Structure of the Y-haplogroup N1c1 updated to 67 markers

Jaakko Häkkinen, 27th December 2011 (updated 17th January 2012)

This is a 67 marker update and addition to the older 9–12 marker haplotype analysis which can still be found in www.mv.helsinki.fi/home/jphakkin/N1c.pdf. Here I will first reconsider the N1c1 founder haplotype and then stratify some of the N1c1 groups according to their shared mutations.

Achtung! Several geneticists mark the locus DYS389I with the values smaller by three than the values used by the majority: their 10 corresponds to the common 13. This is due to the fact that only certain part of this locus varies, while the other part is stable:

Allele* (Repeat #)	Set 1	Set 2	Repeat Structure	Ref.
DYS389I				
9(6)*	235	143	$[TCTG]_3[TCTA]_6$	
10(7)	239	147	$[TCTG]_3[TCTA]_7$	560,782
11(8)	243	151	$[TCTG]_3[TCTA]_8$	560,782
12(9)	247	155	$[TCTG]_3[TCTA]_9$	560,782
13(10)	251	159	$[TCTG]_3[TCTA]_{10}$	560,782
14(11)	255	163	$[TCTG]_3[TCTA]_{11}$	560,782
15(12)	259	167	$[TCTG]_3[TCTA]_{12}$	560,782
16(13)	263	171	$[TCTG]_3[TCTA]_{13}$	560,782

http://www.cstl.nist.gov/strbase/str_y389.htm

Similar occurrences are seen also in DYS437, in which the value can be 6 smaller, the value 8 corresponding to the common 14. Unfortunately these geneticists tend not to inform a reader explicitly about their exceptional habit, which may lead to misinterpretations. Consequently, unlike I stated earlier, the East Asian N1c1 haplotypes seem to be closer to the European haplotypes: the mutation DYS389I = 14 > 11 has not occurred at all, but the confusion is caused by this different marking or repeats.

1. Reconstructing the N1c1 founder haplotype

To reconstruct the N1c1 founder haplotype (FHT) from which all the other N1c1 haplotypes descend, we must take into account both the ancestral (older) and derived (younger) values. The ancestral values can be found by comparing N1c1 to its brother lineage N1b; unfortunately there are not yet any 67 marker N1a haplotypes. The following N1b founder haplotype has been reconstructed by comparing N1b-A and N1b-E, and at the 9–12 marker resolution also some N1a and N* haplotypes. In the N1c1 founder haplotype we can reconstruct those values which are found both in the N1b FHT and in most of the N1c1 groups.

Here is the N1b FHT; mutations differing from the N1c1 FHT are already marked, and the *loci* are in the FamilyTreeDNA order:

393,390,19,391,385a,b,426,388,439,389I,392,389b,458,459a,b,455,454,447,437,448,449,464a,b,c,d,460,Y-gata,YCAIIa,b,456,
607,576,570,CDYa,b,442,438,531,578,395S1a,b,590,537,641,472,406S1,511,425,413a,b,557,594,436,490,534,450,444,481,520,446,
617,568,487,572,640,492,565

N1b FHT (*reconstructed*) GD 20/67

13 23 14 **10** 12 13 11 12 10 **13** 14 16 **16** 09 09 11 12 **28** 14 **18** **27** **13** 14 15 **17** 11 11 18 **19** **15**
14 17 **17** **36** **40** 12 10 11 08 15 **16** 08 11 10 08 **10** 11 12 **20** **23** **13** 10 12 12 17 07 13 **21** 21 **14**
12 11 10 11 11 12 11

The N1c1 founder haplotype:

N1c1 FHT (*reconstructed*)

13 23 14 11 12 13 11 12 10 14 14 16 17 09 09 11 12 24 14 19 28 14 14 15 15 11 11 18 20 14
15 17 19 35 35 12 10 11 08 15 17 08 11 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

N1c1 was born from N1c somewhere in the Southern Siberia, and it seems to have split first into two groups: the Asian (or eastern) group and the European (or western) group.

