 11 | | | | | | | | | | | |
| 12 | - | | | | | | | | | | |
| 13 | 0.0097 | - | | | | | | | | | |
| 14 | 0.0124 | 0.0126 | - | | | | | | | | |
| 15 | 0.0131 | 0.0142 | 0.0146 | - | | | | | | | |
| 16 | 0.0072 | 0.0034 | 0.0085 | 0.0110 | - | | | | | | |
| 17 | 0.0201 | 0.0192 | 0.0115 | 0.0080 | 0.0148 | - | | | | | |
| 18 | 0.0503 | 0.0331 | 0.0577 | 0.0793 | 0.0399 | 0.0656 | - | | | | |
| 19 | 0.0657 | 0.0739 | 0.1047 | 0.1035 | 0.0917 | 0.1099 | 0.0720 | - | | | |
| 20 | 0.0491 | 0.0483 | 0.0719 | 0.0874 | 0.0631 | 0.0810 | 0.0358 | 0.0120 | - | | |
| 21 | 0.0609 | 0.0606 | 0.0949 | 0.0965 | 0.0767 | 0.0923 | 0.0340 | 0.0138 | 0.0129 | - | |
| 22 | 0.0202 | 0.0255 | 0.0266 | 0.0419 | 0.0286 | 0.0292 | 0.0308 | 0.0376 | 0.0174 | 0.0293 | - |

Note. - Populations: 1=Moldavians (Karahasani); 2=Moldavians (Sofia); 3=Gagauzes (Etulia); 4=Gagauzes (Kongaz); 5=Ukrainians (Rashkovo); 6=Romanians (Eastern Romania); 7=Romanians (Ploiesti); 8=Macedonians; 9=Albanians; 10=Aromuns (Romania); 11=Aromuns (Macedonia, Krusevo); 12=Aromuns (Macedonia, Stip); 13=Aromuns (Albania); 14=Turks; 15=Greeks (Northeastern Greece); 16=Greek Cypriots; 17=Turkish Cypriots; 18=Uyghurs; 19=southern Kyrgyzes; 20=northern Kyrgyzes; 21=Kazakhs; 22=Uzbeks.

**Appendix 3** $F_{ST}$ distances among 42 western Eurasian populations based on Y–chromosome haplogroups

| Population | Gagauz [Kongaz] | Gagauz [Etulia] | Ukrainian [Rashkovo] | Moldovian [Sofia] | Karahasani [Moldavian] | Eastern Romanian |
|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | - | | | | | |
| Gagauz [Etulia] | 0.0073 | - | | | | |
| Ukrainian [Rashkovo] | 0.0638 | 0.0203 | - | | | |
| Moldovian [Sofia] | -0.0053 | 0.0066 | 0.0295 | - | | |
| Karahasani [Moldavian] | 0.0303 | 0.0065 | 0.0001 | 0.0059 | - | |
| East Romanian | 0.0183 | 0.0272 | 0.0582 | -0.0008 | 0.0387 | - |
| Constanta [Romanian] | -0.0078 | 0.0109 | 0.0802 | -0.0047 | 0.0465 | -0.0086 |
| Ploiesti [Romanian] | -0.0011 | 0.0312 | 0.1022 | 0.0127 | 0.0617 | 0.0152 |
| Ukrainian | 0.1106 | 0.0564 | 0.0175 | 0.0894 | 0.0352 | 0.1248 |
| Byelorussian | 0.0670 | 0.0311 | -0.0027 | 0.0418 | 0.0081 | 0.0703 |
| Polish | 0.1309 | 0.0642 | 0.0063 | 0.0853 | 0.0257 | 0.1166 |
| Czech and Slovakian | 0.0595 | 0.0247 | 0.0194 | 0.0375 | 0.0272 | 0.0857 |
| Hungrian | 0.1496 | 0.0774 | 0.0286 | 0.1176 | 0.0426 | 0.1683 |
| Croatian [Bosnia] | 0.1423 | 0.2030 | 0.2412 | 0.1231 | 0.1992 | 0.0603 |
| Croatian [Croatia] | 0.0399 | 0.0235 | 0.0171 | 0.0055 | 0.0117 | 0.0054 |
| Bosnian | 0.0175 | 0.0496 | 0.0925 | 0.0141 | 0.0591 | -0.0012 |
| Herzegovinian | 0.1316 | 0.1943 | 0.2283 | 0.1112 | 0.1910 | 0.0522 |
| Serbian [Serbia] | -0.0022 | 0.0298 | 0.0670 | -0.0039 | 0.0301 | 0.0154 |
| Serbian [Bosnia] | -0.0052 | 0.0342 | 0.0797 | 0.0024 | 0.0379 | 0.0216 |
| Macedonian [Macedonia] | 0.0005 | 0.0229 | 0.0774 | 0.0070 | 0.0362 | 0.0267 |
| Albanian [Tirana] | 0.0233 | 0.0352 | 0.0859 | 0.0330 | 0.0470 | 0.0751 |
| Albanian [Kosovo] | 0.1124 | 0.1415 | 0.2143 | 0.1331 | 0.1467 | 0.2015 |
| Greek [Thrace] | 0.0121 | 0.0028 | 0.0414 | 0.0142 | 0.0152 | 0.0520 |
| Greek | 0.0588 | 0.0556 | 0.1005 | 0.0645 | 0.0679 | 0.1244 |
| Greek [Macedonia] | 0.0222 | -0.0062 | -0.0039 | 0.0072 | -0.0199 | 0.0458 |
| German | 0.0782 | 0.0901 | 0.1189 | 0.0495 | 0.0952 | 0.0784 |
| Dutch | 0.2098 | 0.2087 | 0.2282 | 0.1826 | 0.2066 | 0.2390 |
| French | 0.1096 | 0.1143 | 0.1516 | 0.0977 | 0.1266 | 0.1544 |
| Italian | 0.1936 | 0.1717 | 0.2066 | 0.1827 | 0.1945 | 0.2414 |
| Calabrian | 0.0872 | 0.0951 | 0.1527 | 0.1099 | 0.1228 | 0.1795 |
| Andalusian | 0.1989 | 0.1993 | 0.2327 | 0.1937 | 0.2076 | 0.2703 |
| Catalan | 0.2845 | 0.2737 | 0.3062 | 0.2727 | 0.2817 | 0.3435 |
| Spanish Basque | 0.4000 | 0.4012 | 0.4085 | 0.3755 | 0.3735 | 0.4464 |
| Turkish 1 | 0.0637 | 0.0778 | 0.1310 | 0.0857 | 0.0985 | 0.1314 |
| Turkish2 | 0.0753 | 0.0721 | 0.1272 | 0.0894 | 0.1126 | 0.1368 |
| Turkish 3 | 0.0628 | 0.0778 | 0.1300 | 0.1015 | 0.1168 | 0.1582 |
| Turkish 4 | 0.0747 | 0.0700 | 0.1170 | 0.0983 | 0.0939 | 0.1551 |
| Turkish 5 | 0.0517 | 0.0386 | 0.0919 | 0.0743 | 0.0773 | 0.1255 |
| Turkish 6 | 0.0585 | 0.0548 | 0.1037 | 0.0802 | 0.0802 | 0.1402 |
| Turkish 7 | 0.0675 | 0.0693 | 0.1213 | 0.0952 | 0.1079 | 0.1447 |
| Turkish 8 | 0.0369 | 0.0413 | 0.1109 | 0.0711 | 0.0930 | 0.1244 |
| Turkish 9 | 0.0276 | 0.0398 | 0.0885 | 0.0507 | 0.0633 | 0.1015 |

(Contd.)

| | Romanian [Constanta] | Romanian [Ploiesti] | Ukrainian | Byelorussian | Polish | Czech and Slovakian |
|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | | | | | | |
| Gagauz [Etulia] | | | | | | |
| Ukrainian [Rashkovo] | | | | | | |
| Moldovian [Sofia] | | | | | | |
| Karahasani [Moldavian] | | | | | | |
| East Romanian | | | | | | |
| Romanian [Constanta] | - | | | | | |
| Romanian [Ploiesti] | -0.0139 | - | | | | |
| Ukrainian | 0.1539 | 0.1643 | - | | | |
| Byelorussian | 0.0982 | 0.1154 | -0.0056 | - | | |
| Polish | 0.1559 | 0.1857 | 0.0095 | 0.0058 | - | |
| Czech and Slovakian | 0.0701 | 0.0973 | 0.0991 | 0.0647 | 0.0794 | - |
| Hungrian | 0.1940 | 0.2129 | 0.0017 | 0.0188 | -0.0021 | 0.0929 |
| Croatian [Bosnia] | 0.0977 | 0.1165 | 0.3120 | 0.2337 | 0.3078 | 0.2945 |
| Croatian [Croatia] | 0.0316 | 0.0597 | 0.0648 | 0.0245 | 0.0476 | 0.0644 |
| Bosnian | 0.0026 | 0.0002 | 0.1430 | 0.0924 | 0.1546 | 0.1267 |
| Herzegovinian | 0.0876 | 0.1082 | 0.2993 | 0.2230 | 0.2936 | 0.2801 |
| Serbian [Serbia] | 0.0044 | 0.0044 | 0.1159 | 0.0708 | 0.1240 | 0.0772 |
| Serbian [Bosnia] | 0.0067 | -0.0025 | 0.1222 | 0.0783 | 0.1410 | 0.0924 |
| Macedonian [Macedonia] | 0.0055 | -0.0028 | 0.1185 | 0.0811 | 0.1329 | 0.0822 |
| Albanian [Tirana] | 0.0386 | 0.0165 | 0.1371 | 0.1058 | 0.1549 | 0.0588 |
| Albanian [Kosovo] | 0.1445 | 0.1184 | 0.2601 | 0.2282 | 0.2733 | 0.1599 |
| Greek [Thrace] | 0.0294 | 0.0178 | 0.0778 | 0.0542 | 0.1004 | 0.0378 |
| Greek | 0.0791 | 0.0676 | 0.1623 | 0.1336 | 0.1706 | 0.0413 |
| Greek [Macedonia] | 0.0466 | 0.0403 | 0.0153 | 0.0004 | 0.0318 | 0.0282 |
| German | 0.0480 | 0.0937 | 0.2540 | 0.1786 | 0.2134 | 0.0409 |
| Dutch | 0.2025 | 0.2445 | 0.3669 | 0.2964 | 0.3229 | 0.1034 |
| French | 0.1024 | 0.1186 | 0.2744 | 0.2154 | 0.2545 | 0.0417 |
| Italian | 0.1914 | 0.2183 | 0.3263 | 0.2753 | 0.2999 | 0.0739 |
| Calabrian | 0.1160 | 0.1007 | 0.2265 | 0.1928 | 0.2476 | 0.0627 |
| Andalusian | 0.2160 | 0.2382 | 0.3507 | 0.2946 | 0.3297 | 0.0937 |
| Catalan | 0.2993 | 0.3286 | 0.4314 | 0.3699 | 0.4022 | 0.1585 |
| Spanish Basque | 0.4335 | 0.4570 | 0.5365 | 0.4633 | 0.4990 | 0.2647 |
| Turkish 1 | 0.0863 | 0.0444 | 0.1840 | 0.1563 | 0.2201 | 0.0935 |
| Turkish2 | 0.0808 | 0.0687 | 0.2145 | 0.1766 | 0.2334 | 0.0462 |
| Turkish 3 | 0.1085 | 0.0925 | 0.1678 | 0.1468 | 0.2146 | 0.0800 |
| Turkish 4 | 0.1114 | 0.0811 | 0.1497 | 0.1363 | 0.1908 | 0.0781 |
| Turkish 5 | 0.0791 | 0.0607 | 0.1349 | 0.1172 | 0.1744 | 0.0479 |
| Turkish 6 | 0.0900 | 0.0671 | 0.1592 | 0.1346 | 0.1922 | 0.0483 |
| Turkish 7 | 0.0973 | 0.0726 | 0.1655 | 0.1463 | 0.2063 | 0.0724 |
| Turkish 8 | 0.0653 | 0.0639 | 0.1668 | 0.1364 | 0.2031 | 0.0480 |
| Turkish 9 | 0.0591 | 0.0400 | 0.1320 | 0.1068 | 0.1644 | 0.0512 |

(Contd.)

