# Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families

Anne-Mai Ilumäe,[1,15] Maere Reidla,[1,15] Marina Chukhryaeva,[2,11,15] Mari Järve,[1] Helen Post,[1] Monika Karmin,[1] Lauri Saag,[1] Anastasiya Agdzhoyan,[2] Alena Kushniarevich,[1,3] Sergey Litvinov,[4,1] Natalya Ekomasova,[5] Kristiina Tambets,[1] Ene Metspalu,[1] Rita Khusainova,[4] Bayazit Yunusbayev,[1] Elza K. Khusnutdinova,[4,5] Ludmila P. Osipova,[6,7] Sardana Fedorova,[8] Olga Utevska,[9] Sergey Koshel,[10] Elena Balanovska,[11] Doron M. Behar,[1] Oleg Balanovsky,[2,11] Toomas Kivisild,[1,12] Peter A. Underhill,[13] Richard Villems,[1,14] and Siiri Rootsi[1,*]

The paternal haplogroup (hg) N is distributed from southeast Asia to eastern Europe. The demographic processes that have shaped the vast extent of this major Y chromosome lineage across numerous linguistically and autosomally divergent populations have previously been unresolved. On the basis of 94 high-coverage re-sequenced Y chromosomes, we establish and date a detailed hg N phylogeny. We evaluate geographic structure by using 16 distinguishing binary markers in 1,631 hg N Y chromosomes from a collection of 6,521 samples from 56 populations. The more southerly distributed sub-clade N4 emerged before N2a1 and N3, found mostly in the north, but the latter two display more elaborate branching patterns, indicative of regional contrasts in recent expansions. In particular, a number of prominent and well-defined clades with common N3a3'6 ancestry occur in regionally dissimilar northern Eurasian populations, indicating almost simultaneous regional diversification and expansion within the last 5,000 years. This patrilineal genetic affinity is decoupled from the associated higher degree of language diversity.

## Introduction

Northern Eurasia extends from Scandinavia in the west to Beringia in the east, and present-day Siberia has witnessed modern humans for at least 45,000 years.[1] One of the most prevalent paternal lineages in the northern temperate zone of Eurasia is Y chromosome haplogroup (hg) N, which ranges from the Eurasian Beringia and Amur region in the Russian Far East across northern China and Japan to eastern Europe. It occurs in many Eurasian populations with highly variable cultural, linguistic, and autosomal genome-wide backgrounds.[2–6] Its phylogenetic neighbor clade hg O exhibits a remarkably different distribution pattern: covering Southeast Asia at high frequency and extending to Sunda and Oceania.[3,5,7–9] The basal NO-M214* lineage, the predecessor of N and O, co-distributes with hg O in continental Southeast Asia, albeit at a low frequency.[3,5]

Ancient human DNA from archaeological sites provides complementary pre-historical insights into the evolutionary trajectory of hg N. Analyses of prehistoric specimens suggest that it was the predominant paternal hg in northeast China during the Neolithic period 6.5 thousand years ago (kya)[10] and then declined gradually throughout the Bronze Age up to 2.7 kya.[11,12] The earliest finding of hg N in Europe comes from Iron Age Hungary,[13] where this hg is virtually absent today.

It remains unclear as to what demographic processes underlie the present-day distribution of hg N. This study attempts to elucidate these processes in greater detail.

Although hg N studies began in 1997 with the discovery of the first hg-defining marker[2] and have continued thereafter,[4–6,14–24] only recently have efforts in re-sequencing Y chromosomes begun to shed light on the deeper sub-structure of the hg.[25–30]

Here, we resolve successive branching events at a fine scale within Eurasian hg N by using a total of 94 high-coverage Y chromosome sequences from hg N, consisting of 51 sequences from Karmin et al.[29] and 43 new BigY-captured samples sequenced on an Illumina platform. We present a precise time-calibrated hg N phylogeny covering nearly all known hg N sub-clades, defined according to previously proposed nomenclature rules designed to better accommodate the growing number of whole Y chromosome sequences.[29]

We concentrate on reconstructing in detail the important elements regarding the interior lineal sub-structure

[1]Estonian Biocentre and Department of Evolutionary Biology, University of Tartu, Tartu 51010, Estonia; [2]Vavilov Institute of General Genetics, Moscow 119991, Russia; [3]Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk 220072, Republic of Belarus; [4]Institute of Biochemistry and Genetics, Ufa Scientific Center, Russian Academy of Sciences, Ufa 450054, Russia; [5]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa 450074, Russia; [6]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia; [7]Novosibirsk State University, Novosibirsk 630090, Russia; [8]North-Eastern Federal University and Yakut Research Center of Complex Medical Problems, Yakutsk 677010, Sakha Republic, Russia; [9]V.N. Karazin Kharkiv National University, Kharkiv 61022, Ukraine; [10]Faculty of Geography, Lomonosov Moscow State University, Moscow 119991, Russia; [11]Research Centre for Medical Genetics, Moscow 115478, Russia; [12]University of Cambridge, CB2 3QG Cambridge, UK; [13]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA; [14]Estonian Academy of Sciences, Tallinn 10130, Estonia
[15]These authors contributed equally to this work
*Correspondence: sroots@ebc.ee
http://dx.doi.org/10.1016/j.ajhg.2016.05.025.

present within the two most frequent and widespread hg N sub-clades, namely N2a-P43 and N3-M46(TAT) (previously N1b and N1c, respectively[31]). By selectively genotyping binary markers in a large phylogeographic survey involving samples from 56 pertinent populations, we determine patterns consistent with demographic population expansions emerging out of bottlenecks. Our experimental strategy to combine both high-resolution phylogenetic analyses based on many tens of "whole" Y chromosome sequences and a large phylogeographic survey allows us to study prehistoric episodes of hg N diversification in the context of contemporary N2a and N3 geographic distributions, diversities, and divergence times.

## Material and Methods

### Whole Y Chromosome Sequencing and Phylogeny Reconstruction

For reconstructing and rooting the phylogeny of hg N, we used 54 sequences published in Karmin et al.[29] The three hg O and 51 hg N Y chromosome sequences were generated with Complete Genomics (Mountain View) technology at 40× coverage, and data were filtered as described in Karmin et al.[29] The three hg O Y chromosomes were included for more precise rooting of hg N. In order to detect missing or poorly covered sub-lineages, we sequenced 43 additional samples at Gene By Gene by using the commercially available "BigY" service, a target-enrichment design utilizing 67,000 capture probes for sequencing at least 10 Mb on the Illumina HiSeq platform at >60× coverage. The targeted regions lie within the non-recombining male-specific parts of the Y chromosome (the list of regions is available in the Web Resources). After BigY sequencing, the Arpeggi Engine (AEngine) pipeline was used for downstream software analysis. This included short read mapping, alignment post-processing, and variant calling. AEngine's genotype quality score was used for quality filtering with the threshold of 3.02. AEngine's proprietary statistical model considers characteristics of read coverage, individual read-mapping qualities, and base sequencing quality scores by HiSeq. Both variant and reference base calls were filtered for quality. Besides using the AEngine pipeline, we also generated the sequence calls from raw read data by (1) mapping the paired-end reads to the GRCh37 human reference sequence with Burrows-Wheeler Aligner v.0.6.1,[32] (2) removing PCR duplicates with SAMtools v.0.1.19,[33] and (3) calling Y chromosome genotypes (variant and reference) with SAMtools mpileup and BCFtools.[33] Quality filtering was done with vcftools v.0.1.12; the default filter settings were used, except for the following values: base coverage > 4× and < 500, base quality > 20, and distance between SNPs > 5 bp.

To combine our recently generated data with the 54 published sequences, we extracted the overlap between BigY targeted regions and Complete Genomics and applied the "re-mapping filter" on the basis of modeling the poorly mapping regions on the Y chromosome as described in Karmin et al.[29] The overlap was 7.5 Mb of usable sequence of non-recombining male-specific Y chromosome region. This combination and filtering scheme minimizes the possible platform bias between Illumina and Complete Genomics given that the second-generation sequencing errors are mostly due to mismapping of the sequence reads. Because the areas with sufficient coverage and call confidence did not perfectly overlap between all samples, the usable sequence length was further reduced to 6.2 Mb. Sites with a 95% call rate were used in the analyses.

The list with geographic locations and ID labels of all sequences included in the study is provided in Table S1. All DNA samples were obtained from unrelated volunteers who provided informed consent in accordance with the guidelines of the relevant collaborating institutions; in addition, the Research Ethics Committee of the University of Tartu formally approved the study.

We used the software package BEAST v.1.7.5[34] to infer the hg N phylogenetic tree, estimate coalescent times of hgs, and generate Bayesian skyline plots (BSPs). We used a Bayesian skyline coalescent tree prior, the general time reversible (GTR) substitution model with gamma-distributed rates, and a relaxed lognormal clock. Runs were performed with a piecewise-linear coalescent model with 19 groups for 90 million iterations and sampling every 3,000 steps. At the end of each BEAST run, the results were visualized in Tracer v.1.5, and it was confirmed that all effective sample sizes were above 200. An age for hg NO of 41,900 years (95% confidence interval [CI] = 40,175–43,359)[29] was used as as the calibration point for estimating the coalescent times of the phylogenetic structures in hg N.

### Phylogeographic Sampling and Genotyping

We assembled a genotyping panel of 6,521 males from a total of 56 Eurasian populations. This dataset included 3,521 new samples designated as "this study" and 3,000 others updated from earlier studies to the higher level of phylogenetic resolution (Table S2). For genotyping, we chose at least two non-recurrent branch-defining SNPs from sub-clades N2a and N3 and designed primers with Primer3 software.[35,36] Primer specificity was first checked with Primer-BLAST[37] and GenomeTester v.1.3 software[38] in silico and verified by Sanger sequencing. We selected the best markers within each tested pair on the basis of the clearest allele-calling result. Specifications for all markers used for assigning sub-clade status in the populations belonging to hgs N3 and N2a (1,314 and 323, respectively) are given in Table S3. All population samples were genotyped by Sanger sequencing in a hierarchical manner according to the phylogenetic relationships shown in Table S2. Principal-component analysis (PCA) based on N2a and N3 sub-hg frequencies was performed with the freeware popSTR program kindly provided by H. Harpending.

The frequency-distribution maps of hgs N3 and N2a and their sub-branches were created with the data reported here (Table S2) in combination with published data of other Eurasian populations in which the frequency of N3 and N2a was detected to be zero (populations from the literature are indicated in Table S4). The maps were created with GeneGeo software as described previously;[39] the weight function was set to 3, and the radius of influence was set to 10,000 km.

We also typed 671 samples from sub-clades N3 and N2a for 17–23 Y chromosome short tandem repeats (Y-STRs) by using the Y-Filer Kit or the PowerPlex 23 Kit. The amplified fragments were separated with an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems), and lengths were analyzed with the ABI PRISM program GeneMapper 4.0 (Applied Biosystems). The Y-STR loci genotyped were DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439, DYS385a, DYS385b, DYS437, DYS438, DYS448, DYS456, DYS458, DYS635, and Y GATA H4. Because of the duplicated nature of DYS385a and

DYS385b and the uncertainty in their allele assignment, they were not included in any analysis. All other markers were used for constructing phylogenetic networks with Network 4.6.1.1 software (Fluxus-Engineering). The median joining algorithm was applied with the weighting scheme so that the SNP markers defining the sub-clades had higher weight (value 90), and all Y-STR markers were assigned value 10. We calculated hg regional diversity by applying data from Table S2 to the Nei formula.[40]

## Results

### Refined hg N Phylogeny Based on High-Coverage Y Chromosome Data

Based on the SNPs from high-coverage sequencing data, the phylogenetic relationships of the 94 hg N Y chromosomes (Figure 1A; Figure S1) enabled us to identify and expand several previously undetected or poorly resolved sub-clades and lineages (bolder magenta lines in Figure 1A) together with corresponding temporal estimates (Table S5). The total number of SNPs accumulated per branch, including the defining marker, is presented in Figure S2. The 1,967 variant reference-sequence positions, nucleotide changes, and branch assignments corresponding to each split (Figure S1) are presented in Table S6. During our analysis, we recognized that the ~45,000-year-old ancient DNA sample from the Ust'-Ishim region in the western Siberian Plain[1] carried two variants (M2308 and CTS11667), phylogenetically defining the root from which both the extant M214 branch and the ancient individual descended independently (Figure 1A, inset). The average number of mutations from the tree root to the tips was 198, which, given that the analyzed sequence length was 6.2 Mb, yields a mutation rate of 0.76E−9 substitutions per site per year. This is very similar to the rate estimate of 0.74E−9 by Karmin et al.[29] and equal to the one by Fu et al.[1]

Within hg N phylogeny, the deep N5-B482 defined lineage was unexpectedly found in a single sample previously determined to be of mixed origin by admixture analysis involving 730,000 autosomal SNPs to reflect ~80% European and/or Mediterranean and 20% Southeast Asian heritage (Figure 1A).

Although N4-F2930 is the prevalent N hg observed in Cambodia, Vietnam, and China nowadays,[25,27–29] including ancient DNA samples,[10] little is known regarding its structure, other than that it has two branches. Although lacking population-based frequency coverage, we found deep sub-structure by using 11 hg N4 Y chromosome sequences of Chinese and Japanese origin.

The hgs N3 and N2a diverged around 18.0 kya (95% CI = 15.7–20.0 kya) and coalesced around 13.0 kya (95% CI = 11.3–14.6 kya) and 9.0 kya (95% CI = 7.8–10.9 kya), respectively. The split between clades N3c and N3a'b is dated to 13.0 kya (95% CI = 11.3–14.6 kya), whereas the next bifurcation inside N3a'b most likely occurred around the end of the Pleistocene or early Holocene (about 12.0 kya).

For the hg N lineages present in the northern Eurasian mainland, N2a and N3 sub-clade N3a'b, the 95% CIs of coalescence times were distinct from those of N4. The mutation counts in our annotated tree (Figure S2) from the coalescence of N1'4-B481 to nodes N4-F2930, N3-M46(Tat), and N2a-P43 were 19, 35, and 51, respectively, which are quite different given that, on average, there were 97 mutations from the N1'4-B481 node to the tree tips. However, the 95% CIs for the time estimates of N4 and N3 had a slight overlap (Table S5).

### The Phylogeographic Pattern of N3 and N2a Clades

Within the Eurasian circum-Arctic spread zone, N3 and N2a reveal a well-structured spread pattern where individual sub-clades show very different distributions (Figures 1B, 2, 3, and S3). The sub-clade N3b-B187 is specific to southern Siberia and Mongolia, whereas N3a-L708 is spread widely in other regions of northern Eurasia (Figure 3 and Table S2). The deepest clade within N3a is N3a1-B211, mostly present in the Volga-Uralic region and western Siberian Khanty and Mansi populations. Sub-hg N3a2-M2118 is one of the two main bifurcating branches in the nested cladistic structure of N3a2'6-M2110. It is predominantly found in populations inhabiting present-day Yakutia (Republic of Sakha) in central Siberia and at lower frequencies in the Khanty and Mansi populations, which exhibit a distinct Y-STR pattern (Table S7) potentially intrinsic to an additional clade inside the sub-hg N3a2 (Table S2 and Figure S4). The neighbor clade, N3a3'6-CTS6967, spreads from eastern Siberia to the eastern part of Fennoscandia and the Baltic States (Figure 1B).

Perhaps the most striking feature of the geographic and temporal distribution of the N3a3'6 sub-clades is their almost simultaneous regional diversification (Figures 1A, 1B, and 3 and Table S5). These recently expanded regional varieties cover much of northern Asia and eastern Europe and are frequent in geographically distant peoples such as Chukchi and Lithuanians. This pattern mimics the previously described broad distribution of hg N.[5] The split of these lineages occurred ~5.0 kya (95% CI = 4.4–5.7 kya), some two millennia after the N3a2-to-N3a3'6 bifurcation around 7.0 kya (95% CI = 6.1–8.3 kya) (Figure 1A and Figure S1), making the extremely wide distribution of N3a3'6-related lineages all the more remarkable.

In Europe, the clade N3a3-VL29 encompasses over a third of the present-day male Estonians, Latvians, and Lithuanians but is also present among Saami, Karelians, and Finns (Table S2 and Figure 3). Among the Slavic-speaking Belarusians, Ukrainians, and Russians, about three-fourths of their hg N3 Y chromosomes belong to hg N3a3. The only notable exception from the pattern are Russians from northern regions of European Russia, where, in turn, about two-thirds of the hg N3 Y chromosomes belong to the hg N3a4-Z1936—the second west Eurasian clade. Thus, according to the frequency distribution of this clade, these Northern Russians fit better among other non-Slavic populations from northeastern Europe.

