

The role of Southeastern Europe in origins and diffusion of major paternal lineages

Marijana Peričić, Lovorka Barać Lauc, Irena Martinović Klarić, Petra Rajić Šikanjić,
Branka Jančićević, Pavao Rudan

Institute for Anthropological Research, Zagreb, Croatia

mpericic@inantro.hr, lovorka@inantro.hr, irena@inantro.hr, petra@inantro.hr

branka@inantro.hr, pavao.rudan@inantro.hr

ABSTRACT – *The aim of this study is to explore the existing data based on high-resolution phylogenetic studies of Y chromosome variation in populations from Southeastern Europe and elsewhere in Eurasia in order to evaluate the role of the region in the process of the prehistoric colonization of the European continent and the structuring of the modern paternal genetic pool. Even though the distribution and estimated range expansions of major paternal lineages in Southeastern Europe are consistent with the typical European Y chromosome gene pool, the specific role of this region in the process of structuring the European paternal genetic landscape is evident in prehistoric episodes of significant gene flow that diffused from or into the region.*

IZVLEČEK – *Cilj te studije je preučiti postojeće podatke, ki so osnovani na visoko resolucijskih filogenetskih študijah variacije kromosoma Y v populacijah jugovzhodne Evrope in v drugih delih Evrazije, z namenom, da ocenimo vlogo regije v procesu prazgodovinske kolonizacije evropskega kontinenta ter pri strukturiranju moderne moškega genskega sklada. Čeprav sta distribucija in predviden obseg širitve glavnih moških dednih linij v jugovzhodni Evropi skladna s tipičnim Evropskim genskim skladom kromosoma Y, se posebna vloga te regije kaže v procesu strukturiranja evropskega moškega dednega genetskega zemljevida, povezanega z dogodki v prazgodovini in genskim pretokom v regijo in iz regije.*

KEY WORDS – *Human Y chromosome; phylogenetic analysis; Southeastern Europe*

Introduction

The human Y chromosome defines male sex through the action of the sex determining region (SRY). It is an atypical segment of the human genome, since it is haploid in most of its length, escapes recombination with the X chromosome, and undergoes uniparental transmission. These properties make the Y chromosome sequence a valuable tool for the purposes of human history reconstruction and studies focused on the dispersal of anatomically modern humans. The non-recombining region of the Y (NRY) is inherited as a single locus that changes exclusively via mutations accumulating over time, thus allowing the preservation of a relatively simple record of genetic history in comparison to nuclear DNA (autosomes).

There has been an interest in studying paternal genetic history since the mid-80s (e.g. Casanova *et al.* 1985; Hammer 1994; Underhill *et al.* 2000). A constantly growing number of evolutionary informative polymorphisms provide a deeper resolution of human paternal history and evolution. Currently, there are more than 300 known SNPs (single nucleotide polymorphisms) and small indels (YCC 2002; Jobling and Tyler-Smith 2003). In evolutionary genetics terminology, the set of alleles at different biallelic loci along the chromosome is called a haplogroup.

Assuming a 1:1 sex ratio, the effective population size of the Y chromosome in a population would be

about one-quarter of that of any autosome. Consequently, a genetic difference depicted by Y chromosomes, in comparison to autosomes, is more susceptible to the effects of random genetic drift that accelerates geographic clustering and differentiation between different (especially small) populations. The general structure of paternal genealogies is compatible and indicative of the common origin of all non-African contemporary populations from a small subset of Africans. Despite disagreement about the time to the most recent common ancestor (TMRCA) of the Y chromosome, its phylogeny roots in Africa around 100 KYA (e.g. Hammer et al. 1998; Underhill et al. 2001; Underhill 2003).

The first extensive studies of European Y chromosome dispersal by Semino et al. (2000) and Rosser et al. (2000) showed clinal patterns for the most frequent European haplogroups. Moreover, Semino et al. (2000) grouped more than 95% of European Y chromosomes into 10 phylogenetically distinct haplogroups, of which 70–80% of the Y chromosome gene pool was represented by R1a, R1b, I and N3, and the remaining 20% by J2, E3b, and G.

