

Journal: www.joqq.info

Originally Published: Volume 4, Number 1 (Spring 2008)

Reference Number: 41.006

WHERE DID EUROPEAN MEN COME FROM

Author(s): *Kalevi Wiik*

Where Did European Men Come From?

Kalevi Wiik

Abstract

On the basis of variation in the Non-Recombining portion of the Y-chromosome, or Single Nucleotide Polymorphisms (SNPs), European men are usually classified into approximately ten major haplogroups or "clans." The frequencies are thought to give generally reliable information about the migrations of European men during the last about 40,000 years.

The first part of this paper examines Europe as a whole, and presents the frequency isoglosses of ten Y-chromosome haplogroups in seventeen overall maps of the continent. The maps show the essential clinal gradients of the haplogroups and are analyzed to provide a means of better understanding ancient colonization events such as migration. The second part of this paper divides Europe into eight sub-regions and presents frequency tables and isogloss maps as more detailed clinal gradients for each sub-region. Finally, hypotheses are presented about the potential language shifts that have occurred in the European populations.

Introduction

Y Haplogroups in Europe

Nearly all European men can be classified into one of the nine most frequently occurring Y-chromosome haplogroups or clans. Haplogroups are defined by mutations or single nucleotide polymorphisms (SNPs). The nine most common haplogroups, in alphabetical order, are E3b, G, I1a, I1b1-P37, I1b2-M223, J2, N3, R1a, and R1b. Sometimes, two more haplogroups, E3a and (in Northeast Europe) N2, are added to the list, which makes the total number eleven. In addition, there are many other rare haplogroups in Europe.

Figure 1 shows the Y phylogenetic tree for European haplogroups. It includes forty haplogroups attested in Europe, but thirty of these are so rare that their frequencies do not reach one per cent anywhere in Europe. The six thicker branches or groups of branches of the tree, representing Haplogroups R1, I, N, E, J, G, and some of their subgroups, are discussed in this review. The more detailed structure of these haplogroups may be seen in the Y phylogenetic tree maintained by ISOGG (2007). The tree in Figure 1 is a simplification of that in the article by Underhill and Kivisild (2007).

The SNP's presently available provide information only on the large-scale classification of Y chromosomes. Each haplogroup may be further resolved into clusters based on similarity of Y Short Tandem Repeats or Y-STR's. Additional information may be derived from other genetic markers. However, only the SNP classifications will be used in this paper, but this provides information

on the development of the human family, specifically the European part of the family, over the last 40,000 years.

Men's Migrations to Europe

The genetic development and migrations of the ancestors of European men can be presented in the following ten phases; see Map 1.

- (1) About 50 thousand years ago (kya) all the ancestors of the present European men still lived in northeastern Africa and formed only one clan.
- (2) About 45 kya the clan was split into an "African" Clan E and an "Asian" Clan F, and the "Asian" clan moved out of Africa to the Arabic peninsula and the Near East.
- (3) About 40 kya part of the "Middle Eastern" Clan F gave rise to the "Central Asian" Clan K.
- (4) About 35 kya two new clans, R and NO, branched off from K. Clan R moved to western Central Asia and Clan NO to eastern Central Asia.
- (5) About 30 kya Clan R was split into R1 and R2, and Clan R1 moved to the steppe area between the Ural mountains and the Caspian Sea.
- (6) About 25 kya one branch of Clan R1, Clan R1b, reached Iberia and the Atlantic Coast, and somewhat later Clan R1a branched from R1 and became common in the present-day Ukraine.
- (7) About 25 kya the "Middle Eastern" Clan F sent another branch to Anatolia and further to the Balkans, and a new sub-Clan I emerged.

Address for correspondence: Kalevi Wiik, kalevi.wiik@pp1.inet.fi

Received: 10 Oct 2007; accepted: 03 Feb 2008.

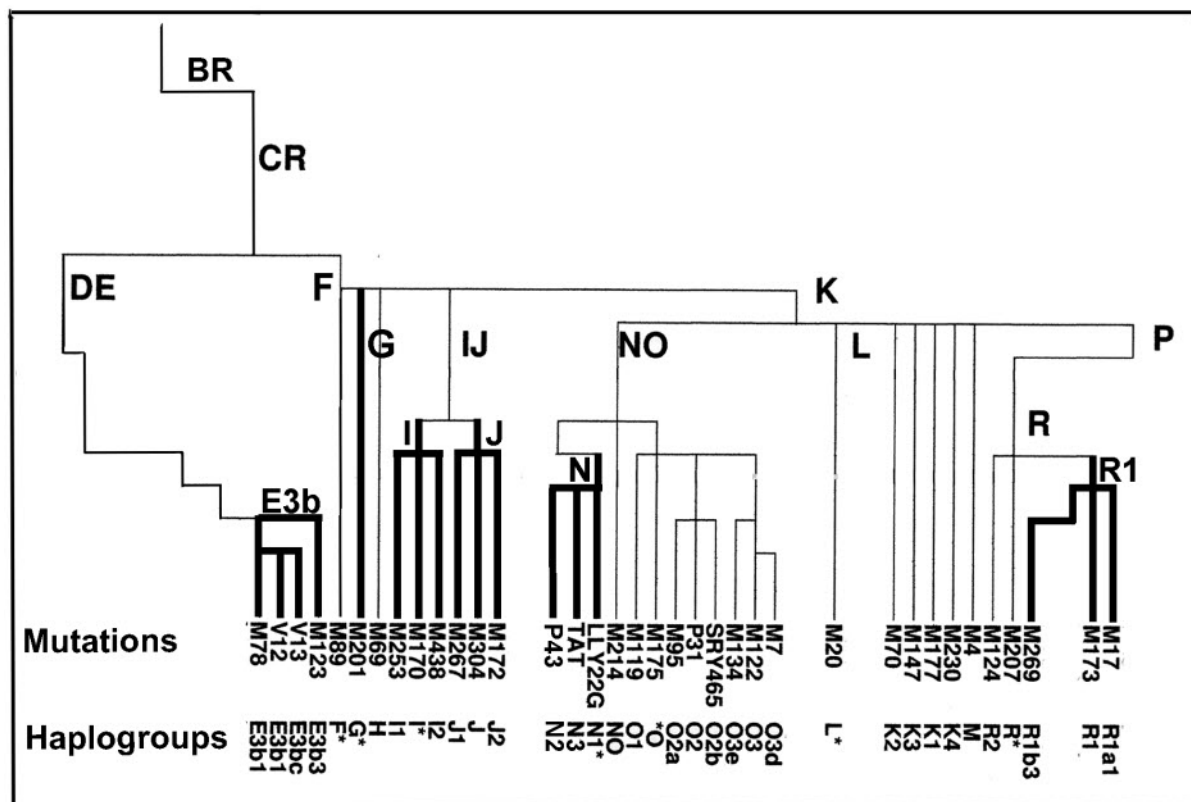
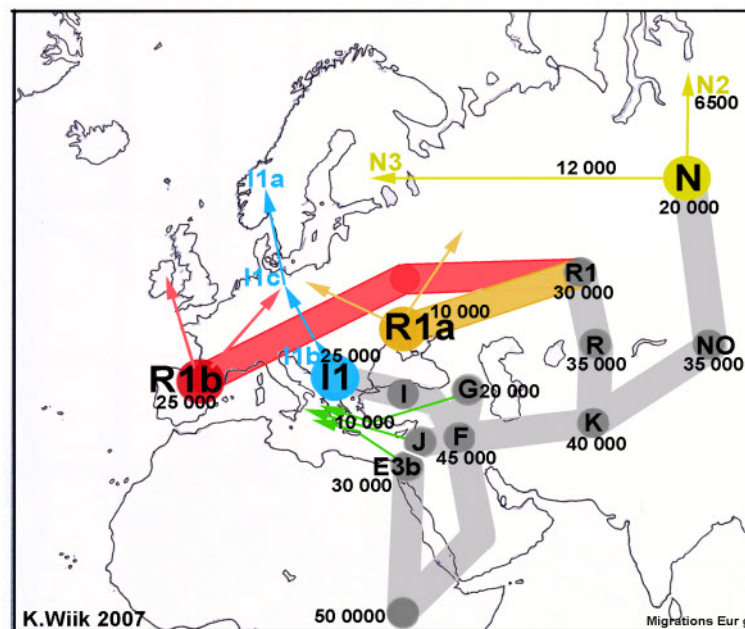


Figure 1 A phylogenetic tree of about forty Y-chromosome haplogroups and subhaplogroups in Europe and its vicinity. The tree is a simplification of the Y-chromosome haplogroups of the entire world. The subhaplogroups of I have been regrouped. Source: Underhill-Kivisild (2007).



Map 1 The migration routes of the ancestors of European men from Africa to Europe. Grey routes show the genetic developments (by mutations) of European Haplogroups E3b, F, G, I, j, N, and R1. Coloured circles = the approximate geographic locations of the main four male populations during the last glacial maximum (LGM) about 20 kya. Red, blue, brown, and yellow arrows show essential population movements during the recolonization of northern Europe during the late glacial period (about 16-10 kya). Green arrows indicate arrival of Early Farmers in the Balkans and Mediterranean coast about 10 kya.

(8) Perhaps at about the same time, Clan NO in eastern Central Asia developed into “Siberian” Clan N and moved towards the north.

(9) Clan N was split into two sub-clans, N3 and N2, and these moved first to northwestern Siberia and later to Eastern and Northeastern Europe.

After these nine phases, Europe experienced a cooling climate and the onset of the Last Glacial Maximum (LGM). During the LGM the ancestors of the European men retreated from northern Europe into four refuges located in Iberia, the Ukraine, the Balkans, and Siberia. These core areas were habitable even during the coldest periods of the LGM.

(10) About 10 kya the farmers of the Middle East, representing African Clan E (its sub-clan E3b) and two sub-clans of F (the “Caucasian” Clan G and the “Near Eastern” Clan J), spread to Anatolia and further to Greece and the Mediterranean coast.

Clans I, E3b, J, and G all originate from the Middle East, but only E3b, J, and G (not I) belong to the group of “Early Farmers.” Clan I had spread into Europe before the emergence of effective domestication of wild plants and animals (i.e. the beginning of agriculture and cattle raising) in the Middle East. Because of their early departure, they were still hunter-gatherers at the time of the arrival of the Early Farmers in the Balkans, and they were taught to cultivate land and raise cattle by their “Middle East brothers” after the “reunion of the family” in the Balkans. Accordingly, Clan I represents the

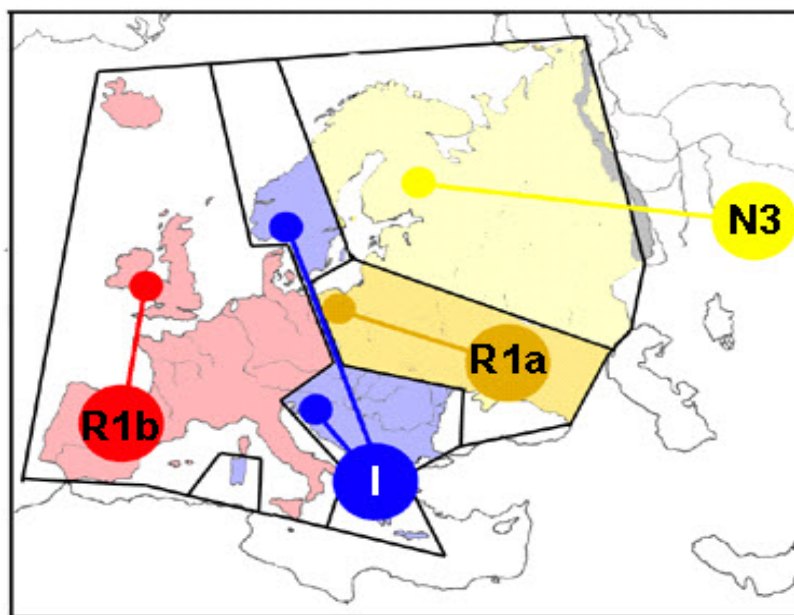
“Old Europeans,” rather the Early Farmers. The rather rare Clan F* in Europe may belong to either group. If we think the men of F* clan left the Middle East for the Balkans before the emergence of effective farming, they are comparable to the men of Clan I and belong to the Old Europeans, but if they left the Middle East only after the beginning of farming, they are comparable to the men of Clans E3b, J and G and represent the Early Farmers.

Two Main Groups of European Men

European men (more precisely, their Y chromosomes) can be classified into two categories:

(a) Those who are Old Europeans in the sense that, at the start of the LGM, their paternal lineages already were in Europe and they came to the four refuges when they were forced out of northern Europe. They were first to repopulate Europe after the LGM and they formed the bulk of the present European male population.

(b) Those who during the Ice Age still were in the warm regions of Asia and Africa. These latter populations came to Europe during the Neolithization of Europe (the arrival of farming) that started about 10 kya. The frequencies of the two groups of men are shown in the tables in columns OE (Old Europeans) and EF (Early Farmers). The frequency of the Old Europeans will be considered here as the sum of R1b + R1a + I + N and that of the Early Farmers is the sum of E3b + J2 + G.



Map 2 Four refuges and their typical Y-chromosome haplogroups. From west to east they are the Iberian, Balkan, Ukrainian, and Siberian refuges. The bigger circles represent the four refuges, and the smaller ones the peak areas of the corresponding populations today.

(a) The Old Europeans

The European populations are generally thought to have been forced southward into three refuges during the Last Glacial Maximum (about 20 kya and thereafter). The refuges were (from the west to the east) in Iberia, the Balkans, and the Ukraine. In the present study, one additional refuge, that of Siberia, is considered. The four refuges are seen in Map 2 as the larger circles; the

smaller circles represent locations of the peak areas of the four populations today (Clan I has two peak areas in the map). Each refuge had its own prominent Y-chromosome haplogroup: the haplogroup was R1b in Iberia, I in the Balkans, R1a in the Ukraine, and N3 in Siberia. Until about 10 kya, the European male population consisted typically of only these four clans: R1b, I, R1a, and N3. There may have been others, but their impact on the European male population was smaller.

Table 1

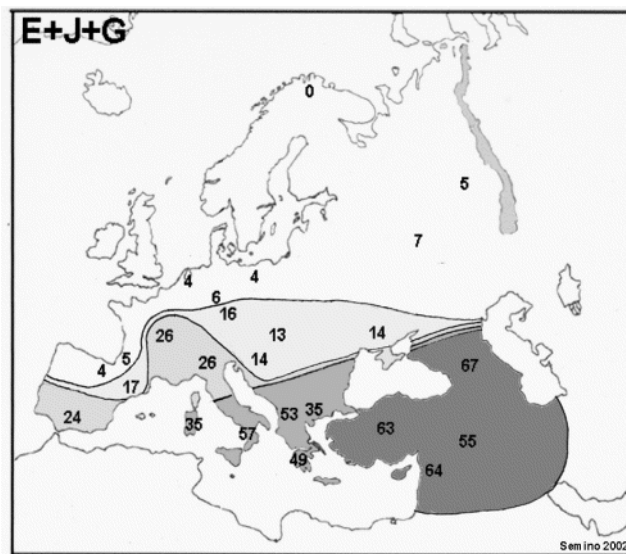
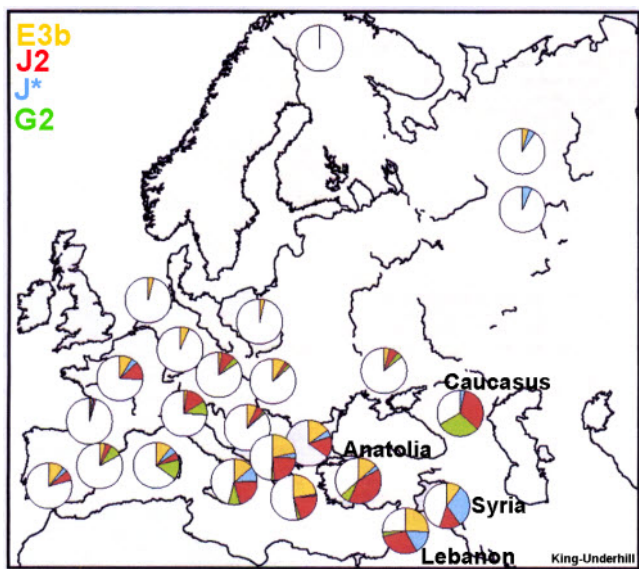
Percentages of the Most Common Haplogroups of 25 Male Populations in Europe and Vicinity. The haplogroups are classified into three groups: those of the Old Europeans, Early Farmers, and Others. The percentage of the Early Farmers can be used as an indicator of the relative number of men bringing agriculture as a demic diffusion to Europe about 10 kya and soon after. "Others" represent haplogroups and clans that may belong to either group. The table is simplified from the equivalent table in King (2002). Others = A, C, H, K, L, P.

	Population	Old Europeans				Total	Early Farmers				Other Total
		R1b	R1a	I	N3		E	J	G	Total	
1	Andalusian	65.5		3.4		68.9	10.3	13.3		23.6	7.5
2	Basque-Sp	88.9		6.6		95.5	2.2	2.2		4.4	0
3	Basque-Fr	86.4		9.1		95.5		4.5		4.5	0
4	Catalan	79.2		4.2		83.4	4.2	4.2	8.3	16.7	0
5	French	52.2		17.4		69.6	8.7	17.3		26.0	4.4
6	Dutch	70.4	3.7	22.2		95.9	3.7			3.7	0.4
7	German	50.2	6.2	37.5		93.7	6.2			6.2	0
8	Czech + Slov	35.6	26.7	15.6	2.2	77.9	2.2	8.9	4.4	15.5	6.6
9	Italian	62.0	4.0	8.0		70.4	2.2	14.0	10.0	26.2	3.4
10	Calabrian	32.4				32.4	16.4	32.4	8.0	56.6	11.0
11	Sardinian	22.1		37.7		59.8	10.4	10.4	14.2	35.0	5.2
12	Croatian	10.3	29.3	44.8		84.4	6.9	5.2	1.7	13.8	1.8
13	Albanian	17.6	9.8	19.6		47.0	23.6	27.5	2.0	53.1	0
14	Greek	27.6	11.8	7.9		47.3	23.7	22.3	2.6	48.6	4.1
15	Macedonian	10.0	35.0	20.0		65.0	15.2	20.0		35.0	0
16	Polish	16.4	56.4	23.6		96.4	3.6			3.6	0
17	Hungarian	13.3	60.0	11.1		74.4	8.9	2.2	2.2	13.3	2.3
18	Ukrainian	2.0	54.0	18.0	6.0	74.0	4.0	6.0	4.0	14.0	12.0
19	Georgian	14.3	7.9			22.2		36.5	30.1	66.6	11.2
20	Turkish	6.6	6.6	3.3	3.3	19.8	13.3	43.3	6.6	63.2	17.0
21	Lebanese	6.4	9.4	3.2		19.3	25.8	35.1	3.2	64.1	16.6
22	Syrian	15.0	10.0	5.0		30.0	10.0	45.0		55.0	15.0
23	Saami	8.3	8.3	41.7	41.7	100				0	0
24	Udmurt	11.6	37.2	7.0	30.2	86.0	4.7	4.7		5.4	8.6
25	Mari		13.0	4.3	69.5	86.8		6.5		6.5	6.7

(b) *The Early Farmers*

The situation changed about 10 kya when agriculture started to expand from the Middle East. The new subsistence system expanded into Europe in two different ways: (a) as demic diffusion, according to which the new subsistence system was brought to new areas by newcomers who moved in and brought agriculture with them, and (b) by cultural diffusion, according to which

agriculture expanded as the indigenous people acquired a new knowledge of how to cultivate land and they were not moving away from their old living areas. According to the demic diffusion model, this new form of subsistence was brought to Europe mainly by three clans: E3b, J2, and G2. Map 3 shows the relative portions of these clans in southwestern Asia (e.g. the Middle East) and Europe. The percentages of the four haplogroups of the Early Farmers (E3b, J2, JxJ2, and G2) in relation to the



Maps 3 and 4 The Early Farmers of four haplogroups: E3b, J2, J x J2 (J, not in J2--probably all J1), and G2 in Europe. The white section of each circle belongs mainly to the Old European populations R1b, I, R1a, and N3, while the sum of the coloured sections show the approximate share of the Early Farmers. The map is based on King and Underhill (2002).

percentages of the haplogroups of the populations from the four refuges (R1b, I, R1a, and N3) can, at least to some extent, be used as an indicator of how large the early farming populations arriving in various parts of Europe were. I call the two groups of haplogroups and clans “Early Farmers” and “Old Europeans.” The percentages of the two groups can be seen in Table 1 and Maps 3 and 4.

The table and maps can be summarized as follows:

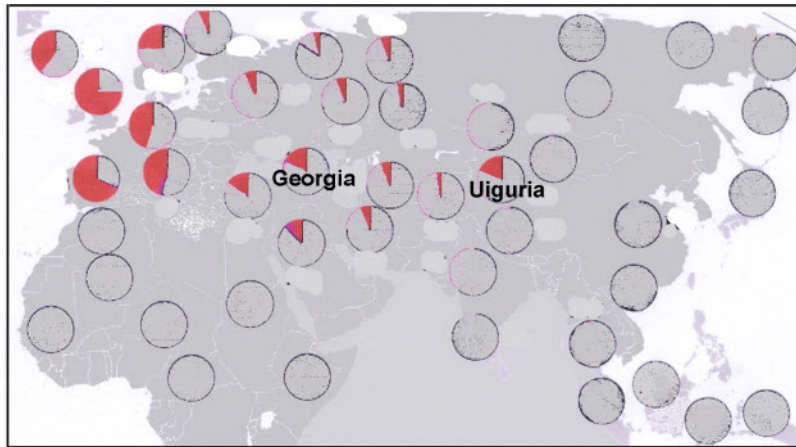
- (1) In the Middle East and Anatolia, the Early Farmers represent the majority (about 55-64%), while in Europe the frequency declines sharply in a clinal gradient from about 35-57% of the male populations in the southern Balkans and southern Italy, to a frequency near zero in far northwest Europe.
- (2) The coloured slices of the easternmost three circles in Map 3 show that the Caucasus was a major source area for Clans J2 and G2 and that Lebanon and Syria were

important sources for Clans E3b, J2, and JxJ2 (probably J1). J2 came in about equal portions from the Caucasus and Lebanon, with slightly less from Syria.

European Men in a Wider Context

In many cases the Wikipedia and Balanovsky maps contain enough data to indicate where the European men have come from. In some cases, wider maps showing the whole of Eurasia are needed for this purpose. Next, I consider four groups from a wider perspective: how far do the distributions of certain haplogroups extend outside Europe?

- (1) Haplogroup R1b has its peak values in West Europe and its total area extends far beyond the eastern border of Europe. In Map 5 the distribution of R1b is seen to extend as far east as Uiguria (in northwestern China). The fact that this haplogroup has two secondary peaks outside Europe (one in Georgia and the other in Uiguria)

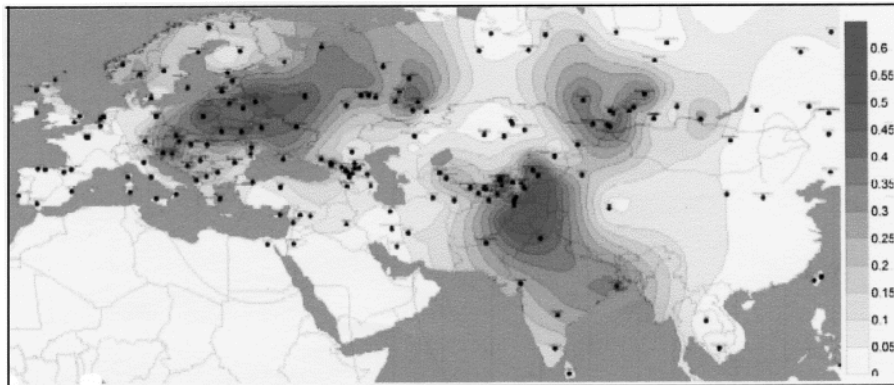


Map 5 The geographic distribution of Haplogroup R1b. Source: Wikipedia.

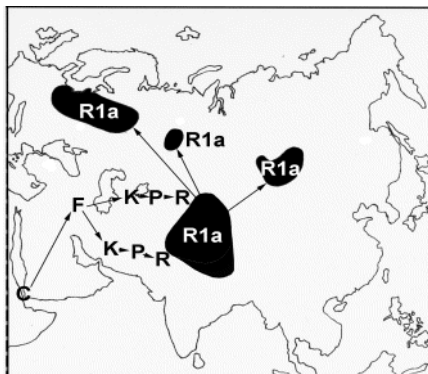
tends to indicate that the R1b men may have arrived in Europe from the Caucasus or Central Asia.

(2) Clan R1a, the “brother clan” of R1b, has a very wide distribution with four peak areas. The East-European

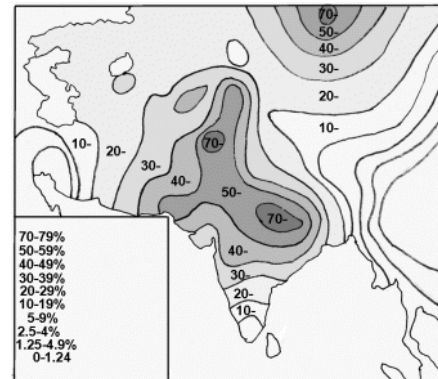
subgroup stayed in the Ukrainian refuge during the cold periods of the Ice Age. The Eurasian total area of Haplogroup R1a is seen in Map 6. The haplogroup has four peak areas: one in northern India, another in Altai, a third one in the Mari area (Central Volga region), and



Map 6 The total geographic distribution of Haplogroup R1a. Source: Kivisild (2005).



Map 7 The main areas of concentration of Haplogroup R1a. The map also shows the development of Haplogroup R1a from Haplogroup F (through intermediate stages K and P).

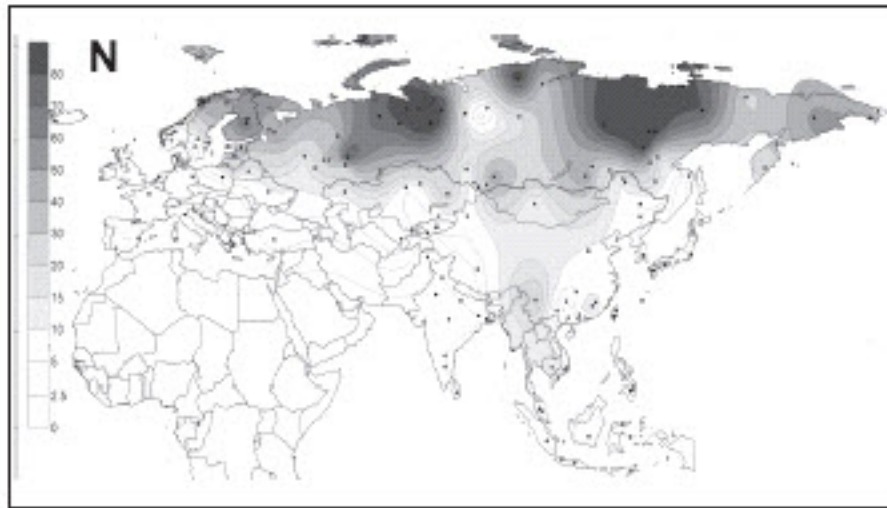


Map 8 The eastern extensions of Haplogroup R1a in India and Mongolia. The map is based on a map in Sahoo et al. (2006).

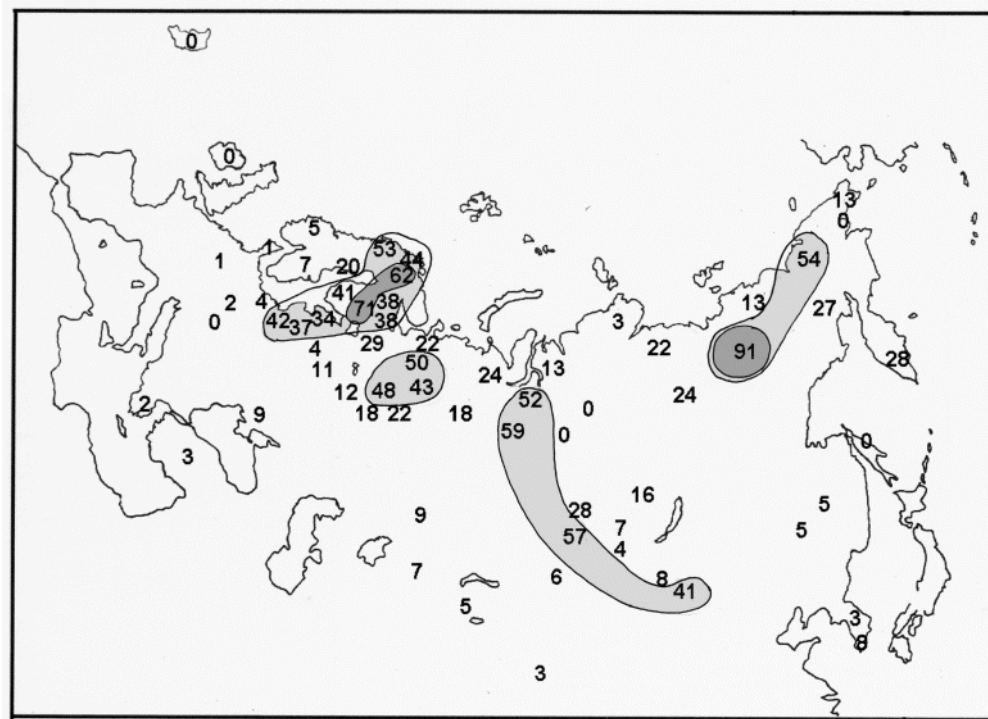
a fourth one in the Polish-Russian region; see also Map 7. The two European maximum areas are also seen in detail in Maps 18 and 19 later in this article.

