

Implications of the role of Southeastern Europe in the origins and diffusion of major Eurasian paternal lineages

Siiri Roots

Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Tartu, EE
sroots@ebc.ee

ABSTRACT – *The aim of this study is to give an overview of the extent and nature of Southeastern Europe (SEE) paternal genetic variation in relation to potential episodes of gene flow during dispersals of the Upper Paleolithic and Neolithic. A survey based on studies of the paternal gene pool of the region revealed consistency with the typical European paternal gene pool, as five major haplogroups E3b1, I1b*, J2, R1a, and R1b contribute more than 70% to the total genetic variation in SEE. Comprehensive characterization and dating of major paternal lineages imply that SEE has been both an important source and recipient of gene flow.*

IZVLEČEK – *Cilj študije je podati pregled o obsegu in naravi jugovzhodno evropske (JVE) očetovske genske variance v povezavi s potencialnimi epizodami genskega toka med širitvama mlajšega paleolitika in neolitika. Pregled, ki temelji na študijah očetovskega genskega fonda v regiji, je razkril skladnost z značilnim evropskim očetovskim genskim fondom, kajti kar pet glavnih haploskupin E3b1, I1b*, J2, R1a, in R1b prispeva več kot 70% k celotni genski varianci v JVE. Obsežna karakterizacija in datiranje glavnih očetovskih linij nakazuje, da je bila JVE tako pomemben vir, kot prejemnik genskega toka.*

KEY WORDS – *Southeastern Europe; human Y-chromosomal gene pool; Paleolithic lineages; Neolithic paternal component*

Introduction

The main biological importance of the Y chromosome is its role in sex determination and male fertility. Understanding its genetics is, therefore, of wide medical importance. That, however, does not exhaust its use as an object of research in population genetics. As we have witnessed recently, the Y chromosome (due to its specific structure and strictly paternal inheritance) has become a powerful instrument in the study of the population genetics of bisexual organisms, including humans. In humans, there have been studies of Y-chromosomal variation for more than 20 years by now. Many polymorphisms in the non-recombining region of Y have been described, including approximately 600 biallelic markers in the last YCC nomenclature (Karafet *et al.* 2008), thereby constantly improving the resolution of the phylogenetic tree of the Y chromosome and

thus proving the usefulness of the Y-chromosome for studying the phylogeny and phylogeography spread of Y-chromosomal lineages worldwide and regionally.

The most comprehensive early picture of the European Y chromosomal landscape was offered by two parallel surveys by Semino *et al.* (2000) and Rosser *et al.* (2000), which both revealed similar clinal patterns for major European haplogroups. Semino *et al.* (2000) found that more than 95% of European Y chromosomes studied could be grouped into 10 phylogenetically defined haplogroups. The geographic distribution and age estimates were interpreted as testifying to two Paleolithic and one Neolithic migratory episode that contributed to the modern European gene pool. The majority of European Y chromo-

somes belong to haplogroups R1a, R1b, I, and N3, which, taken together, cover about 70–80% of the total Y chromosome pool. The remaining 20% of males belong to haplogroups J2, E3b, or G. While the general distribution patterns of European paternal lineages were revealed in the two aforementioned studies, there are numerous studies by different research groups who have focused on detailed region or population-specific studies. In this respect, the Y-chromosomal variation of Southeast Europe has been studied to determine the source regions of the inhabitation of the region, as well as to attempt to indicate the potential episodes of gene flow during the dispersals of the Upper Paleolithic and Neolithic period. These topics have been raised, discussed and debated in many different studies during the last decade.

Results and discussion

The objective of the present study is to give an overview and to attempt to evaluate the extent and nature of Balkan (SEE) paternal genetic variation in relation to potential episodes of gene flow during the dispersals of Upper Paleolithic and Neolithic hunter-gathering and farming populations in light of the knowledge accumulated by recent studies about the paternal heritage of this region (Semino *et al.* 2000; Barac *et al.* 2003; Semino *et al.* 2004; Pericic *et al.* 2005; Marijanovic *et al.* 2005; Martinez *et al.* 2007; King *et al.* 2008; Battaglia *et al.* 2008). The understanding of the temporal aspects of the spread and distribution of paternal lineages in SEE is especially relevant, as the Balkans are located on an important trajectory in the colonization process of Europe, migrations along which have taken place at least twice, in Paleolithic times and during the Neolithic (Mellars 2004; 2006). Moreover, SEE has been stated to be a starting-point of the spread of the European-specific autochthonic paternal lineage I-P37 (Semino *et al.* 2000; Barac *et al.* 2003, Roots *et al.* 2004, Pericic *et al.* 2005; Underhill *et al.* 2007) in

