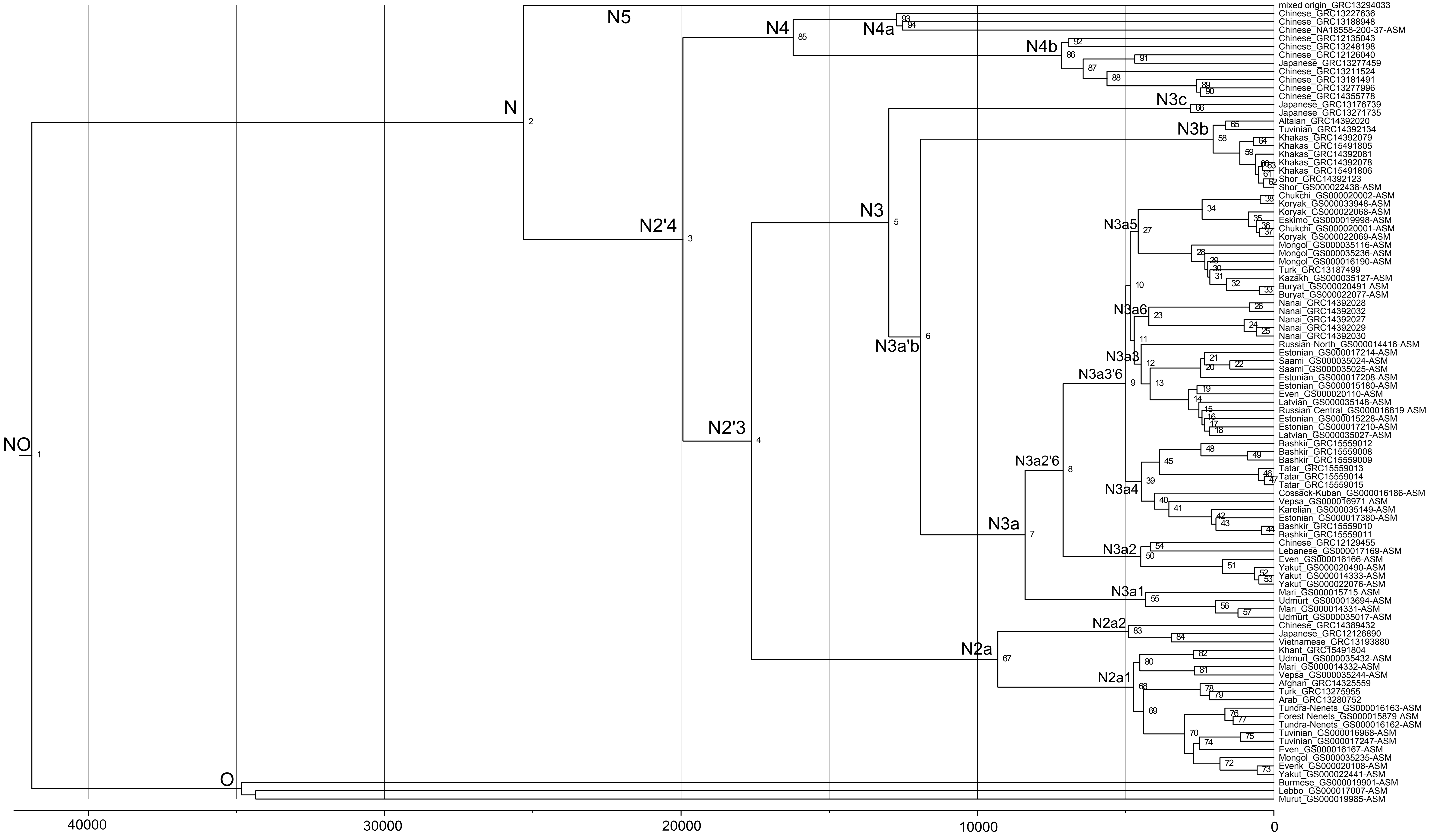
The American Journal of Human Genetics, Volume *99*

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



NO
N
N5
N2'4
N4
N4a
N4b
N2'3
N3
N3c
N3b
N3a'b
N3a5
N3a6
N3a3
N3a3'6
N3a2'6
N3a4
N3a
N3a2
N3a1
N2a
N2a2
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.