(3) Haplogroup I is restricted for the most part to Europe. An indication of the direction of arrival of this clan is seen in Map 25 in which the frequencies are relatively high (4-12%) in Anatolia.

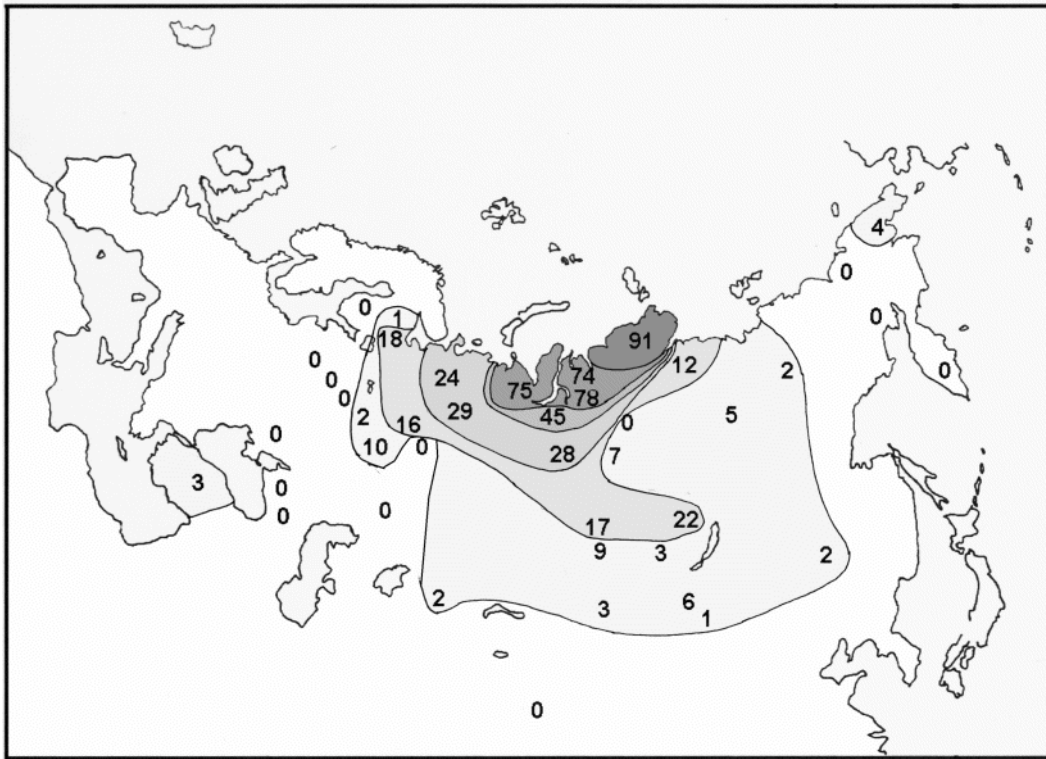
(4) The high frequencies of Haplogroup N in northeastern Europe and practically all Siberia show that the men of this clan came to Europe from the Siberian refuge; see Map 9. This group can also be called “Uralic-Altaic” as many of its present-day representatives speak Uralic (Finno-Ugric) or Altaic languages. Group N was divided into two main parts, N3 and N2; see Maps 10, 11, and 12. Clan N2 was to inhabit northernmost Siberia, while



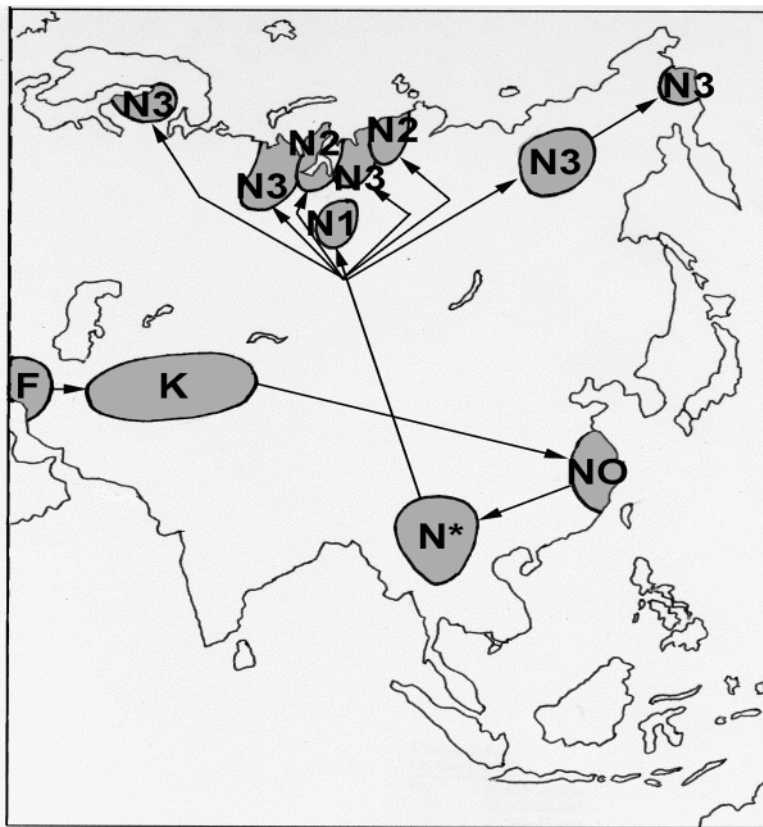
Map 9 The geographic distribution of all the subhaplogroups of N. Source: Rootsi et al. (2006).



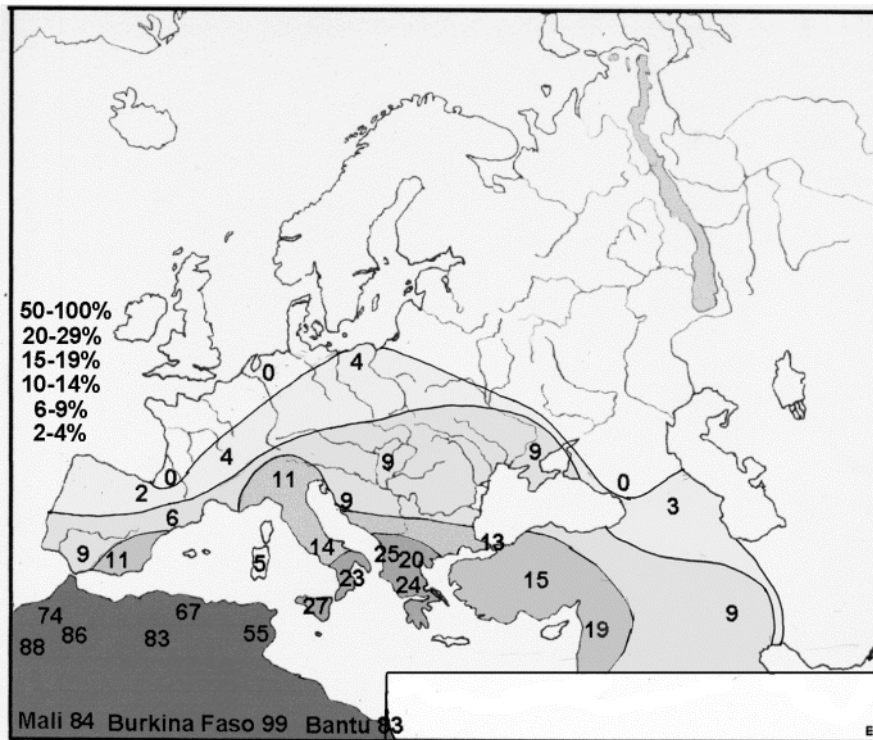
Map 10 Frequencies of Haplogroup N3 in Eurasia.



Map 11 Frequencies of Haplogroup N2 in Eurasia.



Map 12 The development of Haplogroup N from F through K and NO and the potential routes of the clans in question.



Map 13 The geographical distribution of Haplogroup E (including all of its subhaplogroups).

Clan N3 extended to the west and east; today, Clan N2 speaks Samoyedic languages, while the western clans of N3 speak Finno-Ugric languages and eastern clans Altaic and Paleo-Siberian languages.

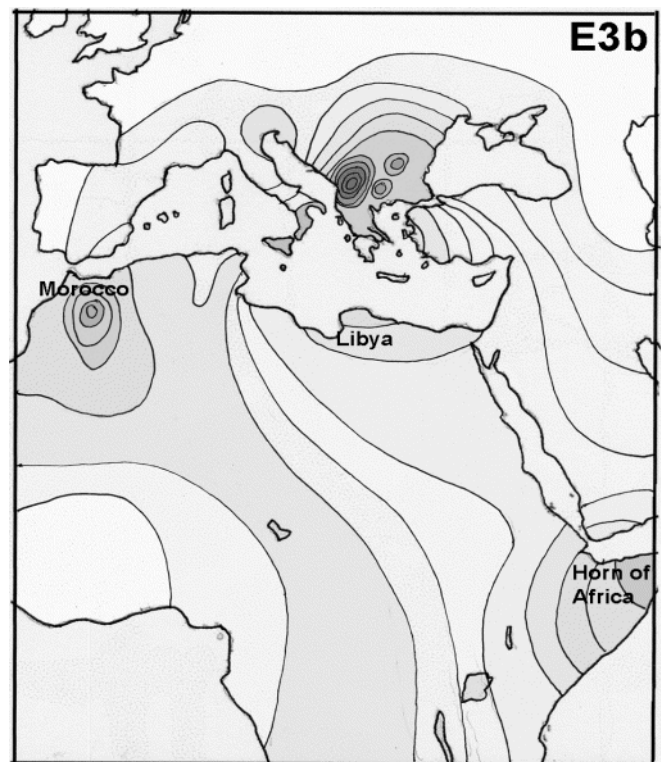
(5) From the European perspective, Haplogroups E3b, J2, and G belonged to the first farmers to arrive in Europe from the Middle East.

(a) Haplogroup E3b is a branch of the "African" Haplogroup E; see Maps 13 and 14.

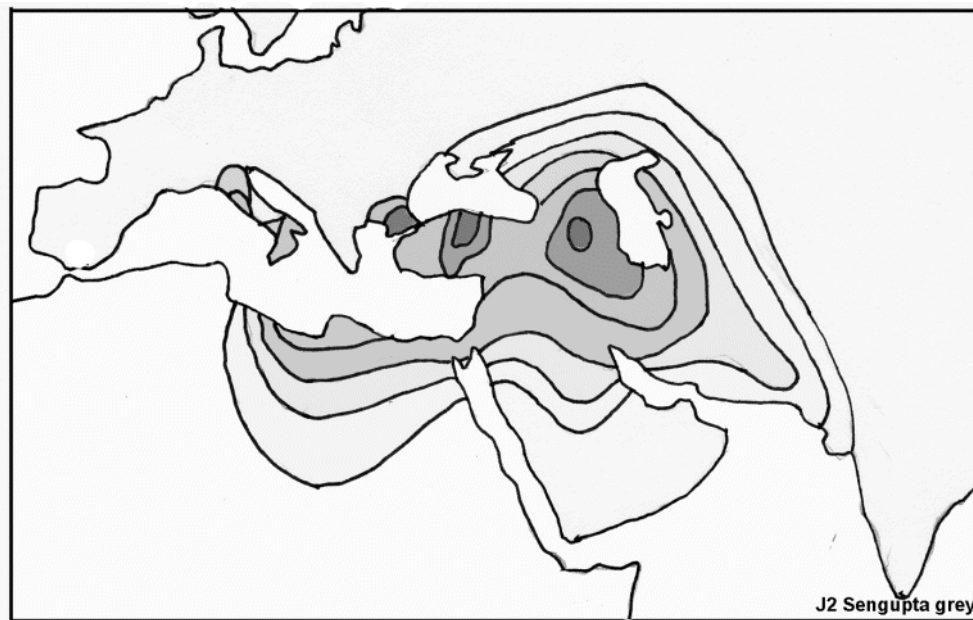
(b) The source area of the "Near Eastern" Haplogroup J was perhaps on the eastern coast of the Mediterranean; its subhaplogroup J2 has its maximum area today in the Caucasus and Anatolia; see Map 15.

(c) The source area of Haplogroup G was perhaps in Pakistan and India; from the European point of view, its peak values are today in the Caucasus; see Map 34.

Haplogroup E originated in Africa. This is true also of its European subhaplogroup E3b. In Africa, this subgroup has four separate peak areas; these are in Southern Africa, Morocco, Northern Libya, and the Horn of Africa. The last two bear witness to the route from northwest Africa through Anatolia to Europe. (The frequencies are relatively high also in Anatolia.) In Europe, Haplogroup E3b has peak values (about 25%) in Albania; about equally high frequencies are also typical



Map 14 Frequency zones of the "European" subhaplogroup E3b in parts of Africa, Asia, and Europe. A southern peak area of E3b in southern Africa is not visible on the map.



Map 15 The geographic distribution of Haplogroup J2. Source: Sengupta et al. (2006).

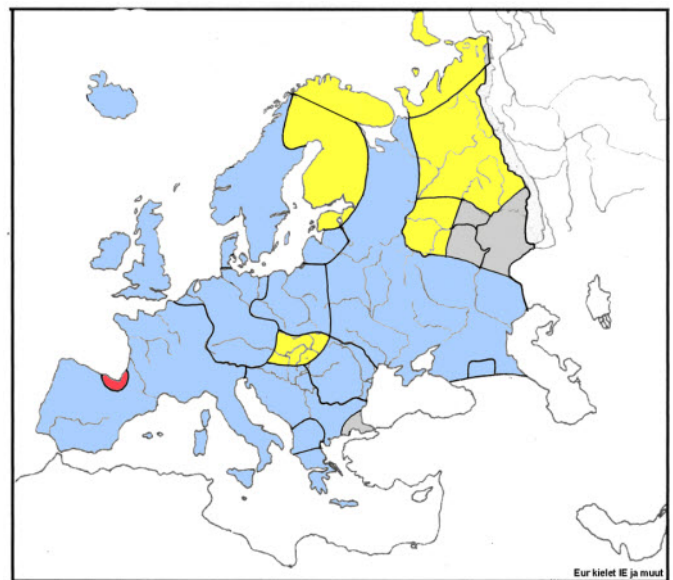
of some other parts of the Balkans and southern Italy. Many of the newcomers representing subclan E3b learnt agriculture on their way through the southwestern part of the Middle East. They represent one group of the typical Early Farmers of Europe.

Languages in Europe

It is likely that there is at least some interdependence between the genome and languages in Europe. The languages spoken in Europe belong to four independent and unrelated phyla: Indo-European, Basque, Finno-Ugric (Uralic), and Turkic (Altaic). The distribution of these phyla is shown in Map 16.

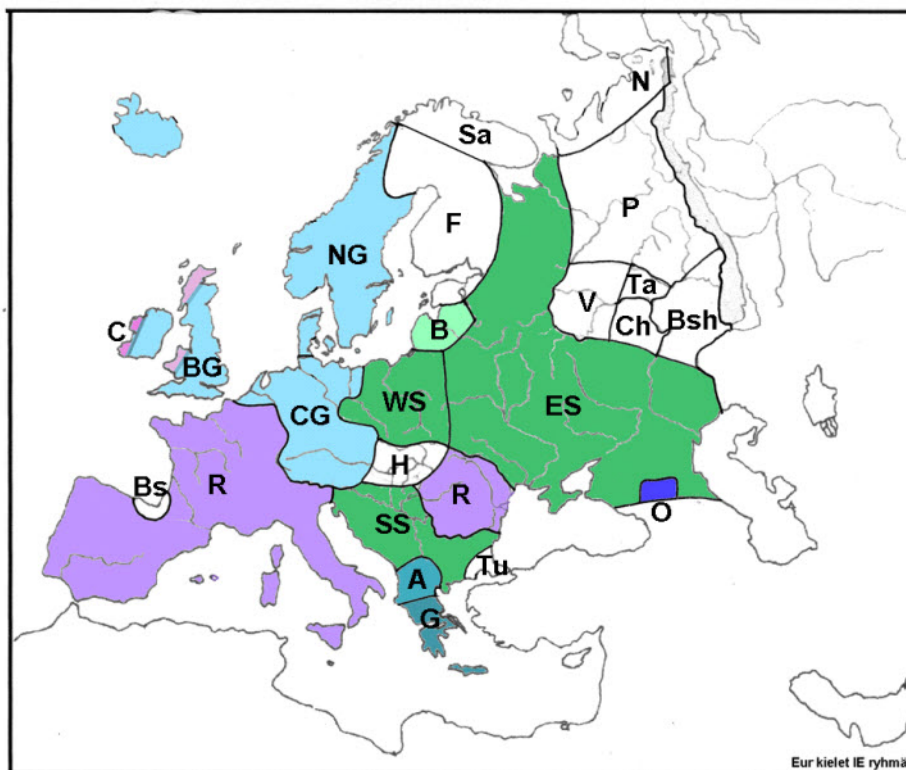
The Indo-European languages represent seven groups: Germanic, Baltic, Slavic, Romance, Celtic, Albanian, and Greek. The Finno-Ugric languages belong to five groups: Finnic, Volgaic, Permic, Ugric, and Samoyedic (represented in Europe only by Nenets). The Turkic languages in Europe are Tatar, Chuvash, Bashkir in the Volga-Ural area, and Turkish in Turkey. The distribution of the language groups or families is shown in Map 17.

The question of which language a population speaks today may be more dependent on the languages spoken in the past by ancient men than by those spoken by ancient women. This statement is supported by the idea that women were more mobile than men—the men have supposedly more often stayed in their original living areas, while women have often moved to the homes of



Map 16 Four language phyla in Europe. Blue = Indo-European, yellow = Finno-Ugric (Uralic), grey = Turkic (Altaic), and red = Basque.

their mates. Accordingly, the languages of men have usually remained unchanged while women have sometimes accepted a new language in their new living area. If this is true, Y-chromosome data being reviewed here may provide more information about the spread of ancient languages than mitochondrial DNA.



Map 17 The approximate areas of European language groups:
 (1) The Indo-European languages: Blue = Germanic (CG = Continental Germanic, BG = Britain Germanic, and NG = North Germanic), B = Baltic, dark green = Slavic (WS = West Slavic, ES = East Slavic, and SS = South Slavic), C = Celtic, R = Romance, A = Albanian, G = Greek, O = Ossetian.
 (2) The Finno-Ugric (Uralic) languages: F = Finnic, Sa = Saami, V = Volgaic, P = Permic, H = Hungarian (a Ugric language), N = Nenets (a Samoyed language).
 (3) The Turkic languages: Ta = Tatar, Ch = Chuvash, Bsh = Bashkirian, Tu = Turkish.
 (4) The Basque language: Bs.

Geographic Distribution of the Haplogroups of European Men

The maps that follow show the approximate frequencies of ten main Y-chromosome haplogroups in various parts of Europe. In some cases haplogroups of more restricted distribution are also considered. The maps are based on the synthetic maps of Wikipedia (in some cases improved by the results of some recent well known genetic studies), and on those in Balanovsky et al. (2008). The next few maps are presented in pairs so that each pair represents one haplogroup with the map on the left based on the Wikipedia version and that on the right on Balanovsky et al. The copyright of the original Wikipedia maps is owned by Relative Genetics, Inc.

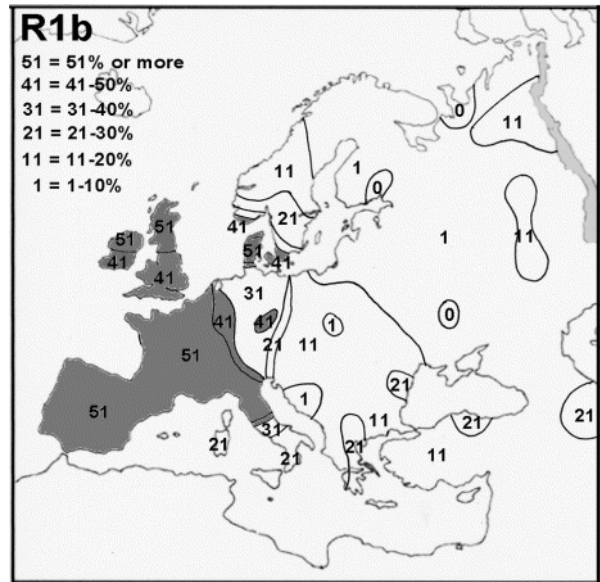
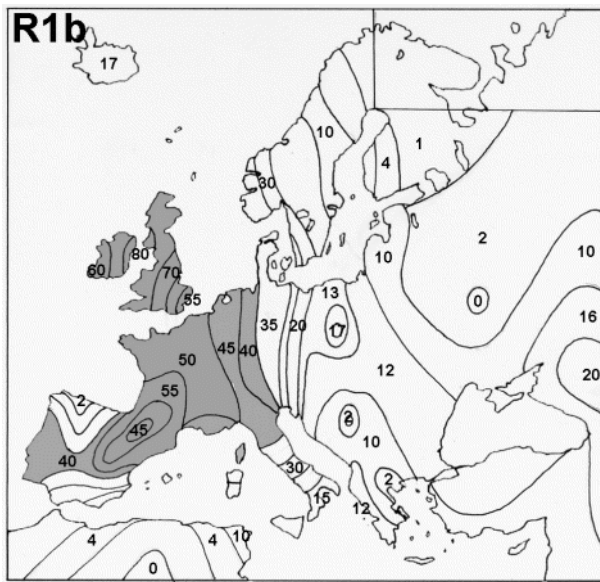
A. Europe as One Entity

Unfortunately, the two groups of maps (Wikipedia and Balanovsky) are not quite analogous in detail. In the Wikipedia maps, the isoglosses and zones of haplogroup frequencies are more regular but less reliable; in some cases, they seem to be caused more by the mathematical

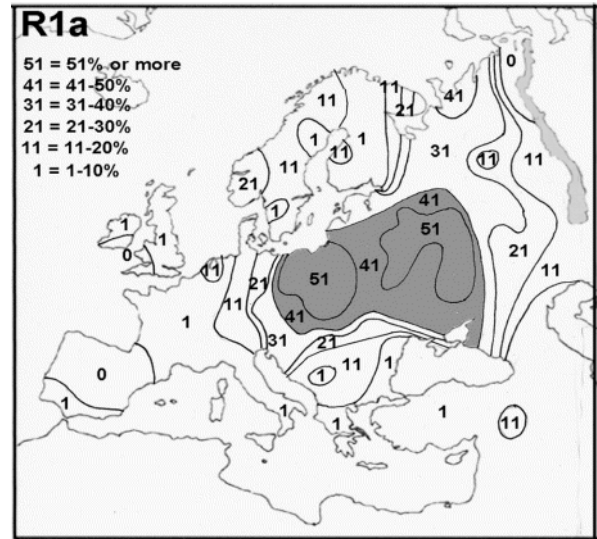
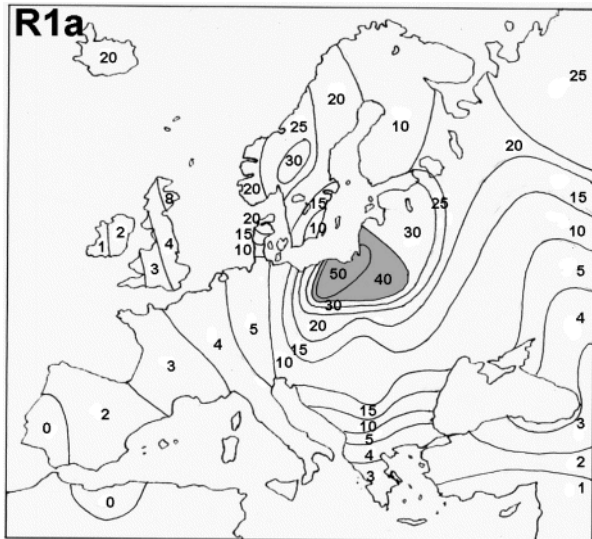
model used than by genetic reality. The maps representing approximate copies of those in Balanovsky et al. are different in that they are more reliably based on real genetic data. The maps are shown in pairs so that the map on the left is a Wikipedia map and that on the right a map from Balanovsky (2008). None of the maps are exact copies from the original sources but have been to some extent reshaped by the present author. To help the reader place each haplogroup in its proper ice age refuge, the main Haplogroups R1b, R1a, I, and N will be referred to respectively, as “Iberian,” “Ukrainian,” “Balkan,” and “Siberian” haplogroups.

B. Europe as Eight Separate Areas

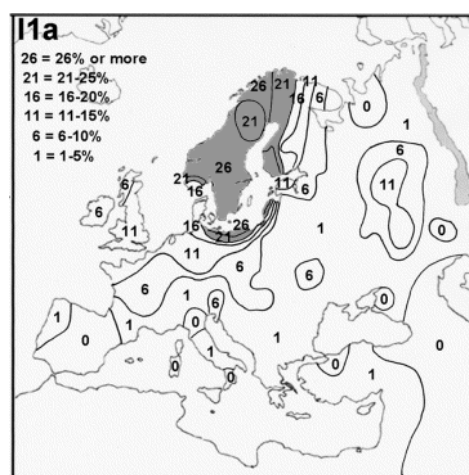
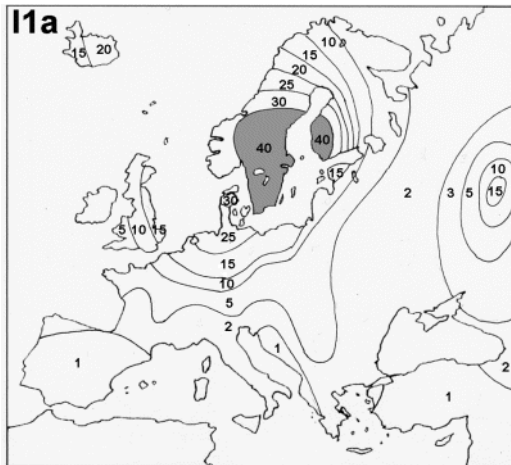
From this point Europe is treated as eight separate regions: (1) Central Europe, (2) North Europe, (3) East Europe, (4) the Balkans, (5) the Italian Peninsula, (6) Iberia, (7) the Atlantic Coast of Europe, and (8) the British Isles; see Map 35. The male populations of each region are mapped from the point of view of the following haplogroups: (1) R1b, (2) R1a, (3) I, (4) N, and (5) E3b+J2+G.



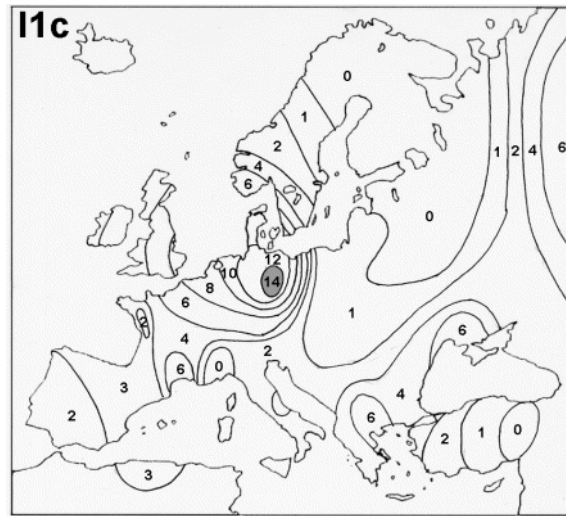
Maps 18-19 Frequencies of the Iberian Haplogroup R1b in Europe.



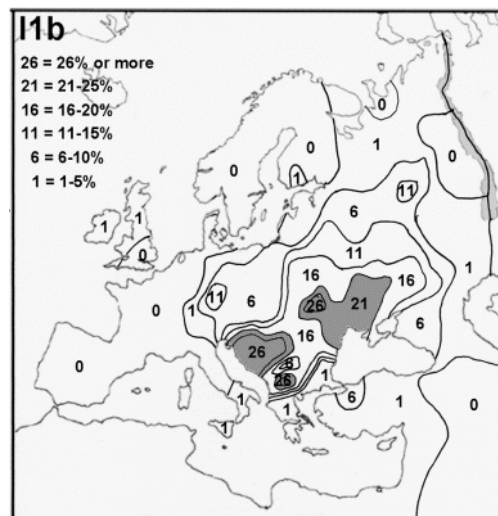
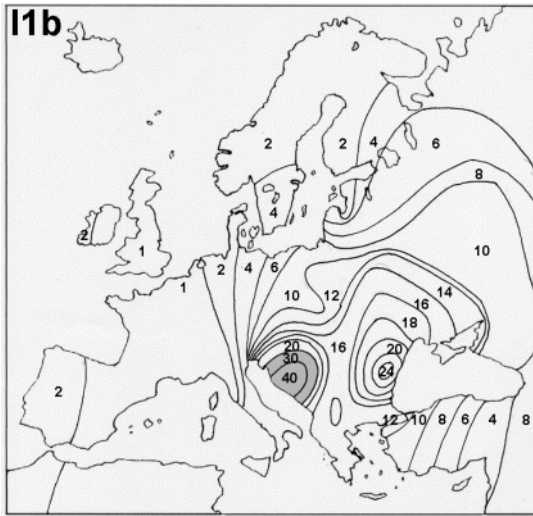
Maps 20-21 Frequencies of the Ukrainian Haplogroup R1a.