| | Hungrian | Croatian [Bosnia] | Croatian [Croatia] | Bosnian | Herzegovinian | Serbian [Serbia] | Serbian [Bosnia] |
|---|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | | | | | | | |
| Gagauz [Etulia] | | | | | | | |
| Ukrainian [Rashkovo] | | | | | | | |
| Moldovian [Sofia] | | | | | | | |
| Karahasani [Moldavian] | | | | | | | |
| East Romanian | | | | | | | |
| Constanta [Romanian] | | | | | | | |
| Ploiesti [Romanian] | | | | | | | |
| Ukrainian | | | | | | | |
| Byelorussian | | | | | | | |
| Polish | | | | | | | |
| Czech and Slovakian | | | | | | | |
| Hungrian | - | | | | | | |
| Croatian [Bosnia] | 0.3797 | - | | | | | |
| Croatian [Croatia] | 0.0922 | 0.1133 | - | | | | |
| Bosnian | 0.1965 | 0.0512 | 0.0330 | - | | | |
| Herzegovinian | 0.3640 | -0.0071 | 0.1056 | 0.0463 | - | | |
| Serbian [Serbia] | 0.1489 | 0.1112 | 0.0326 | 0.0118 | 0.1021 | - | |
| Serbian [Bosnia] | 0.1625 | 0.1152 | 0.0438 | 0.0075 | 0.1068 | -0.0078 | - |
| Macedonian | 0.1494 | 0.1311 | 0.0471 | 0.0183 | 0.1251 | -0.0024 | -0.0044 |
| Albanian [Tirana] | 0.1560 | 0.2268 | 0.0897 | 0.0671 | 0.2209 | 0.0267 | 0.0238 |
| Albanian [Kosovo] | 0.2585 | 0.3520 | 0.2096 | 0.1860 | 0.3445 | 0.1061 | 0.1053 |
| Greek [Thrace] | 0.0999 | 0.2289 | 0.0514 | 0.0532 | 0.2184 | 0.0147 | 0.0159 |
| Greek | 0.1617 | 0.3064 | 0.1279 | 0.1312 | 0.2989 | 0.0672 | 0.0713 |
| Greek [Macedonia] | 0.0330 | 0.2478 | 0.0189 | 0.0526 | 0.2320 | 0.0200 | 0.0228 |
| German | 0.2627 | 0.2570 | 0.0989 | 0.1243 | 0.2325 | 0.0859 | 0.1052 |
| Dutch | 0.3581 | 0.4428 | 0.2318 | 0.2813 | 0.4145 | 0.2169 | 0.2414 |
| French | 0.2756 | 0.3648 | 0.1675 | 0.1836 | 0.3432 | 0.1252 | 0.1384 |
| Italian | 0.3144 | 0.4483 | 0.2363 | 0.2777 | 0.4295 | 0.2136 | 0.2326 |
| Calabrian | 0.2361 | 0.3856 | 0.1919 | 0.1833 | 0.3732 | 0.1195 | 0.1175 |
| Andalusian | 0.3388 | 0.4853 | 0.2623 | 0.2990 | 0.4611 | 0.2198 | 0.2367 |
| Catalan | 0.4199 | 0.5548 | 0.3228 | 0.3736 | 0.5245 | 0.2979 | 0.3204 |
| Spanish Basque | 0.5280 | 0.6218 | 0.3891 | 0.4603 | 0.5822 | 0.3780 | 0.4117 |
| Turkish 1 | 0.2153 | 0.3009 | 0.1519 | 0.1126 | 0.2943 | 0.0813 | 0.0696 |
| Turkish2 | 0.2373 | 0.3432 | 0.1585 | 0.1472 | 0.3287 | 0.1064 | 0.1065 |
| Turkish 3 | 0.1998 | 0.3269 | 0.1709 | 0.1553 | 0.3203 | 0.1087 | 0.1000 |
| Turkish 4 | 0.1709 | 0.3312 | 0.1599 | 0.1467 | 0.3274 | 0.1020 | 0.0930 |
| Turkish 5 | 0.1605 | 0.3246 | 0.1331 | 0.1271 | 0.3159 | 0.0850 | 0.0799 |
| Turkish 6 | 0.1786 | 0.3482 | 0.1474 | 0.1381 | 0.3373 | 0.0902 | 0.0836 |
| Turkish 7 | 0.1944 | 0.3136 | 0.1592 | 0.1412 | 0.3081 | 0.0998 | 0.0933 |
| Turkish 8 | 0.1962 | 0.3263 | 0.1424 | 0.1320 | 0.3135 | 0.0880 | 0.0827 |
| Turkish 9 | 0.1563 | 0.2642 | 0.1138 | 0.0965 | 0.2582 | 0.0517 | 0.0460 |

(Contd.)

| | Macedonian [Macedonia] | Albanian [Tirana] | Albanian [Kosovo] | Greek [Thrace] | Greek | Greek [Macedonia] | German | Dutch |
|---|---|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | | | | | | | | |
| Gagauz [Etulia] | | | | | | | | |
| Ukrainian [Rashkovo] | | | | | | | | |
| Moldovian [Sofia] | | | | | | | | |
| Karahasani [Moldavian] | | | | | | | | |
| East Romanian | | | | | | | | |
| Constanta [Romanian] | | | | | | | | |
| Ploiesti [Romanian] | | | | | | | | |
| Ukrainian | | | | | | | | |
| Byelorussian | | | | | | | | |
| Polish | | | | | | | | |
| Czech and Slovakian | | | | | | | | |
| Hungrian | | | | | | | | |
| Croatian [Bosnia] | | | | | | | | |
| Croatian [Croatia] | | | | | | | | |
| Bosnian | | | | | | | | |
| Herzegovinian | | | | | | | | |
| Serbian [Serbia] | | | | | | | | |
| Serbian [Bosnia] | | | | | | | | |
| Macedonian | - | | | | | | | |
| Albanian [Tirana] | 0.0127 | - | | | | | | |
| Albanian [Kosovo] | 0.0829 | 0.0433 | - | | | | | |
| Greek [Thrace] | 0.0027 | -0.0067 | 0.0708 | - | | | | |
| Greek | 0.0523 | 0.0039 | 0.0358 | 0.0117 | - | | | |
| Greek [Macedonia] | 0.0146 | 0.0123 | 0.1170 | -0.0214 | 0.0368 | - | | |
| German | 0.1013 | 0.0965 | 0.1894 | 0.1080 | 0.0916 | 0.1304 | - | |
| Dutch | 0.2246 | 0.1915 | 0.2582 | 0.2218 | 0.1520 | 0.2654 | 0.0121 | - |
| French | 0.1244 | 0.0742 | 0.1401 | 0.1009 | 0.0438 | 0.1354 | -0.0040 | 0.0118 |
| Italian | 0.2075 | 0.1511 | 0.2170 | 0.1757 | 0.0988 | 0.2180 | 0.0650 | 0.0231 |
| Calabrian | 0.1033 | 0.0332 | 0.0774 | 0.0583 | 0.0069 | 0.0948 | 0.1098 | 0.1512 |
| Andalusian | 0.2181 | 0.1571 | 0.1969 | 0.1908 | 0.0993 | 0.2373 | 0.0778 | 0.0171 |
| Catalan | 0.2939 | 0.2409 | 0.2885 | 0.2811 | 0.1806 | 0.3440 | 0.1418 | 0.0194 |
| Spanish Basque | 0.3753 | 0.3387 | 0.3781 | 0.4073 | 0.2811 | 0.4957 | 0.2609 | 0.0648 |
| Turkish 1 | 0.0606 | 0.0081 | 0.0882 | 0.0277 | 0.0261 | 0.0516 | 0.1471 | 0.2348 |
| Turkish2 | 0.0909 | 0.0326 | 0.1255 | 0.0488 | 0.0198 | 0.0836 | 0.0819 | 0.1431 |
| Turkish 3 | 0.0976 | 0.0549 | 0.1327 | 0.0573 | 0.0527 | 0.0845 | 0.1596 | 0.2240 |
| Turkish 4 | 0.0803 | 0.0234 | 0.0904 | 0.0293 | 0.0235 | 0.0461 | 0.1661 | 0.2314 |
| Turkish 5 | 0.0648 | 0.0174 | 0.1029 | 0.0144 | 0.0157 | 0.0342 | 0.1346 | 0.2118 |
| Turkish 6 | 0.0722 | 0.0116 | 0.0893 | 0.0224 | 0.0062 | 0.0408 | 0.1248 | 0.1964 |
| Turkish 7 | 0.0816 | 0.0312 | 0.1106 | 0.0353 | 0.0295 | 0.0649 | 0.1491 | 0.2144 |
| Turkish 8 | 0.0750 | 0.0384 | 0.1230 | 0.0403 | 0.0344 | 0.0683 | 0.1072 | 0.1853 |
| Turkish 9 | 0.0395 | 0.0043 | 0.0707 | 0.0075 | 0.0098 | 0.0285 | 0.1116 | 0.1907 |

(Contd.)

| | French | Italian | Calabrian | Andalusian | Catalan | Spanish Basque | Turkish 1 |
|---|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | | | | | | | |
| Gagauz [Etulia] | | | | | | | |
| Ukrainian [Rashkovo] | | | | | | | |
| Moldovian [Sofia] | | | | | | | |
| Karahasani [Moldavian] | | | | | | | |
| East Romanian | | | | | | | |
| Constanta [Romanian] | | | | | | | |
| Ploiesti [Romanian] | | | | | | | |
| Ukrainian | | | | | | | |
| Byelorussian | | | | | | | |
| Polish | | | | | | | |
| Czech and Slovakian | | | | | | | |
| Hungrian | | | | | | | |
| Croatian [Bosnia] | | | | | | | |
| Croatian [Croatia] | | | | | | | |
| Bosnian | | | | | | | |
| Herzegovinian | | | | | | | |
| Serbian [Serbia] | | | | | | | |
| Serbian [Bosnia] | | | | | | | |
| Macedonian [Macedonia] | | | | | | | |
| Albanian [Tirana] | | | | | | | |
| Albanian [Kosovo] | | | | | | | |
| Greek [Thrace] | | | | | | | |
| Greek | | | | | | | |
| Greek [Macedonia] | | | | | | | |
| German | | | | | | | |
| Dutch | | | | | | | |
| French | - | | | | | | |
| Italian | -0.0009 | - | | | | | |
| Calabrian | 0.0285 | 0.0733 | - | | | | |
| Andalusian | 0.0004 | -0.0039 | 0.0664 | - | | | |
| Catalan | 0.0577 | 0.0103 | 0.1565 | -0.0053 | - | | |
| Spanish Basque | 0.1632 | 0.0959 | 0.2859 | 0.0712 | 0.0048 | - | |
| Turkish 1 | 0.0918 | 0.1686 | 0.0184 | 0.1786 | 0.2706 | 0.3872 | - |
| Turkish2 | 0.0175 | 0.0614 | -0.0088 | 0.0871 | 0.1681 | 0.3101 | 0.0119 |
| Turkish 3 | 0.1113 | 0.1549 | 0.0214 | 0.1551 | 0.2396 | 0.3439 | 0.0365 |
| Turkish 4 | 0.1031 | 0.1570 | 0.0119 | 0.1640 | 0.2479 | 0.3513 | 0.0004 |
| Turkish 5 | 0.0819 | 0.1310 | 0.0064 | 0.1488 | 0.2346 | 0.3662 | 0.0035 |
| Turkish 6 | 0.0619 | 0.1164 | -0.0126 | 0.1239 | 0.2190 | 0.3575 | -0.0091 |
| Turkish 7 | 0.0911 | 0.1377 | 0.0103 | 0.1512 | 0.2313 | 0.3341 | 0.0084 |
| Turkish 8 | 0.0674 | 0.1131 | 0.0050 | 0.1217 | 0.2093 | 0.3568 | 0.0311 |
| Turkish 9 | 0.0761 | 0.1368 | 0.0115 | 0.1391 | 0.2234 | 0.3240 | 0.0058 |

(Contd.)

| | Turkish2 | Turkish 3 | Turkish 4 | Turkish 5 | Turkish 6 | Turkish 7 | Turkish 8 | Turkish 9 |
|---|---|---|---|---|---|---|---|---|
| Gagauz [Kongaz] | | | | | | | | |
| Gagauz [Etulia] | | | | | | | | |
| Ukrainian [Rashkovo] | | | | | | | | |
| Moldovian [Sofia] | | | | | | | | |
| Karahasani [Moldavian] | | | | | | | | |
| East Romanian | | | | | | | | |
| Constanta [Romanian] | | | | | | | | |
| Ploiesti [Romanian] | | | | | | | | |
| Ukrainian | | | | | | | | |
| Byelorussian | | | | | | | | |
| Polish | | | | | | | | |
| Czech and Slovakian | | | | | | | | |
| Hungrian | | | | | | | | |
| Croatian [Bosnia] | | | | | | | | |
| Croatian [Croatia] | | | | | | | | |
| Bosnian | | | | | | | | |
| Herzegovinian | | | | | | | | |
| Serbian [Serbia] | | | | | | | | |
| Serbian [Bosnia] | | | | | | | | |
| Macedonian | | | | | | | | |
| Albanian [Tirana] | | | | | | | | |
| Albanian [Kosovo] | | | | | | | | |
| Greek [Thrace] | | | | | | | | |
| Greek | | | | | | | | |
| Greek [Macedonia] | | | | | | | | |
| German | | | | | | | | |
| Dutch | | | | | | | | |
| French | | | | | | | | |
| Italian | | | | | | | | |
| Calabrian | | | | | | | | |
| Andalusian | | | | | | | | |
| Catalan | | | | | | | | |
| Spanish Basque | | | | | | | | |
| Turkish 1 | | | | | | | | |
| Turkish2 | - | | | | | | | |
| Turkish 3 | 0.0278 | - | | | | | | |
| Turkish 4 | 0.0197 | 0.0176 | - | | | | | |
| Turkish 5 | -0.0018 | 0.0039 | -0.0101 | - | | | | |
| Turkish 6 | -0.0077 | 0.0057 | -0.0140 | -0.0181 | - | | | |
| Turkish 7 | 0.0022 | 0.0027 | -0.0004 | -0.0127 | -0.0080 | - | | |
| Turkish 8 | 0.0014 | -0.0078 | 0.0159 | -0.0075 | -0.0036 | 0.0001 | - | |
| Turkish 9 | 0.0133 | 0.0082 | 0.0039 | -0.0051 | -0.0079 | 0.0018 | 0.0020 | - |

Note. - $F_{ST}$ values significant at 5% level are shaded.