N1c1 Asian (*reconstructed*) GD 5/67

13 23 14 11 12 13 11 12 **11** 14 14 16 **18** 09 09 11 12 24 14 **20** **29** 14 14 15 15 11 **10** 18 20 14
15 16 19 35 35 12 10 11 08 15 17 08 11 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

The Asian group further divided into two groups: the Altaian group and the Siberian group.

N1c1 Altaian (*reconstructed*) TGD 17/67

13 23 **15** 11 12 **12** 11 12 **11** 14 14 16 **18** 09 09 11 12 24 14 **20** **29** 14 14 15 15 **12** **10** 18 20 14
15 **17** 19 35 35 12 10 11 08 **14** 17 08 **10** 10 08 11 10 12 **24** **24** 14 10 12 12 17 **08** 13 20 **22** **16**
12 11 10 **10** 11 12 11

N1c1 Siberian (*reconstructed*) TGD 16/67

13 23 14 **10** **11** **14** 11 12 **11** 14 14 16 **18** 09 09 11 12 **25** 14 **20** **30** **16** **16** **16** **16** 11 **10** 18 20 14
15 16 **20** 35 35 12 10 11 08 15 17 08 11 10 08 **12** 10 12 21 **21** 14 10 12 12 **15** 07 13 20 21 15
12 11 10 11 11 12 11

The Siberian group can be divided at least into two subgroups: the East Asian group and the Volgan group.

N1c1 East Asian (*reconstructed*) TGD 20/67

13 23 14 **10** **11** **14** 11 12 **11** 14 14 16 **19** 09 09 11 12 **25** 14 **20** **30** **16 16** **16 16** 11 **10** 18 20 14
15 16 **21** 35 35 12 10 11 08 15 17 08 11 10 08 **12** 10 12 21 **21** **15** 10 12 12 **15** 07 13 20 21 15
12 **12** 10 11 11 12 11

N1c1 Volgan (*reconstructed*) TGD 18/67

13 23 14 **10** **11** **14** 11 12 **10** 14 14 16 **18** **10 10** 11 12 **25** 14 **20** **30** **16 16** **16 16** 11 **10** 18 20 14
15 16 **20** 35 35 12 10 11 08 15 17 08 11 10 08 **12** 10 12 21 **21** 14 10 12 12 **15** 07 13 20 21 15
12 11 10 11 11 12 11

There are still so few Asian haplotypes at 67 markers that I will not handle this eastern branch any further. The second main group is the European group, which is almost identical with the N1c1 founder haplotype:

N1c1 European FHT (*reconstructed*) TGD 4/67

14 23 14 11 **11** 13 11 12 10 14 14 16 17 09 09 11 12 **25** 14 19 28 14 14 15 15 11 11 18 20 14
15 16 19 35 35 12 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

The European group has been divided into several groups, of which I have chosen five of the most distinguishable sungroups for the comparison:

N1c1 Spanish (50530)

13 23 **12** 11 11 13 11 12 10 14 14 16 17 09 09 11 12 24 14 19 28 14 14 15 15 11 11 18 20 14
15 16 **17** 36 36 **14** 10 11 08 15 17 08 08 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

N1c1 Scandinavian (120671 etc.)

14 23 14 **10** 11 13 11 12 **09** 14 14 16 **16** 09 09 11 12 **25** 14 19 28 14 14 15 15 11 **10** 18 20 14
15 16 19 38 38 **14** 10 11 08 15 17 08 08 10 08 **10** 10 12 21 22 14 10 12 12 **19** 07 13 **21** **22** **16**
12 11 10 11 11 12 11

N1c1 Balto-Polish (156293, N42695 etc.)

