UDK 902(4)''631/634'':577.2
Documenta Praehistorica XXXIII (2006)

Y-chromosome haplogroup I prehistoric gene flow in Europe

Siiri Rootsi
Estonian Biocentre and Tartu University, Estonia
sroots@ebc.ee

ABSTRACT – *To investigate which aspects of contemporary human Y-chromosome variation in Europe are characteristic of primary colonization, late-glacial expansions from refuge areas, Neolithic dispersals or more recent events in gene flow haplogroup I was analyzed. The analysis of Hg I Y chromosomes revealed several sub-clades with distinct geographic distributions. Sub-clade I1a accounts for most of Hg I in Scandinavia, with a rapidly decreasing frequency towards the East European Plain and the Atlantic fringe; but microsatellite diversity reveals that the Iberian Peninsula/Southern France refugial area could be the source region of the early spread of both I1a and the less common I1c. I1b* extends from the eastern Adriatic to Eastern Europe, and declines noticeably towards the southern Balkans, and abruptly towards North Italy. This clade probably diffused after the Last Glacial Maximum from a homeland in the Balkans or Eastern Europe. In contrast, I1b2 most probably arose in southern France/Iberia, underwent a post-glacial expansion, and marked the human colonization of Sardinia about 9000 years ago.*

IZVLEČEK – *Da bi raziskali, kateri aspekti sodobne človeške variacije kromosoma Y v Evropi so značilni za primarno kolonizacijo, pozno glacialno širitev iz refugijev, neolitske širitve in bolj recentne dogodke v genskem zapisu, smo analizirali haploskupine I. Analize kromosoma Y haploskupine I, so pokazale sub-klade z jasnimi geografskimi porazdelitvami. Sub-klad I1a označuje Hg I v Skandinaviji, s hitrim pojemanjem v smeri proti Vzhodnoevropski ravnici in obrobju Atlantika. Mikrosatelitska raznolikost je pokazala, da je bil Iberijski polotok/južna Francija verjetni refugij in izvorno področje, iz katerega sta se razširila oba sub-klada I1a in manj pogosti I1c. I1b* se razteza od vzhodnega Jadrana do vzhodne Evrope in opazno pojema v smeri proti južnemu Balkanu ter strmo upade v smeri proti severni Italiji. Ta klad se je verjetno razširil po zadnjem glacialnem maksimumu iz domovine na Balkanu ali v vzhodni Evropi. I1b2 najverjetneje izhaja iz južne Francije/Iberije in kaže post-glacialno širitev in človekovo kolonizacijo Sardinije pred okoli 9000 leti.*

KEY WORDS – phylogeny of Y-chromosomal markers; haplogroup I sub-clades; late-glacial expansions; Neolithic dispersals

Introduction

Although there is sound evidence that the majority of present-day European genes descend from indigenous Palaeolithic ancestors, it is to be expected that the pre-LGM landscape of the spread of genetic variation was profoundly re-shaped during and after the LGM – Late Glacial Maximum (*Richards et al. 2000; Semino et al. 2000*). The aim of this investigation was to apply the phylogeographic approach to Y-chromosomal haplogroup I, the only known Y-chromosomal haplogroup that probably arose in Europe in Palaeolithic times, and which is still common and widespread there, whilst being virtually absent elsewhere.

The distribution of haplogroup I was intriguing, with its two high frequency peaks in distant parts of Europe (the Balkan region and Scandinavia); and the further large-scale I haplogroup study by Rootsi et al. (*2004*) concentrated on achieving a better phylo-

© 2006 Oddelek za arheologijo, Filozofska fakulteta - Univerza v Ljubljani, SI

genetic and phylogeographic resolution of this haplogroup, informative for the reconstruction of long-distance gene flows in space and time.

Results and Discussion

It has been shown earlier that the high frequency of hg I is characteristic of two distant and distinct regions – around the Dinaric Alps (*Semino et al. 2000; Barać et al. 2003*) and in Nordic populations of Scandinavia (*Semino et al. 2000; Passarino et al. 2002; Tambets et al. 2004*).