**Appendix 4** Diversity of Y-STR haplotypes based on seven loci in six Dniester-Carpathian and 33 European populations (N=3719 males)

| Population | Nomber of individuals | No. of haplotypes | Haplotype diversity±SD | Average gene diversity±SD | Mean no. of pairwise differences±SD | Reference* |
|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | 47 | 37 | 0.9898±0.0065 | 0.6021±0.3378 | 4.22±2.13 | 1 |
| Etulia [Gagauzia] | 39 | 24 | 0.9636±0.0163 | 0.5645±0.3209 | 3.95±2.02 | 1 |
| Karahasani [Moldova] | 72 | 44 | 0.9804±0.0067 | 0.5552±0.3125 | 3.89±1.97 | 1 |
| Sofia [Moldova] | 50 | 36 | 0.9837±0.0077 | 0.5357±0.3050 | 3.75±1.92 | 1 |
| East Romania | 51 | 34 | 0.9796±0.0082 | 0.5134±0.2940 | 3.59±1.85 | 1 |
| Rashkovo [Ukrain] | 51 | 38 | 0.9875±0.0064 | 0.5739±0.3235 | 4.01±2.04 | 1 |
| Moscow [Russia] | 85 | 55 | 0.9776±0.0079 | 0.5325±0.3009 | 3.73±1.90 | 2 |
| Nowgorod [Russia] | 50 | 37 | 0.9812±0.0094 | 0.4955±0.2853 | 3.47±1.80 | 3 |
| Byelorussia | 69 | 54 | 0.9923±0.0040 | 0.5387±0.3047 | 3.77±1.92 | 3 |
| Kiev [Ukrain] | 82 | 57 | 0.9886±0.0041 | 0.5313±0.3004 | 3.72±1.90 | 3 |
| Albania | 101 | 50 | 0.9471±0.0131 | 0.4932±0.2813 | 3.45±1.78 | 3 |
| Croatia | 457 | 213 | 0.9866±0.0020 | 0.5111±0.2877 | 3.58±1.82 | 4 |
| Turkey | 522 | 327 | 0.9962±0.0005 | 0.6252±0.3422 | 4.38±2.17 | 5 |
| Bulgaria [Turks] | 61 | 52 | 0.9929±0.0050 | 0.6107±0.3402 | 4.28±2.15 | 6 |
| Bulgaria [Bulgarians] | 122 | 85 | 0.9882±0.0044 | 0.5609±0.3134 | 3.93±1.98 | 6 |
| Macedonia | 84 | 58 | 0.9788±0.0077 | 0.5261±0.2978 | 3.68±1.88 | 7 |
| Bosnia | 181 | 90 | 0.9731±0.0054 | 0.4341±0.2515 | 3.04±1.59 | 8 |
| Serbia | 114 | 72 | 0.9898±0.0026 | 0.5393±0.3032 | 3.78±1.92 | 9 |
| Ljubljana [Slovenia] | 121 | 74 | 0.9860±0.0037 | 0.5586±0.3123 | 3.91±1.98 | 3 |
| Gdansk [Poland] | 150 | 92 | 0.9854±0.0038 | 0.5152±0.2910 | 3.61±1.84 | 2 |
| Lublin [Poland] | 134 | 104 | 0.9933±0.0026 | 0.5182±0.2926 | 3.63±1.85 | 2 |
| Wrozlaw [Poland] | 121 | 75 | 0.9825±0.0047 | 0.4994±0.2838 | 3.50±1.79 | 2 |
| Berlin [Germany] | 549 | 301 | 0.9911±0.0012 | 0.5916±0.3261 | 4.14±2.06 | 3 |
| Munich [Germany] | 250 | 155 | 0.9885±0.0024 | 0.5749±0.3188 | 4.02±2.02 | 3 |
| Cologne [Germany] | 135 | 98 | 0.9893±0.0036 | 0.5757±0.3203 | 4.03±2.03 | 3 |
| Vien [Austria] | 66 | 66 | 1.0000±0.0026 | 0.6553±0.3614 | 4.59±2.28 | 3 |
| Denmark | 63 | 43 | 0.9811±0.0076 | 0.5152±0.2937 | 3.61±1.85 | 3 |
| Sweden | 403 | 202 | 0.9798±0.0034 | 0.5594±0.3109 | 3.92±1.97 | 3 |
| Vilnus [Lithuania] | 152 | 101 | 0.9884±0.0031 | 0.5733±0.3189 | 4.01±2.02 | 2 |
| Riga [Latvia] | 145 | 99 | 0.9905±0.0027 | 0.5789±0.3217 | 4.05±2.03 | 2 |
| Tartu [Estonia] | 133 | 93 | 0.9869±0.0038 | 0.5980±0.3311 | 4.19±2.09 | 2 |
| Budapest [Hungary] | 115 | 93 | 0.9951±0.0021 | 0.6067±0.3356 | 4.25±2.12 | 2 |
| Athens [Greece] | 101 | 87 | 0.9964±0.0022 | 0.6195±0.3422 | 4.34±2.16 | 3 |
| Thrace [Greece] | 39 | 30 | 0.9757±0.0145 | 0.5909±0.3338 | 4.14±2.10 | 10 |
| Constanta [Romania] | 31 | 28 | 0.9914±0.0116 | 0.6089±0.3451 | 4.26±2.17 | 10 |
| Ploiesti [Romania] | 36 | 31 | 0.9905±0.0094 | 0.5551±0.3170 | 3.89±2.00 | 10 |
| Lazio [Italy] | 222 | 163 | 0.9947±0.0014 | 0.6017±0.3318 | 4.21±2.10 | 3 |
| Lombardy [Italy] | 182 | 123 | 0.9819±0.0056 | 0.5879±0.3255 | 4.12±2.06 | 3 |
| Sicily [Italy] | 199 | 167 | 0.9978±0.0009 | 0.6199±0.3407 | 4.34±2.16 | 3 |

Note. - *Reference codes: 1, Present study; 2, Ploski *et al.* 2002; 3, Roewer *et al.* 2005; 4, Barac *et al.* 2003; 5, Cinnioğlu *et al.* 2004; 6, Zaharova *et al.* 2001; 7, Pericic *et al.* 2005; 8, Klaric *et al.* 2005; 9, Barac Lauc *et al.* 2005; 10, Bosch *et al.* 2006.

**Appendix 5** Matrix of population pairwise values of $R_{ST}$ (below the diagonal) and probability of identity (above the diagonal), based on microsatellite haplotypes

| Population | Kongaz [Gagauzia] | Etulia [Gagauzia] | Karahasani [Moldova] | Sofia [Moldova] | Eastern Romania | Rashkovo [Ukrain] | Moscow [Russia] |
|---|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | - | 0.0065 | 0.0065 | 0.0094 | 0.0100 | 0.0042 | 0.0048 |
| Etulia [Gagauzia] | 0.0396 | - | 0.0118 | 0.0067 | 0.0101 | 0.0020 | 0.0078 |
| Karahasani [Moldova] | 0.0483 | -0.0154 | - | 0.0081 | 0.0109 | 0.0093 | 0.0119 |
| Sofia [Moldova] | 0.0175 | -0.0052 | -0.0028 | - | 0.0125 | 0.0078 | 0.0089 |
| Eastern Romania | 0.0956 | 0.0062 | 0.0051 | 0.0103 | - | 0.0046 | 0.0069 |
| Rashkovo [Ukrain] | 0.0449 | 0.0033 | 0.0204 | 0.0106 | 0.0355 | - | 0.0150 |
| Moscow [Russia] | 0.0461 | -0.0150 | -0.0049 | -0.0006 | 0.0210 | 0.0073 | - |
| Nowgorod [Russia] | 0.1230 | -0.0002 | 0.0078 | 0.0435 | 0.0309 | 0.0379 | 0.0180 |
| Byelorussia | 0.1261 | 0.0043 | 0.0155 | 0.0346 | 0.0030 | 0.0381 | 0.0189 |
| Kiev [Ukrain] | 0.1216 | 0.0020 | 0.0148 | 0.0326 | 0.0084 | 0.0352 | 0.0167 |
| Albania | 0.0497 | 0.1369 | 0.1381 | 0.1147 | 0.2224 | 0.1091 | 0.1469 |
| Croatia | 0.0853 | 0.0115 | 0.0144 | 0.0077 | -0.0059 | 0.0413 | 0.0198 |
| Turkey | 0.0331 | 0.1157 | 0.1207 | 0.0790 | 0.1562 | 0.0753 | 0.1069 |
| Bulgaria [Turks] | 0.0100 | 0.0745 | 0.0852 | 0.0452 | 0.1213 | 0.0364 | 0.0769 |
| Bulgaria [Bulgarians] | -0.0019 | 0.0525 | 0.0558 | 0.0223 | 0.0890 | 0.0437 | 0.0547 |
| Macedonia | 0.0044 | 0.0323 | 0.0331 | 0.0133 | 0.0601 | 0.0488 | 0.0440 |
| Bosnia | 0.1332 | 0.0186 | 0.0292 | 0.0467 | 0.0038 | 0.0620 | 0.0391 |
| Serbia | 0.0091 | 0.0097 | 0.0145 | 0.0006 | 0.0408 | 0.0228 | 0.0169 |
| Ljubljana [Slovenia] | 0.0494 | -0.0137 | -0.0009 | 0.0042 | 0.0182 | 0.0092 | -0.0075 |
| Gdansk [Poland] | 0.1196 | 0.0108 | 0.0205 | 0.0410 | 0.0211 | 0.0444 | 0.0173 |
| Lublin [Poland] | 0.1386 | 0.0172 | 0.0308 | 0.0553 | 0.0382 | 0.0436 | 0.0220 |
| Wrozlaw [Poland] | 0.1489 | 0.0226 | 0.0356 | 0.0600 | 0.0444 | 0.0492 | 0.0280 |
| Berlin [Germany] | 0.0159 | 0.0337 | 0.0480 | 0.0257 | 0.0914 | 0.0238 | 0.0276 |
| Munich [Germany] | 0.0222 | 0.0826 | 0.0982 | 0.0605 | 0.1558 | 0.0523 | 0.0742 |
| Cologne [Germany] | 0.0413 | 0.1098 | 0.1229 | 0.0804 | 0.1894 | 0.0640 | 0.1012 |
| Vien [Austria] | 0.0489 | 0.1506 | 0.1707 | 0.1138 | 0.2330 | 0.1039 | 0.1550 |
| Denmark | 0.1122 | 0.2158 | 0.2392 | 0.1850 | 0.3214 | 0.1267 | 0.2090 |
| Sweden | 0.0463 | 0.1417 | 0.1581 | 0.1162 | 0.2226 | 0.1044 | 0.1326 |
| Vilnus [Lithuania] | 0.0996 | 0.0194 | 0.0339 | 0.0334 | 0.0595 | 0.0265 | 0.0221 |
| Riga [Latvia] | 0.0924 | 0.0254 | 0.0455 | 0.0418 | 0.0821 | 0.0227 | 0.0254 |
| Tartu [Estonia] | 0.0880 | 0.0908 | 0.1138 | 0.0937 | 0.1778 | 0.0708 | 0.0918 |
| Budapest [Hungary] | 0.0036 | 0.0253 | 0.0364 | 0.0097 | 0.0674 | 0.0104 | 0.0228 |
| Athens [Greece] | 0.0099 | 0.0646 | 0.0725 | 0.0431 | 0.1208 | 0.0373 | 0.0652 |
| Thrace [Greece] | 0.0225 | 0.0005 | 0.0090 | 0.0044 | 0.0321 | 0.0080 | 0.0163 |
| Constanta [Romania] | 0.0437 | 0.0170 | 0.0270 | 0.0040 | 0.0090 | 0.0312 | 0.0304 |
| Ploiesti [Romania] | 0.0126 | 0.0148 | 0.0232 | -0.0020 | 0.0311 | 0.0154 | 0.0207 |
| Lazio [Italy] | 0.0475 | 0.1497 | 0.1596 | 0.1155 | 0.2201 | 0.0974 | 0.1423 |
| Lombardia [Italy] | 0.0614 | 0.1422 | 0.1567 | 0.1152 | 0.2287 | 0.0879 | 0.1345 |
| Sicily [Italy] | 0.0549 | 0.1785 | 0.1852 | 0.1347 | 0.2340 | 0.1268 | 0.1677 |

(Contd.)