Palaeolithic haplogroups

Haplogroup I is the only autochthonous European haplogroup assumed to have arisen in an Epi-Gravettian group among the descendants of people who arrived in Europe from the Near East around 25 KYA

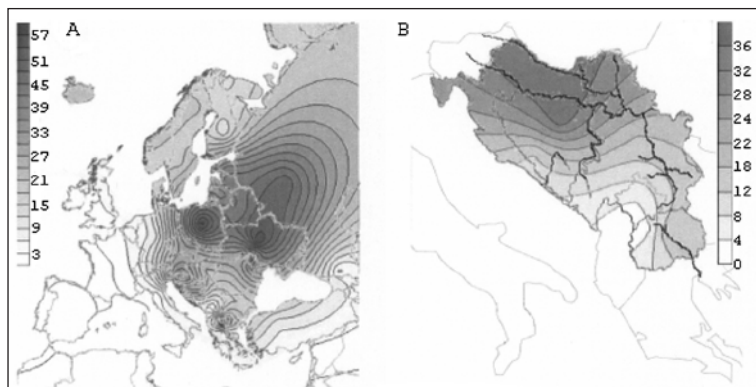


Fig. 2. R1a frequency distribution in Europe, Northern Africa and Asia Minor (panel a) as well as in Southeastern Europe (panel b). Frequency distributions surfaces are taken from Peričić et al. (2005). R1a frequency data for different Eurasian populations were generated from literature, as listed in Table 1 in Peričić et al. (2005).

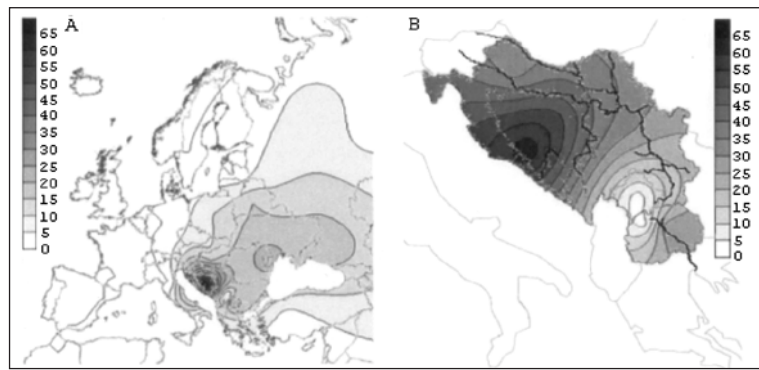


Fig. 1. I1b* frequency distribution in Europe, Northern Africa and Asia Minor (panel a) as well as in Southeastern Europe (panel b). Frequency distributions surfaces are taken from Peričić et al. (2005). I1b* frequency data for different Eurasian populations were generated from literature, as listed in Table 1 in Peričić et al. (2005).

(Semino et al. 2000). This haplogroup is almost entirely restricted to the European continent, where it shows frequency peaks in two areas – Scandinavia and Southeastern Europe (Semino et al. 2000). Further phylogenetic subdivision revealed subclades I1a, I1b*, I1b2, and I1c (Rootsi et al. 2004). The geographical distribution of I1a (the highest frequencies in Northern Europe among Norwegians, Swedes and Saami) is considered to be a result of the recolonization of Europe after the LGM from the Franco-cantabrian refugial area (Rootsi et al. 2004). The origin of the less frequent I1c, that covers a wide range of Europe and peaks in northwest coastal Europe, is in concordance with I1a (Rootsi et al. 2004). A completely different distribution pattern is observed in I1b* Y chromosomes, the most frequent haplogroup I clade in Eastern Europe and on the Balkan Peninsula. I1b* reaches maximum frequencies in Southeastern Europe in Bosnia and Herzegovina (Fig. 1).

Our results indicate that the homogenous distribution of elevated I1b* frequency among different populations in Southeastern Europe could support the hypothesis of their having a common paternal history shared over a long period of time (Peričić et al. 2005). Rootsi et al. (2004) estimated that I1b* diverged from I* at 10.7+4.8 KYA, possibly in relation to the post Younger Dryas (YD) climate amelioration in Europe, and that I1b* expansion occurred around the early Holocene at 7.6+2.7 KYA. Our coalescent estimate of I1b* (Peričić et al. 2005) is substantially older (11.1+4.8 KYA). This finding suggests that the I1b* lineages might