In a study by Rootsi et al. (*2004*), concentrating on haplogroup I phylogeography and phylogeny, more than seven thousand individuals from Europe and the surrounding regions were assessed for the marker M170, which defines hg I.

1104 Y chromosomes from 48 European and 12 populations from surrounding regions which showed the derived M170 C- allele were further genotyped with a set of markers (M253, P37, M26 and M223) that define distinct sub-clades of I, respectively I1a, I1b, I1b2 and I1c.

Thanks to the new informative markers used, the improved resolution of phylogeny of hg I enabled to reveal distinct phylogeographical patterns of sub-clades I1a, I1b and I1c, which jointly cover about 95% of hg I individuals. Sub-clade I1a is widely distributed in northern Europe, with its highest frequencies in Scandinavia: in Norwegians, Swedes and Saami, accounting for 88–100% of hg I individuals in these populations and showing rapidly decreasing frequency towards both the East European Plain and the northwestern coastal areas of Europe. Combined analysis of STR diversity and a relative portion of I1a sub-clade among all I lineages suggests that France or possibly more precisely – the Franco-cantabrian refugial area – could have been the source region of the spread of I1a during the post-LGM re-colonization of Europe. The same may apply to the spread of the less common sub-clade I1c. This scenario is also supported by a high positive correlation (0.75) between the geographic distributions of I1a and I1c. I1c covers a wide range in Europe, with the highest frequencies in north-west coastal Europe, and a lower frequency elsewhere (Fig. 1).

A totally different distribution pattern can be seen for I1b*, which is the most frequent haplogroup I clade in Eastern Europe and the Balkans. It reaches its highest frequencies in Croatian and Bosnian populations, encompassing almost 80–90% of hg I there. When comparing frequencies in different regions of Croatia (*Barać et al. 2003*), clear and significant differences between the three southern islands with higher frequency and the mainland and the northerly island of Krk with lower frequency, became apparent. More than half of Croatian hg I sampled individuals – 126 out of 221 (57%) – share an identical STR haplotype, which was named the Dinaric Modal Haplotype. This haplotype was not present in 102 hg 2 chromosomes (according to Jobling's nomenclature) as reported by Helgason et al. (*2000*); the most frequent among them was labeled as the Nordic Haplotype (*Barać et al. 2003*).

The phylogenetic network of hg I STR haplotypes points to characteristic haplotype patterns in different sub-clades, which allows us to identify possible founder haplotypes for the different sub-clades, and calculate their possible expansion times according to the method described in Zhivotovsky et al. (*2004*).

The estimates for possible expansion times suggest that the expansion phase of I1a and I1b occurred around the early Holocene and only the less frequent sub-clade I1c shows an earlier age for its STR variation, suggesting that the corresponding mutation arose earlier.

High frequency combined with the high diversity of sub-clade I1b in the Croatian population (both mainland and island populations) suggests that during the LGM there might have been a refugium nearby. According to our knowledge, placing of the western Balkans on the list of human refugia during the LGM has not been confirmed so far unambigously by archaeologists. However, the northern part of the Adriatic Sea, including the Dalmatian Islands, was at that time a part of dry land, being covered by water only much later, at the boundary of Holocene (references within *Barać et al. 2003*). Therefore, one may speculate that the wealth of the traces of human occupancy of the area lies submerged. Nevertheless, it is justified to suggest that the present-day western Adriatic was the reservoir of M170 (I1b) lineages, as well as a starting point for the spread of these lineages during the post-glacial re-colonization of Europe. Meanwhile, the star-like pattern of both, I1a and I1b* STR haplotypes, might be explained by simultaneous re-colonization of Europe from different refugia.