| Population | Nowgorod [Russia] | Byelorussia | Kiev [Ukrain] | Albania | Croatia | Turkey | Bulgaria [Turks] | Bulgaria [Bulgarians] |
|---|---|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | 0.0038 | 0.0056 | 0.0067 | 0.0154 | 0.0050 | 0.0030 | 0.0070 | 0.0084 |
| Etulia [Gagauzia] | 0.0092 | 0.0074 | 0.0075 | 0.0180 | 0.0081 | 0.0028 | 0.0063 | 0.0101 |
| Karahasani [Moldova] | 0.0131 | 0.0079 | 0.0080 | 0.0132 | 0.0090 | 0.0031 | 0.0064 | 0.0067 |
| Sofia [Moldova] | 0.0100 | 0.0104 | 0.0093 | 0.0127 | 0.0082 | 0.0021 | 0.0052 | 0.0111 |
| East Romania | 0.0082 | 0.0077 | 0.0084 | 0.0212 | 0.0116 | 0.0028 | 0.0100 | 0.0100 |
| Rashkovo [Ukrain] | 0.0090 | 0.0071 | 0.0110 | 0.0025 | 0.0063 | 0.0021 | 0.0032 | 0.0031 |
| Moscow [Russia] | 0.0148 | 0.0119 | 0.0172 | 0.0030 | 0.0104 | 0.0029 | 0.0060 | 0.0047 |
| Nowgorod [Russia] | - | 0.0107 | 0.0127 | 0.0059 | 0.0086 | 0.0016 | 0.0026 | 0.0043 |
| Byelorussia | 0.0049 | - | 0.0122 | 0.0049 | 0.0087 | 0.0015 | 0.0029 | 0.0064 |
| Kiev [Ukrain] | 0.0014 | -0.0083 | - | 0.0031 | 0.0097 | 0.0019 | 0.0036 | 0.0048 |
| Albania | 0.2227 | 0.2524 | 0.2376 | - | 0.0072 | 0.0062 | 0.0151 | 0.0218 |
| Croatia | 0.0438 | 0.0147 | 0.0189 | 0.1835 | - | 0.0017 | 0.0045 | 0.0070 |
| Turkey | 0.1907 | 0.1934 | 0.1852 | 0.0610 | 0.1474 | - | 0.0037 | 0.0039 |
| Bulgaria [Turks] | 0.1674 | 0.1610 | 0.1586 | 0.0608 | 0.1079 | - 0.0012 | - | 0.0079 |
| Bulgaria [Bulgarians] | 0.1334 | 0.1315 | 0.1273 | 0.0503 | 0.0782 | 0.0196 | -0.0029 | - |
| Macedonia | 0.1080 | 0.1061 | 0.0978 | 0.0698 | 0.0531 | 0.0612 | 0.0267 | 0.0062 |
| Bosnia | 0.0493 | 0.0144 | 0.0201 | 0.2447 | 0.0132 | 0.1843 | 0.1576 | 0.1212 |
| Serbia | 0.0700 | 0.0729 | 0.0627 | 0.0749 | 0.0372 | 0.0626 | 0.0277 | 0.0111 |
| Ljubljana [Slovenia] | 0.0160 | 0.0169 | 0.0103 | 0.1423 | 0.0204 | 0.1140 | 0.0790 | 0.0592 |
| Gdansk [Poland] | 0.0055 | 0.0020 | -0.0011 | 0.2353 | 0.0338 | 0.1860 | 0.1585 | 0.1322 |
| Lublin [Poland] | 0.0078 | 0.0061 | 0.0035 | 0.2562 | 0.0456 | 0.1935 | 0.1749 | 0.1488 |
| Wrozlaw [Poland] | 0.0057 | 0.0069 | 0.0049 | 0.2657 | 0.0514 | 0.2049 | 0.1891 | 0.1605 |
| Berlin [Germany] | 0.0948 | 0.0995 | 0.0934 | 0.0760 | 0.0788 | 0.0450 | 0.0269 | 0.0240 |
| Munich [Germany] | 0.1626 | 0.1690 | 0.1608 | 0.0698 | 0.1305 | 0.0274 | 0.0249 | 0.0298 |
| Cologne [Germany] | 0.1955 | 0.2023 | 0.1988 | 0.0764 | 0.1570 | 0.0281 | 0.0374 | 0.0429 |
| Vien [Austria] | 0.2540 | 0.2522 | 0.2550 | 0.0922 | 0.2078 | 0.0360 | 0.0515 | 0.0599 |
| Denmark | 0.3345 | 0.3185 | 0.3225 | 0.1623 | 0.2531 | 0.0559 | 0.0907 | 0.1113 |
| Sweden | 0.2302 | 0.2389 | 0.2297 | 0.0738 | 0.1912 | 0.0300 | 0.0401 | 0.0531 |
| Vilnus [Lithuania] | 0.0361 | 0.0359 | 0.0421 | 0.1821 | 0.0562 | 0.1549 | 0.1308 | 0.1090 |
| Riga [Latvia] | 0.0526 | 0.0544 | 0.0588 | 0.1734 | 0.0734 | 0.1395 | 0.1199 | 0.1040 |
| Tartu [Estonia] | 0.1465 | 0.1614 | 0.1654 | 0.1346 | 0.1667 | 0.1153 | 0.1135 | 0.1070 |
| Budapest [Hungary] | 0.0903 | 0.0895 | 0.0829 | 0.0698 | 0.0599 | 0.0276 | 0.0038 | 0.0046 |
| Athens [Greece] | 0.1457 | 0.1527 | 0.1503 | 0.0391 | 0.1099 | 0.0084 | -0.0049 | 0.0009 |
| Thrace [Greece] | 0.0580 | 0.0614 | 0.0511 | 0.0801 | 0.0322 | 0.0739 | 0.0330 | 0.0194 |
| Constanta [Romania] | 0.0917 | 0.0423 | 0.0544 | 0.1919 | 0.0023 | 0.1043 | 0.0602 | 0.0416 |
| Ploiesti [Romania] | 0.1008 | 0.0683 | 0.0690 | 0.1311 | 0.0184 | 0.0550 | 0.0175 | 0.0066 |
| Lazio [Italy] | 0.2394 | 0.2487 | 0.2409 | 0.0559 | 0.1936 | 0.0077 | 0.0186 | 0.0393 |
| Lombardia [Italy] | 0.2281 | 0.2414 | 0.2344 | 0.0684 | 0.1937 | 0.0336 | 0.0499 | 0.0611 |
| Sicily [Italy] | 0.2747 | 0.2711 | 0.2666 | 0.1065 | 0.2068 | 0.0063 | 0.0176 | 0.0465 |

(Contd.)

| Population | Macedonia | Bosnia | Serbia | Ljubljana [Slovenia] | Gdansk [Poland] | Lublin [Poland] | Wrozlaw [Poland] | Berlin [Germany] |
|---|---|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | 0.0099 | 0.0059 | 0.0075 | 0.0063 | 0.0060 | 0.0011 | 0.0046 | 0.0037 |
| Etulia [Gagauzia] | 0.0143 | 0.0071 | 0.0079 | 0.0036 | 0.0074 | 0.0029 | 0.0083 | 0.0042 |
| Karahasani [Moldova] | 0.0107 | 0.0105 | 0.0091 | 0.0112 | 0.0111 | 0.0033 | 0.0112 | 0.0061 |
| Sofia [Moldova] | 0.0138 | 0.0124 | 0.0091 | 0.0094 | 0.0079 | 0.0046 | 0.0055 | 0.0041 |
| East Romania | 0.0133 | 0.0109 | 0.0122 | 0.0109 | 0.0069 | 0.0028 | 0.0065 | 0.0067 |
| Rashkovo [Ukrain] | 0.0056 | 0.0078 | 0.0050 | 0.0102 | 0.0093 | 0.0031 | 0.0105 | 0.0042 |
| Moscow [Russia] | 0.0119 | 0.0082 | 0.0064 | 0.0131 | 0.0100 | 0.0042 | 0.0143 | 0.0055 |
| Nowgorod [Russia] | 0.0060 | 0.0087 | 0.0081 | 0.0122 | 0.0116 | 0.0082 | 0.0149 | 0.0065 |
| Byelorussia | 0.0083 | 0.0087 | 0.0051 | 0.0086 | 0.0102 | 0.0042 | 0.0117 | 0.0053 |
| Kiev [Ukrain] | 0.0086 | 0.0094 | 0.0062 | 0.0110 | 0.0124 | 0.0054 | 0.0161 | 0.0056 |
| Albania | 0.0275 | 0.0088 | 0.0169 | 0.0076 | 0.0054 | 0.0007 | 0.0038 | 0.0093 |
| Croatia | 0.0119 | 0.0142 | 0.0083 | 0.0082 | 0.0069 | 0.0031 | 0.0085 | 0.0034 |
| Turkey | 0.0036 | 0.0021 | 0.0030 | 0.0023 | 0.0019 | 0.0005 | 0.0016 | 0.0018 |
| Bulgaria [Turks] | 0.0101 | 0.0048 | 0.0066 | 0.0061 | 0.0037 | 0.0010 | 0.0041 | 0.0037 |
| Bulgaria [Bulgarians] | 0.0164 | 0.0088 | 0.0089 | 0.0047 | 0.0042 | 0.0018 | 0.0035 | 0.0034 |
| Macedonia | - | 0.0130 | 0.0140 | 0.0083 | 0.0051 | 0.0020 | 0.0066 | 0.0038 |
| Bosnia | 0.0820 | - | 0.0102 | 0.0084 | 0.0084 | 0.0040 | 0.0093 | 0.0028 |
| Serbia | -0.0035 | 0.0611 | - | 0.0077 | 0.0051 | 0.0024 | 0.0058 | 0.0040 |
| Ljubljana [Slovenia] | 0.0403 | 0.0297 | 0.0151 | - | 0.0100 | 0.0030 | 0.0113 | 0.0061 |
| Gdansk [Poland] | 0.1043 | 0.0392 | 0.0692 | 0.0143 | - | 0.0042 | 0.0166 | 0.0069 |
| Lublin [Poland] | 0.1274 | 0.0532 | 0.0848 | 0.0191 | -0.0048 | - | 0.0051 | 0.0027 |
| Wrozlaw [Poland] | 0.1411 | 0.0649 | 0.0966 | 0.0264 | -0.0021 | -0.0058 | - | 0.0078 |
| Berlin [Germany] | 0.0445 | 0.1083 | 0.0281 | 0.0360 | 0.0939 | 0.0955 | 0.1029 | - |
| Munich [Germany] | 0.0698 | 0.1803 | 0.0563 | 0.0828 | 0.1603 | 0.1651 | 0.1729 | 0.0068 |
| Cologne [Germany] | 0.0981 | 0.2253 | 0.0829 | 0.1130 | 0.1956 | 0.2026 | 0.2080 | 0.0186 |
| Vien [Austria] | 0.1221 | 0.2824 | 0.1206 | 0.1655 | 0.2571 | 0.2703 | 0.2748 | 0.0525 |
| Denmark | 0.1991 | 0.3401 | 0.1804 | 0.2123 | 0.3072 | 0.3195 | 0.3272 | 0.0682 |
| Sweden | 0.1061 | 0.2393 | 0.0994 | 0.1409 | 0.2293 | 0.2342 | 0.2436 | 0.0345 |
| Vilnus [Lithuania] | 0.1126 | 0.0899 | 0.0786 | 0.0341 | 0.0507 | 0.0464 | 0.0397 | 0.0553 |
| Riga [Latvia] | 0.1158 | 0.1076 | 0.0799 | 0.0372 | 0.0642 | 0.0576 | 0.0532 | 0.0400 |
| Tartu [Estonia] | 0.1450 | 0.2144 | 0.1205 | 0.1088 | 0.1711 | 0.1677 | 0.1640 | 0.0435 |
| Budapest [Hungary] | 0.0183 | 0.0960 | 0.0067 | 0.0265 | 0.0845 | 0.0928 | 0.1029 | 0.0034 |
| Athens [Greece] | 0.0314 | 0.1548 | 0.0289 | 0.0726 | 0.1518 | 0.1636 | 0.1747 | 0.0170 |
| Thrace [Greece] | -0.0034 | 0.0439 | -0.0110 | 0.0091 | 0.0629 | 0.0800 | 0.0931 | 0.0383 |
| Constanta [Romania] | 0.0290 | 0.0220 | 0.0258 | 0.0296 | 0.0668 | 0.0880 | 0.1000 | 0.0602 |
| Ploiesti [Romania] | -0.0005 | 0.0454 | -0.0041 | 0.0178 | 0.0720 | 0.0925 | 0.1087 | 0.0277 |
| Lazio [Italy] | 0.0962 | 0.2495 | 0.0940 | 0.1496 | 0.2391 | 0.2483 | 0.2587 | 0.0469 |
| Lombardia [Italy] | 0.1207 | 0.2582 | 0.1069 | 0.1441 | 0.2326 | 0.2387 | 0.2448 | 0.0365 |
| Sicily [Italy] | 0.1063 | 0.2658 | 0.1116 | 0.1735 | 0.2569 | 0.2701 | 0.2840 | 0.0709 |

(Contd.)