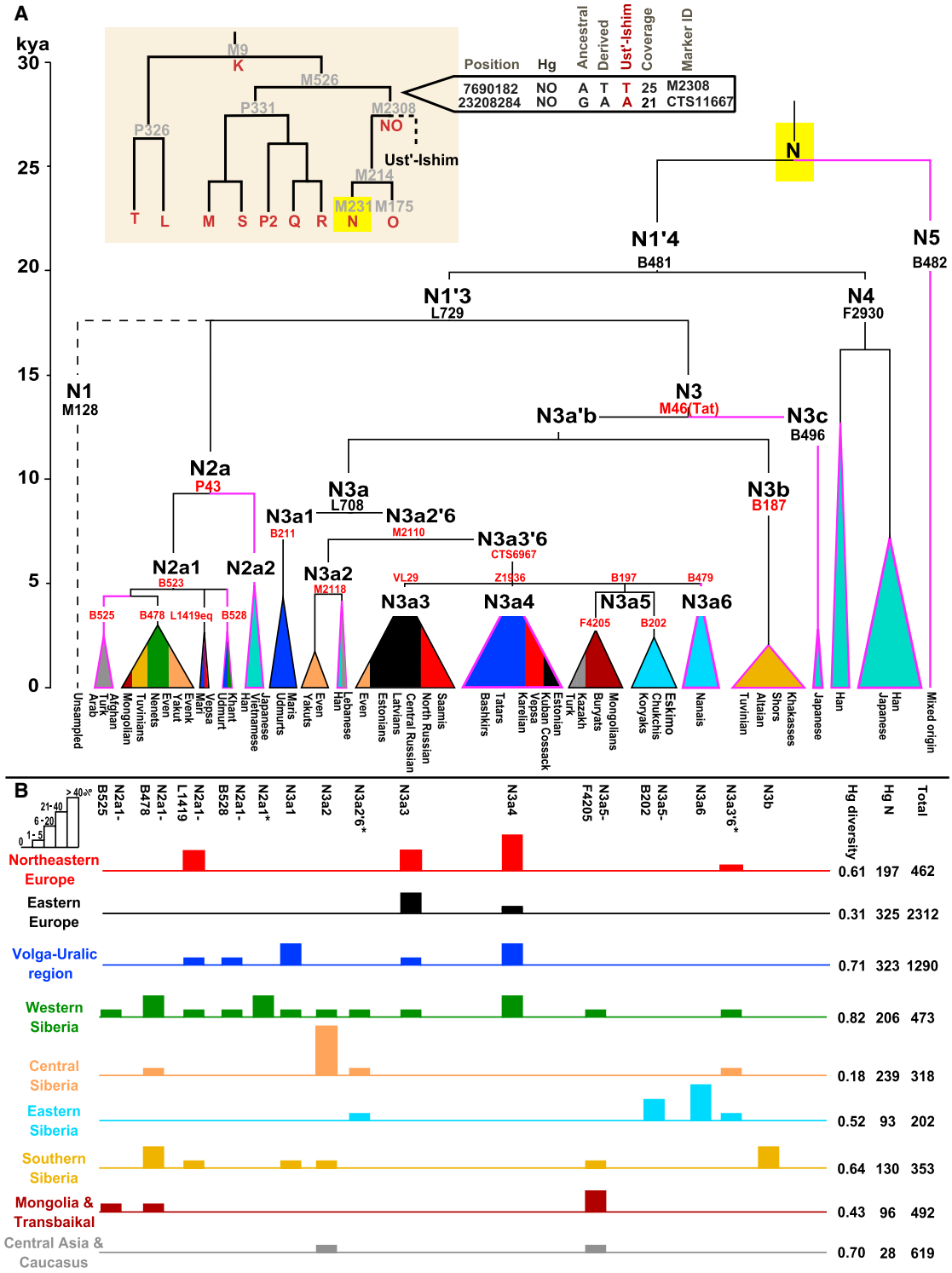



**Figure 1. The Schematic Phylogenetic Tree of hg N and Geographic Distribution of hg N Sub-clades**

(A) Schematic phylogenetic tree of hg N. The inset shows the schematic position of hg N, the ancient-DNA sample from Ust'-Ishim, and its shared hg NO mutations on the Y chromosome phylogenetic tree. The yellow box indicates the branching point of hg N. The phylogenetic tree of 94 high-coverage Y chromosomes from hg N was reconstructed with BEAST software. Unsupported dichotomies resulting from the strict bifurcation requirement of the software were manually reduced to polytomies. The size and fill proportions of each collapsed clade indicate the number of samples from the geographic regions constituting the clade. Clades are colored according to the color code presented in (B), except that turquoise represents clades with no frequency information at the population level. Markers represented in red were used for genotyping. Bold magenta lines indicate new lineages or expanded sub-clades in comparison with the phylogenetic tree from Karmin et al.[29] The temporal scale was obtained with a relaxed lognormal clock. The annotated tree with the number of mutations on each branch and the list of corresponding Y chromosome positions are available in Figure S2 and Table S6, respectively.

*(legend continued on next page)*

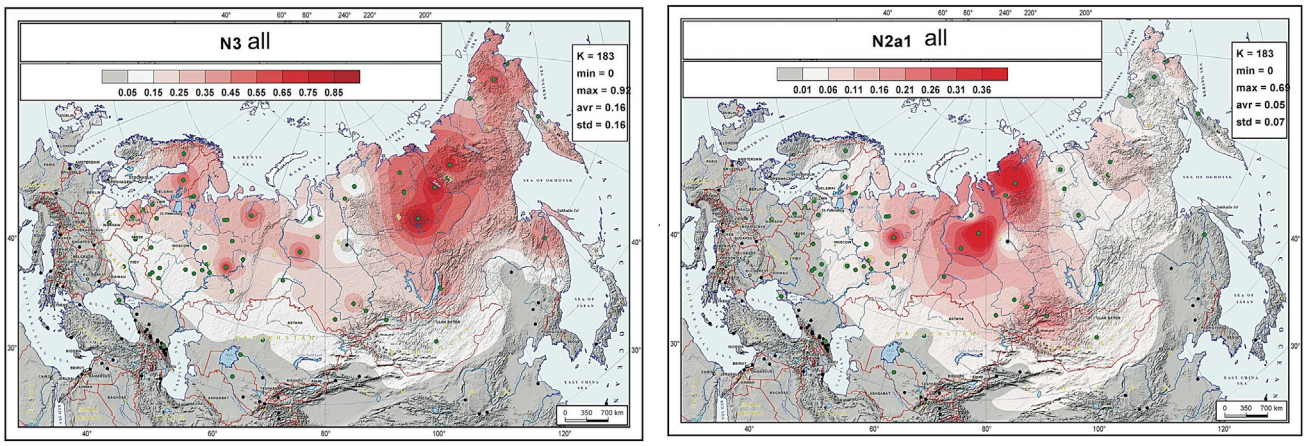



**Figure 2. Geographic-Distribution Maps of hg N3 and N2a1 General Frequency**
Data points from Table S2 and additional data points of zero frequency from the literature were used for creating the plots. All data points are presented on the "N3 all" and "N2a1 all" maps. Data points from this study are indicated by green dots, and data points from the literature are indicated by smaller black dots.

N3a4 tends to increase in frequency toward the north-eastern European regions but is also somewhat unexpectedly a dominant hg N3 lineage among most Turcic-speaking Volga Tatars and South-Ural Bashkirs (Figure 1B and Table S2).

In contrast to the predominantly eastern European sub-clades N3a3 and N3a4, their neighbor clades N3a5-B197 and N3a6-B479 are clearly restricted to eastern Eurasia: N3a5-F4205 is prominent around Lake Baikal among Mongolic-speaking Buryats and Mongols, N3a5-B202 is specific to eastern Siberia and Beringia among Chukchis, Koryaks, and Asian Eskimos, and N3a6 predominates in the Tungusic-speaking Nanai (Hezhe) population of the Amur River Basin in the Russian Far East (Figure 3 and Table S2).

The second widespread sub-clade of hg N is N2a. The split of N2a dates to about 9.3 kya (95% CI = 7.8–10.9 kya). The minor sub-clade, N2a2-B520, is represented in our dataset by individual samples from China, Vietnam, and Japan, indicating the presence of this clade at marginal frequencies in Southeast Asia. The absolute majority of N2a individuals belong to the second sub-clade, N2a1-B523, which diversified about 4.7 kya (95% CI = 4.0–5.5 kya) (Figure 1A and Table S5). Its distribution covers the western and southern parts of Siberia, the Taimyr Peninsula, and the Volga-Uralic region with frequencies ranging from from 10% to 30% and does not extend to eastern Siberia (Table S2 and Figures 2 and S3). Whereas earlier studies presumed two sub-clades on the basis of different Y-STR patterns of N2a1 carriers,[5,15] we now have a total of 19 sequences from N2a1 and reveal three separate sub-clades. The most frequent one splits into a clade represented by nine individuals from Siberian populations (N2a1-B478) and another clade represented by three individuals, a Turk, an Arab, and an Afghani (N2a1-B525) (Figures 1A and S1). The latter sub-clade is occasionally found in populations such as the Mongols, Central Asians, and rarely also the Russians (Table S2). The "European" branch suggested earlier from Y-STR patterns turned out to consist of two clades: N2a1-B528, spread in the southern Volga-Uralic region, and N2a1-L1419, spread mainly in the northern part of that region (Figure S3). Although not all samples fall into any defined N2a1 branches (Table S2), when Y-STR haplotypes (Table S7) are used in network analysis, potential close affinity to the west Siberian N2a1-B528 sub-branch appears (Figure S4).

Using all populations in which at least five hg N Y chromosomes were observed and hg N frequency was ≥5%, we conducted PCA on the basis of comparing the frequencies of hg N sub-groups to that of total hg N. Although passing these criteria, only the Nanai from eastern Siberia were excluded because their unique sub-hg N3a6 uniformity biased the PC plot by compressing all other population variation. In the absence of the Nanai on the PC plot (Figure S5A), the populations clustered mostly according to their relative positions from east to west in Eurasia, except for the Karanogays of the Caucasus region, a population with known east Eurasian affinities. When PCA was done by hg (Figure S5B), PC1, explaining 17.9% of the variation, approximated an east-west axis by separating Siberian populations with a high N3a2 frequency from European and Volga-Uralic populations with a high N3a3 frequency. PC2 (15.7% of the variation) set the north-eastern Siberian populations (Chukchi and Asian Eskimos)

(B) The regional distribution of hg N sub-clades in northern Eurasia. The chart depicts the proportions of hg N sub-clades in studied regions (the percentage of hg N sub-clades in the total number of samples pooled according to their geographic origin). Column heights correspond to the frequency intervals given in the figure. The entire list of populations from each studied region can be found in Table S2.

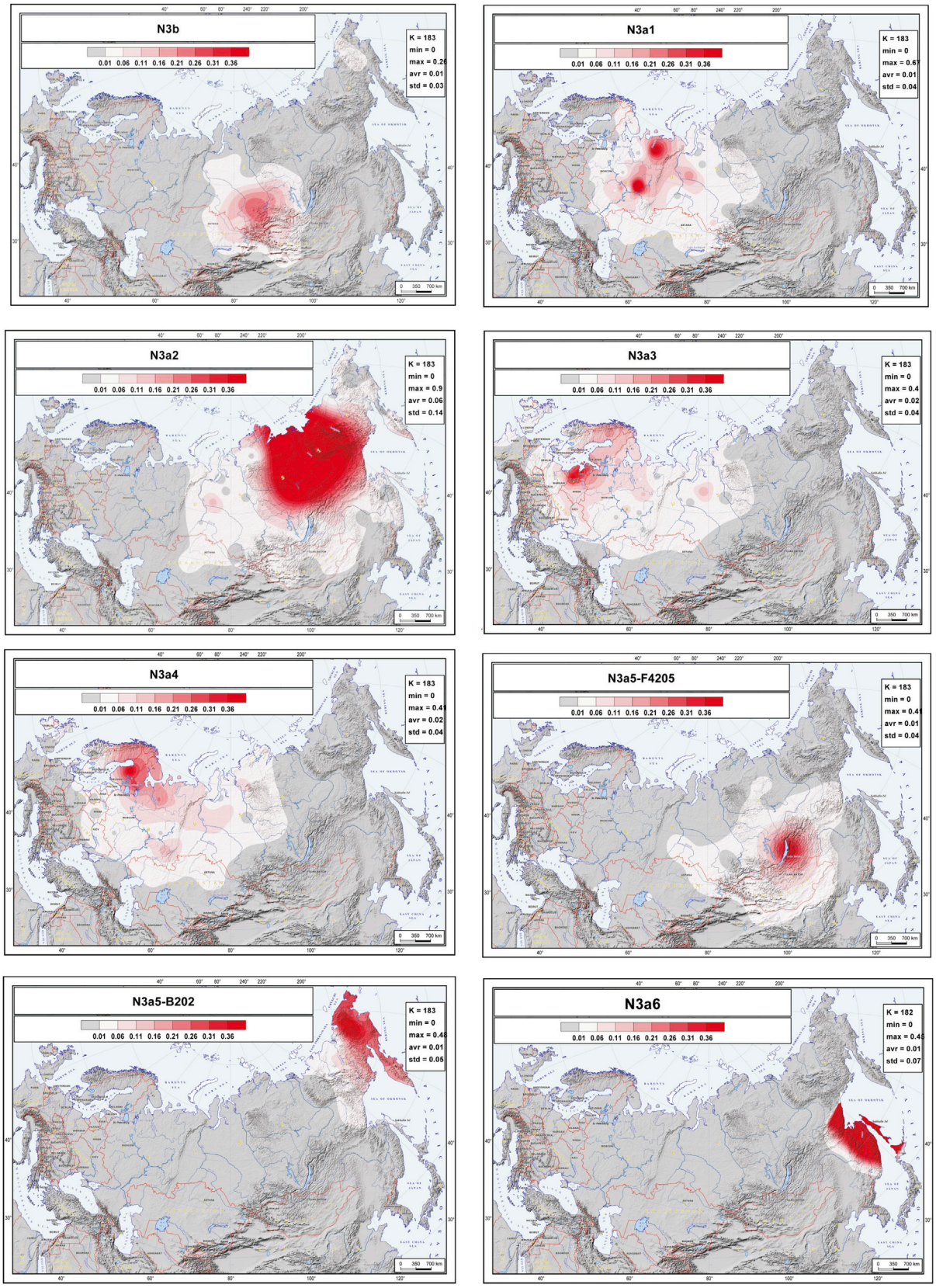



**Figure 3. Frequency-Distribution Maps of Individual Sub-clades of hg N3**
Data points from Table S2 and additional data points of zero frequency from the literature were used for creating the plots of hg N3 sub-clades. All data points are the same as in Figure 2.

apart from the rest because of their high N3a5-B202 content, which was completely absent from all other populations.

We used BSP analysis to infer temporal changes of male effective population size. The BSP of 94 high-coverage Y chromosomes from hg N showed an increase in effective population size starting at about 4.5 kya (Figure S6A). It should be borne in mind that the time to the most recent common ancestor and the time of the fastest population growth for a given hg need not, and often do not, coincide.[41,42] However, for N3 (Figure S6B), the initial expansion dynamics seemed to correspond with the estimated coalescent time of N3a3'6-L392 (Table S5), the most diverse and widespread sub-clade of N3. In contrast, N2a (Figure S6C) showed a relatively slow to moderate expansion.