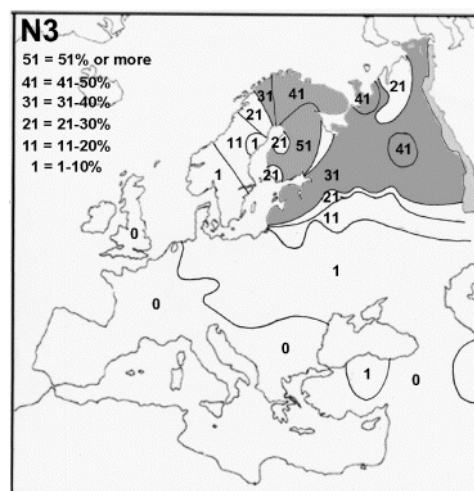
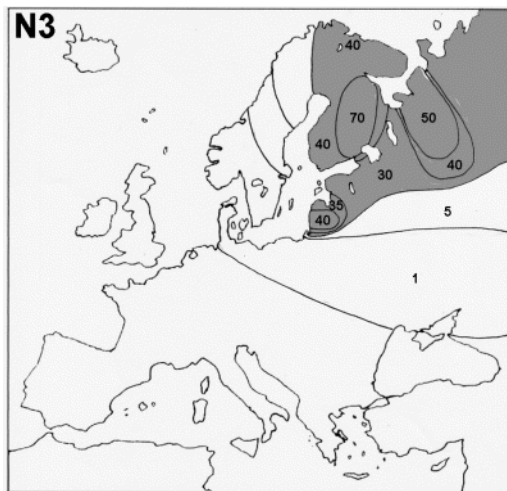
Maps 22-23 Frequencies of the Scandinavian Haplogroup I1a, a subhaplogroup of the Balkan Haplogroup I.



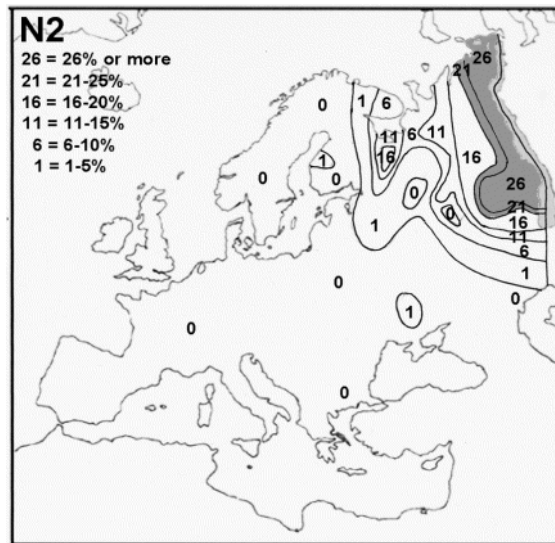
Map 24 Frequencies of the German Haplogroup I1b2-M223, subhaplogroup of the Balkan Haplogroup I.



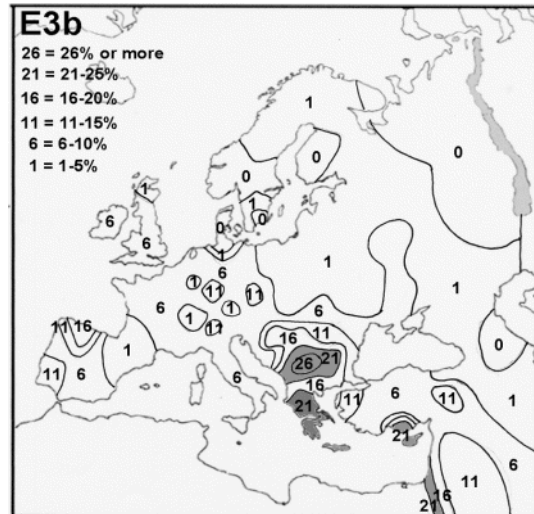
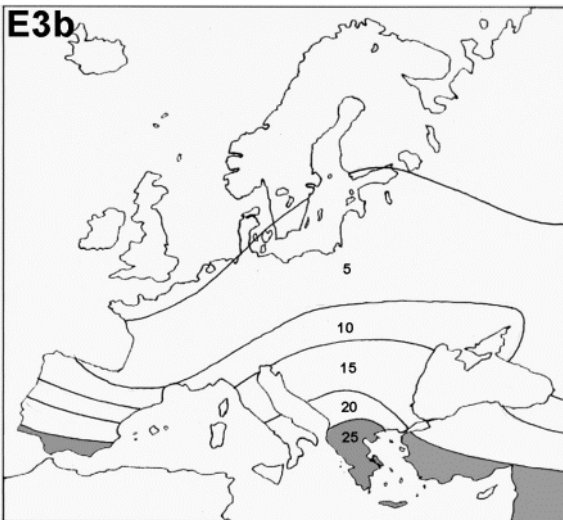
Map 25-26. Frequencies of the West Balkan Haplogroup I1b1-P37, subhaplogroup of the Balkan Haplogroup I.



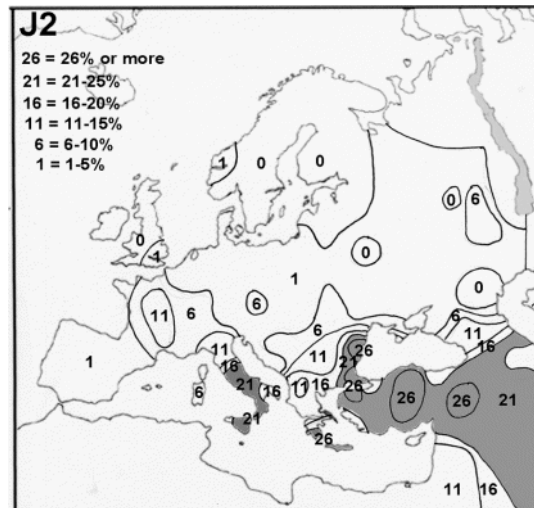
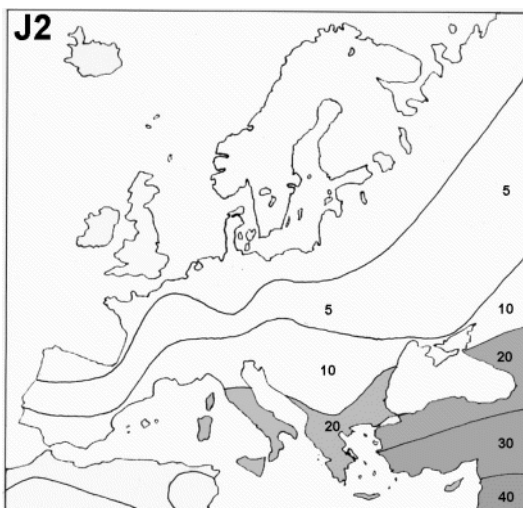
Maps 27-28 Frequencies of the Northeast European Haplogroup N3, subhaplogroup of the Siberian Haplogroup N. The N3 area extends to Siberia and reaches its peak values (about 85%) in the Yakuts near the Pacific Ocean.



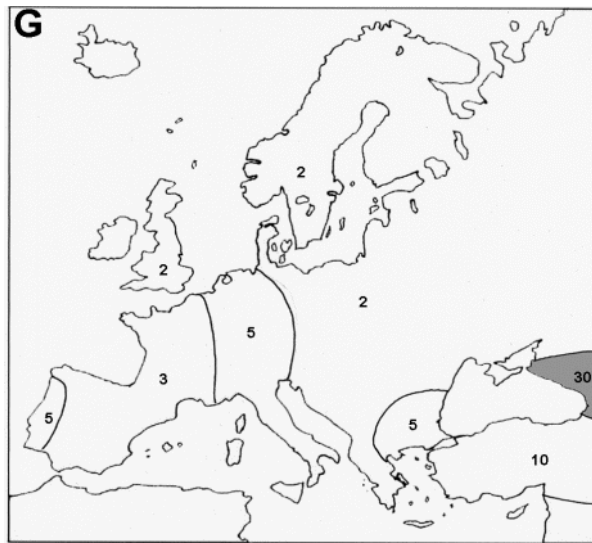
Map 29 Frequencies of the Samoyed Haplogroup N2, subhaplogroup of the Siberian Haplogroup N.



Maps 30-31 Frequencies of the Early Farmers Haplogroup E3b, subhaplogroup of the African Haplogroup E.



Maps 32-33 Frequencies of the Early Farmers Haplogroup J2, subhaplogroup of the Near Eastern Haplogroup J.

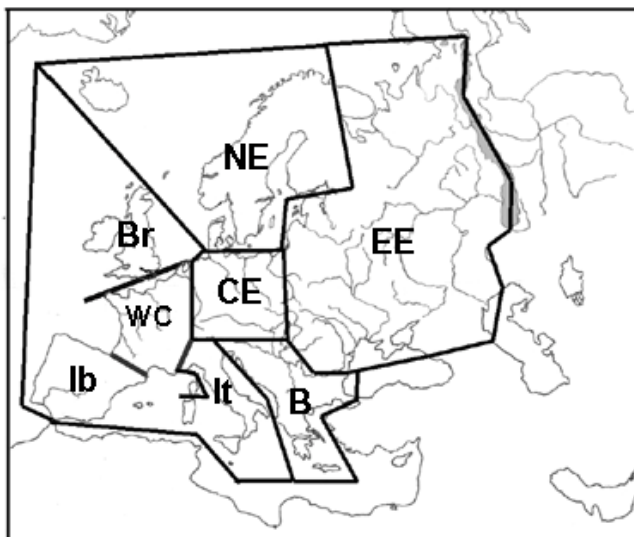


Map 34 Frequencies of the Early Farmers Haplogroup G in Europe.

1. Central Europe

Linguistically, Central Europe consists of a Germanic, Slavic, and Finno-Ugric area. The Germanic language is German (spoken in Germany, Austria and Switzerland); the West Slavic languages are Polish, Czech, and Slovak; the Finno-Ugric language is Hungarian. Slovenian is sometimes (not, however, in this study) included in Central Europe; it then represents a fourth language group, the South Slavic languages.

The total area of Central Europe is shown in Map 36, which also shows the areas of the seven Central European states, as well as the approximate locations of the populations involved in this study. Eleven populations are considered in Germany, five in the Czech Republic, eight in Poland, three in Austria, and two in Switzerland. The other Central-European states (Slovakia and Hungary) are represented only by one overall national population each, while Austria and Switzerland are represented by both an overall national population and two or three regional populations. The haplogroup frequencies of the Germans and Poles are from Kayser et al. (2005), and those of the Czechs from Luca et al. (2007).



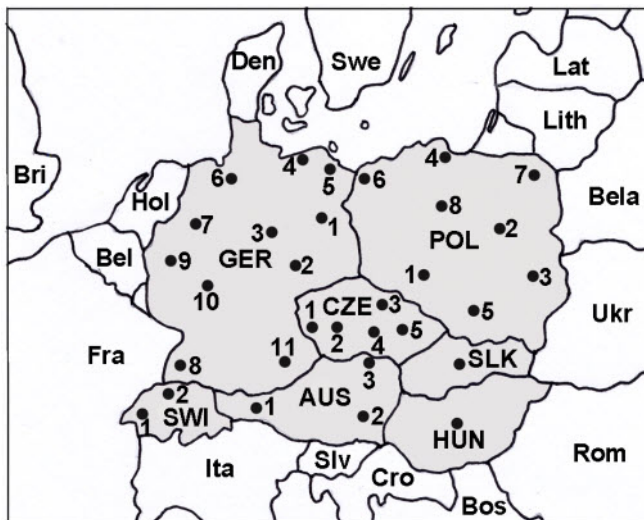
Map 35 The eight European subareas dealt with separately in this article. CE = Central Europe, NE = North Europe, EE = East Europe, B = the Balkans, It = the Italian peninsula, Ib = Iberia, WC = the Western Atlantic area, and Br = the British Isles.

Table 2 is a compilation of haplogroup frequencies for the different populations. The Polish, German, and Czech parts of the table are simplifications of the equivalent tables in Kayser et al. (2005) and Luca et al. (2007). The frequencies of the three Austrian populations and two Swiss populations were calculated by Kari Hauhio using YHRD-data processed through Whit Athey's haplogroup predictor (Athey, 2005; Athey, 2006). The subhaplogroup frequencies from the program were added to correspond to the major haplogroups shown in Table 2.

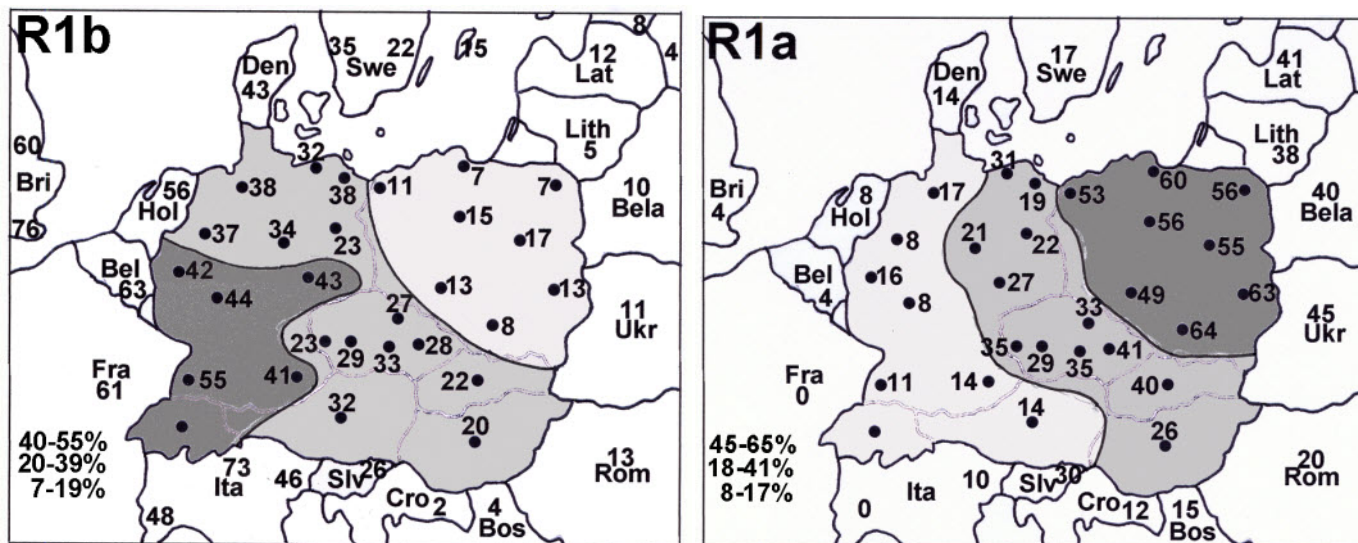
Maps 40 and 41 bear witness to the fact that agriculture arrived in Central Europe from two directions, the Balkans and Western Europe. Agriculture had first arrived from the Middle East and Anatolia in Greece and it continued from there as two branches: one came to Central Europe by land, while the other extended from Greece along the Mediterranean coast and the Atlantic coast of West Europe by boat.

Table 2
Haplogroup Frequencies of the Central European Populations.

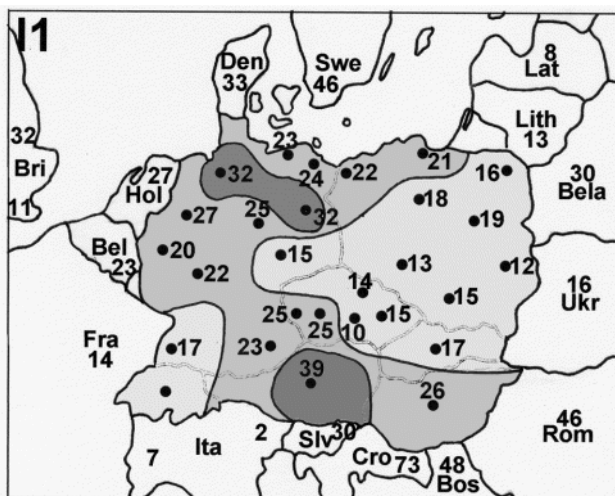
	Polish Population	R1b	R1a1	I	N3	OE	E3b	DE	J2	G	EF	F	P	K	Other	n
1	Wroclaw	12.9	48.5	12.9	5.0	79.3	11.9	0	2.0		13.9	5.9	1.0	0	0	101
2	Warsaw	17.4	54.5	19.0	1.7	92.6	2.5	0.8	3.3		6.6	0.8	0	0	0	121
3	Lublin	12.5	62.5	11.6	0.9	87.5	3.6	0.9	3.6		8.1	2.7	0	1.8	0	112
4	Gdansk	7.3	60.0	21.3	3.3	91.9	3.3	0	2.7		6.0	1.3	0	0.7	0	150
5	Krakow	8.0	64.0	15.0	4.0	91.0	3.0	0	2.0		5.0	2.0	2.0	0	0	100
6	Szczecin	11.4	53.3	21.9	3.8	90.4	6.7	0	1.9		8.6	1.0	0	0	0	105
7	Suwalki	7.3	56.1	15.9	11.0	90.3	2.4	0	2.4		4.8	3.7	0	1.2	0	82
8	Bydgoszcz	14.8	55.6	18.3	2.8	91.5	3.5	2.1	2.1		7.7	0	0	0.7	0	142
	Polish total	11.6	57.0	17.3	3.7	89.6	4.5	0.5	2.5		7.5	2.0	0.3	0.5	0	913
	German Population															
1	Berlin	23.3	22.3	32.0	1.9	79.5	9.7	0	1.9		11.6	3.9	0	3.9	1.0	103
2	Leizig	43.1	27.1	14.6	0.7	85.5	6.9	0	2.8		9.7	3.5	1.4	0	0	144
3	Magdeburg	34.0	21.0	25.0	1.0	81.0	7.0	0	2.0		9.0	6.0	3.0	1.0	0	100
4	Rostock	32.3	31.3	22.9	2.1	88.6	6.3	0	2.1		8.4	2.1	0	1.0	0	96
5	Greifswald	37.5	19.2	24.0	1.0	81.7	2.9	0	2.9		5.8	3.8	5.8	1.9	1.0	104
6	Hamburg	37.9	16.8	31.7	1.9	88.3	0	0.6	5.0		5.6	3.7	1.2	1.2	0	161
7	Muenster	37.3	7.8	26.5	1.0	72.6	9.8	0	4.9		14.7	7.8	1.0	2.9	1.0	102
8	Freiburg	54.9	10.8	16.7	0	82.4	4.9	0	8.8		13.7	2.9	1.0	0	0	102
9	Cologne	41.7	15.6	19.8	6.3	83.4	5.2	0	5.2		10.4	2.1	1.0	3.1	0	96
10	Mainz	44.2	8.4	22.1	1.1	75.8	11.6	2.1	6.3		20.0	3.2	0	0	1.1	95
11	Munich	41.1	14.3	23.2	0.9	79.5	7.1	0	2.7		9.8	8.0	0	2.7	0	112
	Ger. total	38.9	17.9	23.6	1.6	82.0	6.2	0.2	4.0		10.4	4.3	1.3	1.6	0.3	1215
	Czech Population															
1	Klatovy	22.9	35.4	25.1	0	83.4	4.2		2.1	8.4	14.7			0	2.1	48
2	Pisek	29.2	29.2	24.6	3.1	86.1	1.5		4.5	6.2	12.2			1.5	0	65
3	J.Hradec	26.5	32.7	14.3	2.0	75.5	8.2		2.0	6.1	16.3			0	8.1	49
4	Trebic	32.7	34.7	10.2	2.0	79.6	6.1		8.1	4.1	18.3			0	2.0	49
5	Brno	28.3	41.3	15.2	0	84.8	6.5		6.5	0	13.0			0	2.2	46
	Czech total	28.0	34.2	17.9	1.6	81.7	5.1		4.7	5.1	14.9			0.4	2.9	257
	Slovaks	22	40	17	3	82	10		3	0	13		0	2	0	
	Austrians	31.8	14.0	38.8	1.6	86.2	13.2		0		13.2		0	0	0.8	
	Hungarians	20	26	26		72.0	11		8	2	21.0			2		
	Austrian Population															
1	Tyrol	59.4	12.5	6.2		78.1	15.6			6.3	21.9					
2	Graz	14.3	42.9	28.6		85.8	4.7			9.5	14.2					
3	Wien	20.0		50.0		70.0			20.0							
	Austr. total	31.2	18.5	28.3		80.0					6.8					
	Swiss Population															
1	Lausanne	68.8	9.3	12.5		90.6			6.3	3.1	9.4					
2	Bern	32.1	21.4	32.2		85.7	7.1		7.1		14.2					
	Swiss total	50.5	15.4	22.5		88.2	3.6		6.7	1.6	11.8					



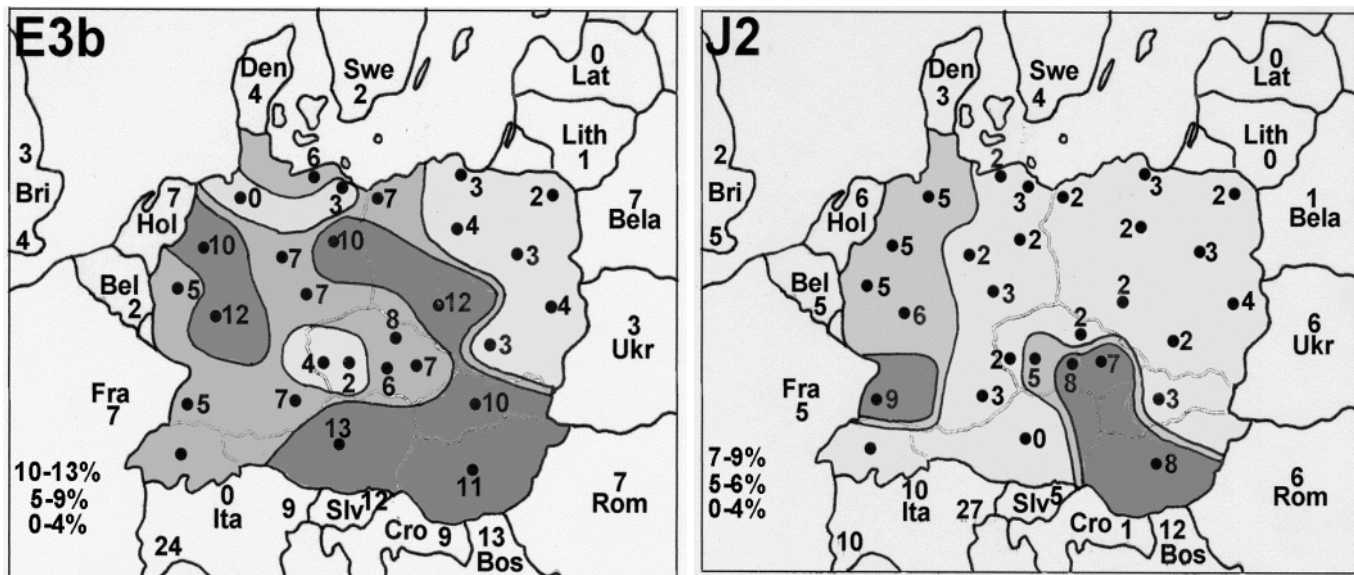
Map 36. The approximate locations of the Central European populations involved in this study.



Map 37 and 38 Frequencies of Haplogroups R1b and R1a in Central Europe.



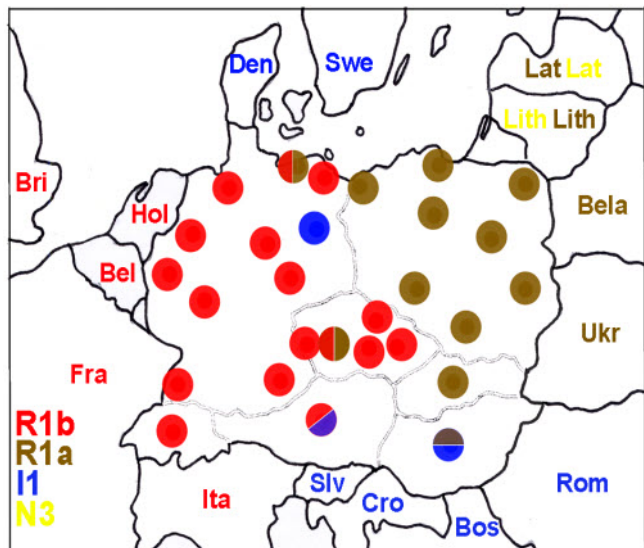
Map 39 Frequency of Haplogroup I in Central Europe.



Map 40 and 41 Frequencies of Haplogroups E3b and J2 in Central Europe.

Map 42 shows the most frequent haplogroup at various places in Central Europe.

Several conclusions can be drawn from the data on the distribution of haplogroup frequencies over the geography of Central Europe:



Map 42 Most frequent haplogroups in various parts of Central Europe. The colours of the four haplogroups are shown at the lower left corner of the map. The circles and country names with two colours represent cases with equal or almost equal frequencies of two haplogroups.

(1) The Iberian Haplogroup R1b forms a west-east gradient in Central Europe: the percentages are high (roughly 50%) in westernmost Germany and low (below 20%) in the east (eastern Poland and Hungary).

(2) The percentages of the Ukrainian Haplogroup R1a show an opposite tendency: They are low (less than 10%) in westernmost Central Europe and high (more than 50%) in Poland; they are slightly lower (about 40% or less) in the Czech Republic and Slovakia and still lower (less than 30%) in Hungary. In the R1b and R1a maps, the sum of the frequencies of the two haplogroups (R1b + R1a) is more or less constant; it is about 70% in Poland, about 60% in the Czech Republic and Slovakia, and about 40-50% in Austria and Hungary.

(3) The frequencies of the Balkan Haplogroup I in Central Europe are reflections of two I-centres, the North-German one (which represents subhaplogroup I1b2-M223, formerly I1c, and to some extent also the Scandinavian subhaplogroup I1aM-253) and the West-Balkan one (which represents subhaplogroup I1b1-P37). The former is situated within Central Europe and has values over 30% in northern Germany, while the reflections of the latter are seen in Central Europe as the relatively high I frequencies (over 20%) in Hungary. The Swiss and Austrian areas with I frequencies less than 10% result from their geographic locations relatively far from both I subgroup peak areas in question.

(4) The Siberian Haplogroup N3 is rare in Central Europe. Its average frequency is 3.7% in Poland, 3.0% Slovakia, 1.6% in Germany, Austria, and the Czech Republic, and 0-0.5% in Hungary.

(5) Central Europe has two separate centres for the Early Farmers' Haplogroups E+J+G (more precisely those of E3b, J2, and G2): The Hungarian centre with frequencies of about 20% is a reflection of the E+J+G centre in Greece where early farming first arrived from Anatolia and the Middle East. The Hungarian centre and its neighbouring areas in Slovakia and the Czech Republic represent the farmers of the Körös (6000-5500 BCE) and Linearbandkeramik (LBK) (4500-3900 BCE) cultures. The other Central European centre is in Holland. This area represents the other main branch of Early Farmers who expanded from Greece along the Mediterranean coast to the west and came to Central Europe along the Atlantic coast through France.

Central European men represent three main types:

(a) The "German type" came originally from two main sources, the Iberian refuge and the Balkan refuge; the number of those coming from the Iberian refuge is slightly higher (about 45%) than those coming from the Balkan refuge (about 40%); see Maps 18, 19, 22, 23, and 24 concerning Haplogroups R1b, I1a and I1b2-M223 (labeled as I1c on May 24). The number of those coming from the Ukrainian refuge (R1a) is lower (about 5%).

(b) The "West-Slavic type" came mostly (about 30-50%) from the Ukrainian refuge; a smaller portion of them (about 25%) came from the Balkan refuge and a still smaller portion (about 15%) from the Iberian refuge; see the R1a, I1a, and I1b1-P37 maps above.

(c) The "Hungarian type" is characterized by the fact that about equal numbers (about 25-30%) of these men arrived from the three refuges, the Iberian, Balkan, and Ukrainian refuges; see the maps concerning R1b, R1a, I1a, and I1b.

In addition, all three main types of the male populations in Central Europe received about 4-21% of their men from the three groups (E, J and G) representing the Early Farmers from the Middle East. The exact percentage of the farmers is dependent on the geographic location of the population in relation to the Middle East: The frequency is highest (21%) in the Hungarians.

As seen in the maps, the three genetic types represent also three linguistic groups. In this paper, I do not, however, deal in detail with the possible explanations for the emergence of the linguistic difference between the Germanic and Slavic languages. One possibility is the difference arose from the original linguistic differences between the Iberian and Balkan refuges, but there are other possible explanations (see, for example, Wiik 2002). The Hungarian language (unlike the majority of the genetically defined modern Hungarian people) arrived from the southern Ural area with the ancient

Hungarians of the Great Migration in 500-895 AD. According to this interpretation, the genome of the modern Hungarian men is typically Central European but their Finno-Ugric language is from the east.

The Czech Republic

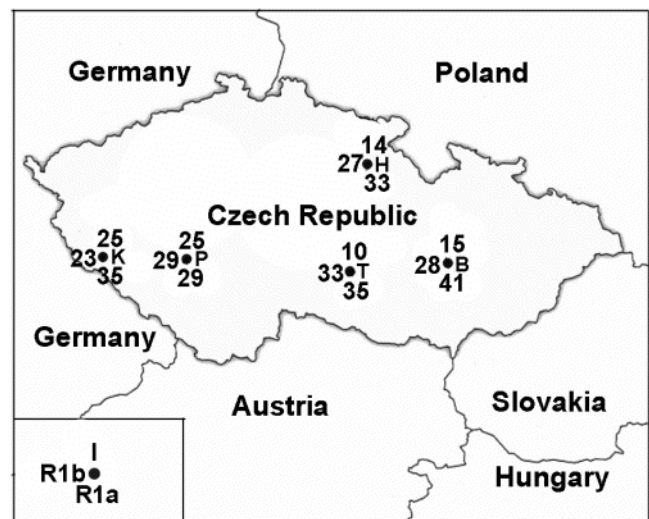
Though the Czechs are included in the treatment above, a more detailed analysis is in order. The haplogroups of the men of the Czech Republic are analysed thoroughly in Luca et al. (2005), and the frequencies are summarized in Map 43.