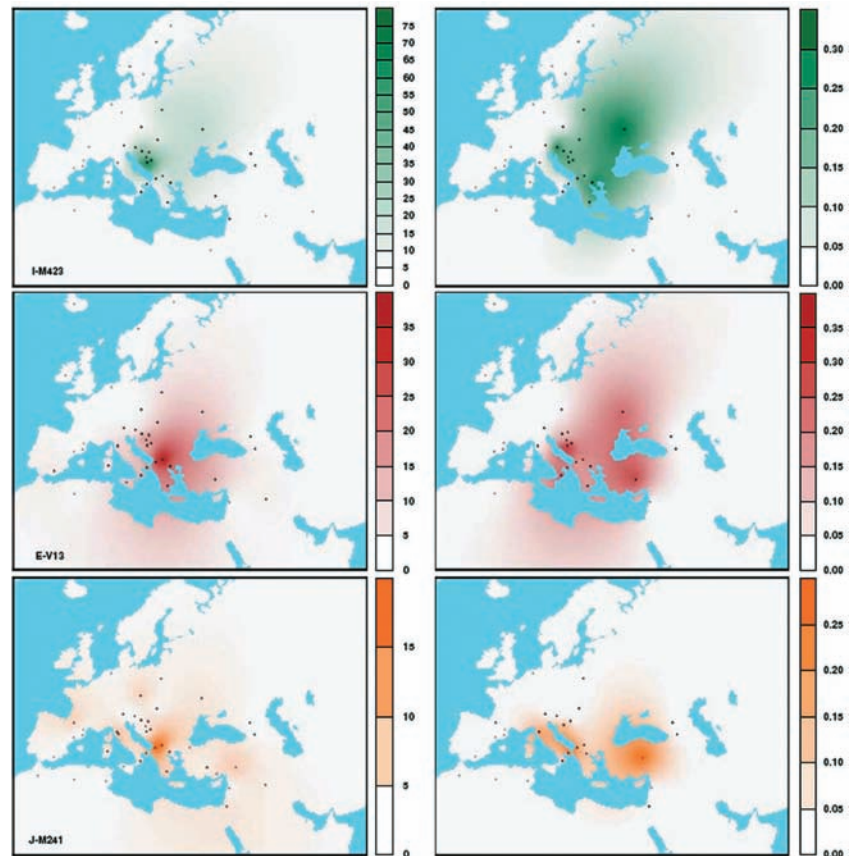


Fig. 1. Frequency (left) and variance (right) distributions of the main Y-chromosome haplogroups, I-M423, E-V13 and J-M241 in SEE region; observed in Battaglia *et al.* 2009.

the re-colonization process of Europe after the LGM in the Early Holocene; and later in the Neolithic period, J-M12(M102) lineages trace the diffusion of people from the southern Balkans to the west (Semino *et al.* 2004).

One of the first studies focusing on the distribution of paternal lineages in North-western Balkan was published by Barac *et al.* (2003), where Y chromosome variation in 457 Croatian samples (mainland and four island populations) was studied using 16 SNPs/indels and eight STR loci. The study was the first to reveal the high frequency of haplogroup I in Croatian populations and to suggest the Adriatic coast as one likely source for the re-colonization of Europe following the Last Glacial Maximum, according to phylogeography and the STR diversity pattern. In contrast, R1a frequency was suggested as a sign of the Slavic impact in the Balkan region. Haplogroups J, G, and E, related to the spread of farming, characterized a minor part (12.5%) of Croatian paternal lineages. Similar conclusions about the spread pattern and proportions of paternal lineages were reached in a study by Marjanovic *et al.* (2005) regarding the peopling of Bosnia-Herzegovina. The vari-