I1b* sub-clade dissipates very rapidly west of the Balkans, being virtually absent among Italian, French and Swiss populations, but extending eastwards at

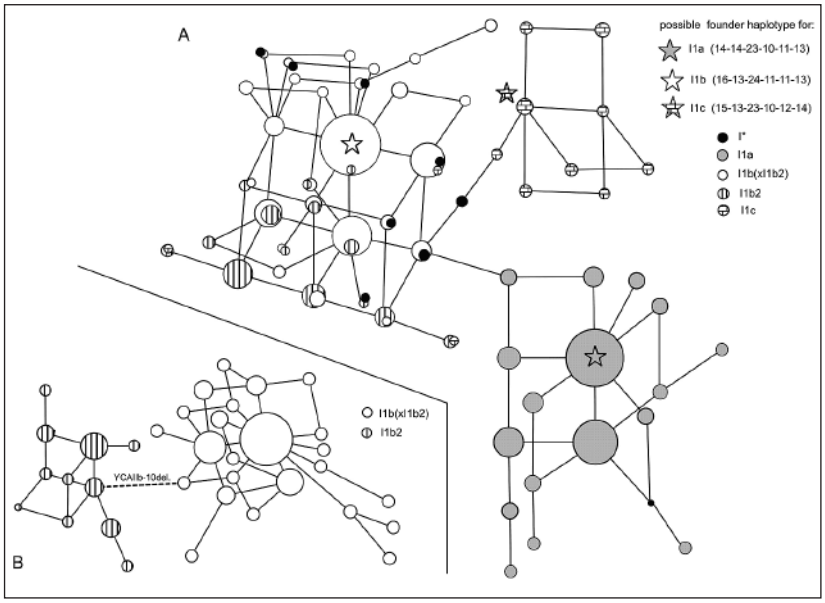


Fig. 1. A – Median-joining network of combined haplotypes of six STR loci (DYS19, 388, 390, 391, 392, 393) in 25 populations/584 individuals (Norwegians, Estonians, Saami, Swedes, Hungarians, Czechs and Slovaks, Poles, Ukrainians, Croats, Bosnians, Macedonians, Albanians, Greeks, Moldavians, Gagauz, Turks, Italians, Sardinians, French, Dutch, Andalusians, Bearnais, Basques and Swiss). Only haplotypes with frequency >1 were used. Nodes indicate haplotypes with sizes proportional to their frequency (smallest node corresponds to 1 individual – only in case of overlap between subclades, otherwise haplotypes with frequency >1 are presented). Haplotypes of different sub-clades are indicated with different patterns.
The most frequent haplotype of I1a sub-clade (14-14-23-10-11-13) corresponds to the earlier named Nordic Haplotype and the dominant in I1b* haplotype is the so-called Dinaric Modal Haplotype (16-13-24-11-11-13) according to Barać et al. (2003). The possible founder lineage for the third sub-clade, I1c (15-13-23-10-12-14), was revealed in Rootsi et al. (2004).
B - Median joining network of I1b(xI1b2) and I1b2 lineges based on seven STRs (DYS 19, 388, 390, 391, 392,393 and YCAIIa,b). I1b* (YCAIIa,b mostly 21,21 alleles) and I1b2 (YCAIIa,b; 21,10 alleles) lineages are clearly separated from each other by the difference inYCAII a,b haplotypes. They differ by 10 repeats in YCAIIb allele length that is probably a result of one single mutational event rather than step-by step deletions - in particular because no intermediate repeat variants have been detected.

notable frequencies, mostly in the north Balkans and among Slavic-speaking populations, including more eastern Ukrainians. This finding suggests that I1b* may have expanded from a glacial refuge area, which may have been located in the Balkans. As indicated above, there is only limited archaeological evidence for such a refugium in this region at present. Nevertheless, data on the re-occupation of northern Europe from the Balkan region by mammals such as the brown bear *Ursus arctos* (*Taberlet and Bouvet 1994*) and European hedgehog Erinaceus europeus (*Hewitt 2000*), birds such as the European great tit *Parus major* (*Kvist et al. 1999; Kvist 2000*) and insects such as the meadow grasshopper *Chortippus parallelus* (*Hewitt 2000*), supports its existence indirectly.