The sum total of the frequencies of the most common three haplogroups cover 74-84% of all haplogroups in the Czech Republic. The most common haplogroup is R1a (about 35%), second is R1b (about 28%), and third is I (about 18%). The Czech area is homogeneous; only two clinal gradients seem to exist: (a) In the southeastern corner of the country (Brno and Trebic), R1a seems to be somewhat more frequent than elsewhere. This is expected on the basis of the Central European centre of R1a in Poland and Slovakia. (b) The frequencies of I are higher in the west and lower in the east.

The northeastern (Siberian and linguistically Finno-Ugric) Haplogroup N3 has frequencies of about 2-3% in the Central areas of the Czech Republic; the average total for the entire country is below 2%.

2. North Europe

By "North Europe" I mean here the area consisting of the Scandinavian peninsula (Norway and Sweden), Den-



Map 43. Frequencies of three haplogroups, R1b, R1a, and I, in five Czech localities: Klatovy (K), Pisek (P), J.Hradec (H), Trebic (T), and Brno (B). The relative positions of the figures for the three haplogroups are shown in the lower left corner.

mark, Iceland, Finland, and Karelia. The continental part of this area is often called Fenno-Scandia.

Linguistically, the people of this area represent two language phyla, Indo-European and Finno-Ugric. The IE languages are represented by the Germanic, more precisely the North Germanic or Scandinavian languages, and the FU languages by two main branches of the “Early Proto-Finnic” languages, more precisely two Finnic languages (Finnish and Karelian) and Saami. The four Scandinavian languages, Icelandic, Danish, Norwegian, and Swedish are spoken in their respective countries. Swedish is spoken, in addition, on the Borthian and Newland (Uusimaa) coasts of Finland as well as in the archipelago between Finland and Sweden. Finnish is spoken outside Finland in northern Sweden and Norway. The Karelian language is spoken only by a minority in the Russian Republic of Karelia.

Genetically the North European male population (like many other European populations) is concise in the

sense that about 96-98% of its Y-chromosomes are members of just four haplogroups: R1b, R1a, I, and N3. Each of these is from a different Ice-Age refuge, which means that the male populations of North Europe came originally from the Iberian, Balkan, Ukrainian, and Siberian refuges. Therefore, they represent the Old Europeans (not the Early Farmers). The percentages of the individual haplogroups are, however, quite different in various parts of North Europe, which makes it possible to draw conclusions about where the North Europeans originally came from. The overall frequencies of the haplogroups in the seven North-European male populations are seen Table 3. Next, the main four haplogroups of Northern Europe will be considered separately.

Haplogroup R1b in Northern Europe

The frequencies of the “Iberian” Haplogroup R1b (generally thought to be the oldest in Europe) are seen in Map 44. The group is typically West European; its highest frequencies (about 80-90%) are found in Ireland

Table 3
Frequencies of the Haplogroups of Y-chromosome DNA of Seven North-European Populations, By Language Group. The table is based on Helgason (2001), Rosser (2000), Sanches (2003), Rootsi (2004), Passarino (2002), Semino (2000), Raitio (2001), and Wells (2001).

	Population	R1b	R1a	I1a	N3	OE	Eb3	J2	G2	EF	P	K
1	Icelandic	43.7	22.4	33.1	0.3	99.5	0	0	0	0	0	0
2	Danish	43.2	13.7	34.2	0.5	91.6	4.3	3.2	0.2	7.7	0.4	0
3	Nowegian	27.5	24.6	41.4	3.9	97.4	2.2	1.3	0	3.5	0	0
4	Swedish	22.4	17.0	47.9	8.2	95.5	1.6	1.4	0.4	3.4	0.9	0.6
	NG* Total	34.2	19.4	39.2	3.2	96.0	2.0	1.5	0.2	3.7	0.3	0.2
5	Saami	6.8	14.2	33.8	42.6	97.4	0	0	0	0	0	0
6	Finnish	1.4	8.9	25.7	61.2	97.2	0.5	0	0	0.5	1.3	1.0
7	Karelian	1.3	40.7	20.5	38.0	100	0	0	0	0	0	0
	FU* total	3.2	21.3	26.7	47.3	98.4	0.2	0	0	0.2	1.4	0.3

* NG = North German Language Group; FU = Finno-Ugric Language Group.

Table 4
Frequencies of Eleven Haplogroups in Seven Swedish Populations. The table is based on Karlsson (2006)

Population	R1b	R1a	I1a-M253	I1b2-M223	N3	OE	E3b	J	G	EF	F	P	K	n
Gotland	17.5	12.5	45.0	5.0	10.0	90.0	1.8	0	0	0	0	5.0	5.0	40
Uppsala	25.4	12.7	38.2	0	14.6	90.9	0	5.5	0	5.5	0	1.8	1.8	55
Österg./Jönköp.	31.7	14.6	29.3	2.4	7.3	85.4	2.4	9.8	0	12.2	0	2.4	0	41
Blek./Kristians.	21.9	17.1	41.5	4.9	2.4	87.8	0	0.0	4.9	4.9	2.4	2.4	2.4	41
Skaraborg	40.0	8.9	35.6	2.2	2.2	88.9	2.2	4.4	0	4.4	0	6.7	0	45
Värmland	16.7	14.3	45.2	4.8	9.5	90.5	0	4.8	2.4	7.1	0	2.4	0	42
Västerbotten	17.1	9.8	24.4	17.1	19.5	87.9	2.4	0	4.9	4.9	2.4	2.4	2.4	41
Sweden total	24.6	11.8	37.0	4.9	9.5	88.5	1.2	3.6	1.6	6.4	0.7	3.3	1.6	305

F = F*(xG,I,J,K), K = K*(xN3,P), P = P*(xR1a,R1b3).

Table 5
Frequencies of Four Haplogroups in Eight Norwegian Populations. The table is based on a map in Dupuy et al. (2005).

Norway Area	R1b	R1a	I	N3	Other
North	26.8	27.1	34.7	10.6	0.8
Middle	27.1	31.5	39.7	~ 1-2	
West	43.2	24.3	30.2	~ 1-2	
East	26.4	26.8	41.4	3.2	2.2
Sogn/Fjord.	43.2	24.3	30.2		2.3
Bergen	35.5	28.0	33.3	~ 2	~ 1
Oslo	35.8	19.3	40.4	~ 4	
South	44.7	13.2	42.1		
Norway Total	35.3	24.3	36.5	2.9	1.0

Table 6
Frequencies of the Haplogroups in Nine Finnish Populations. The table is based on Lappalainen et al. (2006).

Population	R1b	R1a1	I1a	I1c	N3	N2	Other	n
Southern Karelia	0	8.3	16.7	2.0	70.8	0	2.1	48
Northern Karelia	4.5	4.5	18.2	0	68.2	0	4.5	22
Northern Savo	0.9	4.7	14.9	2.0	78.5	0	0	107
Northern Ostrobothnia	4.6	6.2	23.3	0	65.1	0	0.8	129
East Finland Total	2.6	5.9	18.9	0.6	70.9	0	1.9	306
Southern Ostrobothnia	3.4	19.0	46.6	0	25.9	3.4	1.7	58
Swedish-Speaking Ostrobothnia	8.0	12.0	36.0	0	40.0	0	4.0	25
Satakunta	6.2	8.3	52.1	2.0	27.1	0	4.2	48
Häme	6.1	2.0	34.7	0	55.1	0	2.0	49
South-Western Finland	4.0	2.0	28.0	2.0	60.0	0	4.0	50
West Finland Total	5.2	8.7	40.0	0.9	41.3	0.9	3.2	230
Finland Total	3.7	7.1	28.0	0.7	58.2	0.4	2.5	536

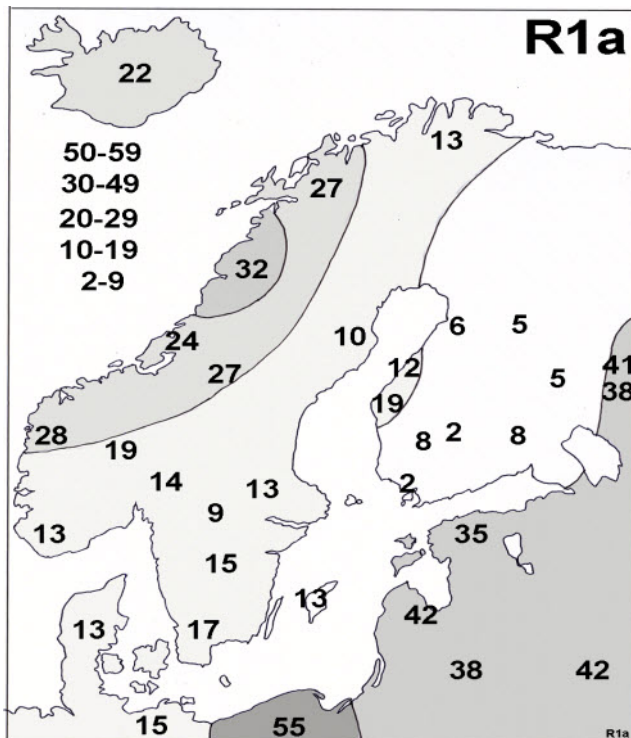
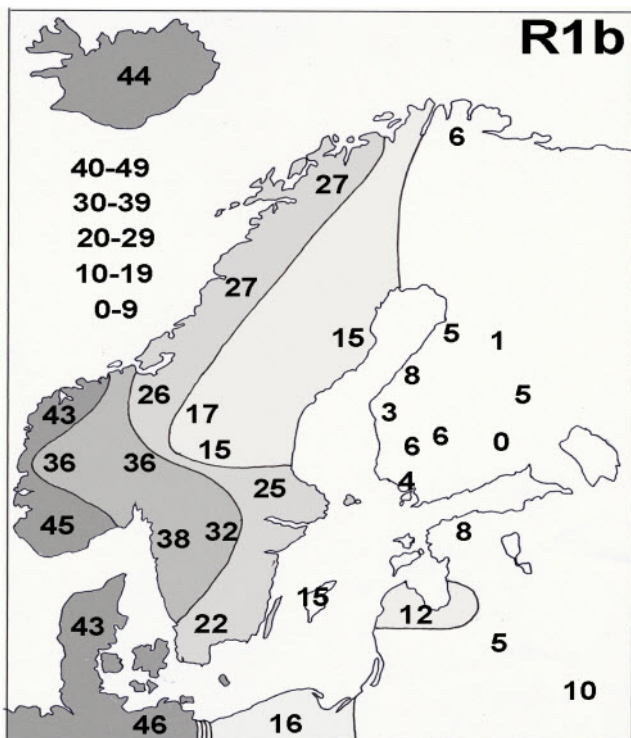
and on the whole in the western parts of the British Isles. In Germany, its percentage is slightly less than 50%, and about the same percentages (43-45%) are also found in the westernmost areas of North Europe, Denmark, western Norway and Iceland. The other extreme in North Europe is represented by the Finns and Saami whose R1b percentages are only 0-8% (average about 4%). Between the two extremes, there are three intermediate zones in the map: (a) In southwestern Scandinavia, the percentages are 32-38%, (b) in northern Norway and southeastern Sweden 22-27%, and (c) in northern Sweden and Gotland 15-17%. Thus, the R1b percentages form a west-east gradient according to which the percentage descends from about 45% to zero from Denmark and Southern Norway to Eastern Finland.

The gradient is a reflection of the migrations from West Europe (the Atlantic Coast) and ultimately from the Iberian refuge to Scandinavia. The migrations were part

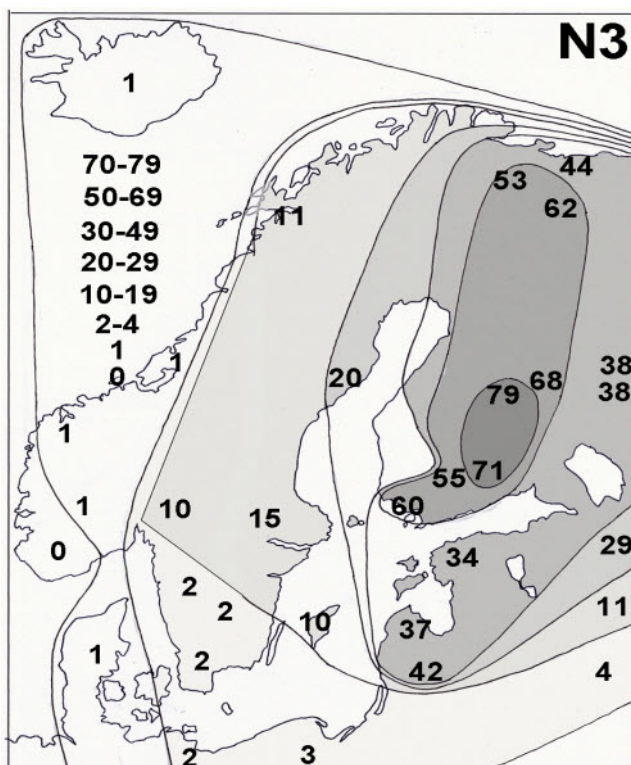
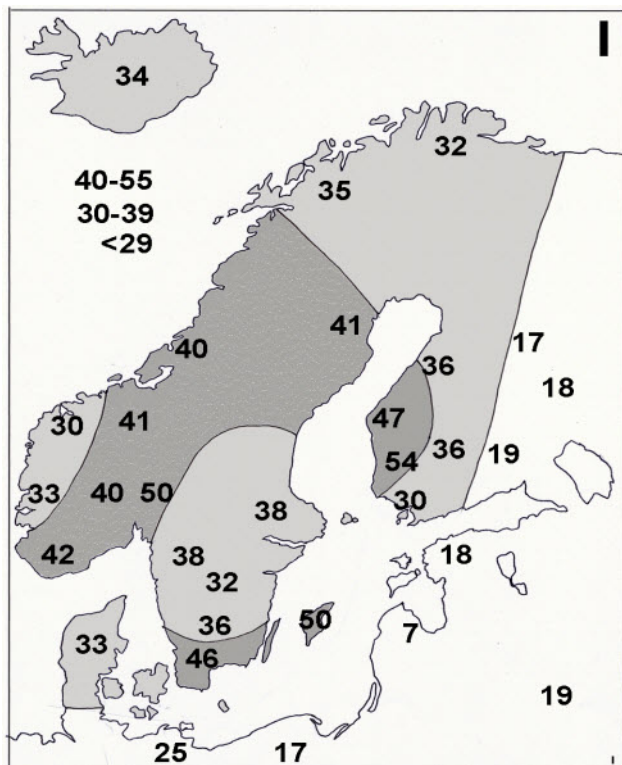
of the Recolonization of Northern Europe that started after the Late Glacial Maximum and has continued in many phases after that.

Haplogroup R1a in Northern Europe

The “Ukrainian” Haplogroup R1a shows, somewhat surprisingly, a west-east gradient in North Europe: The frequency is (a) highest (32%) in west-central Norway, (b) slightly lower (about 24-28%) in many other parts of Norway and in Iceland, (c) about 10-19% in southern and northern Norway, Denmark, Sweden, and Finnish Bothnia, and (d) only about 2-8% in the other parts of Finland. This frequency distribution may seem surprising because the European peak area (frequency about 55%) is in Poland. The frequency distribution makes one believe that there has been a movement of R1a men from Central Europe to the Central-Norwegian coast. This expansion represents the Ahrensburgian culture



Maps 44-45. Frequencies of Haplogroups R1b and R1a in North Europe and its vicinity. The North European parts of the maps are based on Tables 4, 5, and 6.



Maps 46-47. Frequencies of Haplogroups I and N3 in North Europe and its vicinity. The North European parts of the maps are based on Tables 4, 5, and 6.

(about 8500 BC), perhaps also the Hamburgian culture (about 15-13.7 kya) of the northern parts of Central Europe (cf. Saukkonen 2006, p. 72). The route of this expansion was western (through Denmark and southern Scandinavia) rather than eastern (through Balticum and Finland): this is shown by the fact that R1a-frequencies are low in Finland. As a matter of fact, Finland is very much like a vacuum in this respect in North Europe: In Karelia and the Baltic countries (Estonia, Latvia, and Lithuania), R1a-frequencies are of the order of 35-42 (i.e. even higher than on the central Norwegian coast).

Haplogroup I in Northern Europe

Haplogroup I is generally thought to have spread to its modern areas from the Balkan refuge. The haplogroup has many subgroups. The most frequent of these in North Europe is I1a; Haplogroup I1b2-M223 (formerly I1c) is common in North Germany and Haplogroup I1b1-P37 (formerly I1b) in the western Balkans. As seen in Map 46, Haplogroup I (most of which consists of Haplogroup I1a-M253 in North Europe) is common over almost all of North Europe: frequencies of 30-50% are found almost everywhere in North Europe; the only exception is eastern Finland (and evidently also Karelia) where the frequencies are below 20%; there is a rather steep west-east gradient in Finland, the frequencies being about 50-40% in the west and below 20% in the east. As seen from the frequencies of Germany (25%) and Poland (17%), as well as those in East Europe (7-19%), North Europe forms an independent island of Haplogroup I.

In Finland, Satakunta is exceptional in having a frequency as high as 52% for Haplogroup I. In Sweden also, there is one rather exceptional area: The “German” Haplogroup I1b2-M223 is as high as about 14% in Västerbotten. The average total of this haplogroup is below 5% in Sweden as a whole.

Haplogroup N3 in Northern Europe

As seen in Map 47, Haplogroup N3 is typically eastern. Its total area extends as far east as the Pacific Ocean, and it has very high frequencies (85%) in the Yakuts in northeastern Siberia. In North Europe, Haplogroup N3 is commonest in Eastern Finland (71-78%). The percentage diminishes with geographic distance outside Eastern Finland and the percentages are lower (53-68%) elsewhere in Finland. The percentage is slightly lower in the East-Karelians and Vepsians (38%), as well as in the Estonians, Latvians, and Lithuanians (34-42%). In Scandinavia, there is a northeast-southwest gradient, the N3-percentages being about 10-15% in the northeastern and western Scandinavia and very low in the south and west (in Southern Norway it reaches zero). To the south of North Europe the N3-percentage is very low: according to the map, 2% in Germany, 3% in Poland, and 4%

in Belarus. In Russia, there is a south-north gradient (cf. Maps 27-28).

Linguistically, the N3 men are generally thought to represent the speakers of Finno-Ugric languages (in western Siberia; they represent the speakers of Altaic/Turkic languages).

The Haplogroups E3b, J2 and G of the Early Farmers, occur in Scandinavia, while among the Finns, Karelians, and Saami, these haplogroups are practically non-existent. The sum total of the frequencies of these Haplogroups (E3b+J2+G) is highest (4.3+3.2+0.2 = 7.7%) among the Danes and lower among the Norwegians (2.2+1.3+0 = 3.5%) and Swedes (1.6+1.4+0.4 = 3.4%). The Icelanders do not have these haplogroups.

3. East Europe

By “East Europe” is meant here the geographic area covered by most of the European parts of Russia, the Ukraine, Belarus, Romania, Moldova, Lithuania, Latvia, and Estonia. The area is linguistically heterogeneous in that it includes languages of four language phyla: (1) Indo-European, (2) Finno-Ugric (Uralic), (3) Altaic, and (4) North Caucasian. The Indo-European languages belong to East Slavic, Romance and Iranian groups. The individual East Slavic languages are Russian, Belarussian, and Ukrainian. The Finno-Ugric (Uralic) language groups are Finnic (e.g. Estonian), Volgaic (Mordvian and Mari), Permic (Udmurtian and Komi), Ugric (Hungarian) and Samoyed (Nenets). The Altaic languages belong to the Turkic group (Turkish, Tatar, Chuvash, and Bashkirian). The North Caucasian languages are represented, for example, by Chechenian.

Haplogroup R1b in Eastern Europe

The frequency of Haplogroup R1b is very low (below 10%) in Russia, but it rises to about 20-40% in some parts of the Caucasus (cf. the Wikipedia and Balanovsky maps and Map 48). In Map 48 East Europe is seen to consist of three separate areas (cf. the three shades of gray):

(a) The southeastern corner (Baskirs and Ossetians) has R1b values (43-47%) of the “West European” type. The explanation is that a considerable portion of the R1b or R1 men first arriving in eastern Europe from Asia about 40 kya stayed in the steppe and mountainous areas around the Caspian Sea and the Caucasus.

(b) The intermediate zone with R1b frequencies of 10-19% is situated in the southern and eastern parts of East Europe and consists of the following populations: Nenets (19%), Komi (16%), Chuvash (12%), Mordvians (13%), Ukrainians (11%), Belarussians (10%), Lat-

Table 7

Frequencies of the Haplogroups of Y-chromosome DNA in Twenty-Three East-European Populations. "Other" = C, F*, N*, Q, R2. Sources: Rootsi (2002), Rosser (2000), Laitinen (2002), Balanovsky (2008), and Nasidze (2004).

Population	R1b	R1a	I	N3	N2	OE	E3b	DE	J	G	EF	K	other
Estonians	7.9	31.9	20.1	33.9	0	93.8	1.8	0.5	0.7		3.0	3.5	
Latvians	11.8	40.5	8.4	37.8		98.5	0.2		0		0.2		0.2?
Lithuanians	4.5	38.3	13.3	42.2		98.3	1.2		0	0	1.2	0.2	
Belarus a	9.9	40.0	29.7	2.8		82.4	6.7	3.3	1.3		11.3	1.5	
Ukrainians	10.7	45.4	16.0	7.6		79.7	3.1		6.3		9.4		9.8?
Romanians	13.0	20.4	46.3	0	0	79.7	7.4		5.7	5.6	18.7		
Moldovians	16.7	27.6	30.9	1.9	0.7	77.8	12.8		7.7	1.0	21.5	1.0	1.0
Gagauzes	12.5	19.7	27.9	2.1	0	62.2	13.3		5.9	13.8	33.0	3.2	
Ossetians a		4.0	32.0			36.0			29.0	21.0	50.0	7.0	8.0
Ossetians b			13.0			13.0			3.0	74.0	77.0		10.0
N. Russians	5.4	34.2	13.1	35.5	7.5	95.7	0.2		1.8	1.2	3.2	0	1.5
C. Russians	7.5	46.5	15.3	16.3	0.5	86.1	5.0		3.4	0	8.4	1.7	1.7
S. Russians	4.8	55.4	21.0	9.5	0.5	91.2	1.8		3.5	1.0	6.3	1.4	1.1
Mordvians	13.3	26.5	19.2	16.9	2.4	78.3	?	?	?	?	?	?	?
Mari	5.0	21.0	4.2	51.3	0	81.5			6.3		6.3	11.8	
Udmurts	9.2	30.5	3.5	36.6	14.4	94.2	2.4		2.4		4.8	2.3	
Komi	16.0	23.7	4.5	36.2	24.3	104?	0		0		0		0
Hanti	19.1	4.2		38.2	38.2	99.7					0		
Nenets, for	0	0	0	51.7	44.9	96.6	0	0	0	0	0	0	3.4
Nenets, tun	0	0	0	23.7	74.6	98.3	0	0	0	0	0	0	1.7
Tatars, Kaz	6	29	16	25	0	76.0	2		8		10.0	7.0	6.0
Chuvash	12	18	24	18	0	72.0	6		6		12.0	18.0	
Bashkirs	47	26	?	17	0	90	?	?	?	?		?	10?

vians (12%), Moldovians (17%), Gagauzes (13%), and Romanians (13%).

(c) The R1b values are very low (1-9%) in the zone to the northwest of the intermediate zone. This zone consists of the following populations: Northern, Central and Southern Russians (5%, 8% and 5% respectively), Udmurts (9%), Mari (5%), Lithuanians (3%), and Estonians (8%), as well as the North European populations Finns (1%) and Saami (7%).

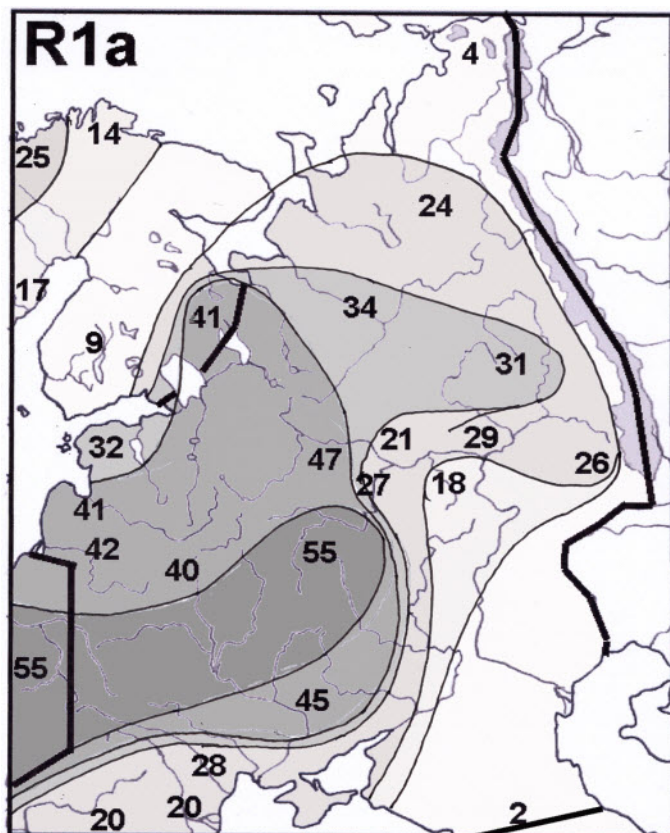
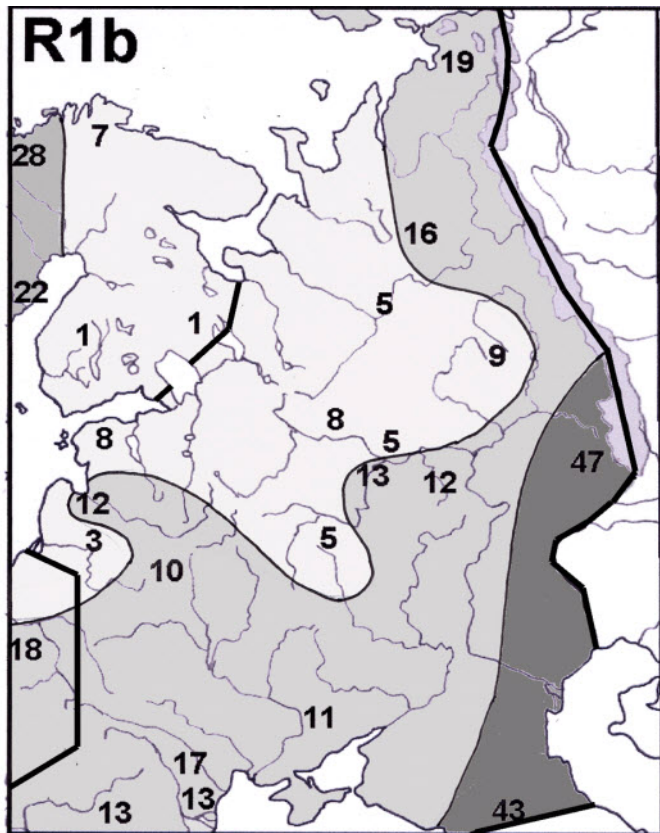
Haplogroup R1a in Eastern Europe

The frequencies of Haplogroup R1a are high (about 30-50 %) in the following East-European populations: Estonians, Latvians, Lithuanians, Belarussians, Ukrainians, Russians, and Tatars. These are evidently the populations that (at least partly) used the Ukrainian refuge during the Ice Age. This means that the Estonians, Latvians, Lithuanians, and North Russians are

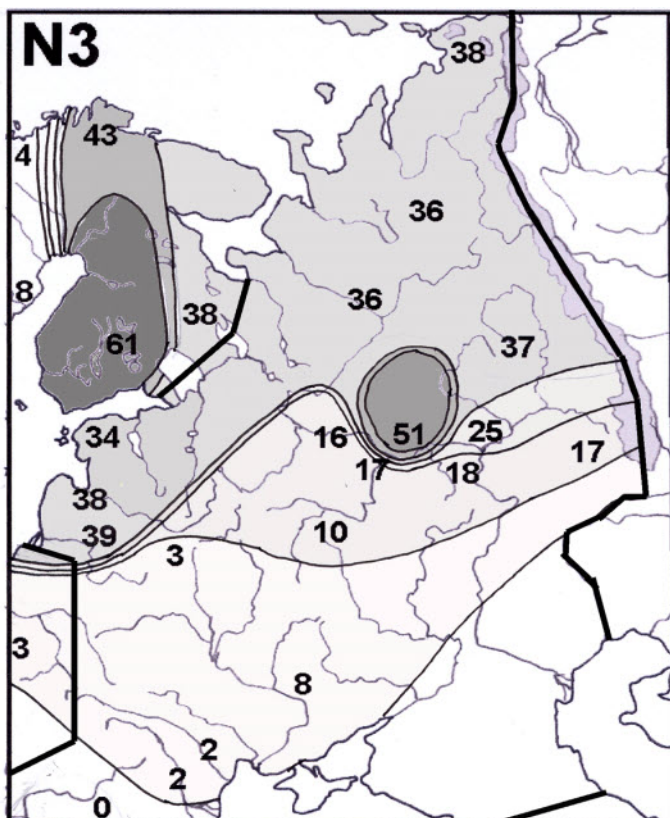
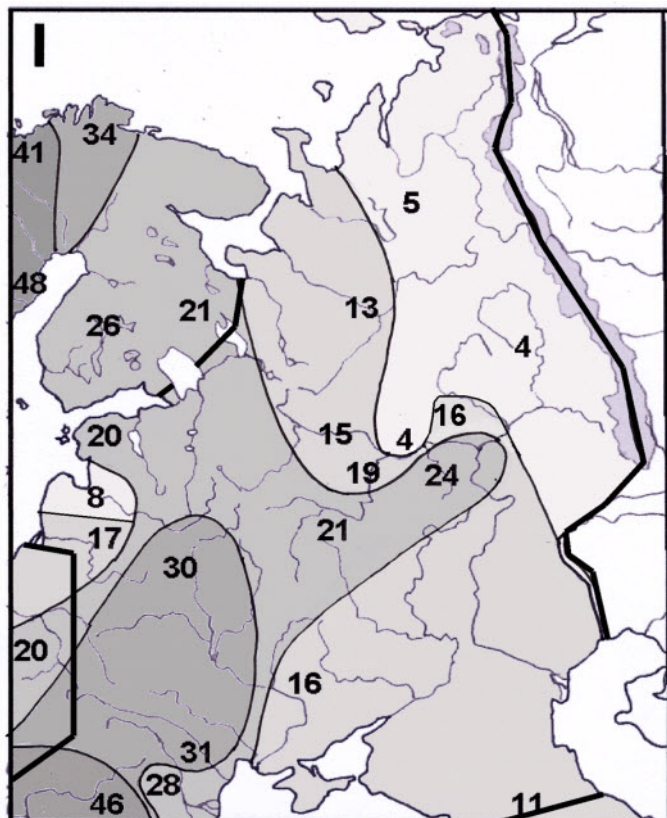
mixed populations in the sense that about half of their men represent the ancient mammoth hunters from the Siberian refuge (representing Haplogroup N3) and another half from the Ukrainian refuge (representing Haplogroup R1a).