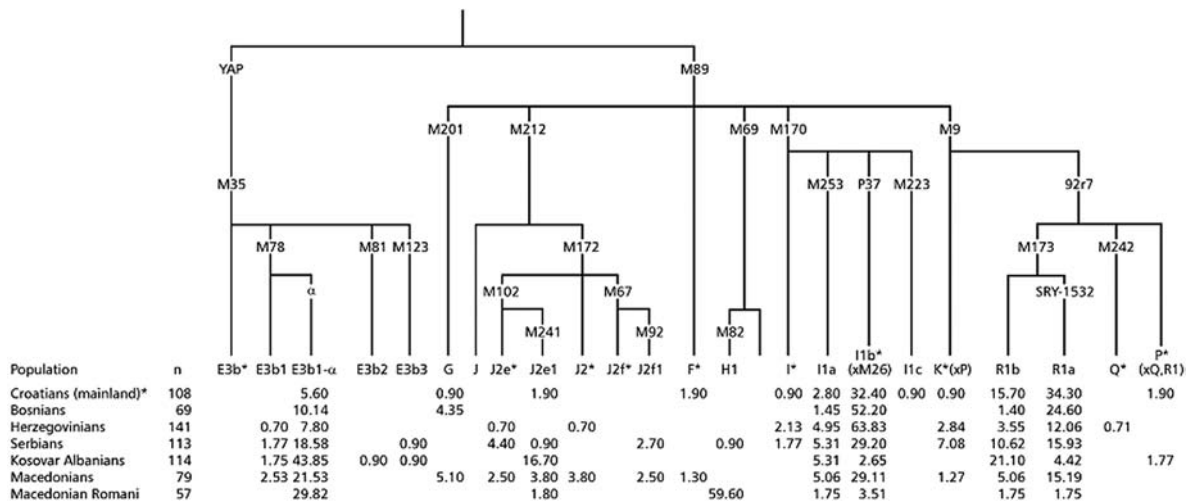
ation at 28 Y-chromosome biallelic markers was analyzed in 256 males (90 Croats, 81 Serbs and 85 Bosnians) from Bosnia-Herzegovina. The three main groups of Bosnia-Herzegovina, in spite of some quantitative differences, share a large fraction of the same ancient gene pool (high frequency of the I-P37 lineage) distinctive for the Balkan area.

In comparison to the north-western area of the Balkans, the populations of the southern part of the Balkan Peninsula, like the Greek, have a somewhat different haplogroup frequency distribution. Two studies, King *et al.* (2007) and Martinez *et al.* (2007), present the distribution patterns of Y-chromosomal lineages in populations from the Southern Balkans, which partly overlap with those in the other Balkan populations described earlier, but partly reveal more similarities to Middle Eastern/Anatolian populations. The main hgs observed in Europe (E, I, J, R1a and R1b) contribute differently to the gene pool of the various SEE regions, Hg I (mostly I-P37 or I-M423 according to more recent nomenclature) and Hg R (both R1a and R1b), being the most represented in the whole Balkan region, while Hg E (V13) and Hg J2 (M12 sub-lineages) are mainly frequent in the southern Balkan populations.