| Population | Munich [Germany] | Cologne [Germany] | Vien [Austria] | Denmark | Sweden | Vilnus [Lithuania] | Riga [Latvia] | Tartu [Estonia] |
|---|---|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | 0.0044 | 0.0027 | 0.0023 | 0.0037 | 0.0068 | 0.0020 | 0.0043 | 0.0030 |
| Etulia [Gagauzia] | 0.0036 | 0.0025 | 0.0023 | 0.0033 | 0.0016 | 0.0078 | 0.0076 | 0.0066 |
| Karahasani [Moldova] | 0.0058 | 0.0057 | 0.0019 | 0.0035 | 0.0026 | 0.0073 | 0.0079 | 0.0050 |
| Sofia [Moldova] | 0.0043 | 0.0024 | 0.0015 | 0.0035 | 0.0041 | 0.0051 | 0.0073 | 0.0062 |
| East Romania | 0.0095 | 0.0051 | 0.0024 | 0.0065 | 0.0070 | 0.0032 | 0.0050 | 0.0034 |
| Rashkovo [Ukrain] | 0.0047 | 0.0045 | 0.0009 | 0.0037 | 0.0062 | 0.0095 | 0.0110 | 0.0044 |
| Moscow [Russia] | 0.0064 | 0.0026 | 0.0007 | 0.0032 | 0.0054 | 0.0151 | 0.0123 | 0.0090 |
| Nowgorod [Russia] | 0.0050 | 0.0031 | 0.0003 | 0.0022 | 0.0028 | 0.0139 | 0.0094 | 0.0092 |
| Byelorussia | 0.0045 | 0.0020 | 0.0009 | 0.0028 | 0.0040 | 0.0092 | 0.0073 | 0.0068 |
| Kiev [Ukrain] | 0.0055 | 0.0034 | 0.0015 | 0.0023 | 0.0061 | 0.0122 | 0.0112 | 0.0069 |
| Albania | 0.0124 | 0.0073 | 0.0054 | 0.0099 | 0.0087 | 0.0017 | 0.0046 | 0.0038 |
| Croatia | 0.0037 | 0.0020 | 0.0006 | 0.0009 | 0.0021 | 0.0059 | 0.0059 | 0.0037 |
| Turkey | 0.0023 | 0.0024 | 0.0008 | 0.0017 | 0.0021 | 0.0014 | 0.0014 | 0.0011 |
| Bulgaria [Turks] | 0.0046 | 0.0027 | 0.0012 | 0.0029 | 0.0049 | 0.0028 | 0.0040 | 0.0037 |
| Bulgaria [Bulgarians] | 0.0036 | 0.0020 | 0.0022 | 0.0022 | 0.0026 | 0.0025 | 0.0034 | 0.0027 |
| Macedonia | 0.0039 | 0.0011 | 0.0023 | 0.0019 | 0.0035 | 0.0046 | 0.0057 | 0.0039 |
| Bosnia | 0.0027 | 0.0016 | 0.0005 | 0.0004 | 0.0014 | 0.0040 | 0.0064 | 0.0038 |
| Serbia | 0.0049 | 0.0035 | 0.0021 | 0.0038 | 0.0055 | 0.0032 | 0.0049 | 0.0033 |
| Ljubljana [Slovenia] | 0.0076 | 0.0047 | 0.0020 | 0.0062 | 0.0087 | 0.0086 | 0.0083 | 0.0058 |
| Gdansk [Poland] | 0.0060 | 0.0040 | 0.0019 | 0.0032 | 0.0035 | 0.0095 | 0.0075 | 0.0049 |
| Lublin [Poland] | 0.0018 | 0.0012 | 0.0005 | 0.0005 | 0.0009 | 0.0046 | 0.0034 | 0.0034 |
| Wrozlaw [Poland] | 0.0069 | 0.0050 | 0.0018 | 0.0035 | 0.0031 | 0.0118 | 0.0099 | 0.0070 |
| Berlin [Germany] | 0.0079 | 0.0055 | 0.0020 | 0.0080 | 0.0056 | 0.0044 | 0.0040 | 0.0050 |
| Munich [Germany] | - | 0.0105 | 0.0032 | 0.0143 | 0.0104 | 0.0047 | 0.0038 | 0.0058 |
| Cologne [Germany] | 0.0009 | - | 0.0035 | 0.0146 | 0.0075 | 0.0023 | 0.0026 | 0.0033 |
| Vien [Austria] | 0.0198 | 0.0091 | - | 0.0051 | 0.0039 | 0.0007 | 0.0010 | 0.0005 |
| Denmark | 0.0351 | 0.0250 | 0.0049 | - | 0.0150 | 0.0018 | 0.0013 | 0.0060 |
| Sweden | 0.0080 | 0.0080 | 0.0051 | 0.0107 | - | 0.0024 | 0.0044 | 0.0084 |
| Vilnus [Lithuania] | 0.0970 | 0.1054 | 0.1575 | 0.1885 | 0.1576 | - | 0.0097 | 0.0069 |
| Riga [Latvia] | 0.0754 | 0.0844 | 0.1358 | 0.1566 | 0.1298 | -0.0021 | - | 0.0074 |
| Tartu [Estonia] | 0.0443 | 0.0348 | 0.0584 | 0.0619 | 0.0653 | 0.0568 | 0.0359 | - |
| Budapest [Hungary] | 0.0150 | 0.0314 | 0.0623 | 0.0944 | 0.0484 | 0.0672 | 0.0582 | 0.0761 |
| Athens [Greece] | 0.0144 | 0.0191 | 0.0410 | 0.0733 | 0.0283 | 0.1085 | 0.0964 | 0.0803 |
| Thrace [Greece] | 0.0733 | 0.1033 | 0.1386 | 0.2121 | 0.1193 | 0.0775 | 0.0810 | 0.1280 |
| Constanta [Romania] | 0.1065 | 0.1379 | 0.1505 | 0.2293 | 0.1598 | 0.0761 | 0.0859 | 0.1475 |
| Ploiesti [Romania] | 0.0617 | 0.0945 | 0.1197 | 0.1945 | 0.1064 | 0.0818 | 0.0837 | 0.1299 |
| Lazio [Italy] | 0.0198 | 0.0146 | 0.0165 | 0.0323 | 0.0094 | 0.1746 | 0.1527 | 0.0963 |
| Lombardia [Italy] | 0.0103 | -0.0010 | 0.0092 | 0.0166 | 0.0056 | 0.1390 | 0.1141 | 0.0467 |
| Sicily [Italy] | 0.0484 | 0.0518 | 0.0423 | 0.0654 | 0.0379 | 0.2159 | 0.1962 | 0.1517 |

(Contd.)

| Population | Budapest | Athens | Thrace | Constanta [Romania] | Ploiesti [Romania] | Lazio [Italy] | Lombardia [Italy] | Sicily [Italy] |
|---|---|---|---|---|---|---|---|---|
| Kongaz [Gagauzia] | 0.0043 | 0.0051 | 0.0033 | 0.0027 | 0.0047 | 0.0050 | 0.0051 | 0.0033 |
| Etulia [Gagauzia] | 0.0051 | 0.0053 | 0.0039 | 0.0041 | 0.0078 | 0.0039 | 0.0039 | 0.0035 |
| Karahasani [Moldova] | 0.0047 | 0.0051 | 0.0068 | 0.0040 | 0.0073 | 0.0048 | 0.0066 | 0.0027 |
| Sofia [Moldova] | 0.0066 | 0.0055 | 0.0103 | 0.0032 | 0.0106 | 0.0028 | 0.0029 | 0.0030 |
| East Romania | 0.0073 | 0.0080 | 0.0101 | 0.0070 | 0.0136 | 0.0060 | 0.0122 | 0.0035 |
| Rashkovo [Ukrain] | 0.0039 | 0.0027 | 0.0060 | 0.0025 | 0.0065 | 0.0027 | 0.0015 | 0.0009 |
| Moscow [Russia] | 0.0068 | 0.0044 | 0.0051 | 0.0034 | 0.0088 | 0.0017 | 0.0010 | 0.0019 |
| Nowgorod [Russia] | 0.0049 | 0.0042 | 0.0051 | 0.0000 | 0.0050 | 0.0022 | 0.0034 | 0.0027 |
| Byelorussia | 0.0063 | 0.0046 | 0.0045 | 0.0028 | 0.0040 | 0.0013 | 0.0012 | 0.0013 |
| Kiev [Ukrain] | 0.0058 | 0.0041 | 0.0050 | 0.0031 | 0.0068 | 0.0016 | 0.0012 | 0.0013 |
| Albania | 0.0073 | 0.0109 | 0.0074 | 0.0035 | 0.0110 | 0.0123 | 0.0219 | 0.0085 |
| Croatia | 0.0049 | 0.0038 | 0.0079 | 0.0068 | 0.0108 | 0.0011 | 0.0019 | 0.0011 |
| Turkey | 0.0024 | 0.0031 | 0.0027 | 0.0012 | 0.0028 | 0.0030 | 0.0032 | 0.0023 |
| Bulgaria [Turks] | 0.0046 | 0.0045 | 0.0042 | 0.0032 | 0.0077 | 0.0043 | 0.0049 | 0.0032 |
| Bulgaria [Bulgarians] | 0.0050 | 0.0050 | 0.0040 | 0.0037 | 0.0082 | 0.0043 | 0.0059 | 0.0040 |
| Macedonia | 0.0056 | 0.0052 | 0.0079 | 0.0061 | 0.0129 | 0.0043 | 0.0045 | 0.0038 |
| Bosnia | 0.0035 | 0.0034 | 0.0152 | 0.0070 | 0.0160 | 0.0005 | 0.0015 | 0.0008 |
| Serbia | 0.0047 | 0.0056 | 0.0065 | 0.0040 | 0.0085 | 0.0040 | 0.0056 | 0.0028 |
| Ljubljana [Slovenia] | 0.0064 | 0.0044 | 0.0074 | 0.0029 | 0.0069 | 0.0028 | 0.0050 | 0.0017 |
| Gdansk [Poland] | 0.0059 | 0.0038 | 0.0051 | 0.0034 | 0.0052 | 0.0019 | 0.0029 | 0.0015 |
| Lublin [Poland] | 0.0025 | 0.0013 | 0.0015 | 0.0005 | 0.0027 | 0.0008 | 0.0006 | 0.0009 |
| Wrozlaw [Poland] | 0.0050 | 0.0034 | 0.0045 | 0.0027 | 0.0055 | 0.0015 | 0.0022 | 0.0014 |
| Berlin [Germany] | 0.0049 | 0.0040 | 0.0029 | 0.0014 | 0.0029 | 0.0039 | 0.0073 | 0.0025 |
| Munich [Germany] | 0.0067 | 0.0057 | 0.0053 | 0.0019 | 0.0040 | 0.0063 | 0.0121 | 0.0028 |
| Cologne [Germany] | 0.0043 | 0.0034 | 0.0034 | 0.0012 | 0.0023 | 0.0059 | 0.0090 | 0.0024 |
| Vien [Austria] | 0.0014 | 0.0017 | 0.0016 | 0.0000 | 0.0013 | 0.0025 | 0.0037 | 0.0008 |
| Denmark | 0.0072 | 0.0046 | 0.0041 | 0.0010 | 0.0018 | 0.0074 | 0.0126 | 0.0028 |
| Sweden | 0.0059 | 0.0046 | 0.0034 | 0.0008 | 0.0025 | 0.0059 | 0.0082 | 0.0019 |
| Vilnus [Lithuania] | 0.0044 | 0.0021 | 0.0020 | 0.0015 | 0.0035 | 0.0011 | 0.0008 | 0.0014 |
| Riga [Latvia] | 0.0034 | 0.0028 | 0.0046 | 0.0013 | 0.0057 | 0.0014 | 0.0014 | 0.0010 |
| Tartu [Estonia] | 0.0056 | 0.0031 | 0.0042 | 0.0005 | 0.0031 | 0.0017 | 0.0021 | 0.0015 |
| Budapest [Hungary] | - | 0.0041 | 0.0047 | 0.0036 | 0.0053 | 0.0034 | 0.0056 | 0.0018 |
| Athens [Greece] | 0.0068 | - | 0.0033 | 0.0019 | 0.0036 | 0.0039 | 0.0067 | 0.0026 |
| Thrace [Greece] | 0.0123 | 0.0348 | - | 0.0157 | 0.0100 | 0.0036 | 0.0045 | 0.0015 |
| Constanta [Romania] | 0.0346 | 0.0740 | 0.0245 | - | 0.0081 | 0.0009 | 0.0012 | 0.0003 |
| Ploiesti [Romania] | 0.0020 | 0.0308 | -0.0018 | -0.0141 | - | 0.0026 | 0.0037 | 0.0014 |
| Lazio [Italy] | 0.0456 | 0.0135 | 0.1099 | 0.1603 | 0.1019 | - | 0.0091 | 0.0034 |
| Lombardia [Italy] | 0.0520 | 0.0289 | 0.1251 | 0.1779 | 0.1259 | 0.0101 | - | 0.0046 |
| Sicily [Italy] | 0.0584 | 0.0313 | 0.1340 | 0.1561 | 0.1004 | 0.0162 | 0.0561 | - |

Note. - $R_{ST}$ values significant at 5% level are shaded.

**Appendix 6** The distribution of Y-STR haplotypes affiliated with binary haplogroups in six samples analyzed

| Haplotype number | Haplogroup | Allele status at | | | | | | | No. of instances | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | Gagauzes K. | Gagauzes E. | Mold. K. | Mold. S. | Romanians | Ukrainians | Total |
| 1 | E3b1-M78 | 13 | 10 | 17 | 23 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 2 | E3b1-M78 | 13 | 10 | 17 | 24 | 10 | 11 | 13 | 2 | 3 | 3 | 3 | 2 | 0 | 13 |
| 3 | E3b1-M78 | 13 | 10 | 17 | 24 | 10 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 4 | E3b1-M78 | 13 | 10 | 17 | 25 | 10 | 11 | 13 | 1 | 1 | 0 | 0 | 0 | 0 | 2 |
| 5 | E3b1-M78 | 13 | 10 | 18 | 24 | 10 | 11 | 13 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| 6 | E3b1-M78 | 13 | 11 | 16 | 24 | 9 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 7 | E3b1-M78 | 13 | 11 | 17 | 24 | 10 | 9 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 8 | E3b1-M78 | 13 | 11 | 17 | 24 | 10 | 11 | 14 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 9 | E3b1-M78 | 13 | 11 | 17 | 25 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 10 | E3b1-M78 | 13 | 12 | 16 | 25 | 10 | 12 | 14 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 11 | E3b1-M78 | 14 | 10 | 17 | 24 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 12 | E3b1-M123 | 13 | 9 | 18 | 24 | 10 | 11 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 13 | E3b1-M123 | 13 | 9 | 18 | 24 | 11 | 11 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | E3b1-M123 | 13 | 10 | 17 | 25 | 9 | 11 | 14 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| 15 | G-M201 | 13 | 11 | 17 | 24 | 10 | 9 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 16 | G-M201 | 14 | 9 | 16 | 23 | 10 | 11 | 15 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 17 | G-M201 | 14 | 9 | 16 | 24 | 10 | 12 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 18 | G-M201 | 15 | 8 | 17 | 22 | 10 | 10 | 14 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 19 | G-M201 | 15 | 9 | 17 | 21 | 10 | 11 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 20 | G-M201 | 15 | 9 | 17 | 23 | 10 | 12 | 14 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 21 | G-M201 | 16 | 9 | 16 | 21 | 10 | 11 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 22 | G-M201 | 16 | 9 | 16 | 22 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 23 | G-M201 | 16 | 9 | 17 | 22 | 10 | 10 | 14 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| 24 | I*-M170 | 15 | 11 | 16 | 22 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 25 | I*-M170 | 17 | 11 | 18 | 24 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 26 | I1a-M253 | 13 | 9 | 17 | 23 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 27 | I1a-M253 | 14 | 9 | 15 | 22 | 10 | 11 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 28 | I1a-M253 | 14 | 9 | 16 | 22 | 10 | 11 | 13 | 0 | 0 | 0 | 2 | 0 | 0 | 2 |
| 29 | I1a-M253 | 14 | 9 | 16 | 23 | 10 | 11 | 13 | 2 | 0 | 0 | 0 | 1 | 2 | 5 |
| 30 | I1a-M253 | 14 | 9 | 16 | 24 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 31 | I1a-M253 | 14 | 9 | 17 | 22 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |
| 32 | I1a-M253 | 15 | 9 | 17 | 22 | 10 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 33 | I1a-M253 | 16 | 9 | 16 | 22 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 34 | I1b-P37 | 14 | 10 | 17 | 24 | 10 | 11 | 13 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| 35 | I1b-P37 | 15 | 9 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 36 | I1b-P37 | 15 | 9 | 19 | 24 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 37 | I1b-P37 | 15 | 10 | 17 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 38 | I1b-P37 | 15 | 10 | 17 | 24 | 11 | 11 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 39 | I1b-P37 | 15 | 10 | 18 | 22 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 40 | I1b-P37 | 15 | 10 | 18 | 24 | 11 | 11 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 41 | I1b-P37 | 15 | 10 | 18 | 24 | 11 | 11 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 42 | I1b-P37 | 15 | 10 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 43 | I1b-P37 | 15 | 10 | 19 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 44 | I1b-P37 | 15 | 10 | 19 | 24 | 11 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 45 | I1b-P37 | 16 | 10 | 16 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| 46 | I1b-P37 | 16 | 10 | 17 | 24 | 11 | 11 | 13 | 1 | 0 | 1 | 4 | 2 | 1 | 9 |
| 47 | I1b-P37 | 16 | 10 | 18 | 24 | 10 | 11 | 13 | 2 | 0 | 0 | 2 | 3 | 0 | 7 |