According to the Wikipedia and Balanovsky R1a maps, the European maximum area (with frequencies over 50%) of Haplogroup R1a is in Poland, and the area of almost equally high frequencies (over 40%) extends to Belarus. In East Europe, there is a west-east gradient and the frequencies descend to about 20-30% in northern Russia and close to zero in the southeastern corner (the Caucasus) of Europe.

According to Map 49, the "Polish" maximum area actually extends to southern Russia (55%) and there is a zone with frequencies over 40% in practicality all the eastern parts of East Europe. This zone contains in addition to the North European Karelians (41%), the



Maps 48-49. Frequencies of Haplogroups R1b and R1a in East Europe. The maps are based on Table 7.



Maps 50-51. Frequencies of Haplogroups I and N3 in East Europe. The maps are based on Table 7.

following East European populations: Latvians (41%), Lithuanians (42%), Belarussians (40%), Central Russians (47%), and Ukrainians (45%). The next zone, with frequencies of 30-39% consists of the Estonians (32%), Northern Russians (34%), and Udmurtians (31%). Still further from the peak area, with frequencies of 20-29%, are the populations of the Komi (24%), Bashkirs (26%), Tatars (29%), Mari (21%), Mordvians (27%), Moldovians (28%), Gagauzes (20%), and Romanians (20%). The R1a frequency is somewhat lower (10-19 %) in the Chuvash population, and still lower (4%) in the Nenets in the northeastern corner of Europe. Also the Finns (considered here to represent North, rather than East Europeans) have a very low R1a frequency (9%). The low value of the Finns tends to show that the route of the R1a men of Norway (25% of Norwegians in the map) followed the western route from Poland to Denmark and finally to Norway, and not to the same extent the eastern route from East Karelia through Lapland to Norway. Supporting this hypothesis is the frequency for the Saami, which is relatively low at 7%.

Haplogroup I in Eastern Europe

All of the maps for Haplogroup I and its subgroups show that the most common subgroup of Haplogroup I in the Balkans is Haplogroup I1b1-P37, which has a maximum frequency of about 40% and even more in some parts of the Western Balkans. This subhaplogroup has a secondary maximum in Romania (maximum frequency about 25-30%). It is particularly the Romanian maximum that (according to the maps used) has influenced the East European populations. There is in East Europe a south-north gradient (from about 24-30% to 4%) extending from Romania to the Nenets territory in the northeastern corner of Europe.

The other two subhaplogroups of I are less frequent in East Europe. The maximum area of I1a-M253 is in southern Scandinavia and that of I1b2-M223 is in Germany. Both have, however, a secondary maximum in the eastern parts of East Europe. I1a reaches as high values as 15% in the Volga area, and I1b2 reaches the frequency of 6% in the Ural area; see the Wikipedia and Balanovsky I1a maps.

In Map 50 all three subhaplogroups of I are put together. It shows the Romanian maximum area (46%) and outside it the zone of about 30% in Moldova (31%) and Belarus (30%). The next zone with frequencies of 20-29% is further away from Romania: Moldovians (28%), southern Russians (21%), Chuvash (24%), and Estonians (20%); also the Karelians (21%) and Finns (26%) in North Europe belong to this zone. The next zone has frequencies 10-19%: Ukrainians (16%), Ossetians (11%), Mordvians (19%), Central Russians (15%), and Northern Russians (13%). Haplogroup I is

rare in the easternmost populations: the frequency is 5% in the Komi and 4% in the Udmurts.

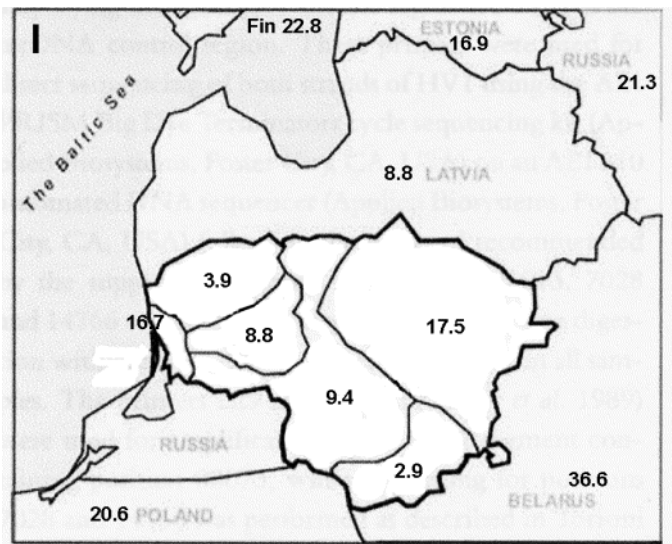
Regarding the I frequencies of the East Europeans, one should keep in mind that these populations have not received their Haplogroup I men only from the I1b1-P37 maxima in the Balkans and Romania. Particularly in the northwestern parts (e.g. in Estonia and Latvia), a majority of the representatives of Haplogroup I have come from the peak areas of Haplogroups I1a and I1b2 in southern Scandinavia and Germany. So, for example, about 82% of the Estonian men of Haplogroup I belong to the "Scandinavian" subgroup I1a and only about 2% of them belong to the "German" Haplogroup I1b2-M223; about 16% of the Estonian men belong to the "Balkan" Haplogroup I1b1-P37; see Table 8. These figures tend to show that about 82+2 = 84% of the Estonian men of this haplogroup came from the west and only about 16% from the south. The equivalent figures for the Latvians are almost identical: 67+17 = 84% from the west and 16% from the south. Quite different ratios are shown by the Romanians: about 8% of the Romanian men of Haplogroup I represent the

Table 8
Frequencies of Subhaplogroups I1a-M253, I1b1-P37, and I1b2-M223 in Thirteen European Populations. The Finnish Values are from Lappalainen et al. (2006); the other values are from Rootsi et al. (2004a).

Population	I1a-M253	I1b1-P37	I1b2-M223	Total Hapl. I
Estonians	14.8	2.9	0.5	18.2
Latvians	4.7	1.1	1.2	7.0
Swedes, South	55.7	0	3.6	59.3
Swedes, North	26.8	0	0	26.3
Norwegians	38.9	0	1.4	40.3
Finns, West	40.0	0.4	0.9	41.3
Finns, East	19.0	0	0.7	19.6
Saami	28.6	0	0	28.6
Germans	25.0	0	12.5	37.5
Dutch	16.7	0	10.0	26.7
Romanians	1.7	17.7	1.9	21.3
Moldaovians	3.3	21.7	3.3	28.3
Gagauzes	2.5	24.1	5.1	31.7

Table 8 gives the frequencies of the commonest three subhaplogroups of I in eleven European areas (Rootsi et al. 2004a). The figures show, for example, that a majority of the Estonian and Latvian men of Haplogroup I represent the “Scandinavian” Haplogroup I1a and only a small minority the “Balkan-Romanian” Haplogroup I1b1-P37.

The men of Haplogroup I have clearly arrived in the Baltic area from two directions: from the west and from the south. The former men represent “Germanic” (i.e. “German” and “Scandinavian”) subhaplogroups I1a-M253 and I1b2-M223 and the latter those of East-European subhaplogroup I1b1-P37. The two directions of



Map 52. Frequencies of Haplogroup I in some East European populations. Lithuania is considered as six separate subareas in the map. Source: Kasperavičiūtė et al. (2004).

arrival are seen also from the Lithuanian I-frequencies shown in Map 52, in which the highest frequencies (16.9% and 17.5%) are in the westernmost and easternmost areas of Lithuania, while in the intermediate areas the frequencies are only 2.9-9.4%. A similar two-peak frequency distribution (high in the west and east and low in the middle) is expected in Latvia.

(4) The following eight populations form a homogenous group on the basis of high (30-50%) N3 values: Estonians, Latvians, Lithuanians, North Russians, Mari, Udmurts, Komi, and Hanti. The ancestors of these populations can be called the ancient “mammoth hunters” of Northeastern Europe. The populations were a genetically and probably also linguistically homogeneous population. The language was, supposedly, Finno-Ugric (Uralic). This, however, implies language shift: the Baltic and north Russian populations shifted their

original FU language to an IE (Baltic and Slavic) one. The language shift took place particularly at the time of the arrival of agriculture and the “Slavic Expansion” more than a thousand years ago.

The peak area of the frequencies of the “Siberian” Haplogroup N3 in Europe is in Eastern Finland (70% in the N3 Maps above). In Map 51, there is a secondary maximum (51%) in the Mari area. Many of the northern populations of East Europe have frequencies of 34-39%: Lithuanians 39%, Latvians 38%, Estonians 34%, Northern Russians 36%, Udmurts 37%, Komi 36%, and Nenets 38%. In the next zone to the south and east of the Volga area, the percentages diminish steeply: Tatars 25%, Central Russians 16%, Mordvians 17%, Chuvash 18%, Bashkirs 17%, and Southern Russians 10%. In the next, more southern zone, the frequencies are even lower: Belarussians 3%, Ukrainians 8%, Moldovians 2%, and Gagauzes 2%. As seen from the map, the frequencies of Haplogroup N3 in East Europe form a regular north-south gradient.

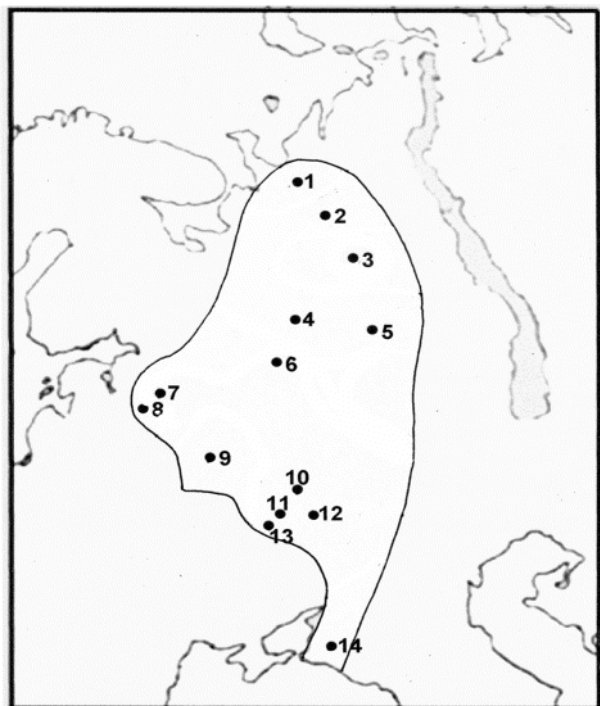
(5) Haplogroups E3b, J2, and G belonged to the first farmers to arrive in Europe. E3b is high in the Moldovians, Turks and Gagauzes; J2 is high in some Caucasian populations like Ossetians as well as Turks; and G is high (in addition to some Caucasus populations like Ossetians) also in the Romanians and Gagauzes. Two haplogroups of the ancient farmers, E3b and J2, show clear south-north gradient: the frequencies of E3b and J2 are about 15 and 10% in Romania and they descend gradually through the Ukraine and Russia and approach zero in the northeastern parts of Europe.

Three East-European Populations

Next three East-European populations are treated separately and in more detail by three language groups: (a) Russians, (b) Turkic-speakers, and (c) Balts and Finno-Ugrians.



Map 53. The three main dialects of European Russian: North Russian, Central Russian, and South Russian.



Map 54. The locations (shown as dots) of the fourteen Russian populations analysed in Balanovsky et al. (2008).

A. Russians

The Russians are treated above as three separate populations: North Russians, Central Russians, and South Russians. This division reflects the three main dialects of Russian, as shown in Map 53.

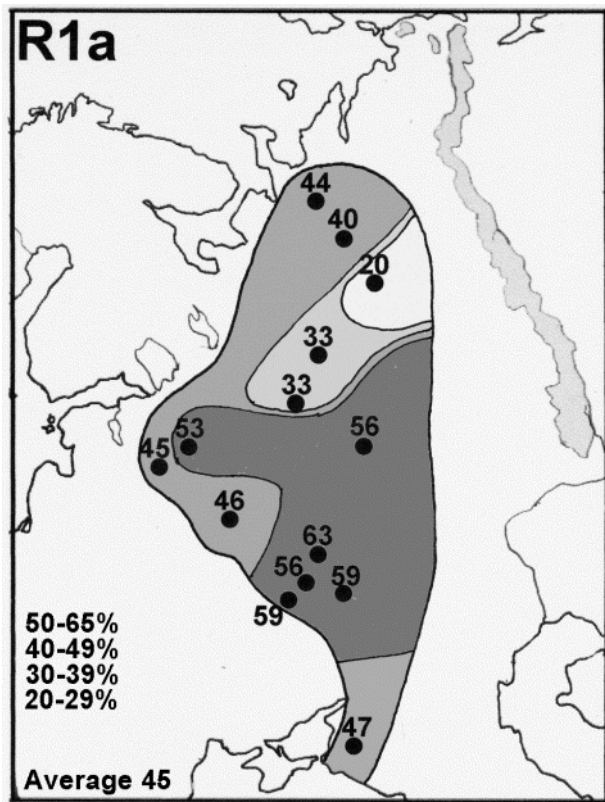
Geographically more detailed data on the frequencies of the Y-chromosome haplogroups in traditional Russia are given in Table 9. The approximate locations of the fourteen localities of the table are shown on Map 54. The table is a simplification of the corresponding table in Balanovsky et al. (2008).

The frequency zones of the eight haplogroups of Table 9 are seen in Maps 55-62. The maps allow the following generalizations about the Russians:

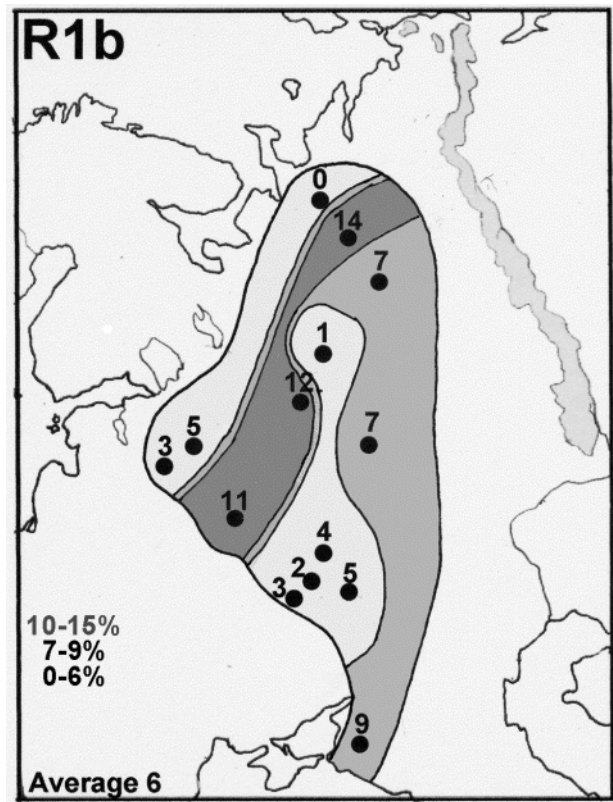
(1) The Russians consist mainly of two populations: (a) almost half (about 45%) of them represent the "Ukrainian" Haplogroup R1a, and (b) about one quarter of them represent the "Siberian" Haplogroup N (more precisely N3 + N2). The former are more frequent in Southern Russia and the latter in Northern Russia. Haplogroup R1a forms a south-north gradient, and, conversely, Haplogroup N3 forms a north-south

Table 9
Frequencies of Y-chromosome Haplogroups in Fourteen Areas of Historical Russia. The individual haplogroups included in "Other" are: C, F*, G1-M285, G2-P15, I*, I1b2-M223, J1, K*, K2, N*, Q, and R2.

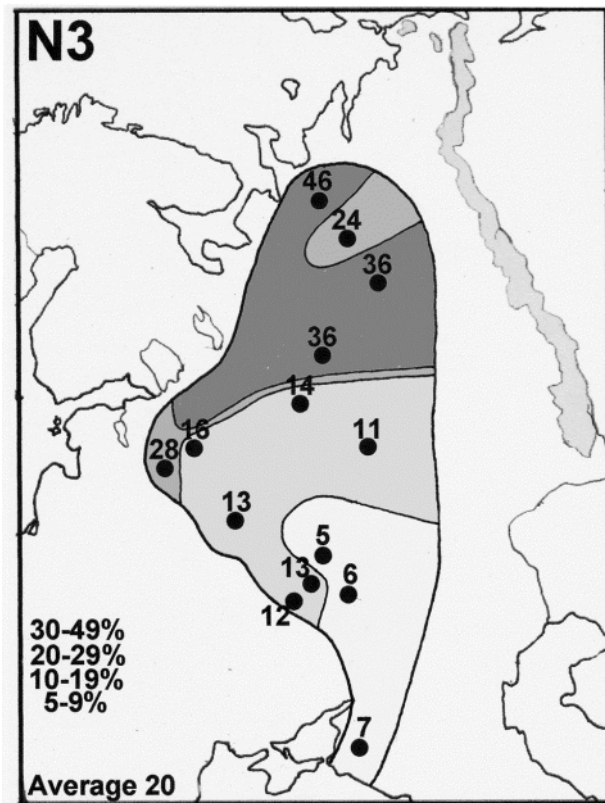
Russian area	R1b	R1a	I1a	I1b	N3	N2	OE	E3b	J2	EF	other	n
Mezen	0	44.4	0	0	46.3	7.4	98.1	0	0	0	1.9	54
Pinega	14.0	39.5	0.9	4.4	23.7	15.8	98.3	0	0	0	1.7	114
Krasnoborsk	6.6	19.8	12.1	9.9	36.3	3.3	88.0	0	5.5	5.5	6.5	91
Vologda	0.8	33.1	11.6	8.3	35.5	3.3	92.6	0.8	0.8	1.6	5.8	121
North Russia	5.4	34.2	6.2	5.7	35.5	7.5	94.3	0.2	1.6	1.8	4.7	380
Unzha	11.5	32.7	11.5	11.5	13.5	0	80.7	5.8	3.8	9.6	9.7	52
Kashin	6.8	56.2	2.7	8.2	11.0	0	84.9	4.1	4.1	8.2	6.9	73
Porhov	5.3	52.6	3.5	10.5	15.8	1.8	89.5	3.6	0	3.6	6.9	57
Ostrov	2.7	45.3	6.7	9.3	28.0	0	92.0	4.0	1.3	5.3	2.7	75
Roslavl	11.2	45.8	1.9	10.3	13.1	0.9	83.2	7.5	2.8	10.3	6.5	107
Central Russia	7.5	46.5	5.3	10.0	16.3	0.5	86.1	5.0	2.4	7.4	6.5	364
Livni	3.6	62.7	8.2	13.6	4.5	0.9	93.5	0.9	0.9	1.8	4.7	110
Pristen	2.2	55.6	4.4	17.8	13.3	0	93.3	2.2	2.2	4.4	2.3	45
Repievka	5.2	59.4	3.1	16.7	6.3	0	90.7	1.0	1.0	2.0	7.3	96
Belgorod	2.8	59.4	3.5	12.6	11.9	0.7	90.9	0.7	4.2	4.9	4.2	143
Kuban Cossacs	8.8	47.3	4.4	16.5	6.6	1.1	84.7	3.3	4.4	7.7	7.6	90
South Russia	4.8	55.4	3.9	15.9	9.5	0.5	90.6	1.8	3.0	4.2	5.2	484
Russian total	5.9	45.4	5.1	10.5	20.4	2.8	90.3	2.3	2.3	4.5	5.5	1228



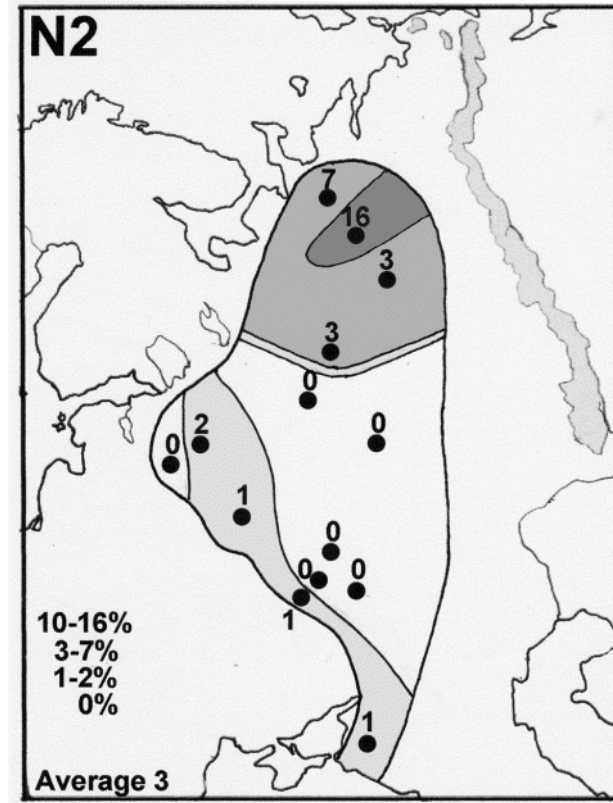
Map 55. Frequencies of Haplogroup R1a in historical Russia.



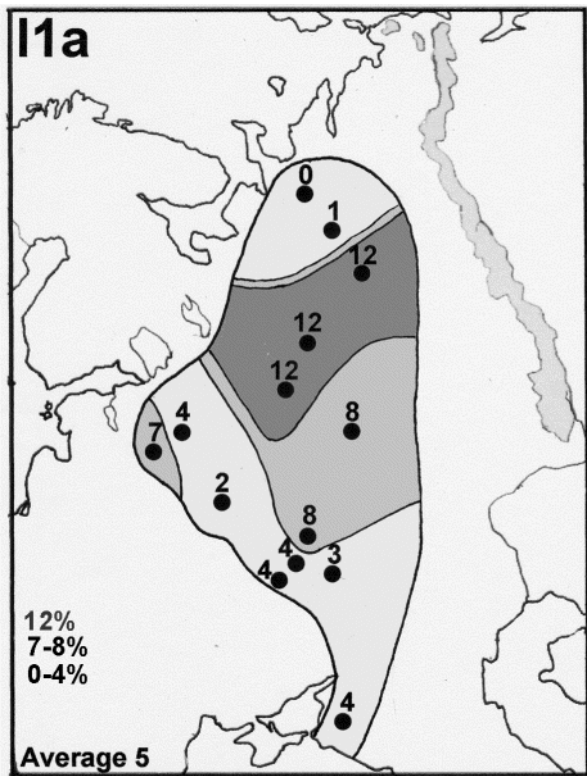
Map 56. Frequencies of Haplogroup R1b in historical Russia.



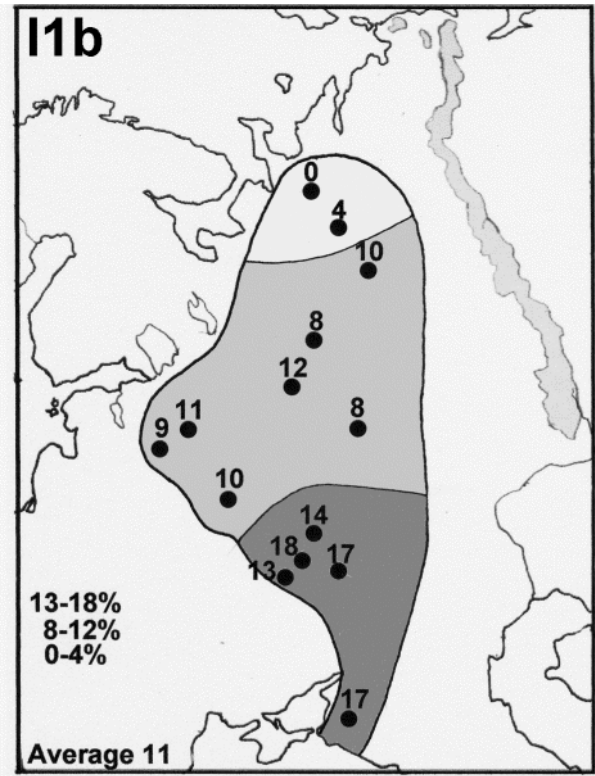
Map 57. Frequencies of Haplogroup N3 in historical Russia.



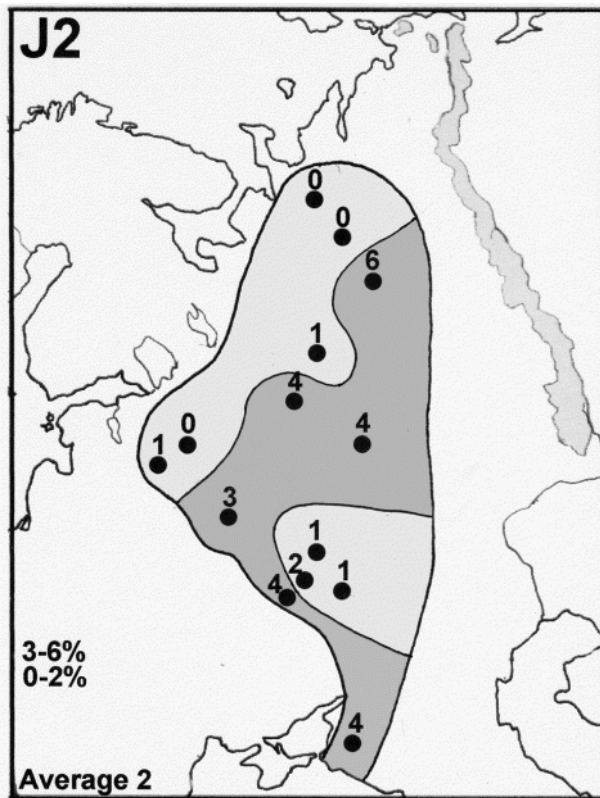
Map 58. Frequencies of Haplogroup N2 in historical Russia.



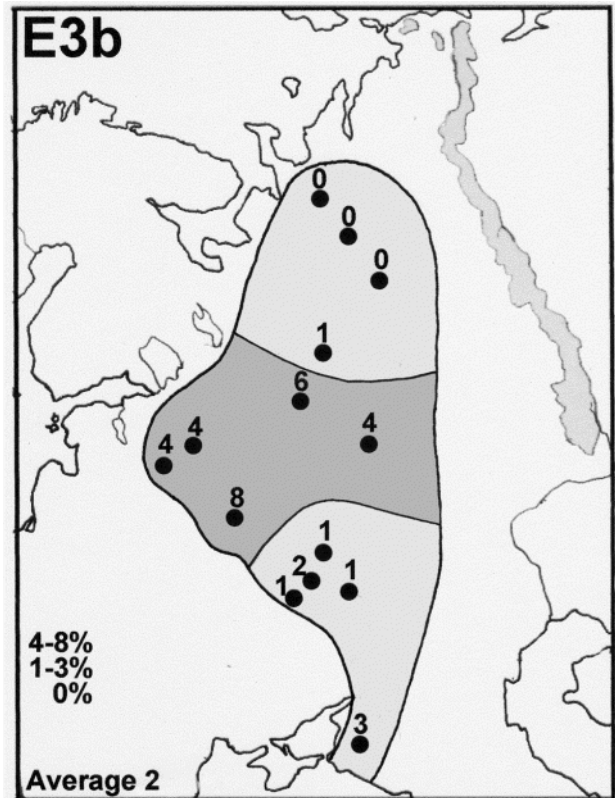
Map 59. Frequencies of Haplogroup I1a-M253 in historical Russia.



Map 60. Frequencies of Haplogroup I1b1-P37 in historical Russia.



Map 61. Frequencies of Haplogroup J2 in historical Russia.



Map 62. Frequencies of Haplogroup E3b in historical Russia.

C. The Balts and Finno-Ugrians

The last language group of the East-European populations to be dealt with consists of two linguistic subgroups. One subgroup speaks Baltic languages and the other, Finno-Ugric (Uralic) languages. The treatment of the two language groups as one is based on the fact that the Balts and Estonians are genetically close to each other.