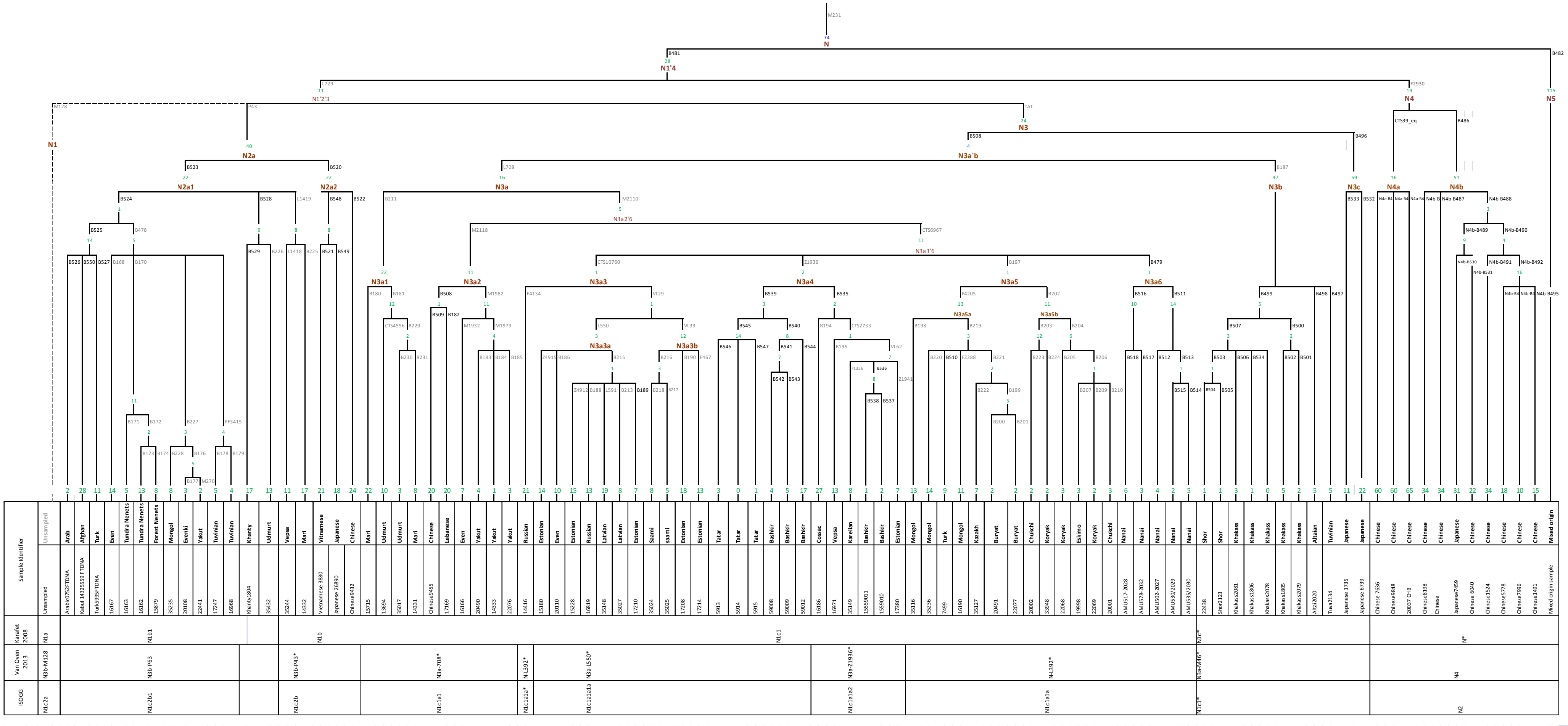


Figure S2. Refined topology of Y-chromosome haplogroup N based on re-sequencing 6.2 Mbp in 94 individuals described in Table S1. Mutation counts and branching defining markers (included in counts) are shown on the branches. Dashed line indicates known branch without sequence data.