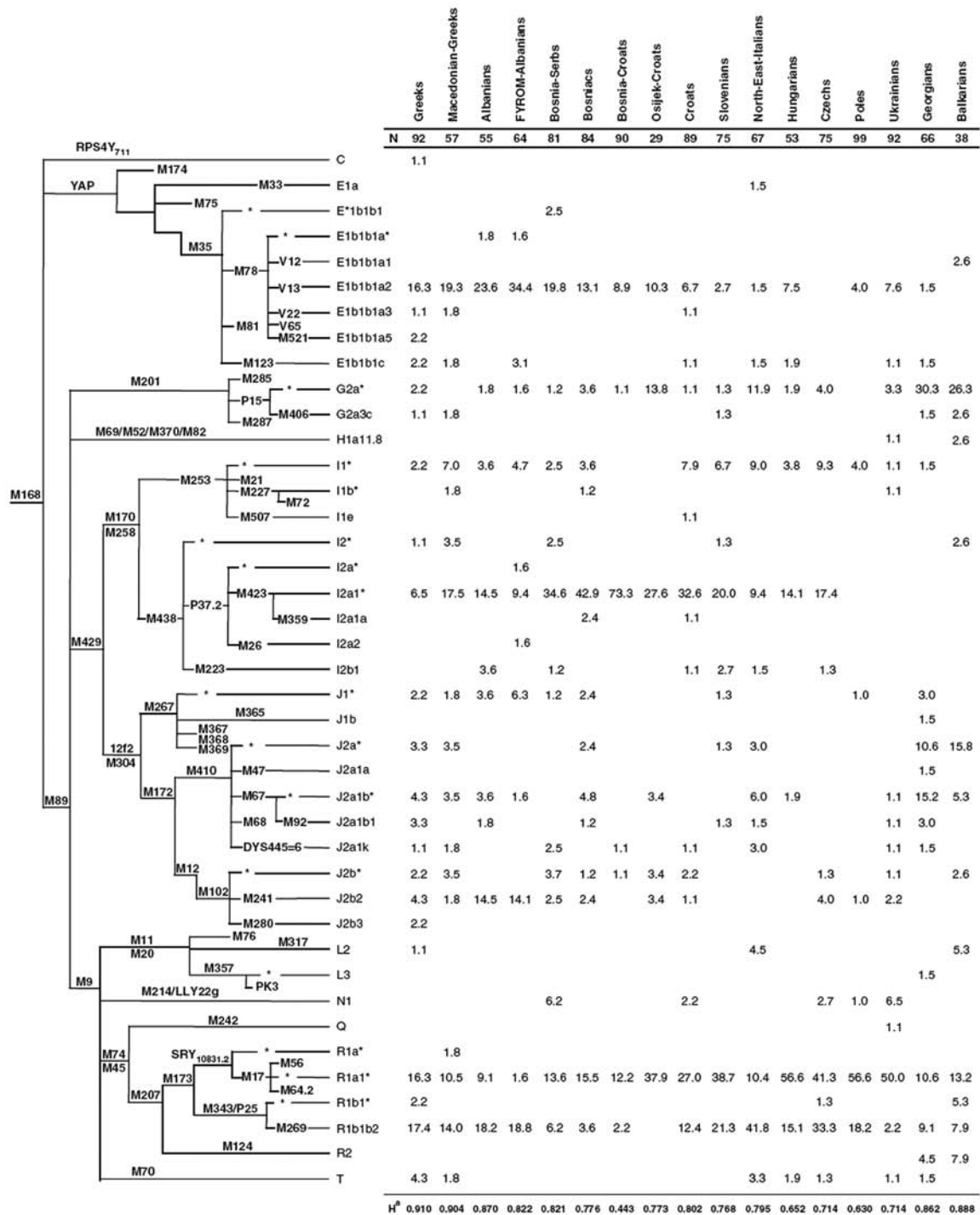
In the study by King *et al.* (2008), 171 samples were collected from areas near three known early Neolithic settlements in Greece together with 193 samples from Crete. An analysis of Y-chromosome haplogroups determined that the samples from the Greek Neolithic sites showed a strong affinity with Balkan data, while Crete showed an affinity with central/Mediterranean Anatolia. Haplogroup J2b-M12 was frequent in Thessaly and Greek Macedonia, while

haplogroup J2a-M410 was scarce. Conversely, Crete, like Anatolia, showed a high frequency of J2a-M410 and low frequency of J2b-M12. The expansion time of Y-STR variation for haplogroup E3b1a2-V13 in the Peloponnese was consistent with an indigenous Mesolithic presence. In turn, two distinct haplogroups, J2a1h-M319 and J2a1b1-M92, had demographic properties consistent with Bronze Age expansions to Crete, arguably from NW/W Anatolia and Syro-Palestine, while a later mainland (Mycenaean) contribution to Crete was indicated by the presence of V13.

Another study, dedicated specifically to elucidating the Cretan paternal gene pool, was published by Martinez *et al.* (2007). The geographic stratification of the contemporary Cretan Y-chromosome gene pool was assessed by high-resolution haplotyping to investigate the potential imprints of past colonization episodes and the population substructure. In addition to analyzing the possible geographic origins of Y-chromosome lineages in relatively accessible areas of the island, the study included samples from the isolated interior of the Lasithi Plateau – a highland plain located in eastern Crete. The potential significance of the results from the latter region was underscored by the possibility that this region was used as a Minoan refugium. Comparisons of Y-haplogroup frequencies among three Cretan populations as well as with published data from additional Mediterranean locations revealed significant differences in the frequency distributions of paternal haplogroups within the island. The most outstanding differences were observed in the cases of haplogroups J2 and R1, with a predominance of haplogroup R lineages in the Lasithi Plateau and that of haplogroup J lin-



Tab. 1. Y-chromosomal SNP tree and haplogroup frequencies in seven SEE populations (from Peričić *et al.* 2005).



Tab. 2. The phylogenetic relationships of Y-chromosome Hgs and their distribution in the examined south-east European populations (from Battaglia et al. 2008).

eages in the more accessible regions of the island. Y-STR-based analyses demonstrated the close affinity that R1a1 chromosomes from the Lasithi Plateau shared with those from the Balkans, but not with those from lowland eastern Crete. In contrast, Cretan R1b microsatellite-defined haplotypes displayed more resemblance to those from Northeast Italy than to those from Turkey and the Balkans.