(Contd.)

| Haplotype number | Haplogroup | Allele status at | | | | | | | No. of instances | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | Gagauzes K. | Gagauzes E. | Mold. K. | Mold. S. | Romanians | Ukrainians | Total |
| 48 | I1b-P37 | 16 | 10 | 18 | 24 | 11 | 11 | 13 | 0 | 1 | 2 | 1 | 3 | 1 | 8 |
| 49 | I1b-P37 | 16 | 10 | 18 | 24 | 11 | 11 | 14 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 50 | I1b-P37 | 16 | 10 | 18 | 24 | 11 | 11 | 15 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 51 | I1b-P37 | 16 | 10 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| 52 | I1b-P37 | 16 | 10 | 18 | 25 | 11 | 13 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 53 | I1b-P37 | 16 | 10 | 19 | 24 | 11 | 11 | 13 | 2 | 1 | 1 | 3 | 2 | 0 | 9 |
| 54 | I1b-P37 | 16 | 10 | 20 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 55 | I1b-P37 | 16 | 10 | 20 | 24 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 56 | I1b-P37 | 16 | 11 | 18 | 24 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| 57 | I1b-P37 | 16 | 11 | 19 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 58 | I1b-P37 | 17 | 10 | 17 | 24 | 11 | 11 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 59 | I1b-P37 | 17 | 10 | 18 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 60 | I1b-P37 | 17 | 10 | 18 | 24 | 11 | 11 | 13 | 1 | 2 | 2 | 0 | 1 | 0 | 6 |
| 61 | I1b-P37 | 17 | 10 | 19 | 24 | 11 | 11 | 13 | 0 | 2 | 2 | 0 | 1 | 0 | 5 |
| 62 | I1b-P37 | 17 | 11 | 19 | 23 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| 63 | I1b-P37 | 18 | 10 | 19 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 64 | I1c-M223 | 15 | 9 | 16 | 23 | 10 | 12 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 65 | I1c-M223 | 15 | 10 | 16 | 23 | 10 | 12 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 66 | I1c-M223 | 15 | 10 | 16 | 23 | 11 | 12 | 14 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 67 | I1c-M223 | 15 | 11 | 18 | 23 | 10 | 12 | 15 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| 68 | I1c-M223 | 16 | 10 | 16 | 25 | 10 | 12 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 69 | I1c-M223 | 16 | 10 | 17 | 23 | 10 | 12 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 70 | J*-12f2 | 15 | 10 | 16 | 23 | 9 | 11 | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 71 | J1-M267 | 14 | 10 | 16 | 23 | 10 | 11 | 12 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| 72 | J1-M267 | 14 | 10 | 16 | 23 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 73 | J1-M267 | 14 | 11 | 16 | 22 | 11 | 11 | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 74 | J1-M267 | 15 | 10 | 17 | 23 | 10 | 11 | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 75 | J2*-M172 | 14 | 10 | 17 | 23 | 10 | 11 | 12 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 76 | J2*-M172 | 14 | 10 | 17 | 23 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 77 | J2*-M172 | 14 | 11 | 17 | 23 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 78 | J2*-M172 | 15 | 10 | 16 | 23 | 9 | 11 | 12 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 79 | J2*-M172 | 15 | 10 | 16 | 24 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 80 | J2*-M172 | 15 | 11 | 17 | 23 | 10 | 11 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 81 | J2*-M172 | 16 | 10 | 16 | 23 | 10 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 82 | J2*-M172 | 16 | 10 | 16 | 24 | 9 | 11 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 83 | J2a1a-M47 | 15 | 9 | 16 | 24 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 84 | J2a1b*-M67 | 14 | 10 | 14 | 22 | 10 | 11 | 12 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 85 | J2a1b*-M67 | 14 | 10 | 16 | 23 | 10 | 11 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 86 | J2a1b1-M92 | 16 | 10 | 18 | 23 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 87 | J2b-M12 | 14 | 10 | 16 | 23 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 88 | J2b-M12 | 14 | 11 | 16 | 24 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 89 | J2b-M12 | 15 | 9 | 16 | 24 | 10 | 11 | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 90 | J2b-M12 | 15 | 10 | 16 | 24 | 10 | 11 | 12 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 91 | K2-M70 | 13 | 11 | 16 | 23 | 10 | 13 | 13 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| 92 | K2-M70 | 14 | 12 | 17 | 23 | 10 | 15 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 93 | K2-M70 | 16 | 10 | 17 | 23 | 10 | 13 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 94 | N2-P43 | 14 | 10 | 16 | 23 | 10 | 15 | 14 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |

(Contd.)

| Haplotype number | Haplogroup | Allele status at | | | | | | | No. of instances | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | Gagauzes K. | Gagauzes E. | Mold. K. | Mold. S. | Romanians | Ukrainians | Total |
| 95 | N3a-M178 | 14 | 11 | 16 | 23 | 10 | 14 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 96 | N3a-M178 | 14 | 11 | 16 | 23 | 11 | 14 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 97 | N3a-M178 | 15 | 10 | 16 | 23 | 11 | 14 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 98 | N3a-M178 | 15 | 10 | 16 | 23 | 11 | 14 | 14 | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| 99 | N3a-M178 | 15 | 12 | 16 | 23 | 11 | 15 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 100 | Q-M242 | 15 | 10 | 17 | 23 | 10 | 16 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 101 | R1a1-M17 | 14 | 10 | 15 | 24 | 11 | 13 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 102 | R1a1-M17 | 15 | 10 | 16 | 24 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 103 | R1a1-M17 | 15 | 10 | 16 | 25 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 3 | 3 |
| 104 | R1a1-M17 | 15 | 10 | 16 | 25 | 10 | 12 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 105 | R1a1-M17 | 15 | 10 | 16 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 106 | R1a1-M17 | 15 | 10 | 16 | 26 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 107 | R1a1-M17 | 15 | 10 | 17 | 24 | 11 | 11 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 108 | R1a1-M17 | 15 | 10 | 17 | 25 | 10 | 11 | 13 | 1 | 0 | 0 | 1 | 1 | 0 | 3 |
| 109 | R1a1-M17 | 15 | 10 | 17 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 3 | 0 | 2 | 5 |
| 110 | R1a1-M17 | 15 | 10 | 17 | 26 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| 111 | R1a1-M17 | 15 | 10 | 18 | 25 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 112 | R1a1-M17 | 16 | 7 | 17 | 25 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 113 | R1a1-M17 | 16 | 9 | 16 | 26 | 10 | 11 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 114 | R1a1-M17 | 16 | 9 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 115 | R1a1-M17 | 16 | 10 | 15 | 25 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 116 | R1a1-M17 | 16 | 10 | 16 | 23 | 10 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 117 | R1a1-M17 | 16 | 10 | 16 | 24 | 10 | 11 | 13 | 0 | 1 | 0 | 0 | 0 | 1 | 2 |
| 118 | R1a1-M17 | 16 | 10 | 16 | 24 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 119 | R1a1-M17 | 16 | 10 | 16 | 25 | 10 | 11 | 13 | 0 | 3 | 3 | 0 | 1 | 0 | 7 |
| 120 | R1a1-M17 | 16 | 10 | 16 | 25 | 10 | 11 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 121 | R1a1-M17 | 16 | 10 | 16 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| 122 | R1a1-M17 | 16 | 10 | 17 | 23 | 10 | 13 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 123 | R1a1-M17 | 16 | 10 | 17 | 24 | 10 | 11 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 124 | R1a1-M17 | 16 | 10 | 17 | 24 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 1 | 2 |
| 125 | R1a1-M17 | 16 | 10 | 17 | 25 | 10 | 11 | 13 | 1 | 0 | 6 | 1 | 0 | 3 | 11 |
| 126 | R1a1-M17 | 16 | 10 | 17 | 25 | 11 | 11 | 13 | 0 | 1 | 1 | 0 | 0 | 2 | 4 |
| 127 | R1a1-M17 | 16 | 10 | 17 | 25 | 11 | 11 | 14 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 128 | R1a1-M17 | 16 | 10 | 17 | 26 | 11 | 11 | 13 | 0 | 6 | 0 | 0 | 0 | 0 | 6 |
| 129 | R1a1-M17 | 16 | 10 | 18 | 25 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 130 | R1a1-M17 | 16 | 11 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 4 | 0 | 5 |
| 131 | R1a1-M17 | 17 | 9 | 17 | 25 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 132 | R1a1-M17 | 17 | 10 | 17 | 24 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 133 | R1a1-M17 | 17 | 10 | 17 | 25 | 10 | 11 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 134 | R1a1-M17 | 17 | 10 | 17 | 25 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 135 | R1a1-M17 | 17 | 10 | 18 | 25 | 11 | 11 | 13 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| 136 | R1a1-M17 | 17 | 10 | 18 | 27 | 10 | 11 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 137 | R1a1-M17 | 17 | 11 | 17 | 25 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 138 | R1a1-M17 | 17 | 11 | 17 | 25 | 11 | 11 | 13 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| 139 | R1a1-M17 | 17 | 11 | 19 | 25 | 10 | 11 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 140 | R1b*-P25 | 14 | 11 | 16 | 19 | 11 | 13 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 141 | R1b*-P25 | 14 | 11 | 16 | 25 | 11 | 13 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

(Contd.)

| Haplotype number | Haplogroup | Allele status at | | | | | | | No. of instances | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | Gagauzes K. | Gagauzes E. | Mold. K. | Mold. S. | Romanians | Ukrainians | Total |
| 142 | R1b*-P25 | 15 | 12 | 14 | 19 | 12 | 14 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 143 | R1b3-M269 | 13 | 10 | 16 | 23 | 11 | 13 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 144 | R1b3-M269 | 13 | 10 | 17 | 25 | 11 | 12 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 145 | R1b3-M269 | 13 | 11 | 17 | 25 | 11 | 12 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 146 | R1b3-M269 | 14 | 10 | 15 | 24 | 10 | 14 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 147 | R1b3-M269 | 14 | 10 | 15 | 24 | 11 | 13 | 12 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| 148 | R1b3-M269 | 14 | 10 | 16 | 23 | 10 | 13 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 149 | R1b3-M269 | 14 | 10 | 16 | 23 | 11 | 13 | 13 | 0 | 0 | 1 | 0 | 0 | 1 | 2 |
| 150 | R1b3-M269 | 14 | 10 | 16 | 23 | 11 | 14 | 13 | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| 151 | R1b3-M269 | 14 | 10 | 16 | 24 | 10 | 12 | 13 | 0 | 0 | 0 | 0 | 2 | 0 | 2 |
| 152 | R1b3-M269 | 14 | 10 | 16 | 24 | 10 | 13 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 153 | R1b3-M269 | 14 | 10 | 16 | 24 | 11 | 11 | 12 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| 154 | R1b3-M269 | 14 | 10 | 16 | 24 | 11 | 12 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 155 | R1b3-M269 | 14 | 10 | 16 | 24 | 11 | 13 | 13 | 1 | 0 | 2 | 0 | 4 | 0 | 7 |
| 156 | R1b3-M269 | 14 | 10 | 16 | 24 | 12 | 13 | 13 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| 157 | R1b3-M269 | 14 | 10 | 16 | 25 | 10 | 13 | 13 | 0 | 0 | 2 | 0 | 0 | 0 | 2 |
| 158 | R1b3-M269 | 14 | 10 | 17 | 23 | 11 | 13 | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 159 | R1b3-M269 | 14 | 10 | 17 | 24 | 10 | 13 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 160 | R1b3-M269 | 14 | 10 | 17 | 24 | 11 | 11 | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 161 | R1b3-M269 | 14 | 10 | 17 | 25 | 10 | 13 | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 162 | R1b3-M269 | 14 | 10 | 17 | 25 | 10 | 13 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 163 | R1b3-M269 | 14 | 11 | 15 | 25 | 10 | 14 | 12 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 164 | R1b3-M269 | 14 | 11 | 16 | 22 | 11 | 13 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 165 | R1b3-M269 | 14 | 11 | 16 | 24 | 11 | 13 | 13 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 166 | R1b3-M269 | 14 | 11 | 16 | 25 | 10 | 13 | 12 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 167 | R1b3-M269 | 14 | 12 | 16 | 24 | 11 | 13 | 13 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 168 | R1b3-M269 | 15 | 10 | 16 | 24 | 10 | 13 | 12 | 0 | 0 | 0 | 2 | 0 | 0 | 2 |
| 169 | R1b3-M269 | 15 | 10 | 17 | 25 | 10 | 13 | 13 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 170 | R1b3-M269 | 16 | 10 | 16 | 24 | 10 | 13 | 12 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 171 | R1b3-M269 | 16 | 10 | 17 | 25 | 10 | 11 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |

Note. - Moldavians: K=Karahasani, S=Sofia; Gagauzes: K=Kongaz, E=Etulia.