## Discussion

In Eurasia, geographically more southerly distributed Y chromosome hgs have, as a rule, retained deeper lineages, as is clearly apparent for hg H in India and hgs C and O in Southeast Asia.[29] In accordance with this, we observed that clade N4-F2930 (n = 11) emerged prior to the more northerly distributed hgs N2a and N3 despite the fact that the latter two had a far larger sample size (n = 63) in our study. It is plausible to assume that lower long-term population size and the influence of harsh climate changes, presumably accompanied by repeated population bottlenecks and subsequent proposed socio-economic changes in the Holocene,[29] have kept Y chromosome variation low in the temperate and arctic parts of Eurasia. The interior branches of hg N3 (formerly N1c), namely N3c, N3b, N3a1, and N3a2 (Figures 1 and 3), are spatially quite distinct from one another. In our dataset, the deeper sub-clade N3c-B496 was absent in the northern part of Eurasian mainland and was only represented by two individuals from Japan, where this lineage has been preserved and previously detected at a frequency of less than 1%.[3,5] Naturally, we cannot exclude the possibility that with denser sampling in mainland Asia, this sub-clade could be found elsewhere. Previous research[24,43] has shown that Y chromosomes of the Turkic-speaking Yakuts (Sakha) belong overwhelmingly to hg N3 (formerly N1c1). We found that nearly all of the more than 150 genotyped Yakut N3 Y chromosomes belong to the N3a2-M2118 clade, just as in the Turkic-speaking Dolgans and the linguistically distant Tungusic-speaking Evenks and Evens living in Yakutia (Table S2). Hence, the N3a2 patrilineage is a prime example of a male population of broad central Siberian ancestry that is not intrinsic to any linguistically defined group of people. Moreover, the deepest branch of hg N3a2 is represented by a Lebanese and a Chinese sample. This finding agrees with the sequence data from Hallast et al.,[27] where one Turkish Y chromosome was also assigned to the same sub-clade. Interestingly, N3a2 was also found in one Bhutan individual[27] who represents a separate sub-lineage in the clade. These findings show that although N3a2 reflects a recent strong founder effect primarily in central Siberia (Yakutia, Sakha) (Figure 3), the sub-clade has a much wider distribution area with incidental occurrences in the Near East and South Asia.

The most striking aspect of the phylogeography of hg N is the spread of the N3a3'6-CTS6967 lineages (Figure 3). Considering the three geographically most distant populations in our study—Chukchi, Buryats, and Lithuanians—it is remarkable to find that about half of the Y chromosome pool of each consists of hg N3 and that they share the same sub-clade N3a3'6. The fractionation of N3a3'6 into the four sub-clades that cover such an extraordinarily wide area occurred in the mid-Holocene, about 5.0 kya (95% CI = 4.4–5.7 kya). It is hard to pinpoint the precise region where the split of these lineages occurred. It could have happened somewhere in the middle of their geographic spread around the Urals or further east in West Siberia, where current regional diversity of hg N sub-lineages is the highest (Figure 1B). Yet, it is evident that the spread of the newly arisen sub-clades of N3a3'6 in opposing directions happened very quickly. Today, it unites the East Baltic, East Fennoscandia, Buryatia, Mongolia, and Chukotka-Kamchatka (Beringian) Eurasian regions, which are separated from each other by approximately 5,000–6,700 km by air. N3a3'6 has high frequencies in the patrilineal pools of populations belonging to the Altaic, Uralic, several Indo-European, and Chukotko-Kamchatkan language families. There is no generally agreed, time-resolved linguistic tree that unites these linguistic phyla. Yet, their split is almost certainly at least several millennia older than the rather recent expansion signal of the N3a3'6 sub-clade, suggesting that its spread had little to do with linguistic affinities of men carrying the N3a3'6 lineages.

Although the initial spread of the hg N3a3'6 clades most likely ignored any existing language barriers, the subsequent diversification often occurred within linguistically defined metapopulations. The split of N3a5-B197 into sub-lineages occurred soon after the emergence of the clade itself (Figure 1), distributing N3a5-F4205 within Mongolic-speaking populations and N3a5-B202 among populations of the Chuckotko-Kamchatka language family (Table S2).

The N3a5-F4205 lineage encompasses the Buryat and Mongol populations, which live about 5,000 km apart from the carriers of the second B202 sub-branch—the Chukchi and Koryaks. Another Siberian sub-clade, N3a6-B479, is dominant among the Nanai (Hezhe) population, which lives in eastern Siberia along the lower reach of the Amur River. This sub-clade includes all of the hg N3 individuals in this population and probably reflects a strong founder effect (Figure 3), but denser sampling from this region, particularly Tungusic-speaking Negidals, Orichs, Udege, and others, could reveal this lineage in neighboring populations as well. However, it is interesting to notice here that we did not detect N3a5-B202 in two more

populous Tungusic-speaking populations, Evenks and Evens in Yakutia (Table S2); instead, like Yakuts, they carry predominantly N3a2-M2118 chromosomes.

The same timescale that describes the fast expansion of Siberian sub-clades also holds true for the expansion of other N3a3′6 lineages in the European direction, where the spread zones of N3a3-VL29 and N3a4-Z1936 partially overlap in northeastern Europe (Figures 1B and 3). Even though our sample of the north Scandinavian (Swedish) Saami is of limited size, the nearly equal presence of both N3a3 and N3a4 Y chromosomes in their hg N3 pool (Table S2) suggests that the frequency pattern of the two lineages has been shaped by random genetic drift in historically small populations dispersed across a wide area.

The mid-Holocene (~4.0–6.0 kya) timing of the rapid spread of sub-clade N3a3′6-CTS6967 coincides with paleoclimatic and archeological evidence that allows us to draw some parallels. Monserud et al.[44] identified a post-glacial warming and precipitation peak in Siberia, accompanied by northward advancement of forests into the current tundra zone, during the mid-Holocene. Similar paleoclimatic conditions occurred locally in the northern Urals,[45] northern Siberia,[46] and central Yakutia.[47] Although calibrated Y chromosome phylogenies can now provide estimates for the duration of bottlenecks and the beginning of population expansions from common ancestral branches, considerable caution should be exercised regarding hypotheses based on drawing parallels between the spread of specific genetic lineages and prehistoric cultural artifacts. Nonetheless, it is still interesting to note intervals of occupational variation in the archaeological record, as well as episodes of common cultural features coinciding in time and space with the spread of hgN sub-clusters. For example, in the proximity of Lake Baikal,[48,49] the collapse of the Kitoi culture in the mid-Holocene was followed by a ~800–1,000 year gap in the archeological record. The subsequent appearance of the ~4,000- to 3,000-year-old Isakovo-Serovo and Glazkovo complex bears evidence of cultural[48]and genetic discontinuity,[50] suggesting large-scale population movement. No clear evidence of the "homeland" of the Isakovo-Serovo and Glazkovo ancestors exists, but western Siberia (Upper Yenisei Basin)[48] and northern China[49] have been proposed as possible regions of origin.

Another pattern involves the similarity in the range of hg N3a3′6, especially in the western part of Eurasia and the distribution of the Seima-Turbino trans-cultural phenomenon during the interval of 4.2–3.7 kya.[51] Extending across northern Eurasia from Mongolia to the Baltic region, this phenomenon encompasses the cultures of nomadic forest and steppe societies with advanced metal-working technology.[51] Taken together, these facts hint at the Seima-Turbino metalsmith-traders as the probable primary carriers of hg N3a3′6 lineages.

N3a1-B211, the early branch of N3a, could have been brought to the eastern fringes of Europe by the same Seima-Turbino groups, but earlier migration(s) cannot be ruled out either, given that a study of ancient DNA[52] revealed a 7,500-year-old influx from Siberia to northeast Europe.

Although such examples of overlapping lines of evidence are notable, they remain speculative. However, given the relatively favorable preservation conditions of ancient DNA and the spatial and temporal characteristics of N3a3′6 distributions in northern Eurasia, it is a propitious time to imagine, given that the emerging field of credible studies of ancient DNA might give us more solid answers in the near future.

The studies of ancient genomes of Eurasia[1,53,54] have revealed that in addition to the Mesolithic hunter-gatherer and Neolithic farmer substrates, contemporary eastern Europeans have genetic heritage in common with the so-called ancient northern Eurasians (ANEs),[53] whose representatives were genetically close to the carriers of the Mal'ta culture of southern Siberia at the Last Glaxial Maximum.[54] It has recently been proposed that the ANE component reached Europe together with the spread of Early Bronze Age steppe-belt Yamna culture, thus making it only distantly related to the Mal'ta.[55] The model of three ancestral components of modern Europeans does not fit well with the genetic profiles of some northeastern European populations that show more East Asian influence than is expected from ANEs.[53] This observation hints that an additional element of Siberian gene flow could be represented by the influx of hg N3a3′4 lineages into the region.

Although the socioeconomic and demographic processes that took place during the Bronze Age along the steppe belt might be linked to the spread of at least some branches of Indo-European languages,[55–57] the expansion of the Uralic linguistic family has been associated with concurrent events along the forest zone of northern Eurasia.[56] The spread of the westernmost hg N branches might be one of the genetic signals of these movements, coinciding mostly, but not entirely, with the present linguistic borders of the Finno-Ugric languages.

Overall, a considerable proportion of men inhabiting much of the Arctic and temperate zones of western and eastern Eurasia share N3a3′6 lineages that date back to the mid-Holocene (4.5–5.0 kya). This common patrilineal ancestry unites widely different linguistic phyla, including Indo-European, particularly Balto-Slavic, branches of the Altaic, such as the Mongolic, Turkic, Tungusic, and Chukotko-Kamchatkan branches, as well as the Balto-Finnic branch of the Finno-Ugric (Table S2).

The improved hg N phylogeny now maps anew the previous homogeneous spread of N3 and N2a into distinctive lineages that in fact arose at various times and spread differently through several independent demographic events. This study demonstrates how Y chromosome re-sequencing instructs and refuels the genotyping projects so that phylogenetic information from individuals is expanded to the population level such that their combination strengthens the understanding of our demographic history.

## Accession Numbers

The complete whole Y chromosome sequences have been deposited in the European Nucleotide Archive under accession number ENA: PRJEB12523. The data are also available at the data repository of the Estonian Biocentre (http://www.ebc.ee/).

## Supplemental Data

Supplemental Data include six figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.05.025.

## Conflicts of Interest

P.A.U. consulted for 23andMe and has 23andMe stock options. D.M.B. is compensated by and serves as the chief science officer of Gene by Gene.

## Acknowledgments

This work was supported by the EU European Regional Development Fund through the Centre of Excellence in Genomics to the Estonian Biocentre, by Estonian institutional research grant IUT24-1, by Estonian personal research grant PUT1217 to K.T., by European Commission grant 205419 ECOGENE to the Estonian Biocentre, by Russian Scientific Foundation grant 14-14-00827 (to O.B., M.C, and A.A), by project no. 6.656 of the Ministry of Education and Science of Russia (to S. F.), and by Russian Foundation for Basic Research grants 14-04-00725-a (to E.K.), 16-06-00303-a (to E.B.), and 14-06-00384-a (to O.B., M.C., and A.A.). L.P.O. was supported by the federal budget under state project no. 0324-2015-0004. M.K. and A.K. acknowledge financial support from Estonian personal grants PUT-766 and PUT1339, respectively. P.A.U. was supported by SAP grant SP0#115016 to Prof. Carlos D. Bustamante.

Received: February 18, 2016
Accepted: May 22, 2016
Published: July 7, 2016

## Web Resources

Targeted regions, https://www.familytreedna.com/documents/bigy_targets.txt

## References

1. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.
2. Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhövel, W., Fretwell, N., Jobling, M.A., Harihara, S., et al. (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am. J. Hum. Genet. *60*, 1174–1183.
3. Hammer, M.F., Karafet, T.M., Park, H., Omoto, K., Harihara, S., Stoneking, M., and Horai, S. (2006). Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J. Hum. Genet. *51*, 47–58.
4. Karafet, T.M., Osipova, L.P., Gubina, M.A., Posukh, O.L., Zegura, S.L., and Hammer, M.F. (2002). High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum. Biol. *74*, 761–789.
5. Rootsi, S., Zhivotovsky, L.A., Baldovic, M., Kayser, M., Kutuev, I.A., Khusainova, R., Bermisheva, M.A., Gubina, M., Fedorova, S.A., Ilumäe, A.M., et al. (2007). A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. Eur. J. Hum. Genet. *15*, 204–211.
6. Shi, H., Qi, X., Zhong, H., Peng, Y., Zhang, X., Ma, R.Z., and Su, B. (2013). Genetic evidence of an East Asian origin and paleolithic northward migration of Y-chromosome haplogroup N. PLoS ONE *8*, e66102.
7. Xue, Y., Zerjal, T., Bao, W., Zhu, S., Lim, S.K., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., et al. (2005). Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am. J. Hum. Genet. *77*, 1112–1116.
8. Kayser, M. (2010). The human genetic history of Oceania: near and remote views of dispersal. Curr. Biol. *20*, R194–R201.
9. Zhong, H., Shi, H., Qi, X.B., Duan, Z.Y., Tan, P.P., Jin, L., Su, B., and Ma, R.Z. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. Mol. Biol. Evol. *28*, 717–727.
10. Cui, Y., Li, H., Ning, C., Zhang, Y., Chen, L., Zhao, X., Hagelberg, E., and Zhou, H. (2013). Y Chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. BMC Evol. Biol. *13*, 216.
11. Gao, S.Z., Zhang, Y., Wei, D., Li, H.J., Zhao, Y.B., Cui, Y.Q., and Zhou, H. (2015). Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age. Am. J. Phys. Anthropol. *157*, 71–80.
12. Zhao, Y.B., Zhang, Y., Li, H.J., Cui, Y.Q., Zhu, H., and Zhou, H. (2014). Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3,000 years ago. Am. J. Hum. Biol. *26*, 813–821.
13. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.
14. Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E.L., Tolk, H.V., Reidla, M., Metspalu, E., Pliss, L., et al. (2004). The western and eastern roots of the Saami–the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am. J. Hum. Genet. *74*, 661–682.
15. Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2007). Y-chromosome haplogroup N dispersals from south Siberia to Europe. J. Hum. Genet. *52*, 763–770.
16. Karlsson, A.O., Wallerström, T., Götherström, A., and Holmlund, G. (2006). Y-chromosome diversity in Sweden - a long-time perspective. Eur. J. Hum. Genet. *14*, 963–970.
17. Lappalainen, T., Koivumäki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M.L., and Lahermo, P. (2006). Regional differences among the Finns: a Y-chromosomal perspective. Gene *376*, 207–215.
18. Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.L., and Lahermo, P. (2008). Migration waves to the Baltic Sea region. Ann. Hum. Genet. *72*, 337–348.
19. Balanovsky, O., Rootsi, S., Pshenichnov, A., Kivisild, T., Churnosov, M., Evseeva, I., Pocheshkhova, E., Boldyreva, M.,

Yankovsky, N., Balanovska, E., and Villems, R. (2008). Two sources of the Russian patrilineal heritage in their Eurasian context. Am. J. Hum. Genet. *82*, 236–250.
20. Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., and Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. Eur. J. Hum. Genet. *16*, 1254–1264.
21. Mirabal, S., Regueiro, M., Cadenas, A.M., Cavalli-Sforza, L.L., Underhill, P.A., Verbenko, D.A., Limborska, S.A., and Herrera, R.J. (2009). Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. Eur. J. Hum. Genet. *17*, 1260–1273.
22. Dulik, M.C., Osipova, L.P., and Schurr, T.G. (2011). Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. PLoS ONE *6*, e17548.
23. Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., and Schurr, T.G. (2012). Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and Indigenous Altaians. Am. J. Hum. Genet. *90*, 229–246.
24. Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S.I., Hooshiar Kashani, B., Olivieri, A., et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. BMC Evol. Biol. *13*, 127.
25. Poznik, G.D., Henn, B.M., Yee, M.C., Sliwerska, E., Euskirchen, G.M., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., Underhill, P.A., and Bustamante, C.D. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science *341*, 562–565.
26. Francalacci, P., Morelli, L., Angius, A., Berutti, R., Reinier, F., Atzeni, R., Pilu, R., Busonero, F., Maschio, A., Zara, I., et al. (2013). Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science *341*, 565–569.
27. Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J.H., Arroyo-Pardo, E., Cavalleri, G.L., de Knijff, P., Destro Bisol, G., Dupuy, B.M., et al. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. Mol. Biol. Evol. *32*, 661–673.
28. Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., Schröder, R., and Stoneking, M. (2014). Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig. Genet. *5*, 13.
29. Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M.A., Järve, M., Talas, U.G., Rootsi, S., Ilumäe, A.M., Mägi, R., Mitt, M., et al. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. *25*, 459–466.
30. Batini, C., Hallast, P., Zadik, D., Delser, P.M., Benazzo, A., Ghirotto, S., Arroyo-Pardo, E., Cavalleri, G.L., de Knijff, P., Dupuy, B.M., et al. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. Nat. Commun. *6*, 7152.
31. Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. *18*, 830–838.
32. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
34. Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.
35. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.
36. Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics *23*, 1289–1291.
37. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.
38. Andreson, R., Reppo, E., Kaplinski, L., and Remm, M. (2006). GENOMEMASKER package for designing unique genomic PCR primers. BMC Bioinformatics *7*, 172.
39. Balanovsky, O., Dibirova, K., Dybo, A., Mudrak, O., Frolova, S., Pocheshkhova, E., Haber, M., Platt, D., Schurr, T., Haak, W., et al.; Genographic Consortium (2011). Parallel evolution of genes and languages in the Caucasus region. Mol. Biol. Evol. *28*, 2905–2920.
40. Nei, M. (1987). Molecular Evolutionary Genetics (Columbia University Press).
41. Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2008). mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. Mol. Biol. Evol. *25*, 468–474.
42. Kitchen, A., Miyamoto, M.M., and Mulligan, C.J. (2008). A three-stage colonization model for the peopling of the Americas. PLoS ONE *3*, e1596.
43. Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., and Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. Hum. Genet. *120*, 334–353.
44. Monserud, R.A., Tchebakova, N.M., and Denissenko, O.V. (1998). Reconstruction of the mid-Holocene palaeoclimate of Siberia using a bioclimatic vegetation model. Palaeogeog. Palaeoclimat. Palaeoecol. *139*, 15–36.
45. Andreev, A.A., Tarasov, P.E., Siegert, C., Ebel, T., Klimanov, V.A., Melles, M., Bobrov, A.A., Dereviagin, A.Y., Lubinski, D.J., and Hubberten, H.W. (2003). Late Pleistocene and Holocene vegetation and climate on the northern Taymyr Peninsula, Arctic Russia. Boreas *32*, 484–505.
46. Anderson, P.M., Lozhkin, A.V., and Brubaker, L.B. (2002). Implications of a 24,000-yr palynological record for a Younger Dryas cooling and for boreal forest development in northeastern Siberia. Quat. Res. *57*, 325–333.
47. Nazarova, L., Lupfert, H., Subetto, D., Pestryakova, L., and Diekmann, B. (2013). Holocene climate conditions in central Yakutia (Eastern Siberia) inferred from sediment composition and fossil chironomids of Lake Temje. Quat. Int. *290*, 264–274.