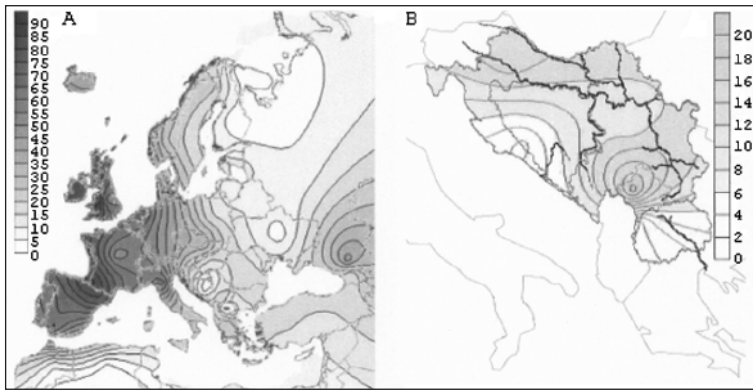


Fig. 3. R1b frequency distribution in Europe, Northern Africa and Asia Minor (panel a) as well as in Southeastern Europe (panel b). Frequency distributions surfaces are taken from Peričić et al. (2005). R1b frequency data for different Eurasian populations were generated from literature, as listed in Table 1 in Peričić et al. (2005).

have expanded from Southeastern to Central, Eastern and Southern Europe in a period not earlier than the YD to Holocene transition and not later than the early Neolithic (Peričić et al. 2005). Although not yet supported by archaeological evidence, the I1b* spread in Europe suggests that Southeastern Europe could have served as an LGM refugium, as previously suggested by Semino et al. (2000) and Barać et al. (2003). This scenario could be indirectly supported by the recolonization of Northern Europe from the direction of Southeastern Europe by at least two species – the brown bear *Ursus arctos* (Taberlet and Bouvent 1994) and the European hedgehog *Erinaceus europaeus* (Hewitt 2000).

Another widespread haplogroup in Europe, R1a, is characteristic of Eastern European populations (Fig. 2a). The age of this haplogroup has been approximated to 15 KYA (Semino et al. 2000; Wells et al. 2001). Kivisild et al. (2003) suggested that Southern and Western Asia might be the source of R1 and R1a differentiation. Present R1a distribution in Europe shows an increasing west-east frequency gradient, with the highest frequencies among Finno-Ugric and Slavic speakers (Fig. 2a). R1a frequency shows a decrease in the north-south direction in Southeastern Europe (Fig. 2b), where its age is estimated at 15.8 ± 2.1 KYA (Peričić et al. 2005). This estimate is consistent with the R1a deep Palaeolithic time depth previously suggested by Semino et al. (2000) and Wells et al. (2001). At

the current level of resolution it is not possible to determine which of three potential episodes of gene flow might have influenced the estimated age in Southeastern Europe: early post-LGM recolonizations from the direction of the Ukrainian refugium, migrations from the northern Pontic steppe in the period between 3000 to 1000 BC, or Slavic migrations between the 5th and 7th centuries AD.

Its sister clade, haplogroup R1b, was introduced by or arose in an Aurignacian group who entered Europe and diffused from east to west about 40 to 35 KYA (Semino et al. 2000).

R1b shows a frequency peak in Western Europe and a decrease in Eastern and Southern Europe (Fig. 3a). Even though R1b frequency decline continues from Western to Southeastern and Southern Europe, two intermediate local peaks are evident in Southeastern Europe (Fig. 3b). According to our data, in Southeastern Europe the coalescent estimate of R1b (11.6 ± 1.4 KYA) closely matches the estimate for the I1b* lineages, pointing to the Younger Dryas to Holocene transition as a possible expansion period of these two major Y chromosome lineages (Peričić et al. 2005).

Neolithic haplogroups

Approximately 20% of European Y chromosomes belong to haplogroups E3b, J2 and G that, due to their decreasing frequency gradients from the Near East to Europe, have been traditionally considered to re-

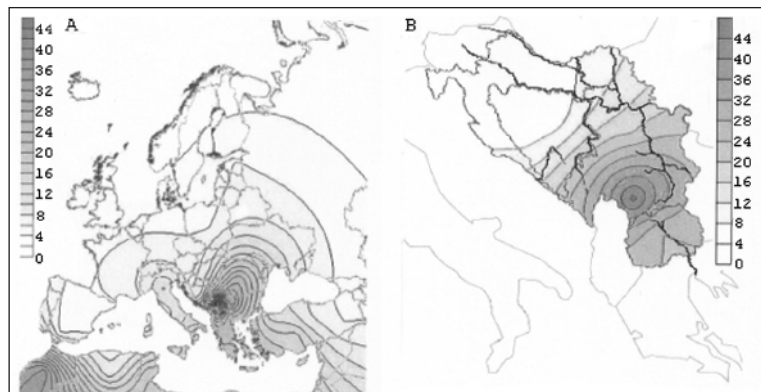


Fig. 4. E3b1 frequency distribution in Europe, Northern Africa and Asia Minor (panel a) as well as in Southeastern Europe (panel b). Frequency distributions surfaces are taken from Peričić et al. (2005). E3b1 frequency data for different Eurasian populations were generated from literature, as listed in Table 1 in Peričić et al. (2005).