# 8 REFERENCES

Altukhov YuP, Korochkin LI, Rychkov YuG (1996) Genetic biochemical diversity in Evolution and individual development. *Russ J Genet* **32**: 1450–1473

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* **28**: 458-472

Ammerman AJ, Cavalli-Sforza LL (1984) *Neolithic transition and the genetics of populations in Europe.* Princeton University Press, Princeton

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 H-L, Puzyrev VP, Herrera RJ (2002) Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* **23**: 3346-3356

Archidiacono N, Storlazzi CT, Spalluto C, Ricco AS, Marzella R, Rocchi M. (1998) Evolution of chromosome Y in primates. *Chromosoma* **107**: 241–246

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA (1996) *Alu* fossil relics – distribution and insertion polymorphism. *Genome Research* **6**: 1084-1092

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

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

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**: 743–753

Barac L, Peričic M, Martinovic B, Klaric I, Janicijevic B, Parik J, Rootsi S, Rudan P (2003) Y-chromosome STRs in Croatians. *Forensic Sci Int* **138**: 127–133

Baskakov NA (1988) *The essays on the history of the functional development of the Turkic languages and their classification.* Ylym, Ashkhabad (in Russian)

Batzer MA, Arcot SS, Phinney JM, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, *et al.* (1996a) Genetic variation of recent Alu insertions in human populations. *J Mol Evol* **42**: 22–29

Batzer MA, Deininger PL (2002) *Alu* repeats and human genome diversity. *Nat Rev Genet* **3**: 370–379

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin C, Schmid CW, Zietkiewicz E, Zuckerkandl E (1996b) Standardized nomenclature for *Alu* repeats. *J Mol Evol* **42**: 3–6

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL (1990) Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res* **18**: 6793–6798

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human specific polymorphic Alu insertions. *Proc Natl Acad Sci USA* **91**: 12288-12292

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

Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y-chromosome. *Am J Hum Genet* **65**: 1623-1638

Bowock AM, Ruiz-Linares A, Tomforde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455–457

Brinkman B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* **62**: 1408–1415

Britten RJ (1994) Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proc Natl Acad Sci USA* **91**: 6148–6150

Buttler JM (2003) Recent Developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci Rev* **7**: 92–111

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* **22**: 231-238

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes.* Princeton University Press, Princeton

Chakraborty K (1986) Gene admixture in human populations: models and predictions. *Yearb Phys Anthropol* **29**: 1-43

Chaplkygina NA (1987) The Dniester-Carpathian lands in the 1st–3rd centuries BC. In: *History of the Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 165–175 (in Russian)

Chetraru NA (1973) *The sites of the Paleolithic and the Mesolithic periods. The Archeological Map of the MSSR.* Ştiinţa, Chişinău (in Russian)

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci USA* **95**: 9053–9058

Childe VC (1968) *The down of European civilization.* Routlege and Kegan Paul, London

Chinnioğlu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman ChC, *et al.* (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* **114**: 127–148

Cimpoieş LS (1997) *Dastan Epic Literature of the Gagauzes.* Acad Sci Moldova, Chişinău (in Russian)

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, Bertranprtit J, Scheil H-G, Huckenbeck W, Efremovska L, Schmidt HD (2004) *Alu* insertion polymorphisms in the Balkans and the origins of the Aromuns. *Ann Hum Genet* **68**: 120–127

Comşa E. (1987) *Neoliticul pe teritoriul României. Consideraţii.* Ed. Acad. RPR. Bucureşti

Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* **5**: 1759-1766

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, *et al.* (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

Dannell R (1983) *European economic prehistory: a new approach.* Academic Press, London

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y-chromosome. *Am J Hum Genet* **67**: 1055-1061

de Knijff P, Kayser M, Caglia A, Corach G, Fretwell N, Gehrig C, Graziosi G, Heidorn A, *et al.* (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Leg Med* **110**: 134

Deininger PL, Batzer MA (1993) Evolution of retroposons. *Evol Biol* **27**: 157–198

Deininger PL, Batzer MA (1999) *Alu* repeats and human disease. *Mol Genet Metab* **67**: 183–193

Dergachev VA (1986) *Moldavia and the neighboring territories in the Bronze Age.* Ştiinţa, Chişinău (in Russian)

Dergachev VA (1990) On the conception of “contact zone”. In: VN Masson (ed) *Archaeological cultures and cultural transformations.* Proceedings of the methodological seminar of the Institute of Archaeology (LB) of the Academy of Sciences of USSR. Institute of Archaeology, Leningrad’s branch, Leningrad, pp 76–82 (in Russian)

Dergachev VA (1997) *The role of metal wares in the genesis of early Hallstat cultures in the Carpathian-Danubian-Northpontic region.* Academy of Sciences of Moldova, Chişinău (in Russian)

Dergachev VA (1999) The particularities of the cultural development of the region between the Carpathians and Dniester. *Stratum plus* **2**: 169–221 (in Russian)

Dergachev VA (2000) The migration theory of Marija Gimbutas. *JIES* **28**: 257–339

Dergachev VA, Bochkarev VC (2002) *The metallic sickles of the Late Bronze Age of Eastern Europe.* High Anthropological School, Chişinău (in Russian)

Dergachev VA, Marchevic VI (1987) Neolithic and Eneolithic. The formation of the Producing Economy. In: *History of the Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 39–107 (in Russian)

Dergachev VA, Sherrat A, Larina OV (1991) Resent results of neolithic research in Moldavia (USSR). *OJA* **10**: 1-16

Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, Cresta M, Cucci F, Di Stasi L, *et al.* (2003) Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol Phylogenet Evol* **28**: 387-395

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, *et al.* (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* **115**: 357-371

Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* **18**: 672-675

Dyachenko VD (1965) *Anthropological structure of the Ukrainians.* Naukova Dumka, Kiev (in Ukrainian)

Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* **87**: 2951-2954

Ecsedi I (1979) *The people of the pit-grave kurgans in Eastern Hungary*. Fontes Archaeologici Hungariae, Budapest

Edwards MC, Gibbs RA (1992) A human dimorphism resulting from less of an Alu. *Genomics* **14**: 590-597

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

Fedorov GB (1999) The ethnogeny of the Volokhs, - the ancestors of the Moldavians (the historical aspect). *Stratum plus* **5**: 14–74 (in Russian)

Felsenstein J (1993) *PHYLIP, Version 3.5.* Seattle: University of Washington Press

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

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**: 935–945

Forster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Teylor-Smith Ch, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y-chromosome. *Am J Hum Genet* **67**: 182–196

Gibson G (2002) Microarrays in ecology and evolution: a preview. *Mol Ecol* **11**: 17-24

Gimbutas M (1970) Proto-Indo-European culture: The Kurgan culture during the fifth, fourth and third millennia BC. In: Cardona G, Hoenigswald, HM, Senn, AM (eds) *Indo-European and Indo-Europeans.* University of Pennsylvania Press, Philadelphia

Gokhman II (1966) *The population of Ukraine in the Mesolithic and Neolithic time.* Nauka, Moscow (in Russian)

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* **92**: 6723–6727

Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y-chromosomes. *Mol Biol Evol* **13**: 1213–1218

Grover D, Kannan K, Brahmachari SK, Mukerji M (2005) *Alu*-ring elements in the primate genomes. *Genetica* **124**: 273–289

Guboglo MN (1967) The ethnic affiliation of the Gagauzes. *Sov Ethnograf* **3**: 160-167 (in Russian)

Gudkova AV (1999) The 1st–7th centuries AD in the Northwest Pontic region: the culture of sedentary population. *Stratum plus* **4**: 235-405 (in Russian)

Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**: 361-372

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* **22**: 239-247

Hammer MF (1995) A recent common ancestry for human Y-chromosomes. *Nature* **378**: 376–378

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

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonne-Tamir B (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes**.** *Proc Natl Acad Sci USA* **97**: 6769-6774

Hansel B (1982) Südosteuropa zwischen 1600 und 1000 v. Chr. In: Helmut Geißlingen (ed) *Prähistorische Archäologie in Südosteuropa.* Berlin, pp 1–38

Houck CM, Rinehart FP, Schmid CW (1979) A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* **132**: 289–306

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, et al. (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

Ilynskaya VA, Terenozhkin AI (1983) *Scythia in the 7th–4th centuries BC.* Naukova Dumka, Kiev (in Russian)

Jin L, Su B (2000) Natives or immigrants: modern human origin in East Asia. *Nature Genet* **1**: 126-133

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes the human Y-chromosome, disease and selection. *Trends Genet* **16**: 356-362

Jobling MA, Tyler-Smith Ch (1995) Fathers and sons: the Y-chromosome and human evolution. *Trends Genet* **11**: 449–456

Jobling MA, Tyler-Smith Ch (2003) The human Y-chromosome: an evolutionary marker comes of age. *Nature Rev Genet* **4**: 598–612

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* **66**: 979-988

Jorde LB, Wooding SP (2004) Genetic variation, classification and 'race'. *Nat Genet* **36**: 28-33

Jurka J, Zuckerkandl E (1991) Free left arms as precursor molecules in the evolution of Alu sequences. *J Mol Evol* **33**: 49-56

Kan YM, Dozy AM (1978) Polymorphism of DNA sequence adjacent to the β-globin structural gene; its relation to the sickle mutation. *Proc Natl Acad Sci USA* **75**: 5631–5635

Kapitonov V, Jurka J (1996) The age of Alu subfamilies. *J Mol Evol* **42**: 59–65

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* **69**: 615-628

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

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, *et al.* (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* **64**: 817–831

Kass DH, Batzer MA (2001) Genome Organization/Human. *Encyclopedia of life science.* Nature Publishing Group (www.els.net)

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

Kayser M, Caglia A, Corach D, Fretwell N, Herzog B, Hidding M, Honda K, Jobling M, *et al.* (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Leg Med* **110**: 125–133

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SO, Rosser Z, *et al.* (2004) A Comprehensive Survey of Human Y–Chromosomal Microsatellites. *Am J Hum Genet* **74:** 1183–1197

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, *et al.* (2001) An extensive analysis of Y chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* **68**: 990–1018

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, *et al.* (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet* **117**: 428-443

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, *et al.* (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 (2005) *Y-chromosome lineages in Siberian populations*. Dissertation in biology. Institute of Medical Genetics, Tomsk Research Center, Russian Academy of Medical Sciences, Tomsk (in Russian)

Kharkov NV, Stepanov VA, Borinskaya SA, Kozhekbajeva JM, Gusar VA, Grechanina EYa, Puzyrev VP, Khusnutdinova EK, Yankovsky NK (2004) Gene pool structure of Eastern Ukrainians as inferred from the Y-chromosomal haplogroups. *Russ J Genet* **40**: 326–331

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

Khit' GL, Dolinova NA (1983) *Dermatoglifics of the USSR peoples*. Moscow: Nauka (in Russian)

Khitrinskaya IIu, 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

Klaric IM, Lauc LB, Pericic M, Janicijevic B, Terzic R, Colak I, Kvesic A, Rudan P (2005) Evaluation of Y-STR variation in Bosnian and Herzegovinian population. *Forensic Sci Int* **154**: 252-256

Kolosovskaya JK (1987) The Roman conquest of the North-Danubian lands and their Romanization. In: *History of the Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 154–165 (in Russian)

Kruts SI (1972) *The population of Ukraine in the Bronze Age and in the Eneolithic period.* Naukova dumka, Kiev (in Russian)

Kruts SI, Buzhilova AP, Varsahr AM (2003) The anthropological materials from the Neolithic cemetery Sakarovka I. *Rossiyskaya Arkheologya* **2**: 119–127 (in Russian)

Lahn BT, Page DC (1997) Functional coherence of the human Y-chromosome. *Science* **278**: 675–680

Lahn BT, Page DC. (1999) Four evolutionary strata on the human X chromosome. *Science* **287**: 877–879

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC (2001) Initial sequencing and analysis of human genome. *Nature* **409**: 860–921

Larina OV (1999) The Linear pottery culture of the area between Pruth and Dniester. *Stratum plus* **2**: 10-140 (in Russian)

Lell JT, Wallace DC (2000) The peopling of Europe from the maternal and paternal perspectives. *Am J Hum Genet* **67**: 1376–1381

Li CC (1976) *First course in population genetics.* Boxwood press, Pacific Grove, California

Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, Garcia O, Giparaki M, *et al.* (2001) A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. *Ann Hum.Genet* **65**: 339-349

Manni FC, Heyer GE (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by “Monmonier’s algorithm”. *Hum Biol* **76**: 173–190

Manzura IV (2000) The lords of the scepters. *Stratum plus* **2**: 237-295

Marchevic VI (1973) *The sites of the Neolithic and the Eneolithic periods. The Archeological Map of the MSSR.* Ştiinţa, Chişinău (in Russian)

Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, *et al.* (2005) The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *An Hum Genet* **69**: 1–7

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

Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. *Science* **254**: 1808–1810

Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M (1995) Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* **57**: 403–414

Melukova AI, Niculiță IT (1987) The Beginning of the Iron Age. The population of the Dniester-Carpathian lands in the 1st millennium BC. In: *History of Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 110–154 (in Russian)