48. Weber, A.W., Link, D.W., and Katzenberg, M.A. (2002). Hunter-gatherer culture change and continuity in the Middle Holocene of the Cis-Baikal, Siberia. J. Anthropol. Archaeol. *21*, 230–299.
49. Weber, A., Katzenberg, M.A., and Schurr, T.G. (2010). Prehistoric Hunter-Gatherers of the Baikal Region, Siberia: Bioarchaeological Studies of Past Life Ways (University of Pennsylvania Press).
50. Mooder, K.P., Schurr, T.G., Bamforth, F.J., Bazaliiski, V.I., and Savel'ev, N.A. (2006). Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. Am. J. Phys. Anthropol. *129*, 349–361.
51. Chernykh, E. (2008). The "Steppe Belt" of stockbreeding cultures in Eurasia during the Early Metal Age. Trab. Prehist. *65*, 73–93. http://dx.doi.org/10.3989/tp.2008.08004.
52. Der Sarkissian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., Zaporozhchenko, V., Gronenborn, D., Moiseyev, V., Kolpakov, E., et al.; Genographic Consortium (2013). Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. PLoS Genet. *9*, e1003296.
53. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
54. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.
55. Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.
56. Anthony, D. (2007). The horse, the wheel and language. How Bronze-Age riders from the Eurasian steppes shaped the modern world (Princeton, NJ: Princeton University Press).
57. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

**The American Journal of Human Genetics, Volume *99***

## Supplemental Data

# Human Y Chromosome Haplogroup N:

# A Non-trivial Time-Resolved Phylogeography

# that Cuts across Language Families

**Anne-Mai Ilumäe, Maere Reidla, Marina Chukhryaeva, Mari Järve, Helen Post, Monika Karmin, Lauri Saag, Anastasiya Agdzhoyan, Alena Kushniarevich, Sergey Litvinov, Natalya Ekomasova, Kristiina Tambets, Ene Metspalu, Rita Khusainova, Bayazit Yunusbayev, Elza K. Khusnutdinova, Ludmila P. Osipova, Sardana Fedorova, Olga Utevska, Sergey Koshel, Elena Balanovska, Doron M. Behar, Oleg Balanovsky, Toomas Kivisild, Peter A. Underhill, Richard Villems, and Siiri Rootsi**

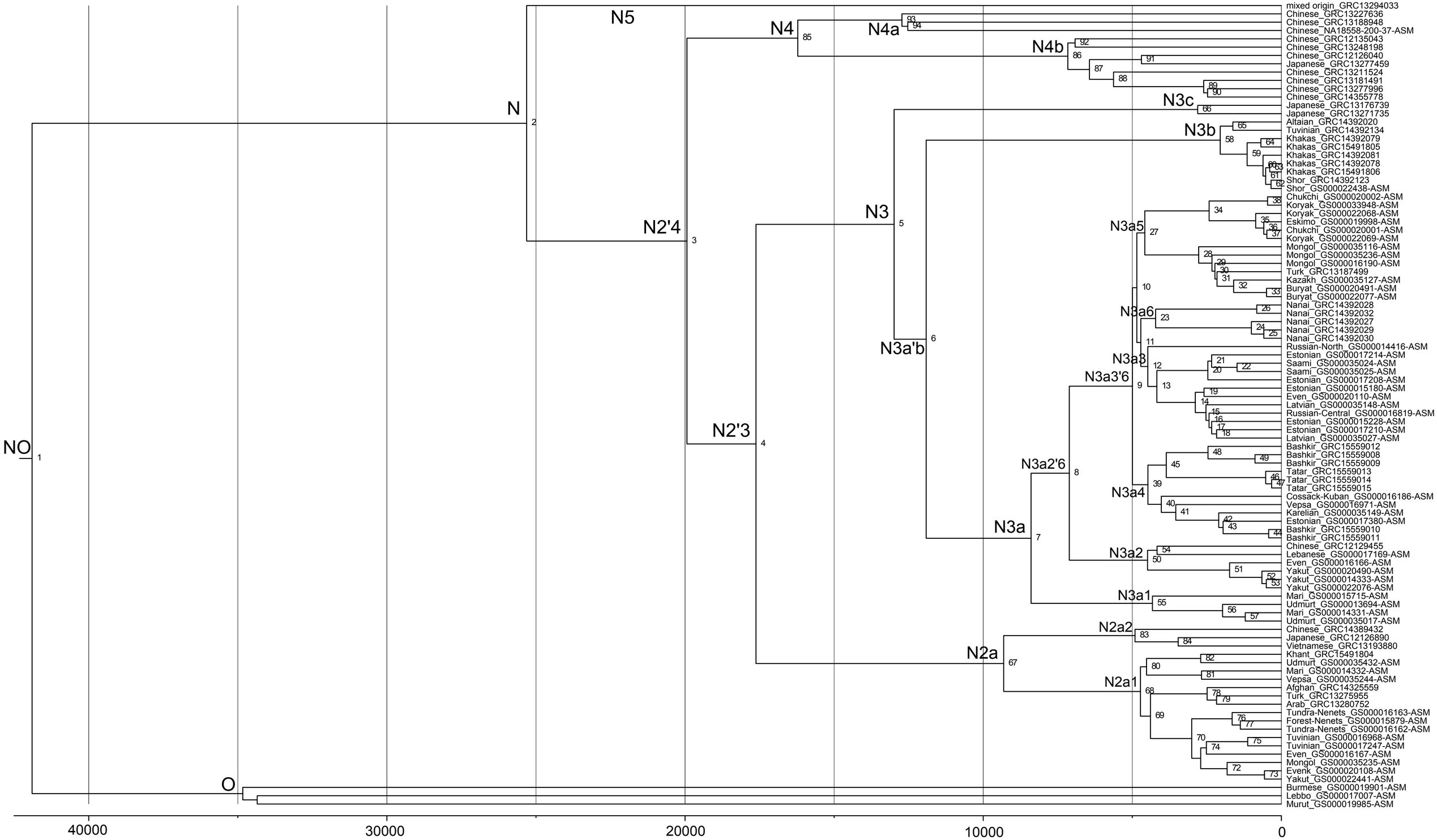

N5
N4
N4a
N4b
N3c
N3b
N
N2'4
N3
N3a5
N3a'b
N3a6
N2'3
N3a3
N3a3'6
NO
N3a2'6
N3a4
N3a
N3a2
N3a1
N2a2
N2a
N2a1
O
mixed origin_GRC13294033
Chinese_GRC13227636
Chinese_GRC13188948
Chinese_NA18558-200-37-ASM
Chinese_GRC12135043
Chinese_GRC13248198
Chinese_GRC12126040
Japanese_GRC13277459
Chinese_GRC13211524
Chinese_GRC13181491
Chinese_GRC13277996
Chinese_GRC14355778
Japanese_GRC13176739
Japanese_GRC13271735
Altaian_GRC14392020
Tuvinian_GRC14392134
Khakas_GRC14392079
Khakas_GRC15491805
Khakas_GRC14392081
Khakas_GRC14392078
Khakas_GRC15491806
Shor_GRC14392123
Shor_GS000022438-ASM
Chukchi_GS000020002-ASM
Koryak_GS000033948-ASM
Koryak_GS000022068-ASM
Eskimo_GS000019998-ASM
Chukchi_GS000020001-ASM
Koryak_GS000022069-ASM
Mongol_GS000035116-ASM
Mongol_GS000035236-ASM
Mongol_GS000016190-ASM
Turk_GRC13187499
Kazakh_GS000035127-ASM
Buryat_GS000020491-ASM
Buryat_GS000022077-ASM
Nanai_GRC14392028
Nanai_GRC14392032
Nanai_GRC14392027
Nanai_GRC14392029
Nanai_GRC14392030
Russian-North_GS000014416-ASM
Estonian_GS000017214-ASM
Saami_GS000035024-ASM
Saami_GS000035025-ASM
Estonian_GS000017208-ASM
Estonian_GS000015180-ASM
Even_GS000020110-ASM
Latvian_GS000035148-ASM
Russian-Central_GS000016819-ASM
Estonian_GS000015228-ASM
Estonian_GS000017210-ASM
Latvian_GS000035027-ASM
Bashkir_GRC15559012
Bashkir_GRC15559008
Bashkir_GRC15559009
Tatar_GRC15559013
Tatar_GRC15559014
Tatar_GRC15559015
Cossack-Kuban_GS000016186-ASM
Vepsa_GS000016971-ASM
Karelian_GS000035149-ASM
Estonian_GS000017380-ASM
Bashkir_GRC15559010
Bashkir_GRC15559011
Chinese_GRC12129455
Lebanese_GS000017169-ASM
Even_GS000016166-ASM
Yakut_GS000020490-ASM
Yakut_GS000014333-ASM
Yakut_GS000022076-ASM
Mari_GS000015715-ASM
Udmurt_GS000013694-ASM
Mari_GS000014331-ASM
Udmurt_GS000035017-ASM
Chinese_GRC14389432
Japanese_GRC12126890
Vietnamese_GRC13193880
Khant_GRC15491804
Udmurt_GS000035432-ASM
Mari_GS000014332-ASM
Vepsa_GS000035244-ASM
Afghan_GRC14325559
Turk_GRC13275955
Arab_GRC13280752
Tundra-Nenets_GS000016163-ASM
Forest-Nenets_GS000015879-ASM
Tundra-Nenets_GS000016162-ASM
Tuvinian_GS000016968-ASM
Tuvinian_GS000017247-ASM
Even_GS000016167-ASM
Mongol_GS000035235-ASM
Evenk_GS000020108-ASM
Yakut_GS000022441-ASM
Burmese_GS000019901-ASM
Lebbo_GS000017007-ASM
Murut_GS000019985-ASM
40000
30000
20000
10000
0

**Figure S1**: Detailed phylogenetic tree of Y-chromosome haplogroup N based on 97 high coverage sequences.

The calibrated tree was constructed using BEAST v1.7.5. Internal node numbers and sample ID-labels on the tips are indicated. Age estimates of hg N clades are reported in Table S5. Number of mutations and marker names are presented in Table S2. All SNPs characterizing the clades (nodes) are presented in Table S6.

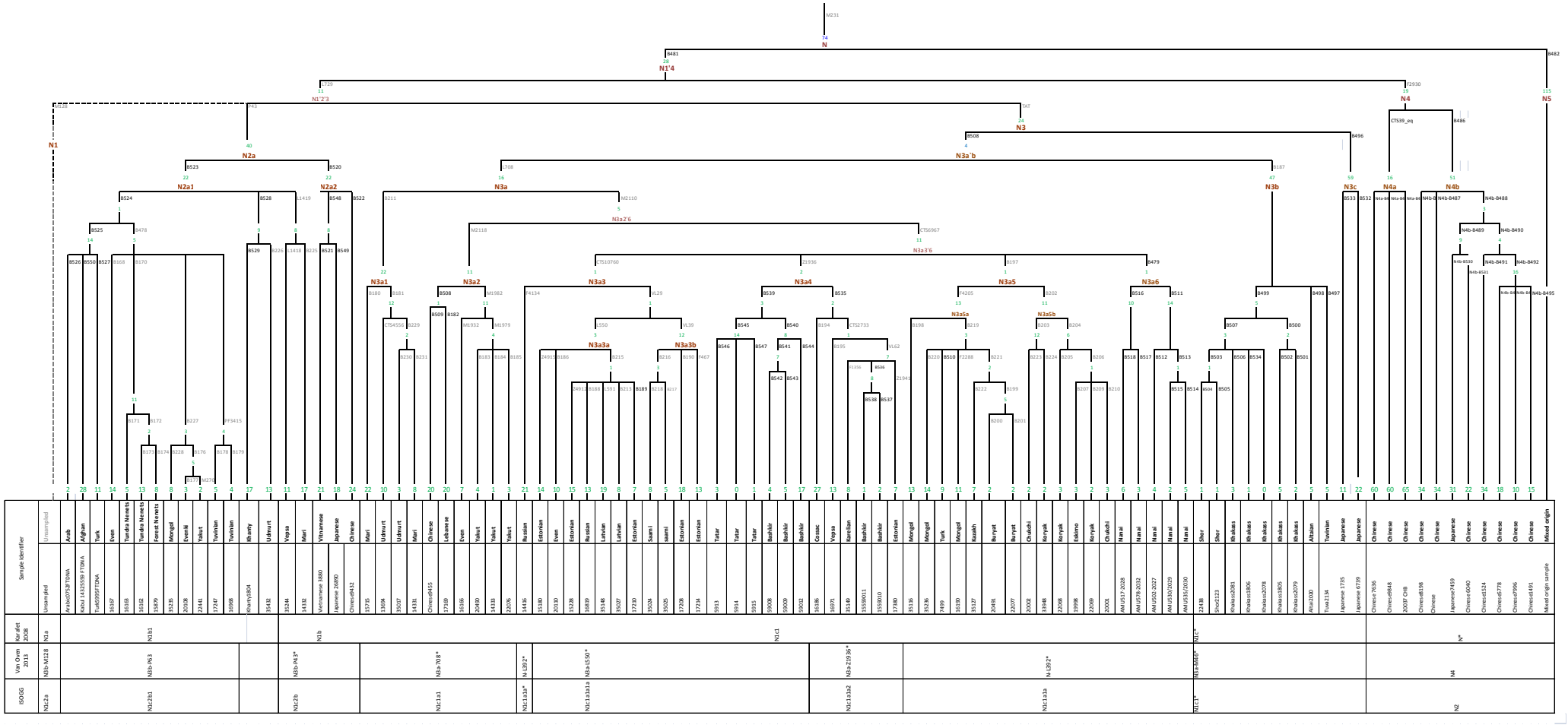
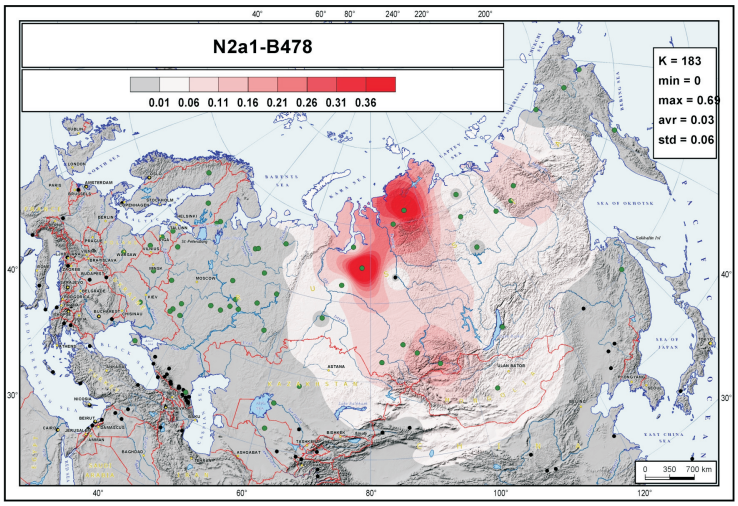
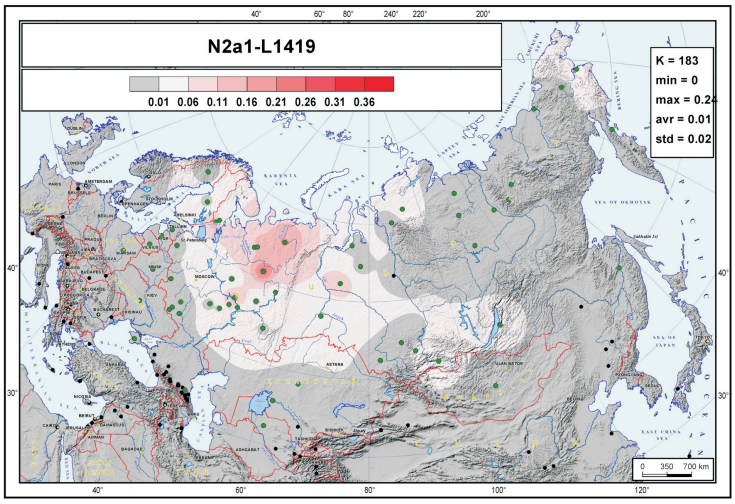
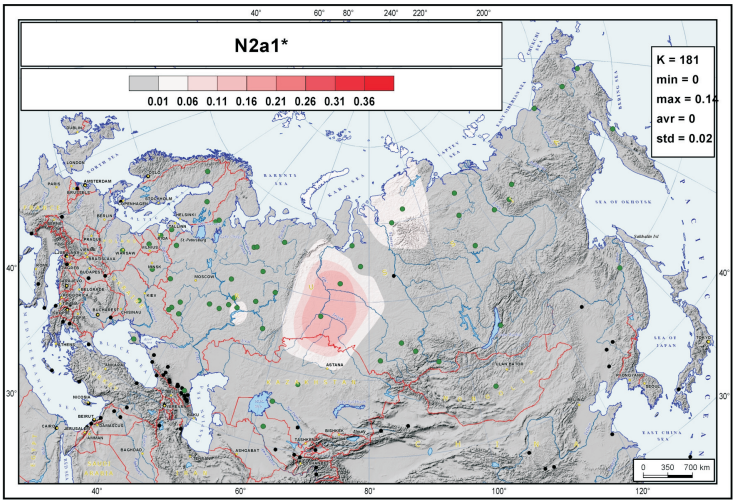
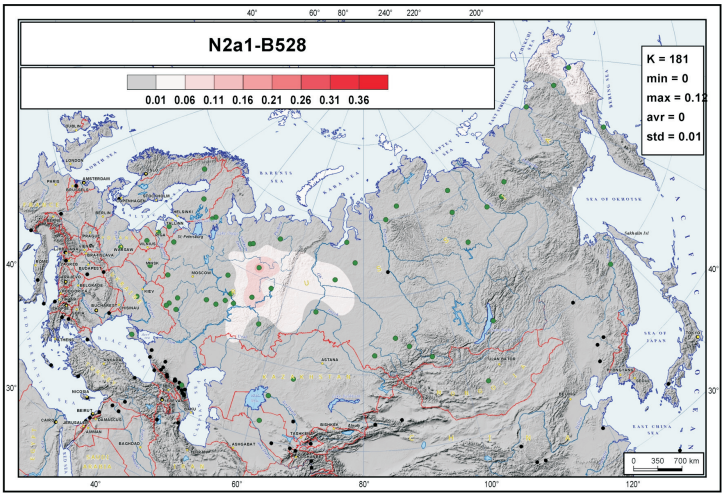