14 23 **15** 11 11 **14** 11 12 10 **13** 14 16 17 09 09 11 12 **25** 14 19 28 14 14 15 15 11 11 18 20 14
15 **17** 19 36 36 **13** 10 11 08 15 17 08 08 10 08 11 10 12 21 22 14 10 12 12 **16** 07 13 20 21 **16**
12 11 10 11 11 12 11

N1c1 Turkic (176969)

14 23 14 11 11 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **29** **15 15** 15 15 11 11 18 20 14
16 **16** **18** 35 35 12 10 11 08 15 17 08 **08** **11** 08 **12** **09** 12 21 22 14 10 12 12 **18** 07 **14** **21** 21 15
12 11 10 11 **12** **13** 11

N1c1 Savonian (141818 etc.)

14 **24** 14 11 11 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 ***31*** **13 13** **14 14** 11 11 18 **18** 14
15 **18** **18** 38 38 12 10 11 08 15 17 08 **08** 10 08 11 **11** 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

Slowly mutating markers:
Orange = Common to all 5 European groups (2 mutations: DYS385a, DYS537)
Red = Common to Scandinavian, Balto-Polish, Turkic and Eastern Finnish (2 mutations: DYS393, DYS447)
Violet = Common to Spanish, Scandinavian, and Balto-Polish (1 mutation: DYS442)
Light red = Common to Scandinavian and Balto-Polish (1 mutation: DYS446)
Green = Common to Turkic and Eastern Finnish (3 mutations: DYS459, DYS449, DYS570 [faster])

Rapidly mutating markers:
Blue = Common to Turkic and Scandinavian (2 mutations: DYS534, DYS481 [slower])
Light blue = Common to Spanish, Scandinavian, Balto-Polish and Eastern Finnish (1 mutation: CDY)
Light green = Common to Scandinavian and Eastern Finnish (1 mutation: CDY)
Light brown = Common to Spanish, Scandinavian and Turkic (1 mutation: DYS576)

If we would ignore the rapidly mutating markers, which usually vary also within a group or even a subgroup, there is only one contradiction against the perfect phylogeny: the Spanish group shares the increasing mutation in DYS442 (12 > 13 ~ 14) with the Scandinavian and Balto-Polish groups. The most economical explanation would be that this single piece of counter-evidence is due to an independent parallel development in the Spanish group, but as the Spanish group shares the SNP-mutation L550+ with the Scandinavian and Balto-Polish groups, we must now (17th January 2012 onwards) derive it from the same founder haplotype than these two groups. It seems that the most economical STR-based solution was not correct this time.

Thus the European founder haplotype seems to be divided into the western (Central European) and eastern (East European) group:

N1c1 Central European FHT (*reconstructed*) TGD 6/67
14 23 14 11 **11** 13 11 12 10 14 14 16 17 09 09 11 12 **25** 14 19 28 14 14 15 15 11 11 18 20 14
15 16 19 **36 36 14** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

It was divided at least into two subgroups, the Southwest European (Spanish) group and the North European group:

N1c1 Spanish (50530) TGD 10/67
13 23 **12** 11 **11** 13 11 12 10 14 14 16 17 09 09 11 12 **24** 14 19 28 14 14 15 15 11 11 18 20 14
15 16 **17** **36 36 14** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

N1c1 North European FHT (*reconstructed*) TGD 7/67
14 23 14 11 **11** 13 11 12 10 14 14 16 17 09 09 11 12 **25** 14 19 28 14 14 15 15 11 11 18 20 14
15 16 19 **36 36 14** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 **16**
12 11 10 11 11 12 11

Descendants of the North European founder haplotype are (among others) the Scandinavian group and the Balto-Polish group:

N1c1 Scandinavian (120671 etc.) TGD 16/67
14 23 14 **10** **11** 13 11 12 **09** 14 14 16 **16** 09 09 11 12 **25** 14 19 28 14 14 15 15 11 **10** 18 20 14
15 16 19 **38 38 14** 10 11 08 15 17 08 **08** 10 08 **10** 10 12 21 22 14 10 12 12 **19** 07 13 **21 22 16**
12 11 10 11 11 12 11