present the male contribution of a demic diffusion of farmers (e.g. *Semino et al. 2000; Semino et al. 2004; Cruciani et al. 2004*). E3b1 shows a frequency peak in Southern and Southeastern Europe (Fig. 4a). In fact, E3b1 shows a rather continuous frequency decline in Southeastern Europe (Fig. 4a). Populations of the Adriatic-Dinaric complex are distinguished from neighboring populations of the Vardar-Morava-Danube river system by a lower frequency of E3b1 (Fig. 4b), possibly due to its different dispersal modes in two proximate geographic regions. Moreover, the Vardar-Morava-Danube river system could have been one of major routes for E3b1 expansion from South and Southeastern to continental Europe, as evidenced in the archeological record (e.g. *Tringham 2000*). The estimated age of this haplogroup of 7.3 ± 2.8 KYA in Southeastern Europe accords with the time of expansion of the Neolithic in Europe (*Cruciani et al. 2004; Semino et al. 2004*).

Haplogroup J is subdivided into two major clades, J1-M267 and J2-M172 (*Cinnioglu et al. 2004*). J2-M172 is more frequent in Europe (*Semino et al. 2004*). In Southeastern Europe the most frequent is haplogroup J2e, which comprises 5% of all chromosomes (*Peričić et al. 2005*), while haplogroup J2, the main J2 cluster among Greeks and Italians (*Di Giacomo et al. 2004*), is present at a frequency of less than 1%. The estimated age of the haplogroup J2e in Southeastern Europe ($2.8+1.6$ KYA), together with its spatial distribution (two frequency peaks positioned in the Balkans and central Italy, Figs. 5a and 5b), may be explained by the maritime spread of J2e lineages from the southern Balkans towards the Apennines later than is traditionally suggested by the demic expansion model (*Peričić et al. 2005*).

Concluding remarks

Even though the distribution and estimated range expansions of major paternal lineages in Southeastern Europe are consistent with the typical European Y chromosome gene pool, the specific role of this region in the process of structuring the European paternal genetic landscape is evident in the following prehistoric episodes of significant gene flow: the post-LGM R1a expansion from Eastern to Western Europe, the YD-Holocene I1b* diffusion out of the

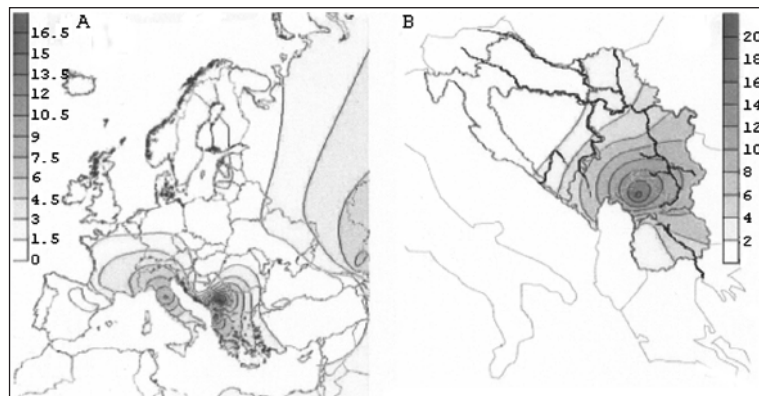


Fig. 5. J2e frequency distribution in Europe, Northern Africa and Asia Minor (panel a) as well as in Southeastern Europe (panel b). Frequency distributions surfaces are taken from Peričić et al. (2005). J2e frequency data for different Eurasian populations were generated from literature, as listed in Table 1 in Peričić et al. (2005).

Balkans, with subsequent R1a and I1b* gene flows between Eastern and Southeastern Europe, and the weaker extent of E3b1 dispersal out of Southern and Southeastern Europe towards Eastern Europe than towards Western (especially Mediterranean) Europe.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Science, Education and Sports of the Republic of Croatia grant for project 0196005 to P.R.