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

Mongait AL (1973) *Archaeology of Western Europe: Stone Age.* Nauka, Moscow (in Russian)

Mongait AL (1974) *Archaeology of Western Europe: Bronze and Iron Age.* Nauka, Moscow (in Russian)

Mountain JL, Knight A, Jobin M, Gignoux C, Miller A, Lin AA, Underhill PA (2002) SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res* **12**: 1766-1772

Mourant AE, Kopec AC, Domanievska-Sobczak K (1976) *The distribution of the human blood groups and other polymorphisms.* Oxford University Press, London

Murdock GP (1967) *Ethnographic atlas.* University of Pittsburgh Press, Pittsburgh

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

Nasidze I, Sarkisian T, Kerimov A, Stoneking M (2003) Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet* **112**: 255–261

Necrasov O (1980) Physical anthropological characteristics of skeletons from the Kurgan graves in Romania. *JIES* **8**: 337-343

Necrasov O, Cristescu M (1963) Contribution a l'etude anthropologiqe des sqelettes neolithiqes la culture Boian. *ASUI*, T9

Nei M (1973) The theory and estimation of genetic distance. In: Morton NE (ed) *Genetic structure of populations.* University of Hawaii Press, Honolulu, pp 45–54

Nei M (1987) *Molecular evolutionary genetics.* Columbia University Press, New York, NY

Nei M, Roychoudhury AK (1988) *Human polymorphic genes: world distribution.* Oxford University Press, New York

Nicolova L. (2000) The pit-grave culture in the Balkans (Dynamics of the structure of the burial rites and its relation to the other early Bronze cultures). *Stratum plus* **2**: 423-458 (in Russian)

Nikitina TV, Nazarenko SA (2004) Human Microsatellites. Mutation and Evolution. *Russ J Hum Genet* **40**: 1301–1318

Novick GE, Batzer MA, Deininger PL, Herrera RJ (1996) The mobile genetic element *Alu* in the human genome. *BioScience* **46**: 32–41

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. *Computer Application in the Biosciences* **12**: 357-358

Paraska PF, Sovetov PV (1987) The formation of the Moldavian early feudal State and the Establishment of its territory. In: *History of Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 318–329 (in Russian)

Pericic M, Klaric IM, Lauc LB, Janicijevic B, Dordevic 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

Pericic M, Lauc LB, Klaric IM, Rootsi S, Janicijevic B, Rudan I, Terzic R, Colak I, *et al.* (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.

Ploski R, Wozniak M, Pawlowski R, Monies DM, Branicki W, Kupiec T, Kloosterman A, Dobosz T, *et al.* (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y-chromosome microsatellite haplotype analysis. *Hum Genet* **110**: 592–600

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

Polevoy LL (1987) The political situation of the Dniester-Carpathian lands in the second half of the 13th – the first half of the 14th centuries. In: *History of Moldavian SSR*. Cartea Moldovenească, Chişinău, pp 316–318 (in Russian)

Potehina ID (1999) *The population of Ukraine in the Neolithic and the Elarly Eneolithic periods on the basis of anthropological data.* Naukova Dumka, Kiev (in Russian)

Przeworski M, Hudson RR, Di Rienzo A (2000) Adjusting the focus on human variation. *Trends Genet* **16**: 296-302

Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002) Y chromosomal DNA variation in Pakistan. *Am J Hum Genet* **70**: 1107–1124

Quentin Y (1992) Origin of the *Alu* family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res* **20**: 3397–3401

Quintana-Murci L, Fellous M (2001) The human Y-chromosome: the biological role of a "functional wasteland". *J Biomed Biotechnol* **1**: 18–24

Radova OK (1997) The Gagauzes of Bessarabia. The settling and quantity in the 19th century. *Etnograf Obozrenie* **1**: 121-128 (in Russian)

Rassamakin JuJa (1994) The main directions of the development of early pastoral societies of northern Pontic zone: 4500-2450. *Baltic-Pontic Studies* **2**: 29-70

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

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **61**: 247-254

Rickman EA (1987) The Chernyakhov culture tribes in the Dniester-Carpathian lands. In: *History of Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 181–201 (in Russian)

Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* **116**: 279-291

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**: 602-12

Rootsi S, Kivisild T, Tambets K, Adojaan M, Parik J, Reidla M, Metspalu E, Laos S, Tolk HV, Villems R (2000) On the phylogeographic context of sex-specific genetic markers of Finno-Ugric populations. In: Künnap A (ed) *The roots of peoples and languages of Northern Eurasia II and III.* University of Tartu, Division of Uralic Languages/Societas Historiae Fenno-Ugricae, Tartu, pp 148–164

Rootsi S, Magri Ch, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, *et al.* (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* **75**: 128–137

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros, *et al.* (2000) Y chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**: 1526–1543

Rowold D, Herrera RJ (2000) *Alu* elements and the human genome. *Genetica 108*: 57–72

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem A.-H., Batzer MA, Deininger PL (2001) *Alu* insertion polymorphisms for the Study of human genomic diversity. *Genetics* 159: 279–290

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y-chromosomes. *Nature* **423**: 873–876

Russev DN (1999) Moldavia in the "Dark Age": trying to understand cultural and historical processes. *Stratum plus* **5**: 379–407 (in Russian)

Rychkov YuG (ed), Zhukova OV, Sheremetyeva VA, Shneider YuG, Balanovskaya EV, Spitsyn VA, Udina IG, *et al.* (2000) *Gene pool and gene geography of populations. Vol. 1: Gene pool of populations of Russia and contiguous countries.* Nauka, St. Petersburg (in Russian)

Saitou N, Nei M (1987) The Neighbor–Joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425

Schmid C, Maraia K (1992) transcriptional regulation and transpositional selection of active SINE sequences. *Curr Opin Genet Dev* **2**: 874–882

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

Sedov VV (2002) *Slavs. Historical and archaeological surveys.* Yazyki slavyanskoi kul'tury, Moscow (in Russian)

Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS (2003) A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* **73**: 700-705

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* **20**: 278–280

Semino O, Magri Ch, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, et al. (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

Semino O, Passarino G, Oefner P, Lin AA, Arbuzova S, Beckman L, Benedictis G, Francalacci P, *et al.* (2000) The Genetic legacy of paleolithic Homo Sapiens in extant Europeans: a Y-chromosome perspective. *Science* **290**: 1155–1159

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, et al. (2006) Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* **78**: 202-221

Sharov OV, Bažan IA (1999) The Chernyakhov Etudes. *Stratum plus* **4**: 19-65 (in Russian)

Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). *J Mol Evol* **33**: 311–320

Shschukin (2005) *The Gothic way. Goths, Rome, and the culture of the Chernjakhov/Sîntana de Mureş.* St Petersburg State University, St. Petersburg (in Russian)

Sitnikova T, Rzhetsky A, Nei M (1995) Interior-branch and bootstrap tests of phylogenetic trees. *Mol Biol Evol* **12**: 319-333

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier LD, Braun LG, Repping S, *et al.* (2003) The male-specific region of the human Y-chromosome is a mosaic of discrete sequence classes. *Nature* **429**: 825–837

Smouse PE, Long JC (1992) Matrix correlation analysis in anthropology and genetics. *Yearb Phys Anthrop* **35**: 187–213

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

Stefan M, Stefanescu G, Gavrila L, Terrenato L, Jobling MA, Malaspina P, Novelletto A (2001) Y-chromosome analyses reveals a sharp genetic boundary in the Carpathian region. *Eur J Hum Genet* **9**: 27–33

Stepanov VA (2002) *Ethnogenomics of the population of North Eurasia*. Pechatnaya Manufaktura, Tomsk (in Russian)

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

Tajima A (1993) Measurement of DNA polymorphism. In: Takahata N, Clark AG (eds) *Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology*. Tokyo, Sunderland, MA: Japan Scientific Societies Press, Sinauer Associates, Inc., pp 37-59

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, Tolk HV, Reidla M, *et al.* (2004) The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y-chromosomes. *Am J Hum Genet* **74**: 661-682

Tishkoff SA, Kidd KK (2004) Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* **36**: 21-27

Todorova Kh. (1986) *The Stone-Cooper Age in Bulgaria*. Nauka i Izkustvo, Sofia (in Bulgarian)

Tsaranov VI, Afteniuk SYa, Dragnev DM, Platon MS, Repida LE, Sovetov PV, Shemeakov DE, *et al.* (1982) *History of the Moldavian SSR.* Ştiinţa, Chişinău (In Russian)

Ullu E, Tschudi C (1984) *Alu* sequences are processed 7SL RNA genes. *Nature* **312**: 171–172

Ullu E, Weiner AM (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature* **18**: 371–374

Underhill PA (2003) Inferring human history: clues from Y-chromosome haplotypes. *Cold Spring Harbor Symposia on Quantitative Biology* **LXVIII**: 487–493

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. *Anthrop 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. *Anthrop Anz* **59**: 11-17

Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, *et al.* (1997) Mutations and sequence variants in the testis-determining region of the Y-

chromosome in individuals with a 46,XY female phenotype. *Hum Genet* **99**: 648–652

Velikanova MS (1975) *Paleoanthropology of the Dniester-Pruth interfluvial.* Nauka, Moscow (in Russian)

Wallace DC (1995) William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet* **57**: 201-223

Walter H (1997) *Populationsgenetik der Blutgruppensysteme des Menschen.* E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* **2**: 1123–1128

Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* **55**: 631-661

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, *et al.* (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* **98**: 10244–10249

Whittle A (1996) Die Ersten Bauern. In: *Illustrierte Vor- und Frühgeschichte Europas* (ed. Barry Cunliffe). Frankfurt/Main: Campus Verlag, pp. 155-190

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

Y-chromosome Consortium (2002) A nomenclature system for the tree of human Y chromosomal binary haplogroups. *Genome Res* **12**: 339–348

Y-STR haplotype reference database, http://www.ystr.org/europe

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

Zelenchuk VS (1987) The settling of the Volochs. In: *History of Moldavian SSR.* Cartea Moldovenească, Chişinău, pp 268–272 (in Russian)

Zerjal T, Dashnymaan B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al. (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

Zhivotovsky LA (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol Biol Evol* **18**: 700–709

Zhivotovsky LA, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* **72**: 1171–1186

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, *et al.* (2004) The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* **74**: 50-61

# Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine andere als die angegebenen Quellen und Hilfsmittel benutzt habe.

Alexander Varzari

# CURRICULUM VITAE

## Personal data

Family name: Varzari
First name: Alexander (Sasha)
Date of birth: 7 December 1976
Place of birth: Kishinev, Moldavia, USSR
Marital status: single

## School education

9/1984 – 6/1994
Secondary School No. 38, Kishinev, Moldavia

## Academic education and experience

9/1994 – 7/1999
Study at the Lomonosov Moscow State University, Faculty of Biology, Department of Anthropology

9/1998 – 7/1999
Diploma degree research “Genetic demographic survey of the Gagauz population of Moldova”. Graduated in July, 1999 cum laude

9/1999 – 8/2002
Work in the Institute of Archaeology and Ethnography of Academy of Sciences of Moldova Republic. Research in the field of paleoanthropology

8/2002 – 8/2003
DAAD (Deutscher Akademischer Austauschdienst) scholarship for the University of Ulm, Institute of Human Genetics and Anthropology. Research in the field of population genetics

10/2003 – 9/2004

Doctorate research work at the Department of Anthropology and Human Genetics at the Ludwig-Maximilians University, Munich

9/2004 – 8/2005

Work in the Institute of Medical Genetics, Tomsk Research Center, Russian Academy of Medical Sciences. Research in the field of population genetics

11/2005 -9/2006

Continuation of the PhD thesis at the Department of Anthropology and Human Genetics at the Ludwig-Maximilians University, Munich.

**Affiliation**

Member of the European Anthropological Association

**Publications**

**Varsahr AM**, Spitsyn VA, Bychcovscaya LS, Kravchuk OI (2001) To the research of the gene pool of the Gagauz population of Moldavia. *Anthropol Anz* **59**:11-17

Tarskaia LA, El'chinova GI, **Varzar' AM**, Shabrova EV (2002) Genetic and demographic structure of the Yakut population: reproduction indices. *Russ J Genet* **38**: 985-991

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

**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

Kruts SI, Buzhilova AP, **Varzar' AM** (2003) Anthropological materials from the Neolithic cemetery Sakarovka I. *Rossiyskaya Arkheologya* **2**: 119–127 (in Russian)

**Varzar' AM**, Pezhemsky DV (2003) Paleoanthropological materil from the late Tripolie complex Tătărăuca Nouă. In: Sava EN, Kashuba MT (eds.) *Interferențe cultural-cronologice în spațiul nord-pontice.* Chişinău, pp. 81-88 (in Russian)

**Varzar' AM**, Pezhemsky DV, Larina OV (2005) Paleoanthropological materil from the late Chernoles complex Tătărăuca Noua XV. *Revista Arheologică. Seria Nouă* **1**: 381-396 (in Russian)

**Varsahr AM**, Scheil HG, Schmidt HD (2006) Blood group and serum protein polymorphisms in a population group of Moldavians. *Anthropol Anz* **64**: 51-58

**Varsahr AM**, Stepanov VA, Schmidt HD, Raicu F, Cojocaru RI, Roschin YuN, Glavce Ch, Dergachev VA, Weiss EH. Gene pool structure of the Dniester-Carpathian population as inferred from insertion/deletion loci [manuscript in preparation]

**Varsahr AM**, Kharkov VN, Weiss EH, Dergachev VA, Raicu F, Cojocaru RI, Roschin YuN, Puzyrev VP, Stepanov VA. Contact zone in the Eastern Trans-Carpathians revealed by Y-chromosome analysis [manuscript in preparation]