Table 11 summarizes the haplogroup frequencies for these eight populations, and allows the following generalizations to be made:

(1) The frequencies of the “Iberian” and “Asian/Caucasian” Haplogroup R1b are 5-16%. Part of the men of this haplogroup are from the west and part from the east.

(2) The frequencies of the “Ukrainian” Haplogroup R1a are about 20-40% in almost all the European FU-speaking populations. The only exception are the Tundra Nenets who do not have this haplogroup. (R1a is also very low or non-existent in the FU-speaking populations (e.g. Hanti) of Siberia.)

(3) The frequencies of Haplogroup I form two categories: Estonians and Mordvians belong to one category with frequencies of about 20% and the Komi, Udmurts and Mari to the other with frequencies of about 4%.

(4) With respect to the frequency of the “Siberian” Haplogroup N3, the Finno-Ugric speaking populations form two categories: this haplogroup is relatively low (about 17%) in the Mordvians, but high (34-51%) in the other populations. The feature makes the Mordvians differ from their linguistic relatives in the direction of more southern populations.

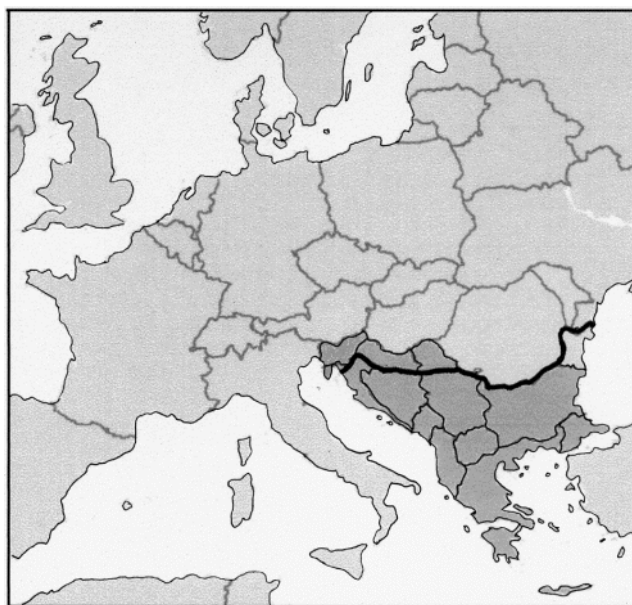
On the basis of the “Samoyedic” Haplogroup N2, the FU (or Uralic) speakers form four categories: N2-frequency is (a) highest (about 45%) in the (Samoyedic)

Tundra Nenets, (b) about equally high (about 38%) in the Hanti of Northwestern Siberia, (c) lower (about 14-24%) in the Komi and Udmurts, and (d) non-existent or almost non-existent in the other FU-speaking populations. (In this respect the Vepsians who are close relatives of the Karelians are an exception: the N2-frequency of the Vepsians is as high as about 17% (Rootsi et al. 2006).

(5) The frequencies of the three haplogroups of the Early Farmers (J+E3b+G) are low (about 2-6%) in all the FU-speaking populations. This results, of course, from the northern habitats of these populations.

4. The Balkans

The region defined here as the home of the “Balkan Europeans” is shown on Map 63. The Balkans are often



Map 63. The Balkan region (south of the dark line).

Table 11

Frequencies of Eight Haplogroups in the Baltic-speaking and Finno-Ugric-Speaking Population of East Europe. The Nenets of the table belong to the Tundra Nenets group (not the Forest Nenets group of Siberia).

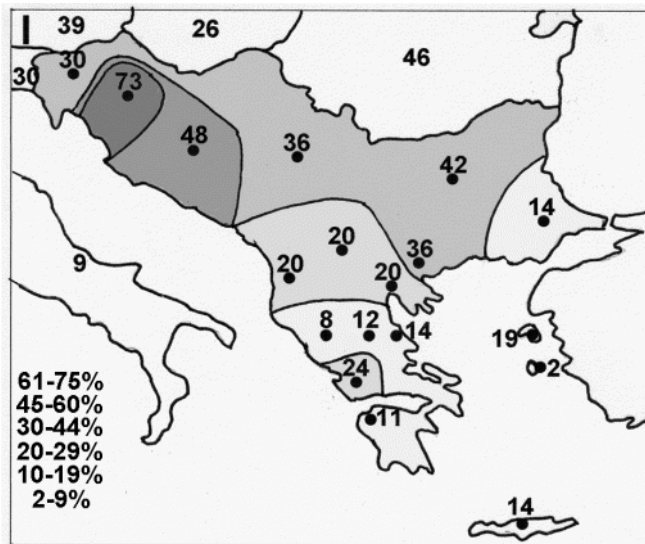
Population	R1b	R1a	I	N3	N2	OE	E3b	DE	J	G	EF	K	other
Latvians	11.8	40.5	8.4	37.8		98.5	0.2		0		0.2		0.2?
Lithuanians	4.5	38.3	13.3	42.2		98.3	1.2		0	0	1.2	0.2	
Estonians	7.9	31.9	20.1	33.9	0	93.8	1.8	0.5	0.7		3.0	3.5	
Mordvians	13.3	26.5	19.2	16.9	2.4	78.3	?	?	?	?		?	?
Mari	5.0	21.0	4.2	51.3	0	81.5			6.3		6.3	11.8	
Udmurts	9.2	30.5	3.5	36.6	14.4	94.2	2.4		2.4		4.8	2.3	
Komi	16.0	23.7	4.5	36.2	24.3	>100?	0		0				0
Nenets, Tundra	0	0	0	23.7	74.6	98.3	0	0	0	0	0	0	1.7



Maps 64-65. Eighteen localities in the Balkans: 1 = Slovenes, 2 = Croats, 3 = Bosniacs (Bosnia-Herzegovians), 4 = Serbs, 5 = Bulgarians, 6 = Macedonians, 7 = Albanians, 8-17 = Greeks, and 18 = Turks.

defined geographically by the Danube-Sava-Kupa line (the black line on Map 63), according to which Slovenia, northern Croatia, and northern Serbia do not belong to the Balkans, but the eastern coast of Romania does belong to it. However, in this study, Slovenia (but not Romania or Moldova), entire Croatia, and entire Serbia are included in the definition.

The Balkan populations represent four language groups: (1) South Slavic, (2) Albanian, (3) Greek, and (4) Turkish. The first three represent IE languages; Turkish is an Altaic language. The three IE languages or language groups and Romanian form a special group often called the “*Balkan Sprachbund*”. The languages share little common vocabulary but they show great similarity in grammar; so, for example, they have very similar case systems and they all have become more analytic. These features can be interpreted as an indication of language shifts having taken place in the area.



Map 66. The frequencies of Y Haplogroup I in eighteen Balkan populations. A large majority of the haplogroup represents subhaplogroup I1b1-P37. Italian averages (in all the Balkan maps) are based on the frequencies of four eastern localities in central Italy: Pescara, Foggia, North Gargano, and Brindisi (Di Giacomo et al. 2003).

The most frequent haplogroup in the Balkans is Haplogroup I, or more precisely, its subhaplogroup I1b1-P37. The fact that R1a and R1b are almost equally frequent in the Balkans indicates that the Balkans lie at the west-east dividing line of Europe or slightly on its eastern side. The high E3b and J2 haplogroup values indicate that the Balkans belong to an area of early farming.

The following gradients can be found in the Balkan area:

(1) There is a north-south gradient of R1b in the north of the Balkans: The frequency of R1b is about 25% in the Slovenes, but is considerably lower elsewhere and only about 2% in some parts of Greece. The gradient is a reflection of the spreading of the West-European or Iberian men to the Balkans (the other alternative explanation is that these men have come to the Balkans directly from the Caucasian/Asian R1b-areas).

(2) Haplogroup R1b exhibits a north-south gradient, with the frequencies of this haplogroup at 10-15% in the northern part of the Balkans and only about 3% in the southern part.

Table 12

Frequencies of the Y Haplogroups of Eight Balkan Populations. The values for Croats, Bosnians, and Serbs are from Marjanovic et al. 2005. Those for the five Croat sub-populations are from Pericic (2005). The Pericic values for I are a sum of I (xI-P37) and I-P37. The Greek data are from DiGiacomo (2003).

Population	R1b	R1a	I	N3	OE	Eb3	J	G	EF	F	P	K	Y*	Other	n
Slovenes	25.5	29.5	30.4	0	85.4	12.0	4.5	0.5	17.0			0		0	
Croats	2.2	12.2	73.3	0	87.7	8.9	1.1	1.1	11.1	1.1		0			90
Mainland	15.7	34.3	37.0		87.0	5.6	1.9	0.9	8.4	1.9	1.9	1.9			108
Krk	16.4	38.4	27.3		82.1	6.8	10.9		17.7						73
Brac	6.4	27.7	53.2		87.3	4.3		6.4	10.7	2.1		2.1			47
Hvar	6.4	8.8	65.9		81.1	4.4	3.3	1.1	8.8	1.1	14.3	1.1			91
Korecula	1.1	20.5	53.0		74.6	3.8	2.3	10.6	16.7	1.5	6.1	1.5			132
Bosnians	3.5	15.3	48.2	0	67.0	12.9	11.9	3.5	28.3	3.5		1.2			85
Serbs	6.2	13.6	35.8	0	55.6	22.3	8.7	1.2	32.2	4.9		7.4			81
Bulgarians	17.0	12.0	42.0	0	71.0	17.0	12.0	0	29.0			0		0	
Macedonians	10.0	35.0	20.0	0	65.0	15.0	20.0	0	35.0			0		0	
Albanians	17.6	9.8	19.6	0	47.0	21.6	27.5	2.0	51.1			0		2.0	
Greeks Total	13.4	11.0	16.0	0	40.4	22.7	21.3	5.9	49.9				7.8	0.4	366
Agrinion	19.0	4.8	23.8	0	47.6	9.5	28.6	4.8	42.9				9.5	0	21
Ioannina	16.7	8.3	8.3	0	33.3	29.2	20.9	4.2	54.3				12.5	0	24
Patrai	11.1	5.6	11.1	0	27.8	44.4	16.7	0	61.1				11.1	0	18
Kardhitsa	8.0	20.0	12.0	0	40.0	28.0	16.0	12.0	56.0				4.0	0	25
Serrai	12.0	8.0	36.0	0	56.0	24.0	16.0	4.0	44.0				0	0	25
Thessalon.	5.0	25.0	20.0	0	50.0	20.0	15.0	5.0	40.0				10.0	0	20
Larissa	19.0	9.5	14.3	0	42.8	14.3	28.6	4.8	47.7				9.5	0	21
Crete	9.8	7.7	14.0	0	31.5	15.4	38.5	7.0	60.9				7.7	0	143
Mitilini	7.4	11.1	18.5	0	37.0	18.5	29.6	7.4	55.5				3.7	3.7	27
Khios	26.2	9.5	2.4	0	38.1	23.8	19.1	9.5	52.4				9.5	0	42
Turks	13.3	5.8	14.2	2.2	35.5	11.7	38.2	3.3	53.2			1.7		9.9	
Balkan Total	12.1	16.0	33.3	0.2	57.1	16.0	16.1	1.9	36.8	0.7	0.9	0.7	3.6	0.7	

Y* = Y (x A, DE, G2, I, J, P) = {e.g., K2, L, NO, GxG2, H}

(3) The frequency of Haplogroup I1b1-P37 is very high (about 40%) in the Western Balkans and it diminishes in all directions, becoming as low as about 10-5% in the northernmost and southernmost parts of the Balkans; there is, however another secondary centre with values of over 20% in Romania to the east.

(4) Haplogroup E3b shows a south-north gradient and its values are about 25% in Greece but only about 10% in the northernmost area of the Balkans.

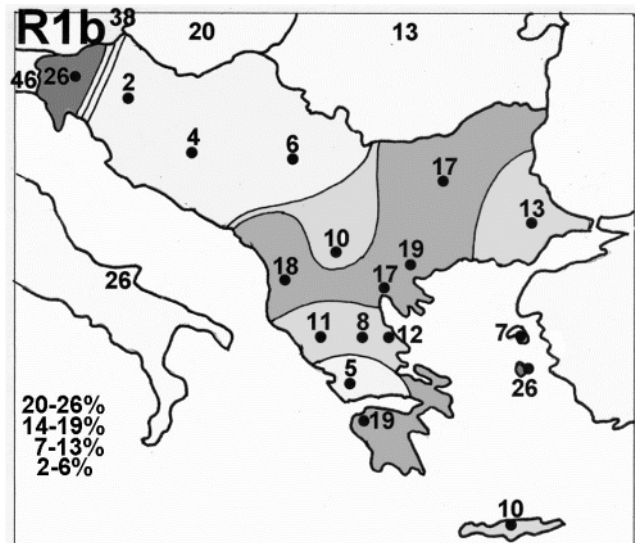
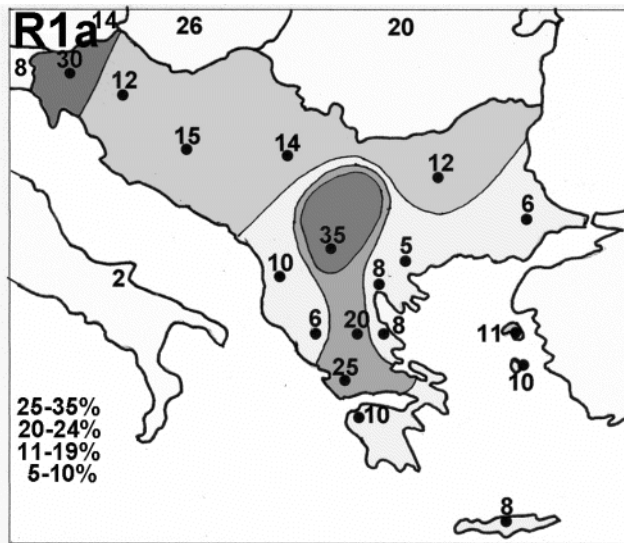
(5) A similar but weaker gradient concerns Haplogroup J2: its maximum value of about 20% in the south diminishes gradually to about 5% in the north.

On the basis of Table 13 and the Balkan maps, the following detailed genetic observations can be made:

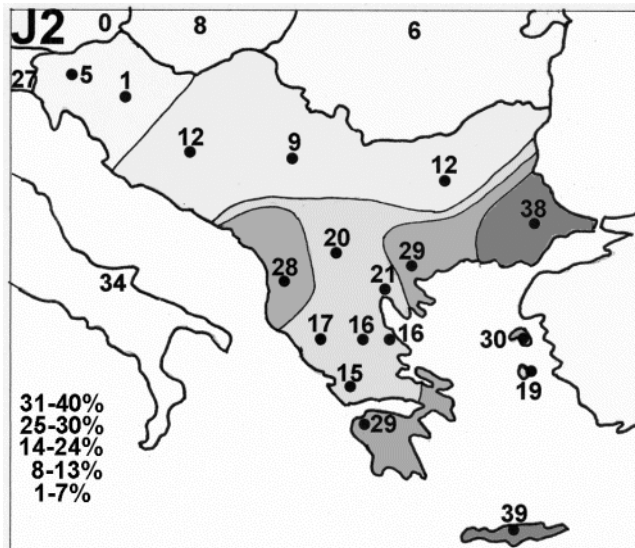
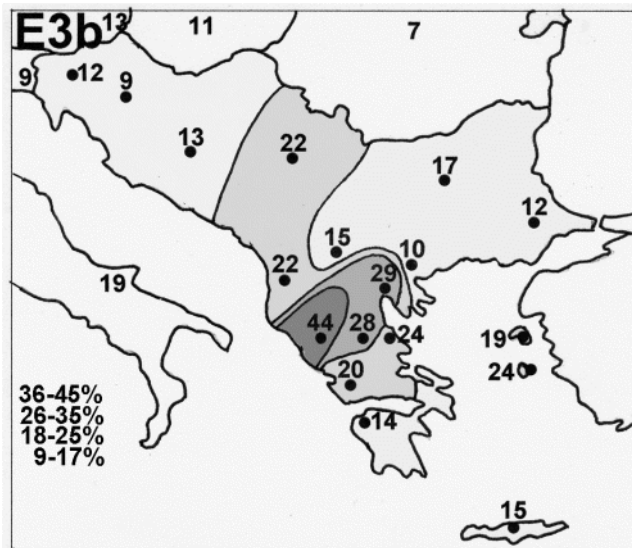
(1) The Slovenes belong to the Central European group with the Hungarians, their immediate geographic neighbours. This is seen in the high R1b value, which is

25.5% for the Slovenes and 20% for the Hungarians; in the other Balkan populations, the R1b values are considerably lower (11-17.6%). This is an example of geographic nearness being a more relevant factor than linguistic relatedness. The Slovenes and Croats are genetically distant from each other, even if they speak related (South-Slavic) languages, while the Slovenes and Hungarians are genetically close to each other, even if they speak unrelated languages. This complicated genetic-linguistic relation may be explained by the fact that there was previously a mostly homogeneous population that spoke a common language, but then part of the population shifted language and the original language was replaced by Hungarian in some areas and by South-Slavic in others.

(2) The Slovenes, Croats, and Macedonians (all of whom are Slavic-speaking) originate more strongly than others from the Ukrainian refuge. This is seen in their relatively high R1a values: 29.5% in the Slovenes, 31.8% in the Croats, and 35% in the Macedonians. In



Maps 67 and 68. The frequencies of Y Haplogroups R1a and R1b in the Balkans and some neighbouring areas.



Maps 69 and 70. The frequencies of Y Haplogroups E3b and J2 in the Balkans and some neighbouring areas.

the other Balkan populations the equivalent frequencies are considerably lower (5.8-16%).

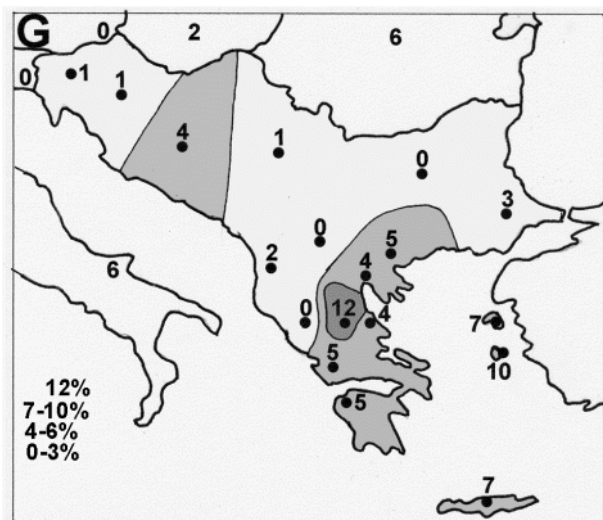
(3) High I-values (mostly I1b1-P37) are typical of the Croats (42%), Bosnians (48%), and Bulgars (42%). In the other Balkan populations, this value is lower (14.2-30.4%). This can be interpreted as a possible indication of the fact that the Croats, Bosnians, and Bulgarians originate from the Balkan refuge more often than their neighbours.

(4) Strong indications of Early Farmers are seen in the high E3b+J2+G values of the Turks (53.2%), Albanians (51.1%), and Greeks (48.3%). These areas are non-

Slavic-speaking coastal areas that were perhaps the first to receive farming in Europe. The lower E3b + J2 + G values of the Macedonians (35%), Bulgarians (29%), Bosnians (28%), Serbs (32%), Slovenes (17.0%), and Croats (11.1%) may indicate that the ancient areas of these populations were not equally suitable for early farming.

Croatians

In Barač et al. (2003) a detailed analysis of the male populations of eleven Croatian localities is presented; see Map 72 and Maps 73-75 for Haplogroups R1a, R1b, and I1b1-P37.



Maps 71. The frequencies of Y Haplogroup G in the Balkans and some neighbouring areas.



Map 72. Geographic locations of the eleven Croatian localities in the study: six localities in the mainland and five on the islands.

The maps allow the following generalizations about the Croatian men:

(1) The R1a men arrived in Croatia mainly from the areas now belonging to Hungary. Ultimately, these men came from the Ukrainian refuge.

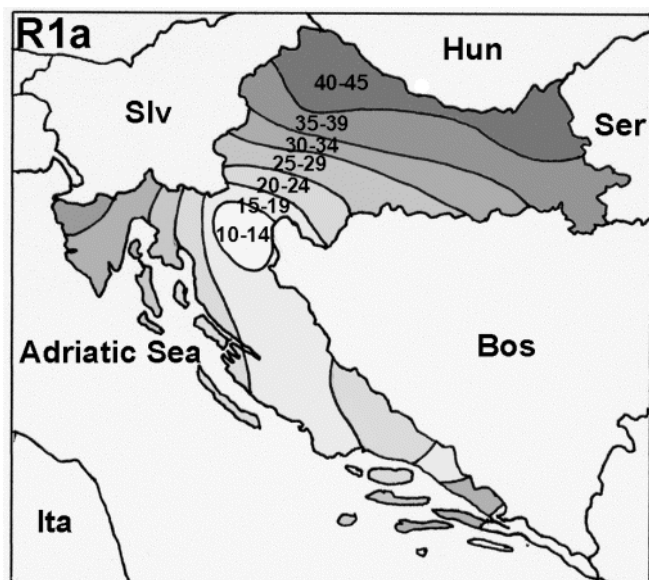
(2) The R1b men arrived in Croatia from the Iberian refuge, mostly through Slovenia and an area now a part of Austria.

(3) The West Balkan peak area (frequencies over 50%) of Haplogroup I1b1-P37 is in central Croatia near Zagreb.

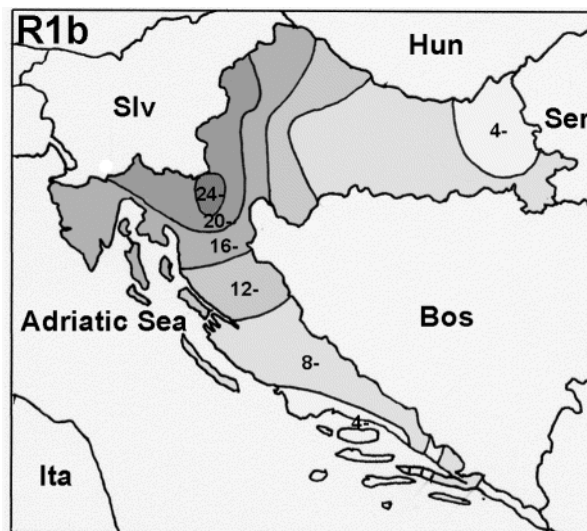
Accordingly, the Croatian men have come to their present habitats from three directions, Slovenia and the Iberian refuge, Hungary and the Ukrainian refuge, and from the south; see Map 76.

5. Italy

The only present language of the Italian peninsula is Italian, a Romance language, although prior to the time of the Romans, there were others, such as Etruscan.

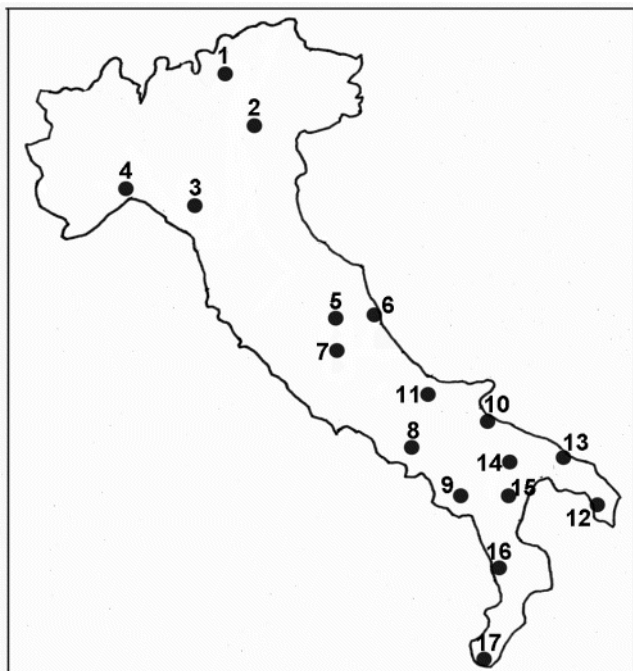


Map 73. Frequency zones of the frequencies of Haplogroup R1a in Croatia.



Map 74. Frequency zones of the frequencies of Haplogroup R1b in Croatia.

Of the three large islands near the peninsula, Sicily and Sardinia belong to Italy and are Italian-speaking. The third island of the area, Corsica, belongs to France and is French-speaking. We first consider the Italian peninsula, and afterward, the three islands.



Map 77. The geographic locations of the seventeen areas of Table 14.

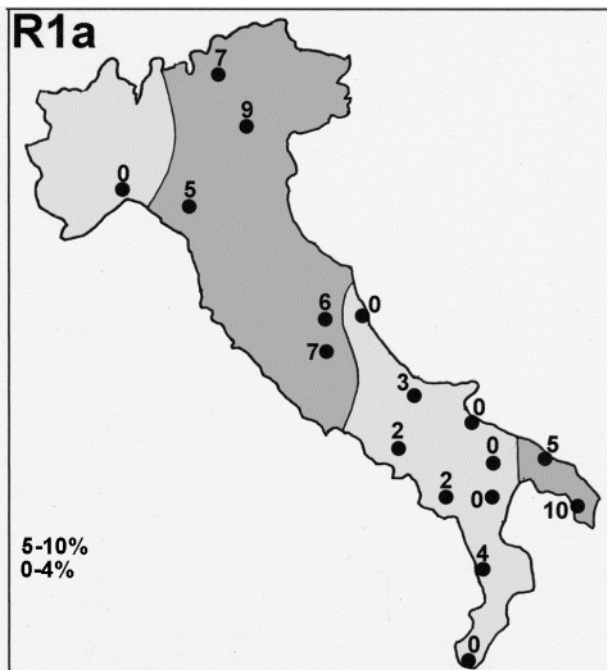
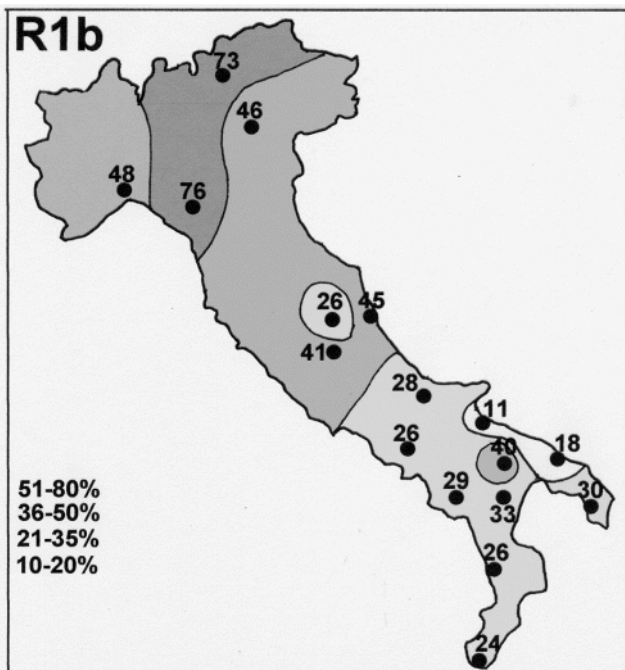
The Italian Peninsula

The Wikipedia and Balanovsky maps show only one strong gradient in the Italian peninsula: The frequencies of R1b are about 40% or more in the north and only about 15-30% in the south. A candidate for a weak gradient is offered by the “West-Balkan” Haplogroup I1b1-P37. For this haplogroup, the frequencies are 1-5% in the south and practically non-existent (0-1%) in the north.

More detailed information about the haplogroup frequencies in various parts of the Italian peninsula is found in Tables 13 and 14. The tables are based on two investigations, one by Scozzari et al. (2001) and the other by Di Giacomo et al. (2003).

The seventeen geographic locations of Table 14 are shown in Map 77. The haplogroup frequencies of the male populations of these localities are shown in Maps 78, 79, 81, 82, 83, 84, and 85.

(1) As show in Map 78, the R1b-frequencies are higher than 70% in northern Italy, but only about 25% in southern Italy. In addition to this north-south gradient, there are exceptionally low frequencies (11-18%) of R1b in the southeast (Foggia and Brindisi). The frequency distribution of Haplogroup R1b from the study by DiGiacomo (2003) is very similar, as shown in Map 80.



Maps 78 and 79. Frequencies of Y Haplogroups R1b and R1a in the Italian peninsula.

Table 13
Frequencies of the Y Haplogroups of Some Italian Populations. "Italian Total"
Refers to Continental Italy. The Table is Based on Scozzari et al. (2001).