It seems somewhat less likely though not impossible that I1b* was preserved during the LGM in an area of much better documented Periglacial refugium in the present-day Ukraine. It appears less likely because (a) not only its frequency, but also diversity is higher in the Adriatic region; (b) a branch of I1b* – I1b2–M26 – has a clearly western pattern of distribution, being totally absent in Ukrainians.

On the other hand, a clearly visible difference that can be observed in the distribution patterns of I1b* and an outshoot of it – I1b2 – suggests that their separation may have occurred even before the LGM, whereas isolation, genetic drift during the LGM, recolonization and an unknown number of putative more recent demographic events created the pattern that one observes among extant populations.

Meanwhile, the extremely high incidence of I1b2 among Sardinians (about 40%) can be explained by the presence of carriers of I1b2 lineage among the first inhabitants of the island early in Holocene and by the influence of genetic drift thereafter.

Certain extent of similarity in distribution patterns of some mtDNA haplogroups, in particular V (*Torroni et al. 1998; Torroni et al. 2001*), with Y-chromosomal hg I sub-branches has been suggested by Rootsi et al. (*2004*). These findings show that distribution patterns characteristic to Y-chromosomal haplogroup I sub-clades are supported by parallel evidence from other genetic markers and probably indicate more general patterns in human past demographic movements.

REFERENCES

BARAĆ L., PERIČIĆ M., KLARIĆ I. M., ROOTSI S., JANIČIJEVIĆ B., KIVISILD T., PARIK J. et al. 2003. Y chromosomal heritage of Croatian population and its island isolates. *European Journal of Human Genetics 11: 535–42*.

HELGASON A., SIGURETH ARDOTTIR S., NICHOLSON J., SYKES B., HILL E. W., BRADLEY D. G., BOSNES V. et al. 2000. Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *American Journal of Human Genetics 67: 697–717*.

HEWITT G. 2000. The genetic legacy of the Quaternary ice ages. *Nature 405: 907–913*.

KVIST L., RUOKONEN M., LUMME J. & ORELL M. 1999. The colonization history and present-day population structure of the European great tit (Parus major major). *Heredity 82: 495–502*.

KVIST L. 2000. *Phylogeny and phylogeography of European Parids*. Dissertation. Oulu University.

PASSARINO G., CAVALLERI G.L., LIN A. A., CAVALLI-SFORZA L. L., BORRESEN-DALE A. L., UNDERHILL P. A. 2002. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *European Journal of Human Genetics 10: 521–9*.

RICHARDS M., MACAULAY V., HICKEY E., VEGA E., SYKES B., GUIDA V., RENGO C. et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *American Journal of Human Genetics 67: 1251–1276*.

ROOTSI S., MAGRI C., KIVISILD T., BENUZZI G., HELP H., BERMISHEVA M., KUTUEV I. et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics 75: 128–37*.

SEMINO O., PASSARINO G., OEFNER P. J., LIN A. A., ARBUZOVA S., BECKMAN L. E., DE BENEDICTIS G. et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science 290: 1155–9*.

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

TAMBETS K., ROOTSI S., KIVISILD T., HELP H., SERK P., LOOGVALI E. L., TOLK H. V. et al. 2004. The Western and Eastern Roots of the Saami-the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. *American Journal of Human Genetics 74: 661–82*.

TORRONI A., BANDELT H. J., MACAULAY V., RICHARDS M., CRUCIANI F., RENGO C., MARTINEZ-CABRERA V. et al. 2001. A signal, from human mtDNA, of post-glacial recolonization in Europe. *American Journal of Human Genetics 69: 844–852*.

TORRONI A., BANDELT H.-J., D'URBANO L., LAHERMO P., MORAL P., SELLITTO D., RENGO C. et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics 62: 1137–52*.

ZHIVOTOVSKY L. A., UNDERHILL P. A., CINNIOGLU C., KAYSER M., MORAR B., KIVISILD T., SCOZZARI R. et al. 2004. The effective mutation rate at y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics 74: 50–61*.