REFERENCES

- BARAĆ L., PERIČIĆ M., MARTINOVIĆ KLARIĆ I., ROOTSI S., JANIĆIJEVIĆ B., KIVISILD T., PARIK J., RUDAN I., VILLEMS R., RUDAN P. 2003. Y chromosomal heritage of Croatian population and its island isolates. *European Journal Human Genetics* 11: 535–542.
- CASANOVA M., LEROY P., BOUCEKKINE C., WEISSENBACH J., BISHOP C., FELLOUS M., PURRELLO M., FIORI G., SINISCALCO M. 1985. A human linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230: 1403–1406.
- CINNIOĞLU C., KING R., KIVISILD T., KALFOGLU E., ATASOY S., CAVALLERI G. L., LILLIE A. S., ROSEMAN C. C., LIN A. A., PRINCE K., OEFNER P. J., SHEN P. D., SEMINO O., CAVALLI-SFORZA L. L., UNDERHILL P. A. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Human Genetics* 114: 127–148.
- CRUCIANI F., LA FRATTA R., SANTOLAMAZZA P., SELLITTO D., PASCONI R., MORAL P., WATSON E., GUIDA V., COLOMB E. B., ZAHAROVA B., LAVINHA J., VONA G., AMAN R., CALI F., AKAR N., RICHARDS M., TORRONI A., NOVELLETTO A., SCOZZARI R. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *American Journal of Human Genetics* 74: 1014–1022.
- DI GIACOMO F., LUCA F., POPA L. O., AKAR N., ANAGNOU N., BANYKO J., BRDICKA R., BARBUJANI G., PAPOLA F., CIAVARELLA G., CUCCI F., DI STASI L., GAVRILA L., KERIMOVA M. G., KOVATCHEV D., KOZLOV A. I., LOUTRADIS A., MANDARINO V., MAMMI' C., MICHALODIMITRAKIS E. N., PAOLI G., PAPPAS K. I., PEDICINI G., TERRENATO L., TOFANELLI S., MALASPINA P., NOVELLETTO A. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Human Genetics* 115: 357–371.
- HAMMER M. F. 1994. A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Molecular Biology and Evolution* 11: 749–761.
- HAMMER M. F., KARAFET T., RASANAYAGAM A., WOOD E. T., ALTHEIDE T. K., JENKINS T., GRIFFITHS R. C., TEMPLETON A. R., ZEGURA S. L. 1998. Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. *Molecular Biology and Evolution* 15: 427–441.
- HEWITT G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- JOBLING M. A., TYLER-SMITH C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature* 4: 598–612.
- KIVISILD T., ROOTSI S., METSPALU M., MASTANA S., KALDMA K., PARIK J., METSPALU E., ADOJAAN M., TOLK H. V., STEPANOV V., GOLGE M., USANGA E., PAPIHA S. S., CINNIOĞLU C., KING R., CAVALLI-SFORZA L., UNDERHILL P. A., VILLEMS R. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics* 72: 313–332.
- PERIČIĆ M., BARAĆ LAUC L., MARTINOVIĆ KLARIĆ I., ROOTSI S., JANIĆIJEVIĆ B., RUDAN I., TERZIĆ R., ČOLAK I., KVESIĆ A., POPOVIĆ D., ŠIJAČKI A., BEHLULI I., ĐORĐEVIĆ D., EFREMOVSKA LJ., BAJEC Đ. D., STEFANOVIĆ B. D., VILLEMS R., RUDAN P. 2005. High-resolution phylogenetic analysis of Southeastern Europe (SEE) traces major episodes of paternal gene flow among Slavic populations. *Molecular Biology and Evolution* 22(10): 1964–1975.
- ROOTSI S., MAGRI C., KIVISILD T., BENUZZI G., HELP H., BERMISHEVA M., KUTUEV I., BARAĆ L., PERIČIĆ M., BALANOVSKY O., PSHENICHNOV A., DION D., GROBEI M., ZHIVOTOVSKY L. A., BATTAGLIA V., ACHILLI A., AL-ZAHERY N., PARIK J., KING R., CINNIOĞLU C., KHUSNUTDINOVA E., RUDAN P., BALANOVSKA E., SCHEFFRAHN W., SIMONESCU M., BREHM A., GONCALVES R., ROSA A., MOISAN J.-P., CHAVENTRE A., FERAK V., FUREDI S., OEFNER P. J., SHEN P., BECKMAN L., MIKEREZI I., TERZIĆ R., PRIMORAC D., CAMBON-THOMSEN A., KRUMINA A., TORRONI A., UNDERHILL P. A., SANTACHIARA-BENERECETTI A. S., VILLEMS R., SEMINO O. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics* 75: 128–137.
- ROSSER Z. H., ZERJAL T., HURLES M. E., ADOJAAN M., ALAVANTIC D., AMORIM A., AMOS W., ARMENTEROS M., ARROYO E., BARBUJANI G., BECKMAN G., BECKMAN L., BERTRANPETIT J., BOSCH E., BRADLEY D. G., BREDE G., COOPER G., CORTE-REAL H. B. S. M., DE KNIJFF P., DECORTE R., DUBROVA Y. E., EVGRAFOV O., GLISSEN A., GLISIC S., GOLGE M., HILL E. W., JEZIOROWSKA A., KALAYDJIEVA L., KAYSER M., KIVISILD T., KRAVCHENKO S. A., KRUMINA A., KUCINSKAS V., LAVINHA J., LIVSHITS L. A., MALASPINA P., MARIA S., MCELREAVEY K., MEITINGER T. A., MIKELSAAR A. V., MITCHELL R. J., NAFA K., NICHOLSON J., NØRBY S., PANDYA A., PARIK J., PATSALIS P. C., PEREIRA L., PETERLIN B., PIELBERG G., PRATA M. J., PREVIDERE C., ROEWER L., ROOTSI S., RUBINSZTEIN D. C., SAILLARD J., SANTOS F. R., STEFANESCU G., SYKES B. C., TOLUN A., VILLEMS R., TYLER-SMITH C., JOBLING M. A. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American Journal of Human Genetics* 67: 1526–1543.
- SEMINO O., PASSARINO G., OEFNER P. J., ARBUZOVA S., BECKMAN L. E., DE BENEDETTIS G., FRANCALACCI P., KOUVATSI A., LIMBORSKA S., MARCIKIĆ M., MIKA A., MIKA