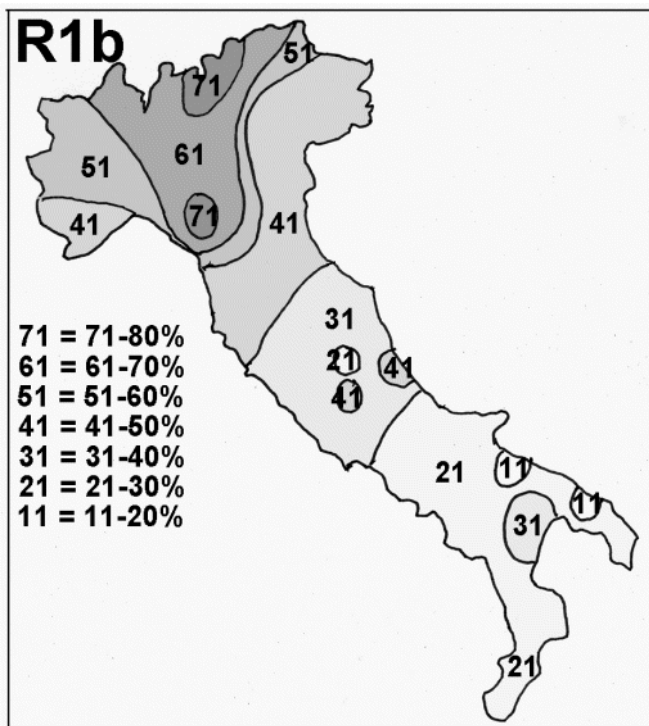
Population	R1b	R1a	I	OE	E3b	J	G2	EF	K	Other	n
Lombardia	61.1	0	11.1	72.2	11.2	11.1	0	22.3	5.6	0	18
Venetia	30.0	10.0	30.0	70.0	25.0	5.0	0	30.0	0	0	20
Liguria	41.2	0	35.3	76.5	5.9	17.7	0	23.6	0	0	17
Latium	33.3	3.0	25.7	62.0	10.6	25.7	0	36.3	1.5	0	66
Campania	20.0	0	26.7	46.7	0	53.4	0	53.4	0	0	15
Calabria	21.1	10.5	31.6	63.2	21.1	15.8	0	36.9	0	0	19
Italian total	34.5	3.9	26.7	65.1	12.3	21.5	0	33.8	1.1	0	155
Sicily	39.1	3.8	0	42.9	20.5	28.8	2.3?	51.6	5.4	0	131
Sardinia	19.6	1.6	55.6	76.8	7.5	12.5	0.3	20.3	0.9	4.1	331

Table 14
The Frequencies of Y Haplogroups in Seventeen Italian Populations.
The table is based on Di Giacomo et al. (2003).

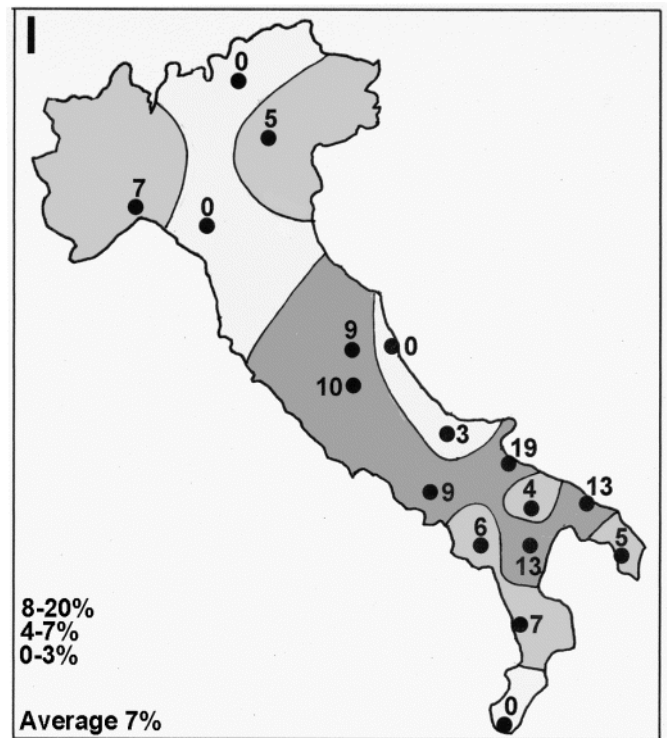
	R1b	R1a	I	OE	DE	J	G2	EF	Other	n
Val di Non	73.3	6.7		80.0		10.0		10.0	10.0	30
Verona	45.5	9.1	4.5	59.1	9.1	27.2		36.3	4.5	22
Garfagnana	76.2	4.8		81.0	2.4	9.5	4.8	16.7	2.4	42
Genoa	48.3		6.9	55.2	24.1	10.3	10.3	44.7		29
L'Aquila	25.7	5.7	8.6	40.0	11.4	34.3	5.7	51.4	8.6	35
Pescara	45.0			45.0	15.0	30.0		45.0	10.0	20
Avezzano	41.4	6.9	10.3	58.6	3.4	20.6	6.9	30.9	10.3	29
Benevento	26.1	2.2	8.7	37.0	17.4	26.1	10.9	54.4	8.7	46
Cilento	29.2	2.1	6.3	37.6	12.5	27.2	14.6	54.3	8.3	28
Foggia	11.1		18.5	29.6	11.1	44.4	14.8	70.3		27
N. Gorgano	27.6	3.4	3.4	34.4	24.1	37.8		61.9	3.4	29
Casarno	30.0	10.0	5.0	45.0	20.0	25.0	10.0	55.0		20
Brindisi	18.4	5.3	13.2	36.9	26.3	23.7	7.9	57.9	5.3	38
Altamura	40.0		4.0	44.0	36.0	12.0	4.0	52.0	4.0	25
Matera	33.3		12.5	45.8	25.0	12.5	4.2	41.7	12.5	24
Paola	25.9	3.7	7.4	37.0	11.1	44.4		55.5	7.4	27
Reggio Cal.	24.2			24.2	27.3	36.4	3.0	66.7	9.1	53
Total	36.4	3.4	6.5	44.3	15.8	25.2	6.3	47.3	6.3	524

Table 15
Frequencies of Seven Haplogroups in the Large Islands Near the Italian
Peninsula. The table is based on Francalacci et al. (2003).

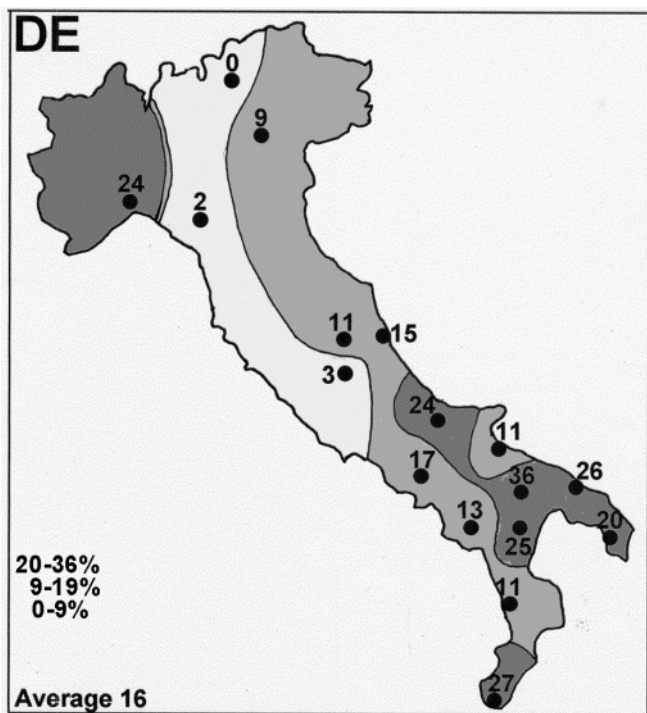
Island	R1b	R1a	I	OE	E3b	J2	G	EF	F	Other	n
Sicily	27.5	3.9	3.9	35.3	5.9	31.4	11.8	49.1	11.8	3.9	51
Sardinia	23.1	0	37.2	60.3	10.3	5.1	14.1	29.5	5.1	5.2	78
Corsica	50.0	0	8.8	58.8	14.7	2.9	11.8	29.4	2.9	8.8	34
Total	33.5	1.3	16.6	51.5	10.3	13.1	12.6	36.0	6.6	6.0	153



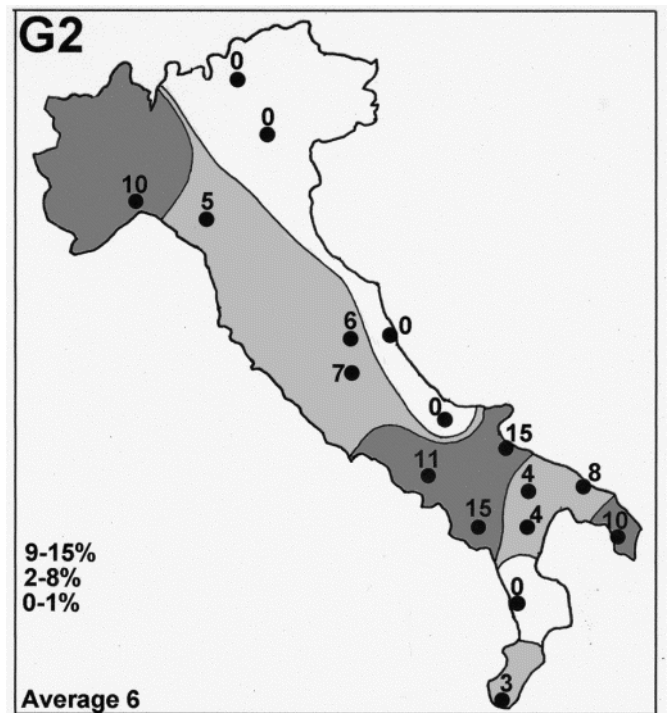
Map 80. Frequencies of Haplogroup R1b in the Italian peninsula. The map is a simplified version of a map in Di Giacomo et al. (2003).



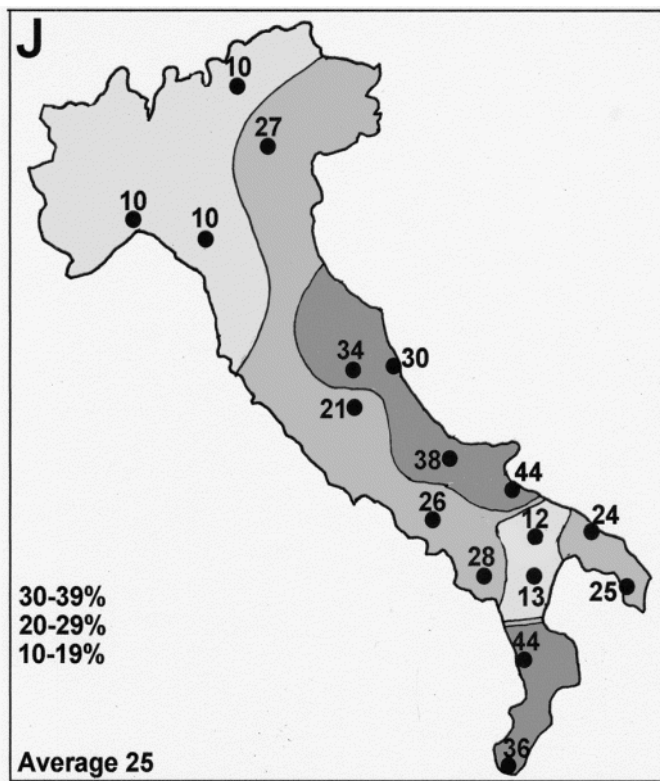
Map 81. Frequencies of Haplogroup I in the Italian peninsula.



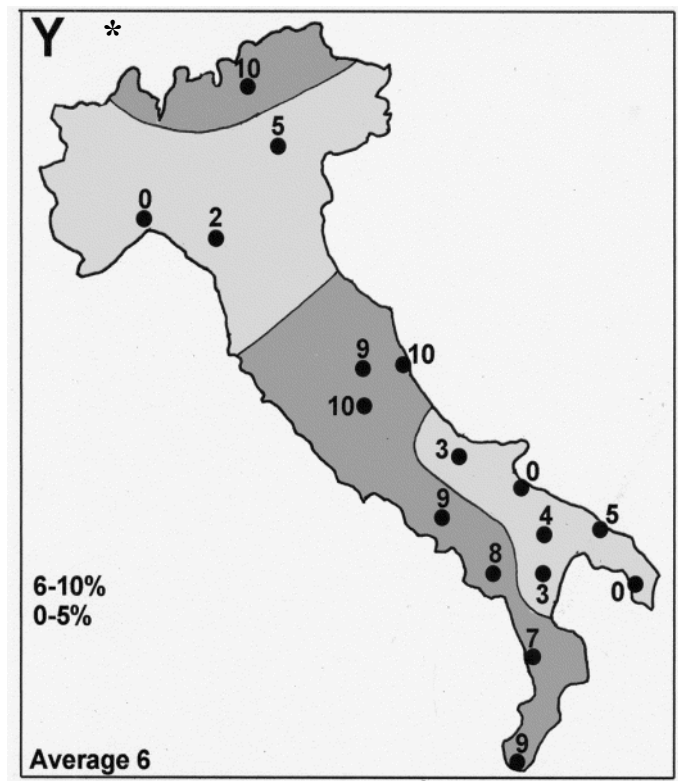
Map 82. Frequencies of Haplogroup DE in the Italian peninsula. Most of Haplogroup DE is represented by E3b in Italy.



Map 83. Frequencies of Haplogroup G2 in the Italian peninsula.



Map 84. Frequencies of Haplogroup J in the Italian peninsula.



Map 85. Frequencies of Haplogroup Y* ("Other") in the Italian peninsula, probably including men of Haplogroups L and K2.

(2) For Haplogroup R1a, no equivalent gradient is found. Instead, there seems to exist two separate frequency peaks (both of the magnitude of about 5-10%); one covers almost the entire northern half of Italy and the other the southeastern corner (Casarno and Brandisi) of the peninsula.

(3) Map 81 shows that the distribution of Haplogroup I is rather patchy. One general observation about the distribution of this haplogroup is that the frequencies are highest (about 20%) in Foggia and zero in the extreme north (Valid Non and Garfognana) and the extreme south (Reggio and Calabria). The Italian men of this haplogroup undoubtedly arrived from the Balkans, but the present uneven distribution seen today does not lend itself to simple explanations.

(4) The haplogroups of the Early Farmers are quite common in Italy. The sum total of Haplogroups DE+G2+J covers almost half (47%) of the Italian men of the peninsula. The areas of the highest frequencies are naturally in the south (as agriculture arrived from the east along the Mediterranean coast), but also the coastal area close to France (Genoa) has quite high percentages of these haplogroups.

The West-Mediterranean Islands: Sicily, Sardinia, and Corsica

The frequencies of seven haplogroups in Sicily, Sardinia, and Corsica are seen in Table 15. The following generalization are possible on the basis of these data:

(1) Sicily has much higher percentages (5.9+31.4 = 37.3%) for the Early Farmers than the other two islands, whose equivalent percentages are only 10.3+5.1 = 15.4% and 14.7+2.9 = 17.6%. Another difference concerning the Early Farmers is that the ratio of the frequencies of two of the haplogroups is different. The "African" Haplogroup E3b is more frequent than the "Near Eastern" Haplogroup J2 in Sardinia and Corsica, but in Sicily J2 is much more frequent than E3b. This difference may indicate a difference in the areas responsible for the arrival of agriculture in the three islands.

(2) Sicily has also higher F* frequencies (about 12%) than the other two islands (about 3-5%). This may be another indication of relatively large number of men arriving in Sicily by boat from the Middle East.

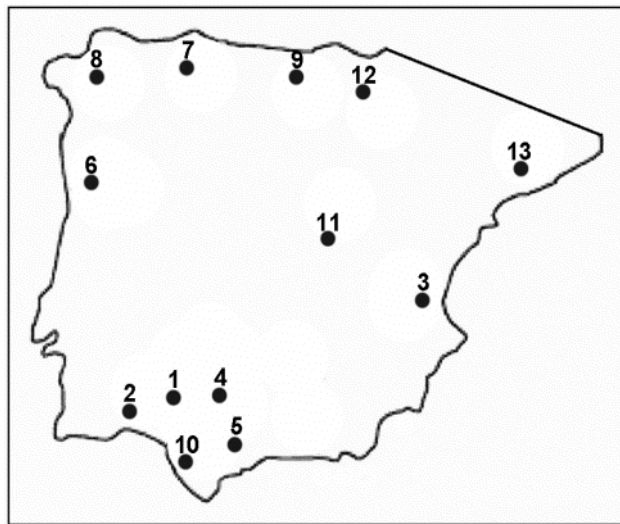
(3) A typical feature of Sardinia is the surprisingly high frequency (almost 40%) of Haplogroup I (more pre-

cisely its subhaplogroup I1b1-P37). This may indicate an early contact with the Western Balkans.

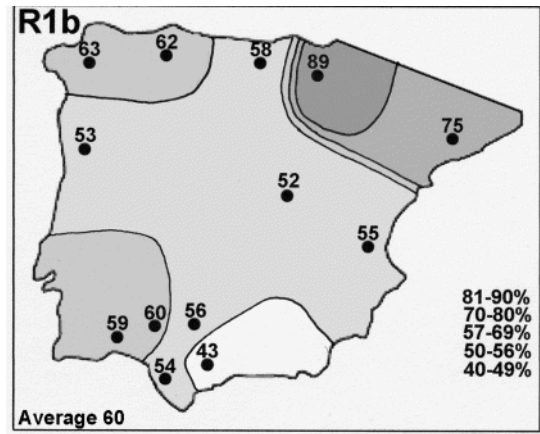
(4) About a half of the Corsican men represent the “Iberian” Haplogroup R1b. In the other two islands, the equivalent frequency is only about 25%. In this respect, the Corsican men are typically West-Europeans and many of them have, no doubt, come to the island from France and Spain.

6. Iberia

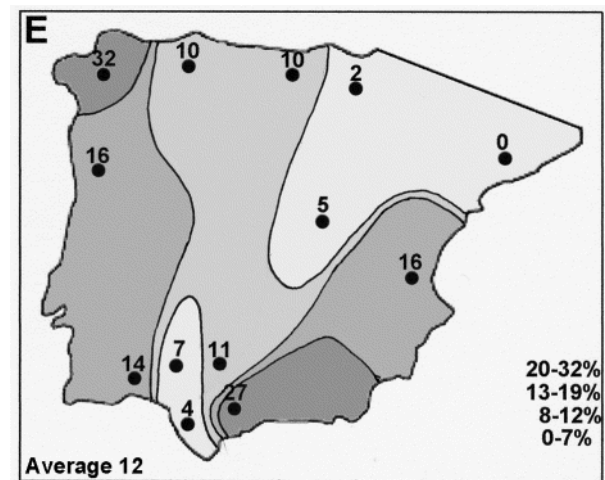
Iberia consists of three areas: Spain, Portugal, and the Basque area. Linguistically, Iberia represents two lan-



Map 86. The locations of the thirteen Iberian populations discussed in the text.



Map 87. Frequencies of Haplogroup R1b in the Iberian peninsula.

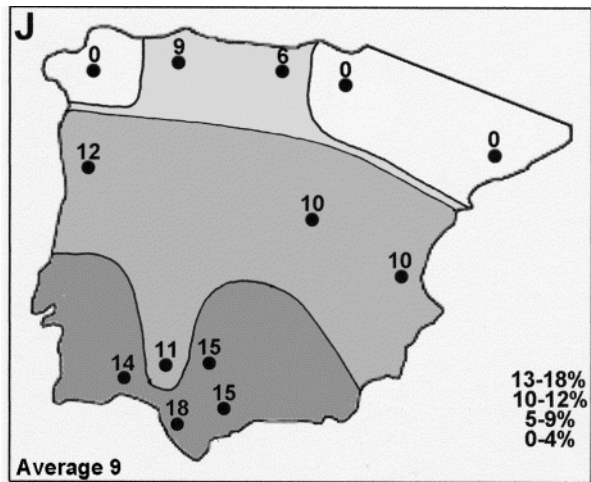


Map 88. Frequencies of Haplogroup E in the Iberian peninsula.

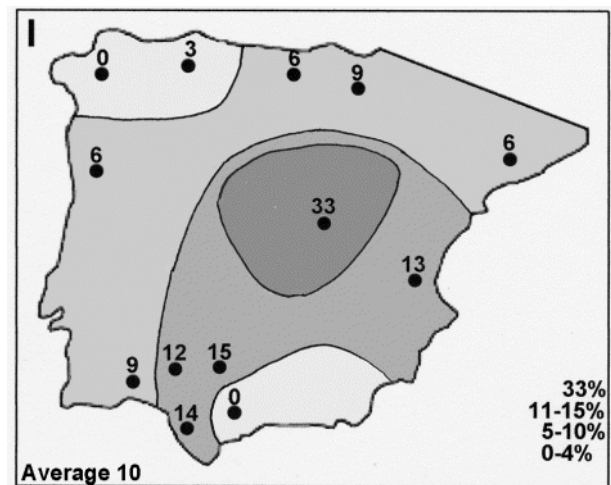
Table 16
Frequencies of the Most Common Haplogroups in Thirteen Iberian Populations.
The table is simplified from Flores et al. (2004).

Population	R1b	I	OE	E	J	EF	K	others	n
Seville	60.0	11.3	71.3	7.0	11.0	18.0	4.5	5.2	155
Huelva	59.1	9.2	68.3	13.5	13.7	27.2	4.5	0	22
Cadiz	53.5	14.3	57.8	3.6	17.9	21.5	10.7	0	28
Cordoba	55.8	14.7	70.5	11.1	14.7	25.8	3.7	0	27
Malaga	42.5	0	42.5	26.8	15.4	42.2	0	15.3	26
N. Portugal	53.3	5.5	58.8	15.5	12.0	27.1	1.8	11.0	109
Leon	61.5	3.4	65.9	10.0	8.7	18.7	3.3	15.1	60
Galicia	63.1	0	63.1	31.6	5.3	36.9	0	0	19
Cantabria	58.4	5.7	64.1	10.0	5.8	15.8	4.3	15.8	70
Valencia	54.8	12.9	67.3	16.2	9.7	25.9	3.2	3.2	31
Castile	52.4	33.3	85.7	4.8	9.5	14.3	0	0	21
Basques	89.0	8.8	97.8	2.2	0	2.2	0	0	45
Catalans	75.1	6.2	81.3	0	?	18.7	0	0	16
Total	59.9	9.6	68.8	11.9	9.5	21.4	2.5	5.0	629

Note: R1b = R1 (xR1a), R1b-M153, R1b-SRY2627; I = I (xI1b2), I1b2; E = E*, E3a, E3b1, E3b2, E3b3*, E3b3a; Others = BC*, DE*(x E3), F*, G, N3a, P*(xR1), R1a.



Map 89. Frequencies of Haplogroup J in the Iberian peninsula.



Map 90. Frequencies of Haplogroup I in the Iberian peninsula.

guage phyla, Indo-European and Basque. The Indo-European languages belong to the Romance group and represent two main languages, Spanish (Castilian) and Portuguese. Spanish is often interpreted as containing two regional languages Galician and Catalan.

The Y-chromosome haplogroups and their frequencies are shown in Table 16. The frequency zones of the most common five haplogroups are seen in Maps 87-91. The maps are based on Table 16.

The subgroups contributing to the total R1b frequencies in Map 87 are as follows: R1 (x R1a, R1b-M153, R1b-SRY2627), (50.1%), R1b-M153 (2.9%), and R1b-SRT2627 (6.9%). Haplogroup R1a is not illustrated in a separate map because it is so small in Iberia (frequency in entire Iberia = 1.7%), but it is included as one of the haplogroups called "other" in Table 16.

The overall frequencies of the subgroups of Haplogroup I in the entire Iberian area are as follows: I (x I-M26) (6.0%) and I-M26 (3.7%). In Castile, the frequency of both components of I are quite high, but the "Sardinian" Haplogroup I-M26 is exceptionally high (19%).

The overall frequencies for all of Iberia for the component subgroups of Haplogroup E are as follows: E3 (x E3a, E3b-M78, E3b-M81, E3b-M34) (0.9%), E3a (0.6%), E3b-M78 (2.7%), E3b-M81 (5.5%), E3b-M123 x E3b-M34 (0.1%), E3b-M34 (1.9%).

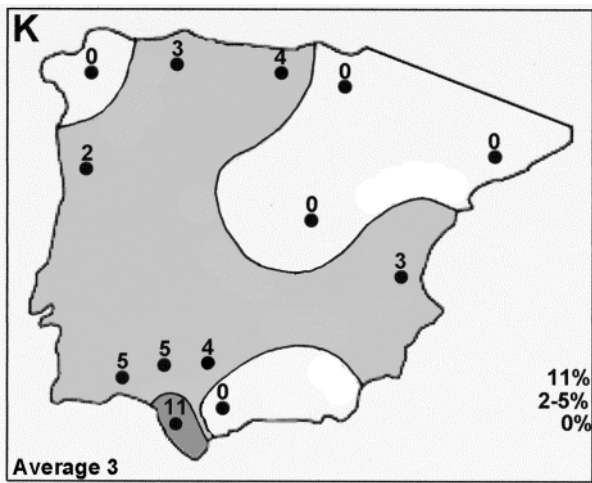
The frequencies in the entire Iberian for the components of Haplogroup J are as follows: J x J2 (1.6%), J2 x J2-M67 (5.6%), and J2-M67 (2.2%).

Several conclusions may be drawn from Table 16 and Maps 87-91 for Iberia:

(1) The average total of the frequencies of Haplogroup R1b in Iberia is about 60%. The centre of this haplogroup (89%) is in the Basque area but the frequencies are quite high (75%) also in Catalonia. In most parts of the peninsula, the R1b-frequencies are about 50-60%. The lowest frequencies (about 43%) are in the Malaga district in the southeastern corner of Iberia. There is a north-south clinal gradient with higher values (over 80%) in the Northeast and lower values (close to 40%) in the Southeast. The high values of Haplogroup R1b reflect the Ice Age Iberian refuge, and the men of this haplogroup can be considered the original inhabitants of the peninsula (after the Neanderthals), the first of whom arrived there about 35 kya.

(2) The average of the frequencies of Haplogroup I in Iberia is about 10%. The geographic distribution of this haplogroup is peculiar in that the peak area (frequency about 33%) is in the middle of the peninsula (Castile) and the frequencies diminish as the distance from this area increases. In the next zone, frequencies vary from about 12% to about 15%; in the zone still further away from Castile, the frequencies are 6-9%, and in the zone furthest away from Castile (in the northwestern and southeastern corners of the peninsula) the frequencies are very low (0-3 %).

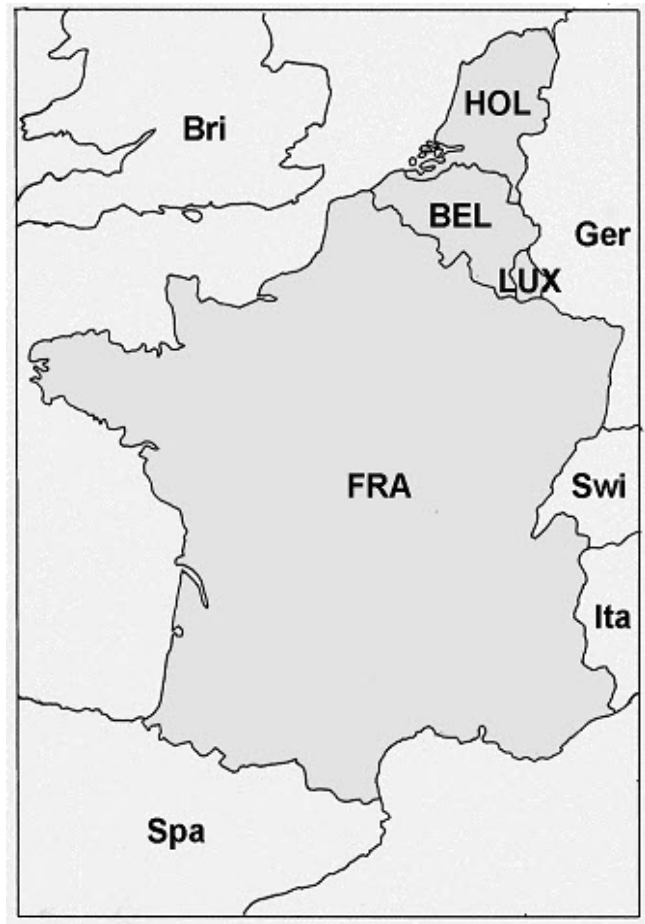
(3) The average total of Haplogroup E in the peninsula is about 12%. The haplogroup has two centres, one in Galicia in the northwest (frequency about 32%) and the other in Malaga in the southeast (frequency about 27%). The men of clan E are ultimately of African origin, but they have come to Iberia by two routes, some directly from Africa across the Mediterranean, and others round the Mediterranean and through the Middle East. The latter men were with the Early Farmers who brought agriculture to Iberia.



Map 91. Frequencies of Y-chromosome haplogroup K in the Iberian peninsula.

Haplogroups R1b and E, are, to some extent, complementary: The areas with high R1b values usually have low E values and vice versa. This dichotomy results from the population being originally R1b, while the E people are more recent immigrants.

(4) The history of Haplogroup J is similar to that of Haplogroup E in that it, too, belonged to the Early Farmers arriving from the Middle East, and, accordingly, Clan J, like Clan E, represents the newcomers to Iberia (not the original inhabitants of the Iberian refuge). The average total of Haplogroup J is about 9% in Iberia,



Map 92. The area defined as "Atlantic Europe" in this article (darker shaded area).

Table 17
Frequencies of the Y Haplogroups in Atlantic Europe.