N1c1 Balto-Polish (156293, N42695 etc.) TGD 13/67
14 23 **15** 11 **11** **14** 11 12 10 **13** 14 16 17 09 09 11 12 **25** 14 19 28 14 14 15 15 11 11 18 20 14
15 **17** 19 **36 36** **13** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 **16** 07 13 20 21 **16**
12 11 10 11 11 12 11

Another European main group is the East European group:

N1c1 East European FHT (*reconstructed*) TGD 7/67
14 23 14 11 **11** 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **29** 14 14 15 15 11 11 18 20 14
15 16 **18** 35 35 12 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

Descendants of the East European founder haplotype are the Turkic group, the Ladogan group and the Finlandian group:

N1c1 Turkic (176969) TGD 17/67
14 23 14 11 **11** 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **29** **15 15** 15 15 11 11 18 20 14
16 16 **18** 35 35 12 10 11 08 15 17 08 **08** **11** 08 **12** **09** 12 21 22 14 10 12 12 **18** 07 **14** **21** 21 15
12 11 10 11 **12** **13** 11

N1c1 Ladogan FHT (*reconstructed*) TGD 12/67
14 **24** 14 11 **11** 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **30** **13 13** 15 15 11 11 18 20 14
15 **18** **18** **37 37** 12 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

N1c1 Finlandian FHT (*reconstructed*) TGD 9/67
14 23 14 11 **11** 13 11 12 10 **13** 14 16 17 **10 10** 11 12 **25** 14 19 **29** 14 14 15 15 11 11 18 20 14
15 16 **18** 35 35 **14** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

Descendants of the Ladogan founder haplotype are the Karelian group and the Savonian group:

N1c1 Karelian FHT (*reconstructed*; see N36698) TGD 15/67
14 **24** 14 11 **11** 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **30** **13 13** **14 14** 11 11 18 20 14
15 **18** **19** **37 37** 12 10 11 08 15 17 08 **09** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

N1c1 Savonian FHT (*reconstructed*; see 159419) TGD 13/67
14 **24** 14 11 **11** 13 11 12 10 14 14 16 17 **10 10** 11 12 **25** 14 19 **30** **13 13** 15 15 11 11 18 **18** 14
15 **18** **18** **37 37** 12 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 15
12 11 10 11 11 12 11

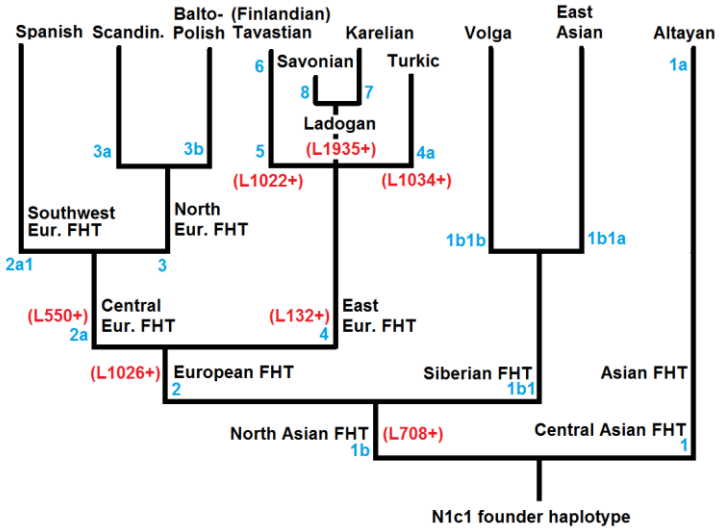
Descendants of the Finlandian group are (among others) the Bothnian group and the Tavastian group:

N1c1 Bothnian FHT (*reconstructed*; see N55552) TGD 14/67
14 23 14 **10** **11** 13 11 12 10 **13** 14 16 17 **10 10** 11 12 **25** 14 19 **28** 14 14 15 15 11 11 18 20 14
15 **15** **18** 35 35 **14** 10 **12** 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 17 07 13 20 21 **14**
12 11 10 11 11 12 11