**Figure S2. Refined topology of Y-chromosome haplogroup N based on re-sequencing 6.2 Mbp in 94 individuals described in Table S1. Mutation counts and branching defining markers (included in counts) are shown on the branches. Dashed line indicates known branch without sequence data.**



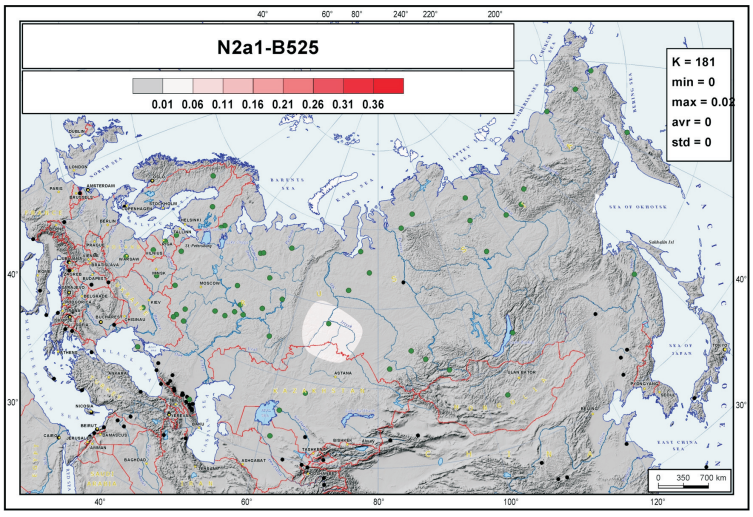




Figure S3. Geographic distribution maps of N2a1 sub-clades in studied regions.
Data points from Table S2 and addtional datapoints with zero frequency values from literature were used for creating the plots. Datapoints from this study are indicated by green dots and datapoints from literature with smaller black dots.

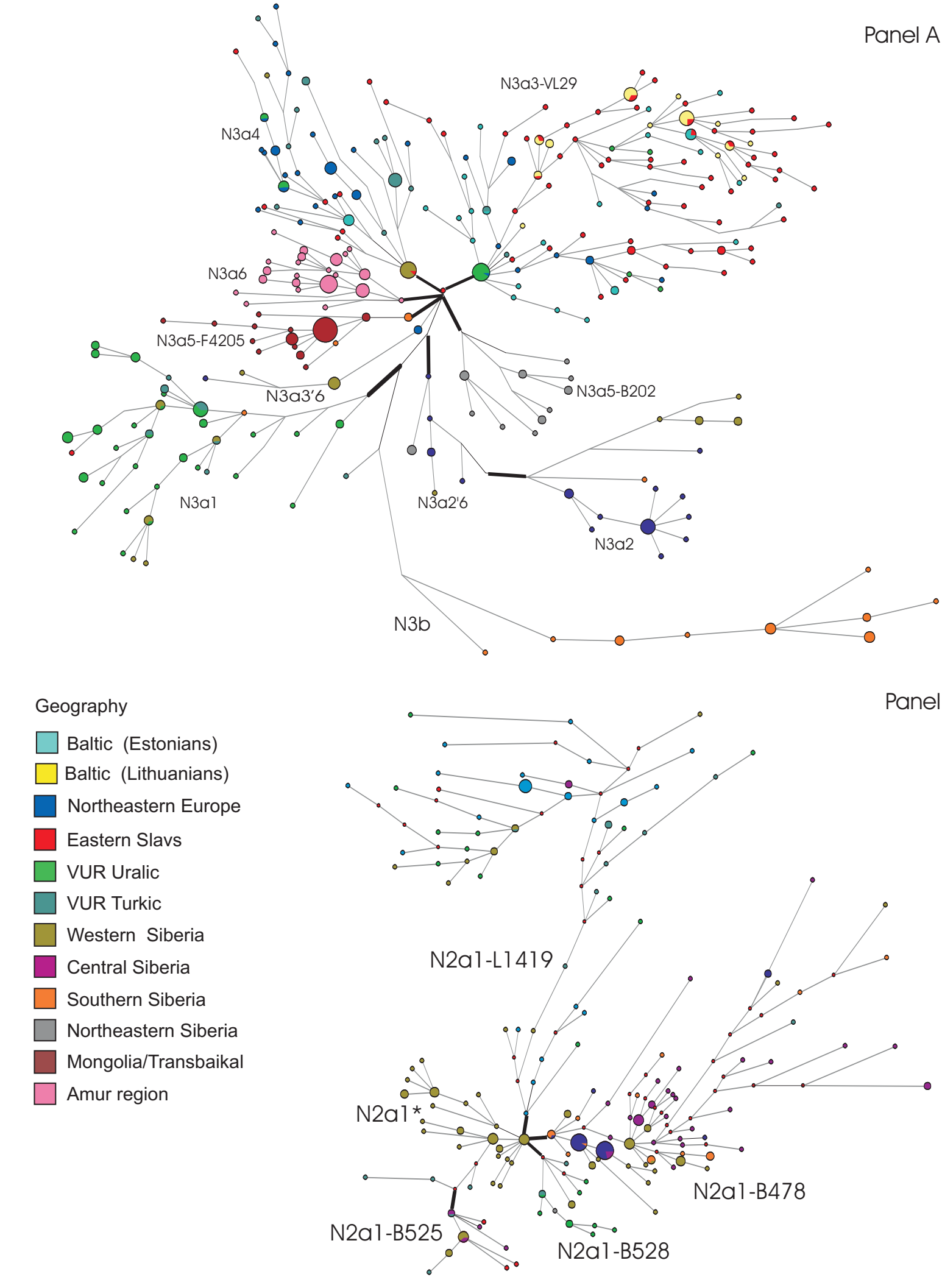

Panel A
N3a3-VL29
N3a4
N3a6
N3a5-F4205
N3a3'6
N3a5-B202
N3a1
N3a2'6
N3a2
N3b
Geography
Baltic (Estonians)
Baltic (Lithuanians)
Northeastern Europe
Eastern Slavs
VUR Uralic
VUR Turkic
Western Siberia
Central Siberia
Southern Siberia
Northeastern Siberia
Mongolia/Transbaikal
Amur region
Panel B
N2a1-L1419
N2a1*
N2a1-B478
N2a1-B525
N2a1-B528

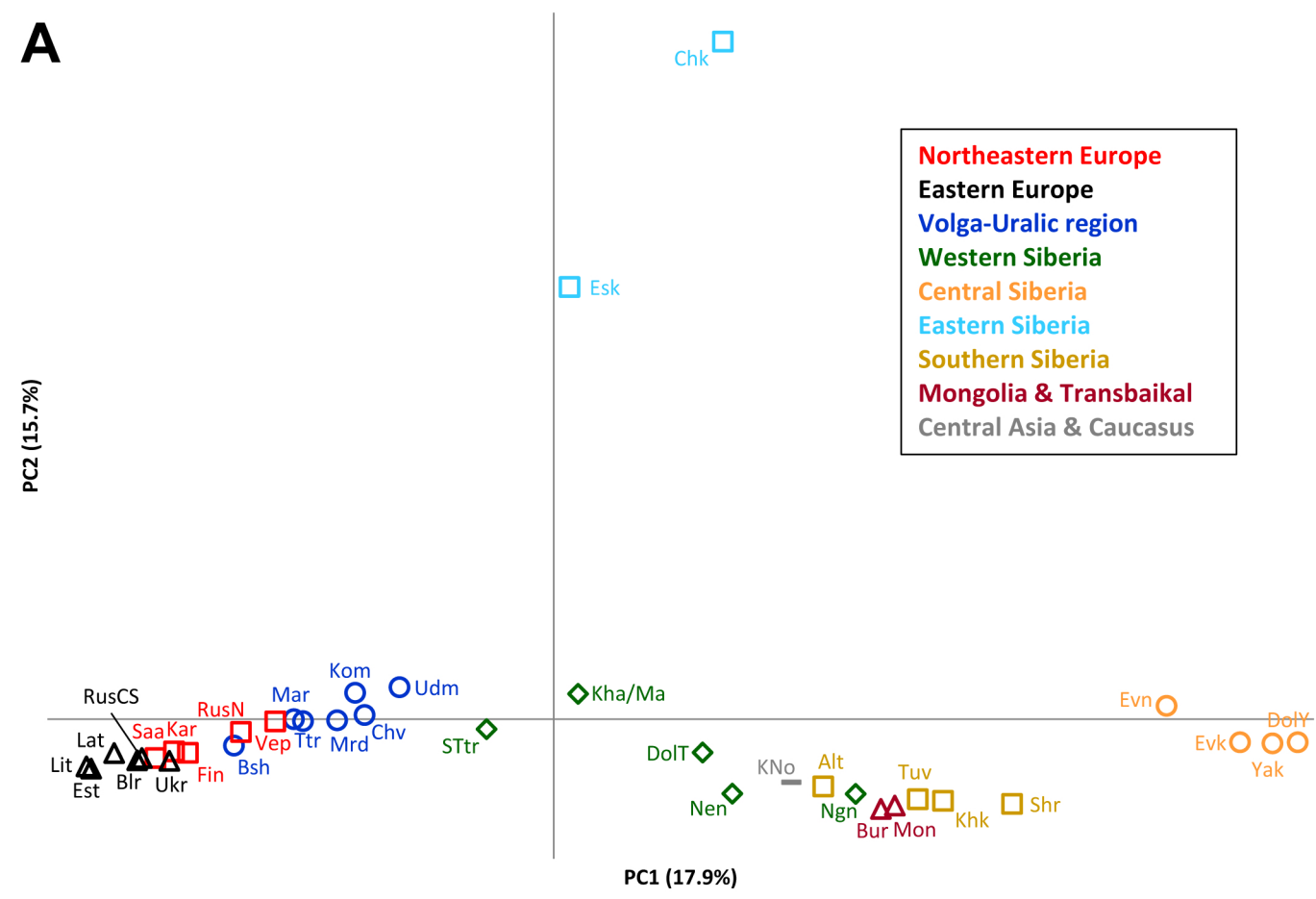
**Figure S4**. Networks of N3 and N2a1 STR haplotypes.

**Panel A)** N3 haplotypes network based on 14 STRs.

Network of 438 samples from the N3 sub-clade (data in Supplementary Table 5). The network combines bi-allelic markers (indicated with thick black lines) defining sub-clades within hg N3 with 16 STR loci (DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS439, DYS437, DYS438, DYS448, DYS456, DYS458, YGATAH4). Datasets are typed with different kits, the overlapping 14 STRs are used for network construction. Circles represent microsatellite haplotypes, the areas of the circles and sectors are proportional to haplotype frequency (the smallest circle corresponds to one individual). The colors indicate groups consisting of the following populations: Baltic region - Uralic-speaking Estonians and Indo-European speaking Lithuanians; Northeastern Europe - Karelians, Vepsas, Northern Russians; Eastern Slavs - Belarusians, Central and South Russians, Ukrainians; populations from Volga-Uralic regions (VUR) are grouped by language - Uralic-speakers (Udmurts, Komis, Maris, Mordvas) and Turkic-speakers (Bashkirs, Chuvashes, Tatars); Western Siberia is represented by Khanties; Central Siberian populations - Yakuts, Evenks, Evens, Yukaghirs; Southern Siberian populations - Tuvans, Altaians, Shors, Khakasses; Northeastern Siberian populations - Koryaks, Eskimos and Chukchis; Amur region is represented by the Nanais.

**Panel B)** N2a1 haplotypes network based on 15 STRs.

Network of 233 samples from the N2a1 sub-clades (data in Supplementary Table 5). The network combines bi-allelic markers (indicated with thick black lines) defining sub-clades within hg N2a1 with 17 STR loci (DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS439, DYS437, DYS438, DYS448, DYS456, DYS458, DYS635,YGATAH4). Datasets are typed with different kits, the overlapping 15 STRs are used for network construction. Circles represent microsatellite haplotypes, the areas of the circles and sectors are proportional to haplotype frequency (smallest circle corresponds to one individual). The colors indicate groups consisting of the following populations: Northeastern European region - Karelians, Vepsas, Northern Russians; Eastern Slavs - Belarusians, Central and Southern Russians; populations from Volga-Uralic regions (VUR) are grouped by language - Uralic-speakers (Udmurts, Komis, Maris) and Turkic-speakers (Bashkirs, Chuvashes, Tatars); Western Siberia is represented by Khanties; Central Siberian populations - Yakuts, Evenks, Evens, Yukaghirs; Southern Siberian populations - Tuvans, Altaians, Shors, Khakasses; Northeastern Siberian populations - Koryaks, Eskimos and Chukchis; Amur region is represented by the Nanais.