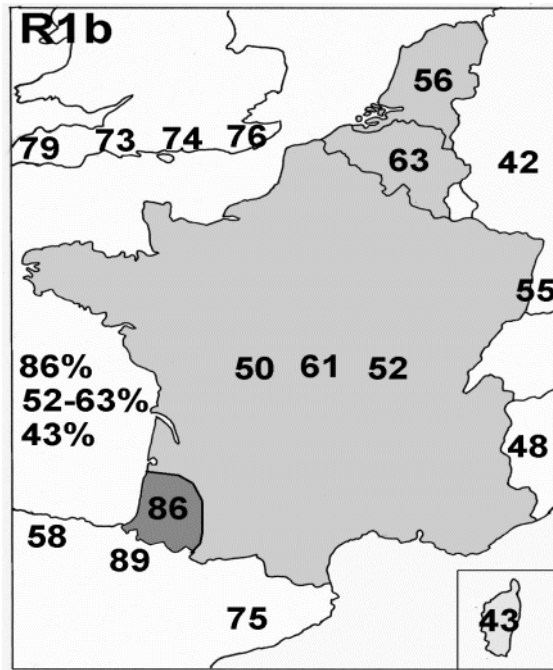
Population	R1b	R1a	I	OE	Eb3	J2	G2	EF	K	P	Q	Source
French a	61.0	0	13.6	74.6	6.7	10.1	3.4	20.2	0	3.4		Kivisild (1999)
French b	50.0	5.0	25.0	80.0	11.0	5.0	5.0	21.0	0	0	0	Rosser (2000)
French c	52.2	0	17.2	69.4	8.7	17.3	0	26.0	0	0	4.3	Semino (2000)
Basque, France	86.4	0	9.1	95.5	0	4.5	0	4.5	0	0	0	Semino (2000)
Belgians	63.0	4.0	23.0	90.0	2.0	5.0	1.0	8.0	1			Rosser (2000)
Dutch	60.0	4.8	25.7	90.5	2.4	6.2	0.8	9.4				Athey (2008)
Total	57	3	22	82	6	9	2	17				

Table 18
Frequencies of the Y Haplogroups in Three French Areas. The frequencies were determined by Kari Hauhio using YHRD data processed through Athey's Haplogroup Predictor (Athey, 2005, 2006).

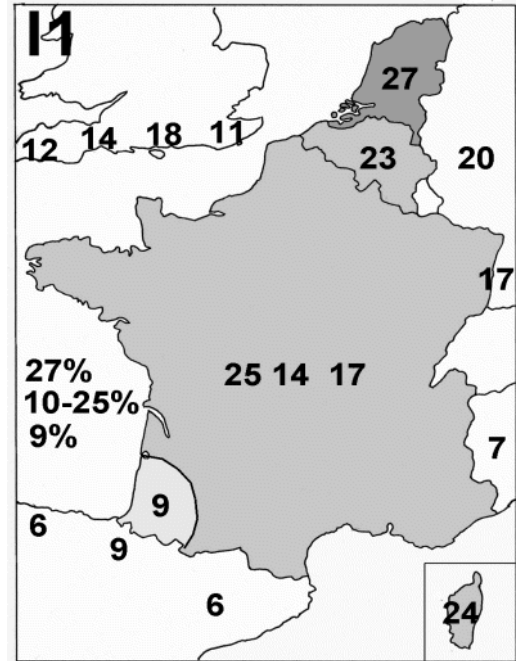
French Population	R1b	R1a	I	OE	E3b	J2	G2	EF	Q	J1	n
Strasbourg	67.6	11.8	17.6	97.0	2.9			2.9			99
Lyon	66.7	5.1	10.2	82.0	5.1	2.6	5.2	12.9	5.1		125
Paris	30.8	15.4	11.4	57.6	19.2	15.4		34.4		7.7	109
Total	55.0	10.8	13.1	78.9	9.1	6.0	1.7	16.7	1.7	2.6	333

and the frequencies form a south-north gradient. Map 89 has three zones: the frequencies of the southern zone are 14-18%, those of the middle zone 10-12% and those of the northern zone 0-9%. The gradient suggests that agriculture spread in Iberia from the south to the north.

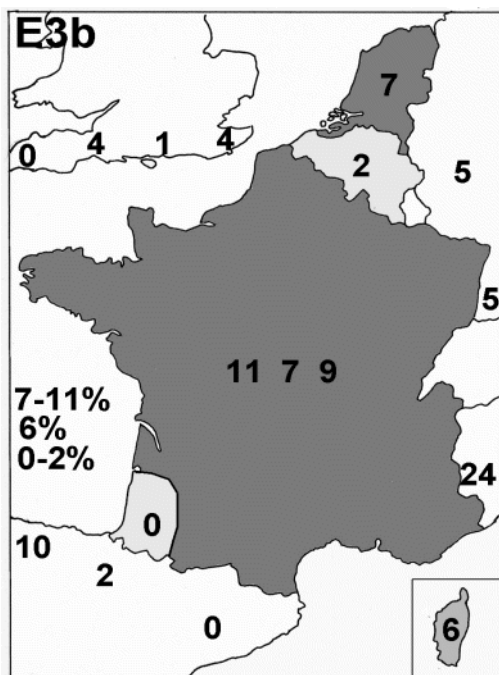
(5) Haplogroup K (x NO, P) is not very common in Iberia as shown in Map 91. Its average frequency in the whole peninsula is about 3%, most of which is K2. The maximum area is in the Cadiz area (west of Gibraltar) which may mean that the men of this haplogroup first arrived in southern Spain.



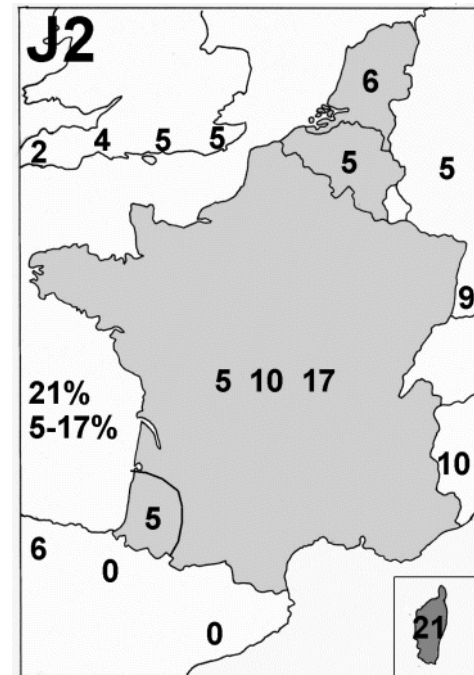
Map 93. Frequencies of Y Haplogroup R1b in Atlantic Europe.



Map 94. Frequencies of Y Haplogroup I in Atlantic Europe.



Map 95. Frequencies of Y Haplogroup E3b in Atlantic Europe.



Map 96. Frequencies of Y Haplogroup J2 in Atlantic Europe.

(1) The "Iberian" Haplogroup R1b is very frequent in Atlantic Europe. Its frequency is about 86% in the French Basque area and about 50-63% in the other continental areas of France and Belgium/Holland.

(2) In the Atlantic Europe area, the "Balkan" Haplogroup I is highest (about 27%) in Holland and slightly lower (14-25%) in Belgium and France; the I-frequency is lowest (below 10%) among the Basques of France. There is a north-south gradient from Holland through Belgium and France to the Basque area.

(3) The frequencies of Haplogroups E3b and J2 of the Early Farmers are spread more or less uniformly across the entire area under consideration. The average total of E3b is about 6% and that of J2 about 8%. The only exception seems to be the very low (0-2%) frequency of E3b in the Basque area.

8. The British Isles

The data for Britain and Ireland, considered here as "The British Isles," are from Capelli, et al. (2003), and

are summarized in Table 19. Maps 97-101 are based upon the table. The following generalizations about the British Isles data can be made:

(1) The European maximum area of the "Iberian" Haplogroup R1b is in Ireland. From there starts a west-east gradient that goes through Britain and continues on the Continent. So, for example, frequencies are about 95% in far northwest Ireland, 60% in the eastern parts of England, about the same in Belgium and Holland, about 30-45% in Germany, and about 10-20% in Poland.

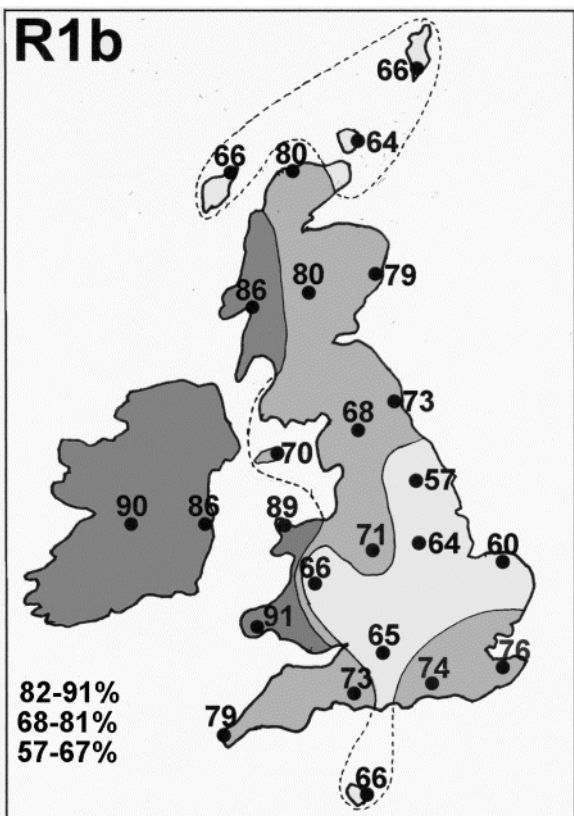
(2) The "Ukrainian" Haplogroup R1a is relatively rare in the British Isles: its frequencies are usually below 10%. There is, however, one exception: the frequencies are about 10-20% in the northern islands. This can be seen as an indication of Scandinavian influence: in Norway and Iceland, for example, the frequencies of R1a are usually above 20%.

(3) In the central parts of eastern England, the frequencies of Haplogroup I are higher (about 22-32%) than elsewhere in the British Isles (about 5-20%). This per-

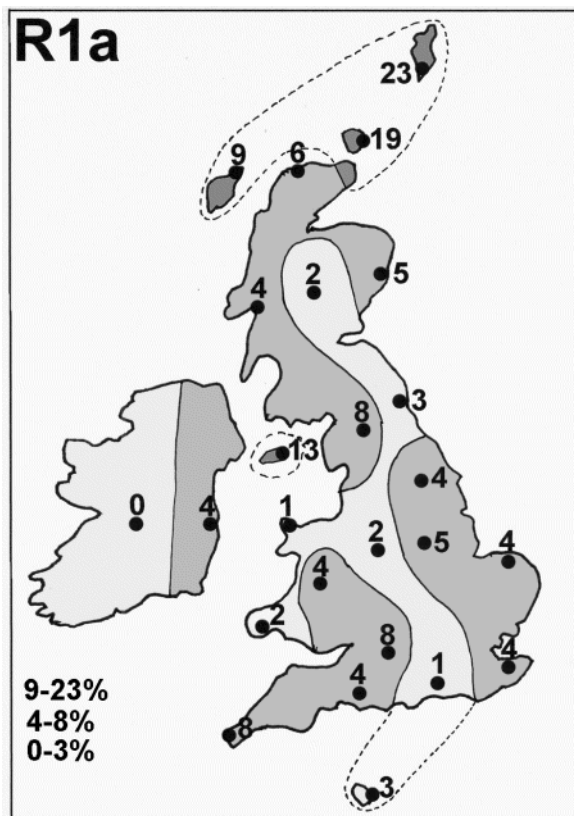
Table 19
Frequencies of Y Haplogroups in Twenty-Five Localities in the British Isles.
 The table is based on a simplification of Capelli et al. (2003).

Locality	R1b	R1a	I	OE	E3b	J	EF	G	P	K	n
Shetland	66	23	10	99			0				83
Orkney	64	19	15	98			0		2		121
Dumess	80	6	14	100			0				51
West. Isles	66	9	25	100			0				99
Stanehaven	79	5	13	97			0	2			44
Pitlochry	80	2	10	92		7	7				41
Oban	86	4	7	97			0	2			42
Morpeth	73	3	18	94		4	4	2			96
Penrith	68	8	18	94	3	2	5	1			90
Isle of Man	70	13	16	99	2		2				62
York	57	4	32	93	4		4	2			46
Southwell	64	5	28	87	6	6	12				70
Uttoxeter	71	2	18	91	4	4	8	1			84
Llanidloes	66	4	19	89	5	2	7	4			57
Llangefni	89	1	4	94	4	1	5			1	80
Rush	86	4	11	101			0				76
Castlerea	90	0	9	99			0				43
Norfolk	60	4	32	95	3	2	5				121
Haverfordw.	91	2	4	97	3		3				59
Chippenham	65	8	22	95		4	4	2			51
Faversham	76	4	11	91	4	5	9				55
Midhurst	74	1	18	93	1	5	6	1			80
Dorchester	73	4	14	91	4	4	8	1			73
Cornwall	79	8	12	99		2	2				52
Channel Isl.	66	3	24	93	4	2	6	2			128
Total	70.9	5.8	16.2	95.1	1.9	2.0	3.9	0.8	0.1	0.0	1863

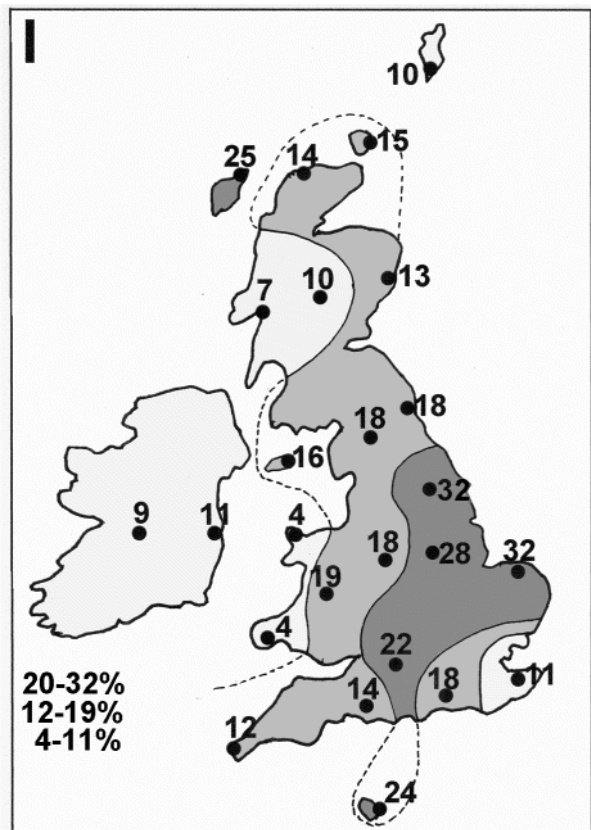
Note: Simplifications: JxJ2 + J2 = J; N3 removed (always zero); PxR removed (only Orkney 2); KxPNO removed (only Llangefni 1); R1a1 = R1a; R1xR1a1 = R1b; FxIJK =G.



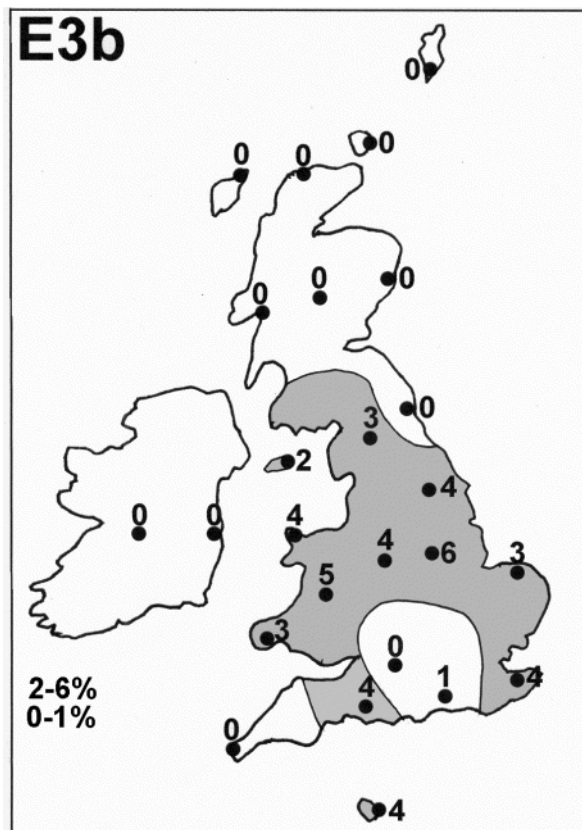
Map 97. Frequencies of Y Haplogroup R1b in the British Isles.



Map 98. Frequencies of Y Haplogroup R1a in the British Isles.



Map 99. Frequencies of Y Haplogroup I in the British Isles.



Map 100. Frequencies of Y Haplogroup E3b in the British Isles.

have we come from?" The special part played by languages was largely based on the concept of the nation state, according to which nations were decided on the basis of the languages they spoke: the French were primarily those who spoke French, the Estonians those who spoke Estonian, etc. The *point de départ* in looking for a common origin of peoples and languages, often even unquestioningly self-evident, was the idea that *gentem facit lingua*, or "language determines the nation." However, the methods used by linguists have their limitations, especially when it comes to time. The farthest back in time that linguists can go is usually regarded as 6,000-10,000 years. In other words, the study of language takes us back no farther than Mesolithic time; the Palaeolithic era remains, as far as the roots of peoples are concerned, completely unstudied.

(2) Later, especially in the 1970's and 1980's, archaeologists have joined the numbers of those interested in the origin of peoples. Archaeologists are able to gain reliable information from much earlier periods than 10,000 years ago but they have been cautious in pronouncing upon the origins of peoples. They often point to the fact that it is difficult to link archaeological cultures to languages. They lived at a time when one had to know what language a people spoke in order to be able to say who the ancestors of the present population were.

(3) A decisive change came about when geneticists started in the 1980's to seriously study peoples' roots. Now came the time when people's origins were decided according to their genes rather than the language they spoke. The geneticists, then, could construct two-dimensional trees for people, the one dimension of which was the degree of relativity between peoples and the other dimension time. These **phylogenetic trees** are in principle the same as the traditional **linguistic trees** of the linguists; the difference is that instead of languages genes are used to identify peoples.

In this article, it has been my purpose to define populations exclusively in genetic terms. To allow a comparison of the old linguistic way of defining populations and the new genetic way of defining them possible, I add some concluding remarks about the assumptions concerning the languages spoken by ancient European populations. The time depth in many cases goes beyond the limits of linguistic facts, which means that the statements in most cases are closer to assumptions than verified facts.

Hypotheses About the Ancient Languages in Europe

Very little is known about the languages of the ancient Europeans. Nevertheless, some hypotheses can be made about the languages. According to one view, each of the four Ice Age refuges had its own language; in addition, there were, of course, the languages of the southern populations of the Middle East and Africa that may have

exerted an influence on the language situation in Europe. In ancient times, there may have been many languages in Europe, now extinct, about which we know nothing.

Iberia

The most plausible candidates for the ancient languages of the Iberian refuge are the Basque languages still spoken by about half a million people in the Basque area of Spain and France. Earlier, there were several languages belonging to this language group, but mainly because of the intensive spread of IE languages in Western Europe, the area of the Basque languages has shrunk ever since. It is probable that the entire Atlantic Coast was linguistically Basque during the Last Glacial Maximum (LGM) and the millennia after it. The area was homogeneous also in respect to subsistence system and genetics: the men were reindeer hunters and their main Y-chromosome haplogroup was R1b.

Siberia

It is a commonly accepted idea that the languages of the "ancient mammoth hunters" of northeastern Europe and northwestern Siberia were Finno-Ugric. It is possible that all these men occupied the entire northernmost zone of Europe during the LGM and the period after it. The populations had a common subsistence system and they were genetically homogeneous: they were mammoth hunters and their main Y-chromosome haplogroup was N3.

The Ukraine

The men of the Ukrainian refuge, like those of the Siberian refuge, were mammoth hunters. They are commonly known for their houses made of mammoth bones. The language of these men may have the same Finno-Ugric language as that of the Siberian refuge; another alternative is that it was Indo-European. The IE branch in question was the East-European group consisting of the GBS (Germanic, Baltic, and Slavic) languages. These language spread later (during the recolonization of Northern Europe) to the northern zone then occupied by the FU speakers from the Siberian refuge. The result was a rather strong FU substratum in all the GBS languages. These men had arrived in Europe through the steppe area between the Ural Mountains and the Caspian Sea; before that, they had occupied Central Asia and the Middle East/India. The main Y-chromosome haplogroup of these men was R1a. Haplogroup R1a was derived from the Middle East Haplogroup F through mutations that gave rise to F > K > P > R > R1 > R1a.

The Balkans

The men of the Balkan refuge were more likely than those of any other to have spoken an early form of the Indo-European language. The IE language in question

would have given rise to the West-European group consisting mainly of the Greek-Italic-Celtic languages. One hypothesis is that IE languages were first brought to Europe by the Early Farmers, displacing what had previously been all non-IE languages, but a more probable scenario is that IE came much earlier with the Haplogroup I men. In either case, the languages of the European Haplogroup I men shifted to the IE languages of the Early Farmers during the Neolithic expansion out of Anatolia. Genetically, the men of the Balkans represent Haplogroup I, which is a further development from the Middle Eastern Haplogroup F.

Language Shifts

Contrary to the general way of thinking among traditional linguists, it is apparent that language shifts have been common during the time of modern man in Europe, and by comparing the genome and languages one can make detailed assumptions about the language shifts having taken place in Europe. At least the following eleven major language shifts seem to have occurred in Europe:

- (1) The men of the South-Slavic populations of the Balkans are genetically from the Balkan refuge with high frequencies of Haplogroup I1b1-P37, but linguistically they are from the Slavic group. In this case, the Balkan populations (whatever their original language) seem to have shifted their original language to a Slavic one. A strong indication of language shifts is offered by the existence of the "Balkan Sprachbund" consisting of a number of languages with unrelated vocabularies, but with similar grammatical and phonological features.
- (2) The emergence of the Romance languages is based on the language shift of the original local languages to Latin. Depending on the original local language, the resulting language was, for example, French, Spanish, Portuguese, Italian, or Romanian.
- (3) In Central and Northern Russia, the original Finno-Ugric languages were replaced by Russian. The original FU-speaking people learnt to speak Russian as their native language.
- (4) An equivalent language shift took place in the Baltic area. The Latvian and Lithuanian men are genetically partly from the Siberian refuge with a high N3 frequency and partly from the Ukrainian refuge with high R1a. As the languages of this area are today Baltic, a language shift or, more precisely, a linguistic assimilation, must have taken place. The men who came originally from the Siberian refuge must have shifted their Finno-Ugric language to a Baltic one: The Baltic area consisted earlier of two genetic types (N3 and R1a) and two linguistic types (Finno-Ugric and Indo-European/Baltic); later the two linguistic types were assimilated but the two Y types, of course, remained separate.
- (5) A similar language shift may have taken place in northern Central Europe in the area of Proto-Germanic. At least part of this area was earlier inhabited by men representing Haplogroup N3, and the language may have been Finno-Ugric. The present Germanic languages (such as Slavic and Baltic) have a strong Finno-Ugric substratum (Wiik 2002).
- (6) Modern Hungarian men are genetically similar to other Central Europeans with high R1b, R1a, and I1b frequencies, but their language is Finno-Ugric. The genetic-linguistic discrepancy can be solved by assuming language shifts in which the local Pannonian men accepted the Hungarian language of the newcomers as their native language. The newcomers were the horse-riding hordes that came from the southern Ural Mountains to Pannonia in 500-895 AD. Hungarian men came from three different refuges: Iberia, resulting in high R1b frequencies, the Balkans, resulting in significant I1b1-P37 frequencies, and the Ukraine, resulting in high R1a frequencies. Linguistically, all speak the same Hungarian language and cannot be distinguished on this basis.
- (7) Before the arrival of the Angles and Saxons, the language of most of those living in the British Isles was Celtic. Today, Celtic languages are spoken only in the most remote areas of Ireland, Wales, and Scotland. A wave of language shift Celtic > English has swept over the British Isles during the last approximately sixteen hundred years.
- (8) A similar language shift that wiped off the Celtic language from Central Europe was more effective than the one in the British Isles. According to the language shift in question, a majority of the Central European Celts learned to speak a Germanic language and a minority learned to speak a West-Slavic (Polish, Czech, and Slovak) language.
- (9) In central and northern Finland, the speakers of the Saami language shifted their language to Finnish; only the most northern Saami retained their original language and still today speak Saami.
- (10) The Samoyeds of Northeastern Europe are genetically different from all other Europeans, but their language is Uralic and related, for example, to Finnish. The complicated genetic-linguistic situation is probably a result of a language shift in which the Samoyeds came into close contact with populations speaking a Finno-Ugric language. The result was a new language group, the Samoyedic languages, that are related to the Finno-Ugric languages. Traditionally, the Finno-Ugric and Samoyedic languages are regarded as "Uralic."

(11) In the Volga area, genetic and linguistic assimilation and mingling has been common in the Finno-Ugric and Turkic populations. The FU populations of the area are Mari, Mordvians, and Udmurtians and the Turkic ones are Tatars and Chuvash, and to a lesser extent Bashkirians.

Web Resources

<http://www.hprg.com/hapest5/>
Haplogroup Predictor Program

References

- Athey TW (2005) Haplogroup prediction using an allele-frequency approach. *J Genet Genealogy*, 1:1-7.
- Athey TW (2006) Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J Genet Genealogy*, 2:34-39.
- Athey TW (2008) Private communication. Dutch YHRD data from five localities were pooled (n = 275) and analyzed using the Haplogroup Predictor program.
- Balanovsky O, Rootsi S, Pshenichnov A, Kivisild T, Chumosov M, Evseeva I, Pocheshkhova E, Boldyreva M, Yankovsky N, Balanovska E, Villems R (2008) The sources of the Russian patrilineal heritage in their Eurasian context. *Am J Hum Genet*, 82:236-250.
- Barač L, Peričić M, Klarić IM, 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. *Eur J Hum Genet*, 11:535-542.
- Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MP, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N, Goldstein DB (2003) A Y chromosome census of the British Isles. *Curr Biol*, 13:979-984.
- Cordaux R, Deepa E, Vishwanathan H, Stoneking M (2004) Genetic Evidence for the Demic Diffusion of Agriculture to India. *Science*, 304:1125.
- DiGiacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, Cresta M, Cucci F, DiStasi L, Agostiano V, Giparaki M, Loutradis A, Mammi C, Michalodimitrakis EN, Papola F, Pedicini G, Plata E, Terrenato L, Tofanelli S, Malaspina P, Novelletto A (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.
- Dupuy BM, Stenersen M, Lu TT, Olaisen B (2005) Geographical heterogeneity of Y-chromosomal lineages in Norway. *Foren Sci Int*, 164: 10-19.
- Flores, C et al (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Gen*, 12, 855-863.
- Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calo CM, Ghianni ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalliforza LL (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthropol*, 121:270-279.
- Karlsson AO, Wallerström T, Götherström A, Holmlund G (2006) Y-chromosome diversity in Sweden – A long-time perspective. *Eur J Hum Genet*, 14:863-970.
- Kasperavičiūtė D, Kucinkas V, Stoneking M (2004) Y chromosome and mitochondrial DNA variation in Lithuanians. *Ann Hum Genet*, 68:438-452.
- Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawłowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet*, 117:428-443.
- King R, Underhill PA (2002) Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. *Antiquity*, 76:707-714.
- Kivisild T et al (1999) An Indian Ancestry: a Key for Understanding Human Diversity in Europe and Beyond. *Archaeologies: DNA and the population prehistory of Europe.* McDonald Institute Monographs, Cambridge, UK.
- Laitinen V, Lahermo P, Sistonen P, Savontaus ML (2002) Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers. *Hum Hered*, 53:68-78.
- Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, Lahermo P (2006) Regional differences among the Finns: A Y-chromosomal perspective. *Gene*, 376:207-215.
- Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, Achilli A, Drobnic K, Andjelinovic S, Torroni A, Santachiara-Benerecetti AS, Semino O (2005) The Peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *Ann Hum Genet*, 69:757-763.
- Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M (2004) Mitochondrial and Y-chromosome variation in the Caucasus. *Ann Hum Genet*, 68:205-221.
- Periè M, Baraè LL, Martinović KI, Janićijević B, Rudan P (2005) Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. *Croat Med J*, 46:502-513.

[Rootsi, S et al \(2002\) The Roots of Peoples and Languages of Northern Eurasia IV. *Societas Historiae Fenno-Ugricae*. Oulu.](#)

[Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, VILLEMS R, Semino O \(2004\) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet*, 75:128-137.](#)

[Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumäe AM, Khusnutdinova EK, Voevoda MI, Osipova LP, Stoneking M, Lin AA, Ferak V, Parik J, Kivisild T, Underhill PA, VILLEMS R \(2006\) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeastern Asia towards Europe. *Eur J Hum Genet*, 15:204-11.](#)

[Rosser ZH, Zerjal T, Hurler ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côté-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskis V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, VILLEMS R, Tyler-Smith C, Jobling MA \(2000\) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J Hum. Genet*, 67:1526-1543.](#)

[Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M,](#)

[VILLEMS R, Kashyap VK \(2006\) A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proc Natl Acad Sci \(USA\)*, 103:843-848.](#)

[Saukkonen P \(2006\) *Suomalais-ugrialaisten kansojen ja kielten alkuperäongelma*. Yliopistopaino, Helsinki.](#)

[Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, Latini V, Varesi L, Memmi MM, Romano V, De Leo G, Gennarelli M, Jaruzelska J, VILLEMS R, Parik J, Macaulay V, Torroni A \(2001\) Human Y-Chromosome Variation in the Western Mediterranean Area: Implications for the Peopling of the Region. *Human Immunology*, 62:871-884.](#)

[Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA \(2000\) The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science*, 90:1155-1159.](#)

[Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CT, Lin AA, Mitra M, Sil SK, Ramesh A, Rani MVU, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA \(2006\) 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 Gen*, 78:202-221.](#)

[Underhill PA, Kivisild T \(2007\) Use of Y Chromosome and Mitochondrial DNA population structure in Tracing Human migrations. *Annu Rev Genet*, 41:539-564.](#)

[Wells RS, Yuldashева N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF \(2001\) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci \(USA\)*, 98:10244-10249.](#)