N1c1 Tavastian FHT (*reconstructed*; see N61541) TGD 13/67

14 24 14 11 **11** 13 11 12 10 **13** 14 16 17 **09 09** 11 12 **25** 14 19 **29** 14 14 15 15 11 11 18 20 14 15 16 **19** 35 35 **14** 10 11 08 15 17 08 **08** 10 08 11 10 12 21 22 14 10 12 12 **18** 07 13 20 21 15 12 11 10 11 11 12 11

This structure can be taken as the basis for the phylogenetic tree of the haplogroup N1c1 – other groups should be easily positioned in the relevant point of the tree.



I will be constantly updating my phylogenetic haplotree in http://www.mv.helsinki.fi/home/jphakkin/N1c1Eng67.xps (or .pdf), so I will not go into details here; suffice to mention that the codes (2b2a etc.) are not final but only serve to illustrate the inner structure of the haplogroup N1c1; they may change when the resolution grows more and more accurate. This updated structure agrees with the new SNP results: L550+ seems to be restricted to the Central European group, and L132+ seems to be restricted to the East European group; Turkic group is yet to be tested, but it can already be predicted positive, because Ladogan, Finlandian and ancestral East European haplotypes also have L132+.

2. N1c1 migrations in time and place

This kind of structure can naturally be interpreted also geographically: it seems that the first split of the haplogroup N1c1 occurred somewhere between the Ural Mountains and the Altai–Sayan Mountains – around the southern part of West Siberia. The eastern group is mainly restricted to Asia, while the western group is mainly restricted to Europe.

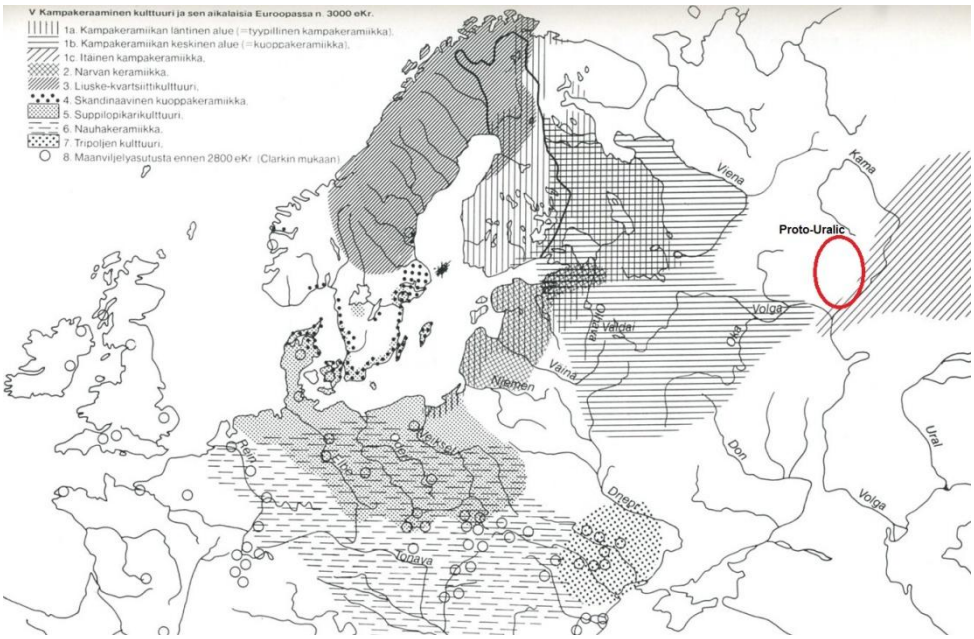
There are so far only very few 67 marker haplotypes from the Asian group, so I will not handle it here any further. Instead, the multitude of European haplotypes allows a more accurate sketch. The first split seems to have occurred between the western Central European group, and the eastern East European group. This split probably occurred somewhere near the Upper Volga basin or in any case east of Baltia.