A
Chk
Esk
Northeastern Europe
Eastern Europe
Volga-Uralic region
Western Siberia
Central Siberia
Eastern Siberia
Southern Siberia
Mongolia & Transbaikal
Central Asia & Caucasus
PC2 (15.7%)
Kom
Udm
RusCS
Mar
RusN
Kha/Ma
Evn
DolY
Lat
Saa
Kar
Vep
Ttr
Chv
Mrd
STtr
Evk
Yak
Lit
Blr
Ukr
Fin
Bsh
DolT
KNo
Alt
Tuv
Est
Nen
Ngn
Bur
Mon
Khk
Shr
PC1 (17.9%)

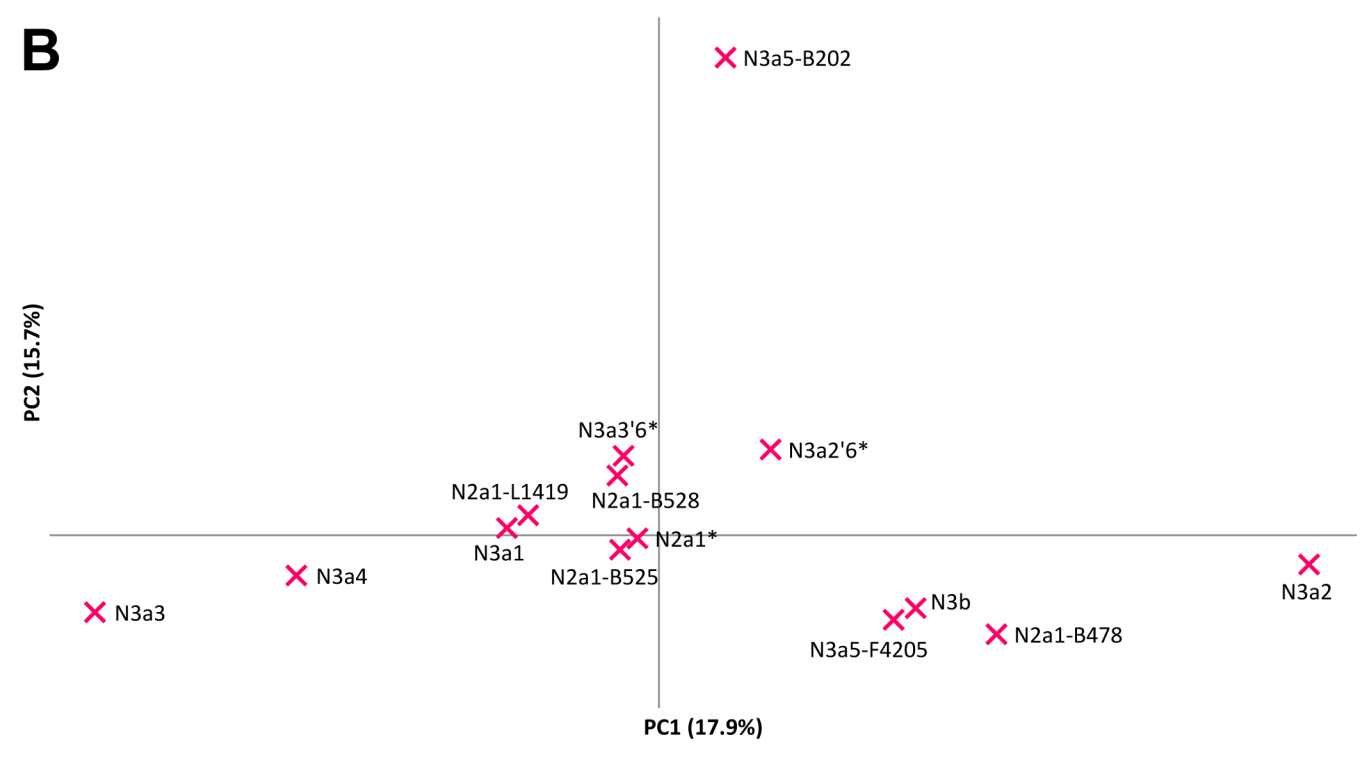

B
N3a5-B202
PC2 (15.7%)
N3a3'6*
N3a2'6*
N2a1-L1419
N2a1-B528
N3a1
N2a1*
N2a1-B525
N3a4
N3a2
N3a3
N3b
N3a5-F4205
N2a1-B478
PC1 (17.9%)

**Figure S5**

**A. PC plot of the populations based on N2a1 and N3 sub-haplogroup frequencies relative to total hg N.**

Colors correspond to the geographic regions of Figure 1. The codes of the populations included in PCA are given in Table S2, populations with fewer than 5 hg N Y-chromosomes or hg N frequency <5% were excluded. All sampled Nanai belong to the sub-hg N3a6, not found in any other population, and were therefore excluded. In cases where several populations were pooled for PCA, they are denoted by the same code in Table S2.

**B.** PC plot of N2a and N3 sub-haplogroups based on their frequencies in the populations relative to total hg N.

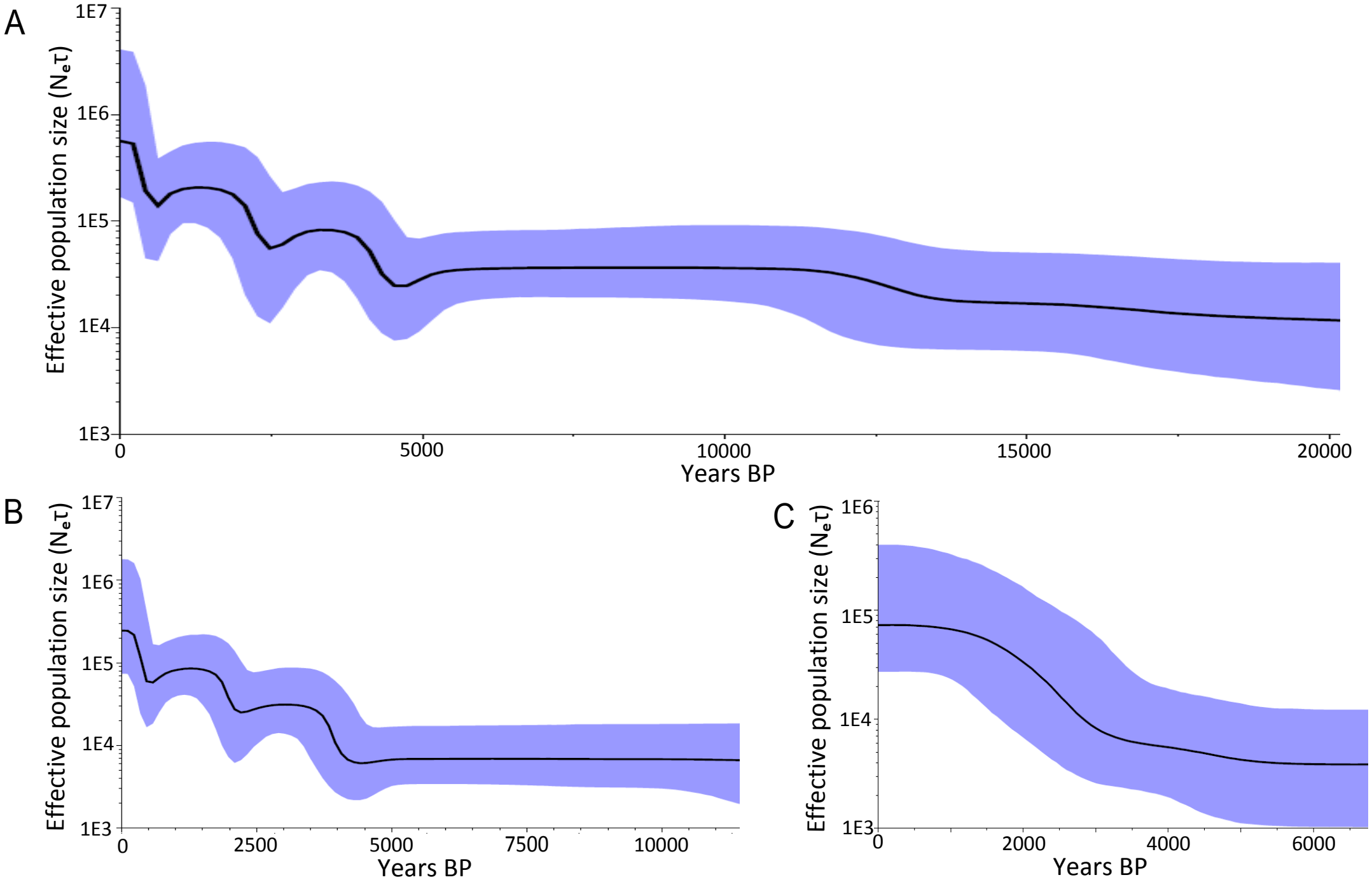



**Figure S6.** Bayesian Skyline Plots for Y-chromosome haplogroup N.
The BSPs show the variation of the effective population size through time for hg N (Panel A), N3 (Panel B) and N2a (Panel C). The BSPs were created using a piecewise-linear coalescence model and sequence data of chromosome Y. The y-axis is equal to $N_eT$ (the product of the effective population size and the generation length in years). The thick black line is the median estimate, and the blue area shows the 95% highest posterior density (HPD) limits

**Table S1. List of high coverage samples and their geographic affiliations**

| | Population | Country of origin | Region of origin | Haplogroup | latitude | longitude | Sequence ID | Source |
|---|---|---|---|---|---|---|---|---|
| 1 | Karelian | Russia | Northeastern Europe | N3a4 | 63,1146 | 33,0139 | GS000035149-ASM | (Karmin et al. 2015) |
| 2 | North-Russian | Russia | Northeastern Europe | N3a3 | 64,7026 | 43,3933 | GS000014416-ASM | (Karmin et al. 2015) |
| 3 | Saami | Norway | Northeastern Europe | N3a3 | 69,8850 | 25,1917 | GS000035024-ASM | (Karmin et al. 2015) |
| 4 | Saami | Norway | Northeastern Europe | N3a3 | 69,8850 | 25,1917 | GS000035025-ASM | (Karmin et al. 2015) |
| 5 | Vepsa | Russia | Northeastern Europe | N2a1 | 61,2588 | 35,5441 | GS000035244-ASM | (Karmin et al. 2015) |
| 6 | Vepsa | Russia | Northeastern Europe | N3a4 | 61,4493 | 34,8212 | GS000016971-ASM | (Karmin et al. 2015) |
| 7 | Kuban-Cossack | Russia | Eastern Europe | N3a4 | 45,0114 | 38,7323 | GS000016186-ASM | (Karmin et al. 2015) |
| 8 | Estonian | Estonia | Eastern Europe | N3a3 | 59,1400 | 24,5400 | GS000015180-ASM | (Karmin et al. 2015) |
| 9 | Estonian | Estonia | Eastern Europe | N3a3 | 58,2100 | 26,2100 | GS000015228-ASM | (Karmin et al. 2015) |
| 10 | Estonian | Estonia | Eastern Europe | N3a3 | 58,9536 | 22,8365 | GS000017208-ASM | (Karmin et al. 2015) |
| 11 | Estonian | Estonia | Eastern Europe | N3a3 | 57,7301 | 26,9168 | GS000017210-ASM | (Karmin et al. 2015) |
| 12 | Estonian | Estonia | Eastern Europe | N3a3 | 58,0930 | 25,1815 | GS000017214-ASM | (Karmin et al. 2015) |
| 13 | Estonian | Estonia | Eastern Europe | N3a4 | 59,2415 | 27,4885 | GS000017380-ASM | (Karmin et al. 2015) |
| 14 | Latvian | Latvia | Eastern Europe | N3a3 | 56,8910 | 24,5078 | GS000035027-ASM | (Karmin et al. 2015) |
| 15 | Latvian | Latvia | Eastern Europe | N3a3 | 56,6234 | 23,2828 | GS000035148-ASM | (Karmin et al. 2015) |
| 16 | Central-Russian | Russia | Eastern Europe | N3a3 | 57,8208 | 28,3365 | GS000016819-ASM | (Karmin et al. 2015) |
| 17 | Bashkir | Russia | Volga-Uralic region | N3a4 | 54,4500 | 55,1300 | GRC15559008 | This study |
| 18 | Bashkir | Russia | Volga-Uralic region | N3a4 | 54,0600 | 53,5100 | GRC15559009 | This study |
| 19 | Bashkir | Russia | Volga-Uralic region | N3a4 | 54,4500 | 55,1300 | GRC15559010 | This study |
| 20 | Bashkir | Russia | Volga-Uralic region | N3a4 | 52,5800 | 58,0300 | GRC15559011 | This study |
| 21 | Bashkir | Russia | Volga-Uralic region | N3a4 | 56,0400 | 54,4900 | GRC15559012 | This study |
| 22 | Mari | Russia | Volga-Uralic region | N2a1 | 55,4692 | 56,0898 | GS000014332-ASM | (Karmin et al. 2015) |
| 23 | Mari | Russia | Volga-Uralic region | N3a1 | 55,4435 | 55,9757 | GS000014331-ASM | (Karmin et al. 2015) |
| 24 | Mari | Russia | Volga-Uralic region | N3a1 | 55,4435 | 55,9757 | GS000015715-ASM | (Karmin et al. 2015) |
| 25 | Tatar | Russia | Volga-Uralic region | N3a4 | 55,0500 | 52,4600 | GRC15559013 | This study |
| 26 | Tatar | Russia | Volga-Uralic region | N3a4 | 54,4500 | 53,0100 | GRC15559014 | This study |
| 27 | Tatar | Russia | Volga-Uralic region | N3a4 | 56,0900 | 50,1100 | GRC15559015 | This study |
| 28 | Udmurd | Russia | Volga-Uralic region | N2a1 | 57,2969 | 52,7563 | GS000035432-ASM | (Karmin et al. 2015) |
| 29 | Udmurd | Russia | Volga-Uralic region | N3a1 | 57,2733 | 54,0578 | GS000013694-ASM | (Karmin et al. 2015) |
| 30 | Udmurd | Russia | Volga-Uralic region | N3a1 | 57,2969 | 52,7563 | GS000035017-ASM | (Karmin et al. 2015) |
| 31 | Khant | Russia | Western Siberia | N2a1 | 63,4800 | 67,3900 | GRC15491804 | This study |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 32 | Forest-Nenets | Russia | Western Siberia | N2a1 | 64,9202 | 77,7779 | GS000015879-ASM | (Karmin et al. 2015) |
| 33 | Tundra-Nenets | Russia | Western Siberia | N2a1 | 67,0045 | 78,2216 | GS000016162-ASM | (Karmin et al. 2015) |
| 34 | Tundra-Nenets | Russia | Western Siberia | N2a1 | 67,0045 | 78,2216 | GS000016163-ASM | (Karmin et al. 2015) |
| 35 | Even | Russia | Central Siberia | N2a1 | 62,6615 | 135,5425 | GS000016167-ASM | (Karmin et al. 2015) |
| 36 | Even | Russia | Central Siberia | N3a2 | 63,2568 | 143,2094 | GS000016166-ASM | (Karmin et al. 2015) |
| 37 | Evenk | Russia | Central Siberia | N2a1 | 63,5954 | 103,9512 | GS000020108-ASM | (Karmin et al. 2015) |
| 38 | Yakut | Russia | Central Siberia | N2a1 | 65,7612 | 105,3406 | GS000022441-ASM | (Karmin et al. 2015) |
| 39 | Yakut | Russia | Central Siberia | N3a2 | 62,3496 | 131,9678 | GS000014333-ASM | (Karmin et al. 2015) |
| 40 | Yakut | Russia | Central Siberia | N3a2 | 68,5162 | 102,1610 | GS000020490-ASM | (Karmin et al. 2015) |
| 41 | Chukchi | Russia | Eastern Siberia | N3a5b | 64,7337 | 177,4916 | GS000020001-ASM | (Karmin et al. 2015) |
| 42 | Chukchi | Russia | Eastern Siberia | N3a5b | 64,7337 | 177,4916 | GS000020002-ASM | (Karmin et al. 2015) |
| 43 | Eskimo | Russia | Eastern Siberia | N3a5b | 64,5041 | 172,8773 | GS000019998-ASM | (Karmin et al. 2015) |
| 44 | Even | Russia | Eastern Siberia | N3a3 | 59,6727 | 150,1240 | GS000020110-ASM | (Karmin et al. 2015) |
| 45 | Koryak | Russia | Eastern Siberia | N3a5b | 61,9661 | 160,3686 | GS000022068-ASM | (Karmin et al. 2015) |
| 46 | Koryak | Russia | Eastern Siberia | N3a5b | 61,9661 | 160,3686 | GS000022069-ASM | (Karmin et al. 2015) |
| 47 | Koryak | Russia | Eastern Siberia | N3a5b | 61,9383 | 159,2329 | GS000033948-ASM | (Karmin et al. 2015) |
| 48 | Nanai | Russia | Eastern Siberia | N3a6 | 49,3600 | 136,3600 | GRC14392027 | This study |
| 49 | Nanai | Russia | Eastern Siberia | N3a6 | 49,3600 | 136,3600 | GRC14392028 | This study |
| 50 | Nanai | Russia | Eastern Siberia | N3a6 | 51,1200 | 138,1800 | GRC14392029 | This study |
| 51 | Nanai | Russia | Eastern Siberia | N3a6 | 51,1200 | 138,1800 | GRC14392030 | This study |
| 52 | Nanai | Russia | Eastern Siberia | N3a6 | 51,1200 | 138,1800 | GRC14392032 | This study |
| 53 | Yakut | Russia | Eastern Siberia | N3a2 | 62,9310 | 152,3849 | GS000022076-ASM | (Karmin et al. 2015) |
| 54 | Altaian | Russia | Southern Siberia | N3b | 51,2200 | 87,1600 | GRC14392020 | This study |
| 55 | Khakas | Russia | Southern Siberia | N3b | 54,0200 | 89,3600 | GRC14392078 | This study |
| 56 | Khakas | Russia | Southern Siberia | N3b | 53,2900 | 91,2100 | GRC14392079 | This study |
| 57 | Khakas | Russia | Southern Siberia | N3b | 54,0300 | 90,2200 | GRC14392081 | This study |
| 58 | Khakas | Russia | Southern Siberia | N3b | 53,0300 | 90,2700 | GRC15491805 | This study |
| 59 | Khakas | Russia | Southern Siberia | N3b | 53,0300 | 90,2700 | GRC15491806 | This study |
| 60 | Shor | Russia | Southern Siberia | N3b | 53,1100 | 88,0800 | GRC14392123 | This study |
| 61 | Shor | Russia | Southern Siberia | N3b | 52,7797 | 87,8641 | GS000022438-ASM | (Karmin et al. 2015) |
| 62 | Tuvinian | Russia | Southern Siberia | N2a1 | 51,4728 | 92,8482 | GS000016968-ASM | (Karmin et al. 2015) |
| 63 | Tuvinian | Russia | Southern Siberia | N2a1 | 51,1621 | 89,4727 | GS000017247-ASM | (Karmin et al. 2015) |
| 64 | Tuvinian | Russia | Southern Siberia | N3b | 51,2400 | 92,3600 | GRC14392134 | This study |
| 65 | Buryat | Russia | Mongolia and Transbaikal | N3a5a | 56,2678 | 112,9980 | GS000020491-ASM | (Karmin et al. 2015) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 66 | Buryat | Russia | Mongolia and Transbaikal | N3a5a | 56,2678 | 112,9980 | GS000022077-ASM | (Karmin et al. 2015) |
| 67 | Mongol | Mongolia | Mongolia and Transbaikal | N2a1 | 47,9090 | 100,8215 | GS000035235-ASM | (Karmin et al. 2015) |
| 68 | Mongol | Mongolia | Mongolia and Transbaikal | N3a5a | 45,7148 | 106,2927 | GS000016190-ASM | (Karmin et al. 2015) |
| 69 | Mongol | Mongolia | Mongolia and Transbaikal | N3a5a | 43,5485 | 104,2822 | GS000035116-ASM | (Karmin et al. 2015) |
| 70 | Mongol | Mongolia | Mongolia and Transbaikal | N3a5a | 49,9971 | 106,5070 | GS000035236-ASM | (Karmin et al. 2015) |
| 71 | Afgan | Afghanistan | Central Asia, Caucasus & West Asia | N2a1 | 34,2400 | 69,1700 | GRC14325559 | This study |
| 72 | Arabic | unknown | Central Asia, Caucasus & West Asia | N2a1 | . | . | GRC13280752 | This study |
| 73 | Kazakh | Kazakhstan | Central Asia, Caucasus & West Asia | N3a5a | 51,1242 | 71,5430 | GS000035127-ASM | (Karmin et al. 2015) |
| 74 | Lebanese | Lebanon | Central Asia, Caucasus & West Asia | N3a2 | 33,8430 | 35,6177 | GS000017169-ASM | (Karmin et al. 2015) |
| 75 | Turkish | Turkey | Central Asia, Caucasus & West Asia | N2a1 | 38,3100 | 29,0900 | GRC13275955 | This study |
| 76 | Turkish | Turkey | Central Asia, Caucasus & West Asia | N3a5a | 39,3500 | 32,5800 | GRC13187499 | This study |
| 77 | Chinese | China | East Asia | N2a2 | 41,4100 | 124,0100 | GRC14389432 | This study |
| 78 | Chinese | China | East Asia | N3a2 | 34,4100 | 109,2500 | GRC12129455 | This study |
| 79 | Chinese | China | East Asia | N4a | 31,5500 | 105,4500 | GRC13188948 | This study |
| 80 | Chinese | China | East Asia | N4a | 28,4400 | 105,2500 | GRC13227636 | This study |
| 81 | Chinese | China | East Asia | N4a | 39,4600 | 116,1800 | NA18558-200-37-ASM | (Karmin et al. 2015) |
| 82 | Chinese | China | East Asia | N4b | 34,3500 | 112,3000 | GRC12126040 | This study |
| 83 | Chinese | China | East Asia | N4b | 29,5100 | 107,0300 | GRC12135043 | This study |
| 84 | Chinese | China | East Asia | N4b | 38,5300 | 115,5300 | GRC13181491 | This study |
| 85 | Chinese | China | East Asia | N4b | 27,5600 | 116,3600 | GRC13211524 | This study |
| 86 | Chinese | China | East Asia | N4b | 38,2100 | 106,3000 | GRC13248198 | This study |
| 87 | Chinese | China | East Asia | N4b | 31,1100 | 111,1700 | GRC13277996 | This study |
| 88 | Chinese | China | East Asia | N4b | 22,2400 | 110,1600 | GRC14355778 | This study |
| 89 | Japanese | Japan | East Asia | N2a2 | 35,4800 | 139,3000 | GRC12126890 | This study |
| 90 | Japanese | Japan | East Asia | N3c | 35,0300 | 137,4100 | GRC13176739 | This study |
| 91 | Japanese | Japan | East Asia | N3c | 43,2400 | 142,2600 | GRC13271735 | This study |
| 92 | Japanese | Japan | East Asia | N4b | 42,5500 | 141,4800 | GRC13277459 | This study |
| 93 | Vietnamese | Vietnam | East Asia | N2a2 | 11,0100 | 106,5500 | GRC13193880 | This study |
| 94 | Mixed origin | unknown | unknown | N5 | . | . | GRC13294033 | This study |
| 95 | Lebbo | Indonesia | Southeast Asia | O2a2 | 1,6553 | 117,1572 | GS000017007-ASM | (Karmin et al. 2015) |
| 96 | Murut | Brunei | Southeast Asia | O1c | 4,6204 | 115,1415 | GS000019985-ASM | (Karmin et al. 2015) |
| 97 | Burmese | Myanmar | East Asia | O3a1 | 19,7361 | 96,2089 | GS000019901-ASM | (Karmin et al. 2015) |