B., PRIMORAC D., SANTACHIARA-BENERECETTI A. S., CAVALLI-SFORZA L. L., UNDERHILL P. A. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290: 1155–1159.

SEMINO O., MAGRI C., BENUZZI G., LIN A. A., AL-ZAHERY N., BATTAGLIA V., MACCIONI L., TRIANTAPHYLIDIS C., SHEN P. D., OEFNER P. J., ZHIVOTOVSKY L. A., KING R., TORRONI A., CAVALLI-SFORZA L. L., UNDERHILL P. A., SANTACHIARA-BENERECETTI A. S. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *American Journal of Human Genetics* 74: 1023–1034.

TABERLET P., BOUVENT J. 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of brown bear (*Ursus arctos*) in Europe. *Proceedings of the Royal Society of London Series B* 255: 195–200.

TRINGHAM R. 2000. Southeastern Europe in the transition to agriculture in Europe: bridge, buffer or mosaic. In T. D. Price (ed.), *Europe's first farmers: 19–56*. Cambridge University Press, Cambridge.

UNDERHILL P. A. 2003. Inferring Human History: Clues from Y-Chromosome Haplotypes. Cold Spring Harbor Symposium on Quantitative Biology. Vol. LXVIII. *Cold Spring Harbor Laboratory Press*: 487–493.

UNDERHILL P. A., SHEN P., LIN A. A., JIN L., PASSARINO G., YANG W. H., KAUFFMAN E., BONNE-TAMIR B., BERTRAN-PETT J., FRANCALACCI P., IBRAHIM M., JENKINS T., KIDD J. R., MEHDI S. Q., SEIELSTAD M. T., WELLS R. S., PIAZZA A., DAVIS R. W., FELDMAN M. W., CAVALLI-SFORZA L. L., OEFNER P. J. 2000. Y chromosome sequence variation and the history of human populations. *Nature Genetics* 26: 358–361.

UNDERHILL P. A., PASSARINO G., LIN A. A., SHEN P., MIRAZON LAHR M., FOLEY R., OEFNER P. J., CAVALLI-SFORZA L. L. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *American Journal of Human Genetics* 65: 43–62.

WELLS R. S., YULDASHEVA N., RUZIBAKIEV R., UNDERHILL P. A., EVSEEVA I., BLUE-SMITH J., JIN L., SU B., PITCHAPPAN R., SHANMUGALAKSHMI S., BALAKRISHNAN K., READ M., PEARSON N. M., ZERJAL T., WEBSTER M. T., ZHOLOSHVILI I., JAMARJASHVILI E., GAMBAROV S., NIKBIN B., DOSTIEV A., AKNAZAROV O., ZALLOUA P., TSOY I., KITAEV M., MIRRAKHIMOV M., CHARIEV A., BODMER W. F. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of National Academy of Science USA* 98: 10244–10249.

Y CHROMOSOME CONSORTIUM. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12: 339–348.