The spread of Comb Ceramics Culture to the Fenno-Baltia about 6 000 years ago was accompanied by many changes in the material culture and dwellings, which hints to the considerable movement of people. Remarkably, the two westernmost Comb Ceramic areas seem to correlate rather well with the division of the European N1c1: the Comb-Pit Ware (or Typical Combed Ware, **1a**) was present in Baltia, Finland and Karelia, while the Pit-Comb Ware (**1b**) was present in Karelia and Northwest Russia up to the Northern Dvina (in

map: Viena), and in the Upper Volga area, from Lake Peipus in the west to the Volga–Kama fork in the east and Upper Dnieper in the south.

The Balto-Polish group can be derived from the southwestern most part of the Comb-Pit Ware area, while the Scandinavian group can be derived from the Scandinavian Pitted Ware (**4**) area; this ceramic style is a bit younger and seem to have possibly received influence from the Comb-Pit Ware of Baltia or Finland. Perhaps it will be found also some Finnish subgroup deriving from the Central European group.

The true Rurikids (members of noble Russian families) belong to the Scandinavian group, and their closest relatives are found in the coastal Finland, among the Swedish-speaking Finns. Their brother group (clan of Tawast–Räihä) is found among the Western Finns.



(**Map modified from:** Huurre, Matti 2004: *9000 vuotta Suomen esihistoriaa*)

Of the descendants of the East European group (in the **1b** area), the Turkic group could have been born in the eastern end of the Pit-Comb Ware area – it spread to the east, behind the Urals and back to Asia, where it developed into the Turkic group. This group was again present in the Altai–Sayan area when the Turkic expansion started about 2 000 years ago. The Ladogan group was probably born near Lake Ladoga. The exact route of expansion for the Finlandian group is not yet known; it may have been spread through Ingermanland and Estonia, or through the Karelian Isthmus. The latter option seems so far improbable, because all the old subgroups are found exclusively in Western Finland. Hopefully we will get more Estonian haplotypes which would enlighten the situation.

It must be noted that there are no traces of any Uralic language in Northeast Poland, and also the new datings for the expansion of Proto-Uralic (ca. 2000 BC) are about two millennia too late to be connected to the spread of the Comb-Pit Ware. Therefore we must exclude the westernmost area, the Comb-Pit Ware (**1a**) from the area of the Uralic languages, and even in the Pit-Comb Ware area (**1b**) the Uralic languages seem to be later newcomers: the original area of Proto-Uralic is on the linguistic basis located on the northern side of the Volga bend and Lower Kama – just in the gap between the central and eastern area of the Comb Ceramics (Häkkinen, Jaakko 2009: Kantauralin ajoitus ja paikannus: perustelut puntarissa. http://www.sgr.fi/susa/92/ hakkinen.pdf).

Only at a later stage can we with slightly greater assurance connect the westernmost Uralic languages to the Finlandian and Ladogan groups: Proto-Finnic to the former, if it is found in Estonia, and Proto-Saami perhaps to certain subgroups of the latter. More about these Finnish-dominated groups can be found in http://www.mv.helsinki.fi/home/jphakkin/ N1c1Eng67.xps (or .pdf).

It must be emphasized that there is no more reason to connect the Comb Ceramic Culture to the spread of the Uralic languages, but it still can be connected to the spread of N1c1 to the Baltic Sea region. No earlier than one millennium later the Corded Ware Culture spread to Baltia, Southwest Finland and Southern Sweden, supposedly spreading the Northwest Indo-European dialect. Only after this wave the N1c1 men of the area (as well as those of any other haplogroup) could have begun to speak an Indo-European language. Yet it is possible that the North European group of N1c1 participated also in the Corded Ware expansion from Poland to the more northern areas, even though the main bulk of the Corded Ware men seem to have been R1a1.