**Table S3. Specifications for SNPs used to genotype population samples (unless otherwise indicated).**

| SNP Position Build37 | Marker name | Clade | Forward Primer 5' → 3' | Reverse Primer 5' → 3' | Product Length (bp) | SNP Position Product | Ancestral | Derived |
|---|---|---|---|---|---|---|---|---|
| 2880546 | B202 | N3a5-B202 | AGTTTAGTATTTTATGGCTGCAAC | GATTTATTCTGAGCTGATTTTCTG | 299 | 146 | T | C |
| 21978781 | B479 | N3a6 | GACAGGGTTCTTATCATTAACAC | GTCTTCTTTTCAGTCTCATGTTG | 462 | 221 | C | A |
| 7870037 | B523_eq [1] | N2a1 | ACCTGACAAACTTTAAGAGAAGAAA | CCCCAGAGAACATTTTGAAATATCA | 294 | 158 | T | G |
| 7207924 | B478 | N2a1-B478 | CAAGAGCAAGGTCTAGTGTA | ACTCTCTACCCTCTGCAAAA | 212 | 121 | T | A |
| 21843827** | B478_eq (**P63**)* | N2a1-B478 | GTCCTATATCTGAAACCAAAAGC | GCAGAATTACCATCCTTAACAAG | 399 | 207 | A | G |
| 7926029 | B524 | N2a1-B524 | CATTGATGTTTTCTCTAGGCTTG | ATTGACTACAGGATCTCAATATGAA | 397 | 224 | C | A |
| 21977460 | B525 | N2a1-B525 | ATACTGTCTCATTTGCTCCCTCTAT | CTTTGTAAACCCTCCCTTGATTA | 460 | 213 | C | A |
| 16041509 | B528 | N2a1-B528 | CTGTTTCTATTTCTATTTCGGGTT | GAAGCTACTATTGTTTGTTTCGAG | 385 | 231 | C | T |
| 14001197 | B187 | N3b | AGTAACACAAAGTAATACAAAGCAG | AGCTCAATTATCCAGTTTTAAGA | 347 | 159 | C | T |
| 14570424 | VL29(CTS2929)[1] | N3a3 | TGGACATATACACCCACACT | GATTGGGGAAAAGTTGGTCA | 228 | 106 | T | C |
| 14827819 | B197[1] | N3a5-B197 | TTATAACTTGTTCTTGGCGTAGATT | TATGATGTTTGCAAGACAGTGAAAT | 382 | 141 | T | C |
| 17090704 | B211[1] | N3a1 | TGTTTCTAGTTGCCCTGATG | AGATGACAGACGGACCTTAA | 196 | 166 | G | A |
| 17216441 | CTS6967(L392_eq) | N3a3'6 | AGAGTGTGTTTCTTTACTGCT | GACACAGGAAGCATGACAA | 195 | 97 | G | C |
| 18973691 | L1419_eq[1] | N2a1-L1419 | ATTATGTCTGCGTTGTCTTATTGAA | GAGACTCAAGCCTGTAATTTTGAA | 391 | 163 | A | T |
| 19282064 | F4205 | N3a5-F4205 | AACTTATGCTGAACTTGAAGATG | GCACCCTAACCAATCCCTTG | 240 | 120 | G | A |
| 21463326 | Z1936 | N3a4 | CTAAACTCGTCCCTCAGTCA | GCTTAACTCTGCCTGACTTC | 232 | 150 | C | T |
| 22762208 | M2110 (CTS10761) | N3a2'6 | AAGAAACCTAAGAAAGCCTGC | ATGTAAGAACGTGCTATCTG | 248 | 142 | T | A |
| 23259624 | M2118 | N3a2 | CCTCTCACTAGCAAAGAACC | CCATGGTACTCTGTTTTCTCA | 187 | 137 | A | G |
| 19080602** | F2930 | N4 | CTGTACTCTCCCTATAATTTCAGTA | CACAACACAGTCAGGAATTCTCA | 400 | 158 | G | A |

* Position 21843827 is indicated as marker P63 in ISOGG with insertion of C in this position, but sequencing confirmed A/G mutation in that position

* *marker not typed in the phylogeographic study

[1] the following markers used in the population survey results in TableS2 were typed but are not listed in Table S6 since they fall outside the 6.2 Mbp sequence lenght.

**Table S4. Datapoints from literature for frequency maps.**

Sample points

| POPULATIONS | References | LATITUDE | LONGITUDE |
|---|---|---|---|
| Italians from South Apulia | 1 | 39,89 | 18,34 |
| Italians from West Campania | 1 | 41,07 | 14,71 |
| Saudi Arabians | 2 | 24,70 | 46,70 |
| Avars | 3 | 42,47 | 46,88 |
| Chechens (Chechnya) | 3 | 43,20 | 45,98 |
| Chechens (Ingushetia) | 3 | 43,20 | 45,20 |
| Darghins | 3 | 42,18 | 47,22 |
| Ingushes | 3 | 43,12 | 45,04 |
| Kubachins | 3 | 42,08 | 47,58 |
| Ossets Digora | 3 | 43,12 | 43,55 |
| Shapsugs | 3 | 44,15 | 39,12 |
| Croatians | 4 | 45,48 | 15,57 |
| Albanians | 5 | 41,19 | 19,49 |
| Bosnians | 5 | 43,85 | 18,42 |
| Greeks Macedonian | 5 | 41,05 | 23,33 |
| Slovenians of Battaglia | 5 | 46,03 | 14,30 |
| Norway | 6 | 59,90 | 10,80 |
| Denmark | 7 | 55,43 | 12,34 |
| UAE | 8 | 24,28 | 54,22 |
| Chinese Blang (Chine) | 9 | 21,40 | 100,20 |
| Chinese Kimmun-Mountain (Chine) | 9 | 22,20 | 101,20 |
| Chinese Miao Guizhou (Chine) | 9 | 26,50 | 108,30 |
| Chinese Miao Yunnan (Chine) | 9 | 24,20 | 103,40 |
| Chinese Mien-N (Chine) | 9 | 23,50 | 106,20 |
| Chinese Mien-Nativer (Chine) | 9 | 24,00 | 111,40 |
| Chinese Mien-W (Chine) | 9 | 22,50 | 101,50 |
| Chinese Pahng (Chine) | 9 | 25,30 | 109,10 |
| Chinese She-N (Chine) | 9 | 27,60 | 119,30 |
| Chinese Yao-Lowland (Chine) | 9 | 24,50 | 110,50 |
| North-West Sicily | 10 | 38,03 | 12,06 |
| Taiwan | 11 | 24,00 | 121,00 |
| Turks North-Eastern | 12 | 40,80 | 38,60 |
| Turks North-Western | 12 | 40,90 | 28,10 |
| Turks South-Eastern | 12 | 37,50 | 39,10 |
| Hungarian | 13 | 46,63 | 20,27 |
| Kalmyks | 14 | 46,00 | 45,38 |
| Egypt | 15 | 30,00 | 31,25 |
| Syrians NE (Ar-Raqqah) | 15 | 36,00 | 38,90 |
| Syrians NW (Aleppo) | 15 | 36,20 | 37,60 |
| Tibet | 16 | 29,58 | 91,12 |
| Lebanese Maronite (Mount) | 17 | 33,80 | 35,50 |
| Okinawa | 18 | 26,32 | 127,78 |
| Evenks (China) | 19 | 50,10 | 125,90 |
| Han Southern | 19 | 24,70 | 115,00 |
| Manchus | 19 | 44,20 | 126,90 |
| Miao | 19 | 28,20 | 107,30 |
| She | 19 | 22,80 | 110,30 |
| Tibetians | 19 | 33,30 | 86,70 |
| Tujians | 19 | 29,30 | 112,30 |
| Uygurs | 19 | 43,90 | 86,90 |
| Yao | 19 | 23,80 | 107,60 |
| Yizu | 19 | 26,50 | 100,50 |
| Miao | 19 | 26,00 | 113,00 |
| She | 19 | 22,00 | 111,50 |
| Tujia | 19 | 27,00 | 115,00 |
| Yao | 19 | 23,00 | 107,00 |
| Chinese Korean | 20 | 42,00 | 124,00 |
| Japanese | 20 | 35,70 | 139,70 |
| Manchurian | 20 | 43,90 | 125,20 |
| Germans (Freiburg i.Br.) | 21 | 48,00 | 7,83 |
| Japanese (Yamaguchi) | 18 | 34,10 | 131,50 |
| Greeks (Crete) | 22 | 35,23 | 25,83 |
| Greeks (Nea Nikomedia) | 22 | 40,58 | 22,25 |
| French (Provence "Neolithic") | 23 | 43,60 | 7,10 |
| Antwerpen | 24 | 51,24 | 4,68 |
| Iranians Muslim (Uromia) | 25 | 37,60 | 45,10 |
| Czech (Brno) | 26 | 49,20 | 16,60 |
| Albanians Kosovar | 27 | 42,70 | 21,20 |
| Dutch | 28 | 52,30 | 4,90 |
| Romanians | 28 | 44,40 | 26,10 |
| Sardinians | 29 | 40,00 | 9,00 |
| Kets | 30 | 63,20 | 87,00 |
| Etulia (Gagauz) | 31 | 45,31 | 28,27 |
| Tajiks (Samarkand) | 32 | 39,60 | 67,60 |
| Uzbeks (Bukhara) | 32 | 39,80 | 64,40 |
| Uzbeks (Fergana Valley) | 32 | 40,60 | 71,00 |
| Yagnobs | 32 | 39,40 | 68,60 |

| | | | |
|---|---|---|---|
| Buyei | 33 | 26,55 | 106,28 |
| Han (Chendu) | 33 | 30,70 | 104,05 |
| Han (MeiXian) | 33 | 34,30 | 107,55 |
| Hui | 33 | 36,98 | 105,92 |
| Li | 33 | 19,17 | 110,03 |
| Qiang | 33 | 31,97 | 102,30 |
| She | 33 | 29,27 | 121,40 |
| Tibetans | 33 | 29,45 | 90,48 |
| Yao (Bama) | 33 | 24,10 | 106,93 |
| Yao (Liannan) | 33 | 24,72 | 112,27 |
| Abazins (Stavropol province) | 34 | 44,68 | 42,00 |
| Abkhazians (Georgia) | 34 | 43,00 | 40,95 |
| Andis (Dagestan) | 34 | 42,60 | 46,20 |
| Balkars (Kabardino-Balkaria) | 34 | 43,33 | 43,33 |
| Dargins (Dagestan) | 34 | 42,08 | 47,58 |
| Georgians (Georgia) | 34 | 42,00 | 44,10 |
| Ingush (Ingushetia | 34 | 43,32 | 45,00 |
| Karachays (Karachay-Cherkessia) | 34 | 43,92 | 42,13 |
| Kumyks (Dagestan) | 34 | 42,73 | 47,37 |
| Tabasarans (Dagestan) | 34 | 42,00 | 47,67 |
| Cypriote | 35 | 35,10 | 33,50 |
| Maltese | 35 | 35,90 | 14,50 |
| Lebanese Shiite Muslim (South) | 36 | 33,20 | 35,42 |
| Lebanese Sunnite Muslim (North) | 36 | 34,50 | 36,27 |
| Buyi (Guizhou) | 37 | 26,20 | 107,50 |
| Buyi (Guizhou) | 37 | 26,10 | 106,20 |
| Han (Guangdong) | 37 | 27,30 | 116,10 |
| Han (Jiangsu) | 37 | 34,50 | 119,10 |
| Han (Shannxi) | 37 | 34,10 | 108,90 |
| Hazak (Xingjiang) | 37 | 43,50 | 82,10 |
| Jing (Guangxi) | 37 | 23,20 | 106,10 |
| She (Fujian) | 37 | 26,20 | 117,50 |
| Tibetan (Qinghai) | 37 | 34,00 | 100,10 |
| Tibetan (Xizang) | 37 | 29,90 | 89,50 |
| Yao (Guangxi) | 37 | 25,70 | 108,70 |