An important study of the SEE paternal heritage from a more general aspect was published by Pericic *et al.* (2005), extending the number of analyzed populations (7 populations) and sample sizes and setting the obtained data in a wider phylogenetic context. The extent and nature of southeastern Europe (SEE) paternal genetic contribution to the European genetic landscape were explored based on a high-resolu-

tion Y chromosome analysis involving 681 males from seven populations in the region. Paternal lineages present in SEE were compared with previously published data from western Eurasian populations. The finding that five major haplogroups (E3b1, I1-P37 (xM26), J2, R1a, and R1b) comprise more than 70% of SEE total genetic variation is consistent with the typical European Y chromosome gene pool. However, the distribution of major Y chromosomal lineages and estimated expansion signals clarify the specific role of this region in structuring European, and particularly Slavic, paternal genetic heritage. The contemporary Slavic paternal gene pool, mostly characterized by the predominance of R1a and I-P37 (xM26) and the scarcity of E3b1 lineages, is a result of several major prehistoric gene flows with different directions: the post-Last Glacial Maximum R1a expansion from east to west, the Younger Dryas-Holocene I-P37(xM26) diffusion out of SEE, in addition to subsequent putative R1a and I-P37(xM26) gene flows between eastern Europe and SEE, and a rather weak diffusion of E3b1 toward regions nowadays occupied by Slavic-speaking populations. To illustrate the proportions of the main components of SEE paternal lineages, Table 1 is presented here.

One more recent study focusing on the topic of SEE paternal heritage was published by Battaglia *et al.* (2008). To investigate the possible involvement of indigenous people in the transition to agriculture in the Balkans, patterns of Y-chromosome diversity in 1206 subjects from 17 population samples, mainly from Southeast Europe, were analyzed in the study. The main conclusions from the study are as follows: evidence from three Y-chromosome lineages – I-M423, E-V13 and J-M241 – makes it possible to distinguish between Holocene Mesolithic forager and subsequent Neolithic range expansions from the eastern Sahara and the Near East. In particular, while the Balkan microsatellite variation associated with J-M241 correlates with the Neolithic period, those related to E-V13 and I-M423 Balkan Y chromosomes are consistent with a late Mesolithic time frame. In addition, the low frequency and variance associated with I-

M423 and E-V13 in Anatolia and the Middle East support a European Mesolithic origin of these two clades. The ensuing range expansions of E-V13 and I-M423 parallel the diffusion of Neolithic Impressed Ware in space and time, thereby supporting a case of cultural diffusion. Illustrating the statements of this study, Figure 1 and Table 2 is presented.

Conclusions

The paternal heritage of SEE is consistent with the typical European paternal gene pool, as five major haplogroups E3b1, I1b*, J2, R1a, and R1b comprise over 70% of the genetic variation in SEE. Comprehensive characterization and dating of major paternal lineages suggest that SEE has been both an important source and recipient of gene flow. Estimated expansion signals related to the major Balkan Y-chromosomal lineage I-M423 (earlier known as I-P37) and E-V13 (more common in the Southern Balkans) are consistent with a late Mesolithic time frame. In addition, the low frequency and variance associated with I-M423 and E-V13 in Anatolia and the Middle East support a European Mesolithic origin of these two clades. Thus, these Balkan Mesolithic foragers, with their own autochthonous genetic signatures, became the earliest to adopt farming when it was subsequently introduced by migrating farmers from the Near East. These converted indigenous farmers became the principal agents who spread this economy by using maritime leapfrog colonization strategies in the Adriatic and transmitting the Neolithic cultural package to other adjacent Mesolithic populations. The Neolithic component in the SEE paternal gene pool is most clearly marked by the presence of the J-M241 (more frequent in the Southern Balkans) lineage, and its expansion signals associated with Balkan microsatellite variation correlate with the Neolithic period.

ACKNOWLEDGEMENTS

This research was partly supported by the Estonian Science Foundation Grant No. 7445 (to SR).

∴

REFERENCES

- BARAĆ L., PERIČIĆ M., KLARIĆ MARTINOVIĆ I., ROOTSI S., JANIČIJEVIĆ B., KIVISILD T., PARIK J., RUDAN I., VILLEMS R. and RUDAN P. 2003. Y chromosomal heritage of Croatian population and its island isolates. *European Journal of Human Genetics* 11: 535–542.
- BATTAGLIA V., FORNARINO S., AL-ZAHERY N., OLIVIERI A., PALA M., MYRES N. M., KING R. J., ROOTSI S., MARJANOVIĆ D., PRIMORAC D., HADZISELIMOVIC R., VIDOVIC S., DROBNIC K., DURMISHI N., TORRONI A., SANTACHIA-RA-BENERECETTI A. S., UNDERHILL P. A., SEMINO O. 2009.

- Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. *European Journal of Human Genetics* 17: 820–830.
- KARAFET T. M., MENDEZ F. L., MEILERMAN M. B., UNDERHILL P. A., ZEGURA S. L., HAMMER M. F. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830–838.
- KING R. J., ÖZCAN S. S., CARTER T., KALFOĞLU E., ATASOY S., TRIANTAPHYLLIDIS C., KOUVATSI A., LIN A. A., CHOW C.-E. T., ZHIVOTOVSKY L. A., MICHALODIMITRAKIS M., UNDERHILL P. A. 2008. Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. *Annals of Human Genetics* 72: 205–214.
- MARJANOVIC D., FORNARINO S., MONTAGNA S., PRIMORAC D., HADZISELIMOVIC R., VIDOVIC S., POJSKIC N., BATTAGLIA V., ACHILLI A., DROBNIC K., ANDELINOVIC S., TORRONI A., SANTACHIARA-BENERECETTI A. S. and SEMINO O. 2005. The Peopling of Modern Bosnia-Herzegovina: Y-chromosome Haplogroups in the Three Main Ethnic Groups. *Annals of Human Genetics* 69: 757–763.
- MARTINEZ L., UNDERHILL P. A., ZHIVOTOVSKY L. A., GAYDEN T., MOSCHONAS N. K., CHOW C. E., CONTI S., MAMOLINI E., CAVALLI-SFORZA L. L., HERRERA R. J. 2007. Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *European Journal of Human Genetics* 5: 485–493.
- MELLARS P. 2004. Neanderthals and the modern human colonization of Europe. *Nature* 432(7016): 461–5.
2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313(5788): 796–800.
- PERIČIĆ M., BARAC LAUC L., MARTINOVIĆ KLARIĆ I., ROOTSI S., JANIČIJEVIĆ B., RUDAN I., TERZIĆ R., ČOLAK I., KVESIĆ A., POPOVIĆ D., ŠIJAČKI A., BEHLULI I., ĐORĐEVIĆ D., EFREMOVSKA L., BAJEC Đ. D., STEFANOVIĆ B. D., VILLEMS R. and RUDAN P. 2005. High-Resolution Phylogenetic Analysis of Southeastern Europe 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., BARAC 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., FÜREDI 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. *The 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., BERTRANPETTIT J., BOSCH E., BRADLEY D. G., BREDE G., COOPER G., CÔRTE-REAL H. B. S. M., DE KNIJFF P., DECORTE R., DUBROVA Y. E., EVGRAFOV O., GILISSEN A., GLISIC S., GÖLGE M., HILL E. W., JEZIOROWSKA A., KALAYDJIEVA L., KAYSER M., KIVISILD T., KRAVCHENKO S. A., KRUMINA A., KUČINSKAS 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., JOÃO PRATA M., PREVIDERÉ C., ROEWER L., ROOTSI S., RUBINSZTEIN D. C., SAILLARD J., SANTOS F. R., STEFANESCU G., SYKES B. C., TOLLUN 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. *The American Journal of Human Genetics* 67: 1526–1543.
- SEMINO O., PASSARINO G., OEFNER P. J., LIN A. A., ARBUZOVA S., BECKMAN L. E., DE BENEDICTIS 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., TRIANTAPHYLLIDIS C., SHEN P., 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. *The American Journal of Human Genetics* 74: 1023–1034.
- UNDERHILL P. A., MYRES N. M., ROOTSI S., CHOW C.-E. T., LIN A. A., OTILLAR R. P., KING R., ZHIVOTOVSKY L. A., BALANOVSKY O., PSHENICHNOV A., RITCHIE K. H., CAVALLI-SFORZA L. L., KIVISILD T., VILLEMS R., WOODWARD S. R. 2007. New phylogenetic relationships for Y chromosome haplogroup I: Reappraising its phylogeography and prehistory. In P. Mellars, K. Boyle, O. Bar-Yosef, C. Stringer (eds.), *Rethinking the Human Revolution: New Behavioural and Biological Perspectives on the Origin and Dispersal of Modern Humans*. McDonald Institute Monographs. McDonald Institute for Archaeology, Cambridge: 33–42.