1. Capelli, C. et al. Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. Mol Phylogenet Evol 44, 228-239, (2007).
2. Abu-Amero, K. K. et al. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. Bmc Genet 10, (2009).
3. Balanovsky, O. et al. Parallel Evolution of Genes and Languages in the Caucasus Region. Mol Biol Evol, (2011).
4. Barac, L. et al. Y chromosomal heritage of Croatian population and its island isolates. European Journal of Human Genetics 11, 535-542, (2003).
5. Battaglia, V. et al. Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. Eur J Hum Genet 17, 820-830, (2009).
6. Bosch, E. et al. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Human Genetics 112, 353-363, (2003).
7. Brion, M. et al. A collaborative study of the EDNAP group regarding Y-chromosome binary polymorphism analysis. Forensic Science International 153, 103-108, (2005).
8. Cadenas, A. M., Zhivotovsky, L. A., Cavalli-Sforza, L. L., Underhill, P. A. & Herrera, R. J. Y-chromosome diversity characterizes the Gulf of Oman. Eur J Hum Genet 16, 374-386 (2008).
9. Cai, X. Y. et al. Human Migration through Bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum Revealed by Y Chromosomes. PLoS One 6, e24282, (2011
10. Capelli, C. et al. Population structure in the Mediterranean basin: A Y chromosome perspective. Annals of Human Genetics 70, 207-225, (2006).
11. Kayser, M. et al. Melanesian and asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Molecular Biology and Evolution 23, 2234-2244, (2006).
12. Cinnioglu, C. et al. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114, 127-148 (2004).
13. Csanyi, B. et al. Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. Annals of Human Genetics 72, 519-534, (2008).
14. Derenko, M. et al. Contrasting patterns of Y-chromosome variation in south Siberian populations from Baikal and Altai-Sayan regions. Human Genetics 118, 591-604, (2006).
15. El-Sibai, M. et al. Geographical Structure of the Y-chromosomal Genetic Landscape of the Levant: A coastal-inland contrast. Annals of Human Genetics 73, 568-581, (2009).
16. Gayden, T. et al. The Himalayas as a directional barrier to gene flow. American Journal of Human Genetics 80, 884-894, (2007).
17. Haber, M. et al. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. European Journal of Human Genetics 19, 334-340, (2011).
18. Hammer, M. F. et al. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J Hum Genet 51, 47-58 (2006).
19. Karafet, T. et al. Paternal population history of east Asia: Sources, patterns, and microevolutionary processes. American Journal of Human Genetics 69, 615-628, (2001).
20. Katoh, T. et al. Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. Gene 346, 63-70, (2005).
21. Kayser, M. et al. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Gen 117, 428-443, (2005).
22. King, R. J. et al. Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. Ann Hum Genet 72, 205-214 (2008).
23 King, R. J. et al. The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. BMCEvolBiol11,69, (2011).
24. Larmuseau, M. H. D. et al. Temporal differentiation across a West-European Y-chromosomal cline: genealogy as a tool in human population genetics. European Journal of Human Genetics 20, 434-440, (2012).
25. Lashgary, Z. et al. Y chromosome diversity among the Iranian religious groups: A reservoir of genetic variation. Ann Hum Biol 38, 364-371, (2011).
26. Luca, F. et al. Y-chromosomal variation in the Czech Republic. Am J Phys Anthropol 132, 132-139, (2007).
27. Pericic, M. et al. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among slavic populations. MBE 22, 1964-1975, (2005).
28. Rosser, Z. H. et al. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67, 1526-1543 (2000).
29. Semino, O. et al. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290, 1155-1159 (2000).
30. Karafet, T. M. et al. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum Biol 74, 761-789 (2002).
31. Varzari, A. et al. Searching for the Origin of Gagauzes: Inferences from Y-Chromosome Analysis. Am J Hum Biol 21, 326-336, (2009).
32. Wells, R. S. et al. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci U S A 98, 10244-10249 (2001).
33. Xue, Y. L. et al. Male demography in East Asia: A north-south contrast in human population expansion times. Genetics 172, 2431-2439, (2006).
34. Yunusbayev, B. et al. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. Mol Biol Evol, (2011).
35. Zalloua, P. A. et al. Identifying Genetic Traces of Historical Expansions: Phoenician Footprints in the Mediterranean. American Journal of Human Genetics 83, 633-642, (2008).
36. Zalloua, P. A. et al. Y-chromosomal diversity in Lebanon is structured by recent historical events. Am J Hum Genet 82, 873-882 (2008).
37. Zhong, H. et al. Extended Y Chromosome Investigation Suggests Postglacial Migrations of Modern Humans into East Asia via the Northern Route. MBE 28, 717-727, (2011).

Table S5. Age estimates of hg N clades reported in Figure S1

| No | Name | Post | Age | Lower[a] | Upper[a] | Lower[b] | Upper[b] |
|---|---|---|---|---|---|---|---|
| 1 | NO calibration point | 1.00 | 41,900 | | | 40,175 | 43,591 |
| 2 | N | 1.00 | 25,313 | 22,764 | 27,934 | 21,722 | 28,956 |
| 3 | N2'4 | 1.00 | 19,937 | 17,954 | 21,988 | 17,134 | 22,793 |
| 4 | N2'3 | 1.00 | 17,621 | 15,677 | 19,570 | 14,952 | 20,282 |
| 5 | N3 | 1.00 | 12,989 | 11,336 | 14,648 | 10,802 | 15,173 |
| 6 | N3a'b | 1.00 | 11,914 | 10,365 | 13,579 | 9,875 | 14,060 |
| 7 | N3a | 1.00 | 8,395 | 7,126 | 9,658 | 6,781 | 9,997 |
| 8 | N3a2'6 | 1.00 | 7,113 | 6,076 | 8,252 | 5,783 | 8,539 |
| 9 | N3a3'6 | 1.00 | 4,995 | 4,353 | 5,700 | 4,147 | 5,902 |
| 10 | | 0.26 | 4,848 | 4,208 | 5,477 | 4,009 | 5,673 |
| 11 | | 0.28 | 4,714 | 4,119 | 5,417 | 3,925 | 5,607 |
| 12 | N3a3 | 1.00 | 4,480 | 3,816 | 5,156 | 3,632 | 5,337 |
| 13 | | 0.99 | 4,177 | 3,484 | 4,857 | 3,312 | 5,026 |
| 14 | | 1.00 | 2,887 | 2,282 | 3,624 | 2,163 | 3,741 |
| 15 | | 1.00 | 2,530 | 2,024 | 3,067 | 1,920 | 3,169 |
| 16 | | 0.22 | 2,426 | 1,971 | 2,957 | 1,871 | 3,055 |
| 17 | | 0.13 | 2,338 | 1,817 | 2,873 | 1,721 | 2,967 |
| 18 | | 0.25 | 2,173 | 1,476 | 2,758 | 1,387 | 2,846 |
| 19 | | 0.48 | 2,595 | 2,004 | 3,278 | 1,897 | 3,383 |
| 20 | | 1.00 | 2,466 | 1,954 | 3,038 | 1,853 | 3,138 |
| 21 | | 0.40 | 2,339 | 1,769 | 2,863 | 1,673 | 2,957 |
| 22 | | 1.00 | 1,486 | 818 | 2,173 | 757 | 2,233 |
| 23 | N3a6 | 1.00 | 4,217 | 3,359 | 5,004 | 3,185 | 5,174 |
| 24 | | 1.00 | 1,006 | 530 | 1,633 | 489 | 1,674 |
| 25 | | 1.00 | 589 | 271 | 1,004 | 247 | 1,028 |
| 26 | | 1.00 | 826 | 332 | 1,544 | 298 | 1,577 |
| 27 | N3a5 | 0.99 | 4,580 | 3,943 | 5,284 | 3,755 | 5,469 |
| 28 | | 1.00 | 2,774 | 2,226 | 3,415 | 2,112 | 3,527 |
| 29 | | 1.00 | 2,328 | 1,798 | 2,842 | 1,702 | 2,936 |
| 30 | | 0.22 | 2,230 | 1,684 | 2,750 | 1,592 | 2,840 |
| 31 | | 0.23 | 2,159 | 1,566 | 2,685 | 1,477 | 2,772 |
| 32 | | 1.00 | 1,602 | 1,019 | 2,252 | 953 | 2,317 |
| 33 | | 1.00 | 498 | 182 | 872 | 162 | 892 |
| 34 | | 1.00 | 2,423 | 1,825 | 3,118 | 1,725 | 3,216 |
| 35 | | 1.00 | 862 | 470 | 1,370 | 435 | 1,405 |
| 36 | | 0.99 | 589 | 324 | 951 | 300 | 975 |
| 37 | | 0.35 | 487 | 227 | 780 | 207 | 800 |
| 38 | | 1.00 | 465 | 153 | 801 | 134 | 820 |
| 39 | N3a4 | 1.00 | 4,476 | 3,790 | 5,156 | 3,606 | 5,337 |
| 40 | | 1.00 | 4,027 | 3,197 | 4,762 | 3,031 | 4,925 |
| 41 | | 0.99 | 3,542 | 2,586 | 4,391 | 2,440 | 4,534 |
| 42 | | 1.00 | 2,102 | 1,313 | 2,709 | 1,227 | 2,794 |
| 43 | | 1.00 | 1,955 | 1,211 | 2,584 | 1,131 | 2,663 |
| 44 | | 1.00 | 428 | 107 | 754 | 89 | 771 |
| 45 | | 1.00 | 3,860 | 3,073 | 4,629 | 2,914 | 4,785 |
| 46 | | 1.00 | 526 | 206 | 893 | 184 | 914 |
| 47 | | 0.56 | 328 | 25 | 618 | 12 | 631 |
| 48 | | 1.00 | 2,461 | 1,836 | 3,122 | 1,735 | 3,221 |
| 49 | | 1.00 | 886 | 419 | 1,487 | 383 | 1,523 |
| 50 | N3a2 | 1.00 | 4,490 | 3,594 | 5,394 | 3,409 | 5,575 |
| 51 | | 1.00 | 1,737 | 990 | 2,539 | 919 | 2,609 |
| 52 | | 1.00 | 653 | 338 | 1,103 | 311 | 1,129 |
| 53 | | 0.43 | 506 | 213 | 855 | 192 | 875 |
| 54 | | 0.96 | 4,170 | 3,206 | 5,052 | 3,034 | 5,220 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 55 | N3a1 | 1.00 | 4,325 | 3,263 | 5,285 | 3,085 | 5,460 |
| 56 | | 1.00 | 1,972 | 1,18 | 2,701 | 1,099 | 2,781 |
| 57 | | 1.00 | 1,211 | 561 | 1,952 | 511 | 2,001 |
| 58 | N3b | 1.00 | 2,051 | 1,301 | 2,664 | 1,217 | 2,747 |
| 59 | | 1.00 | 1,149 | 705 | 1,657 | 658 | 1,703 |
| 60 | | 1.00 | 615 | 352 | 927 | 327 | 952 |
| 61 | | 0.21 | 527 | 262 | 794 | 240 | 815 |
| 62 | | 0.59 | 347 | 91 | 596 | 77 | 610 |
| 63 | | 0.20 | 392 | 96 | 645 | 80 | 661 |
| 64 | | 1.00 | 688 | 354 | 1,138 | 326 | 1,166 |
| 65 | | 0.58 | 1,627 | 901 | 2,491 | 834 | 2,557 |
| 66 | N3c | 1.00 | 2,807 | 1,999 | 4,114 | 1,884 | 4,227 |
| 67 | N2a | 1.00 | 9,314 | 7,802 | 10,888 | 7,419 | 11,264 |
| 68 | N2a1 | 1.00 | 4,727 | 4,018 | 5,502 | 3,824 | 5,693 |
| 69 | | 0.99 | 4,391 | 3,720 | 5,109 | 3,539 | 5,286 |
| 70 | | 1.00 | 3,007 | 2,295 | 3,849 | 2,171 | 3,970 |
| 71 | | 0.52 | 2,706 | 2,12 | 3,500 | 2,009 | 3,609 |
| 72 | | 1.00 | 1,819 | 1,142 | 2,514 | 1,067 | 2,587 |
| 73 | | 1.00 | 567 | 244 | 1,019 | 221 | 1,042 |
| 74 | | 0.27 | 2,515 | 1,951 | 3,229 | 1,848 | 3,331 |
| 75 | | 1.00 | 1,130 | 481 | 1,922 | 435 | 1,968 |
| 76 | | 1.00 | 1,652 | 1,049 | 2,367 | 981 | 2,434 |
| 77 | | 1.00 | 1,378 | 834 | 2,088 | 777 | 2,144 |
| 78 | | 1.00 | 2,491 | 1,947 | 3,113 | 1,845 | 3,214 |
| 79 | | 0.70 | 2,174 | 1,355 | 2,812 | 1,266 | 2,900 |
| 80 | | 0.43 | 4,523 | 3,846 | 5,266 | 3,660 | 5,449 |
| 81 | | 1.00 | 2,678 | 2,012 | 3,615 | 1,902 | 3,723 |
| 82 | | 1.00 | 2,707 | 2,034 | 3,595 | 1,923 | 3,704 |
| 83 | N2a2 | 1.00 | 4,909 | 3,864 | 6,158 | 3,662 | 6,356 |
| 84 | | 1.00 | 3,462 | 2,438 | 4,624 | 2,296 | 4,764 |
| 85 | N4 | 1.00 | 16,220 | 14,400 | 18,196 | 13,733 | 18,851 |
| 86 | N4b | 1.00 | 7,162 | 6,041 | 8,363 | 5,746 | 8,652 |
| 87 | | 1.00 | 6,437 | 5,419 | 7,574 | 5,154 | 7,834 |
| 88 | | 1.00 | 5,631 | 4,575 | 6,740 | 4,343 | 6,967 |
| 89 | | 1.00 | 2,606 | 2,016 | 3,282 | 1,909 | 3,387 |
| 90 | | 0.34 | 2,475 | 1,951 | 3,123 | 1,849 | 3,223 |
| 91 | | 1.00 | 4,694 | 3,778 | 5,729 | 3,585 | 5,919 |
| 92 | | 0.39 | 6,919 | 5,841 | 8,174 | 5,556 | 8,453 |
| 93 | N4a | 1.00 | 12,726 | 11,042 | 14,685 | 10,518 | 15,199 |
| 94 | | 0.32 | 12,530 | 10,802 | 14,496 | 10,286 | 15,002 |

No - node number,
Name - node name,
Post - posterior support to the clade,
Age - average age estimate (years),
Lower - lower 95% boundary, Upper - upper 95% boundary of the age estimate,
a - considering only the variance of branch length estimation in BEAST,
b - considering the uncertainty from the confidence intervals of the calibration point age 40,175 - 43,591.