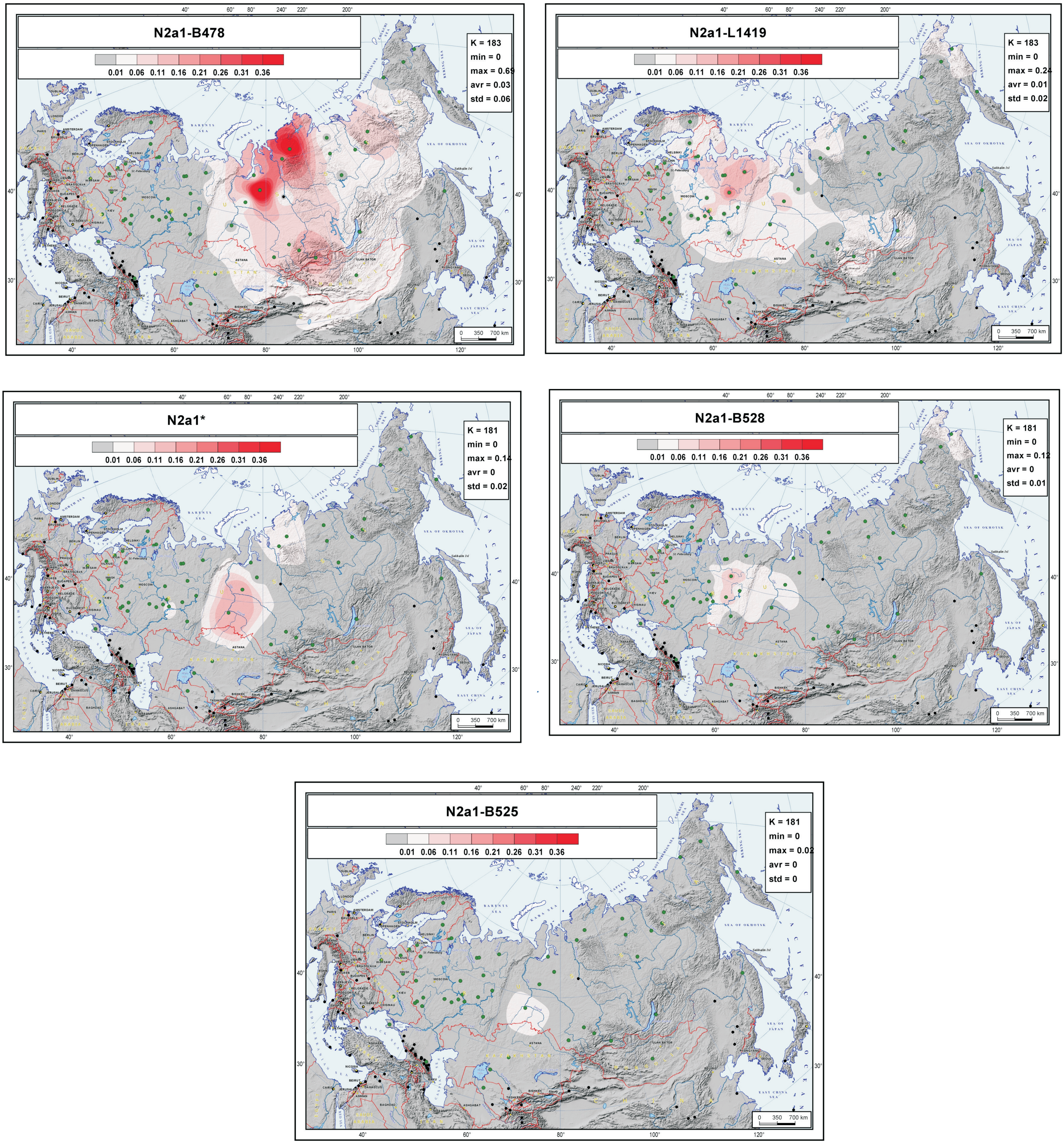
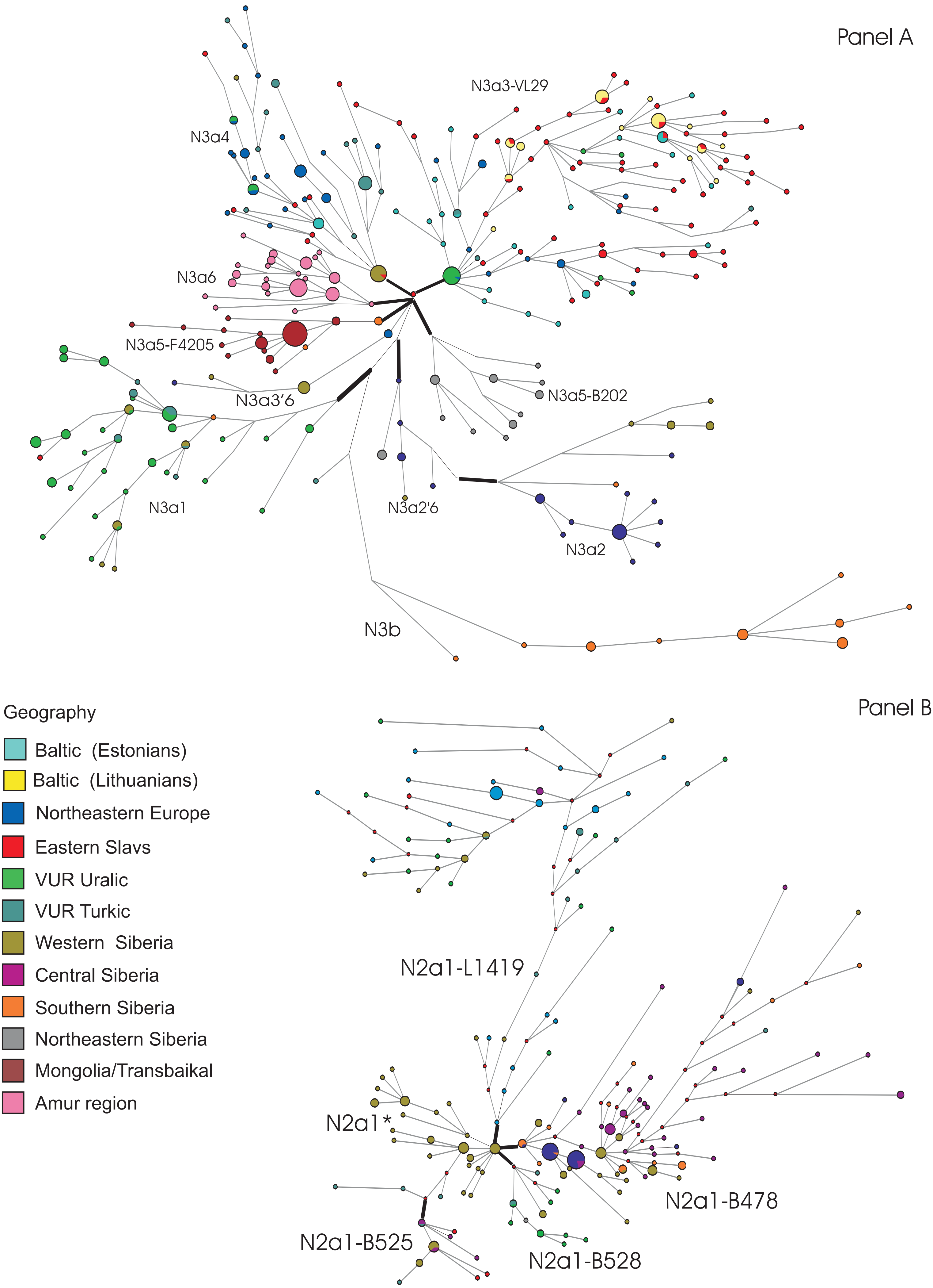


Figure S3. Geographic distribution maps of N2a1 sub-clades in studied regions.
Data points from Table S2 and addtional datapoints with zero frequency values from literature were used for creating the plots. Datapoints from this study are indicated by green dots and datapoints from literature with smaller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



Panel A
N3a3-VL29
N3a4
N3a6
N3a5-F4205
N3a3'6
N3a5-B202
N3a1
N3a2'6
N3a2
N3b
Panel B
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
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.

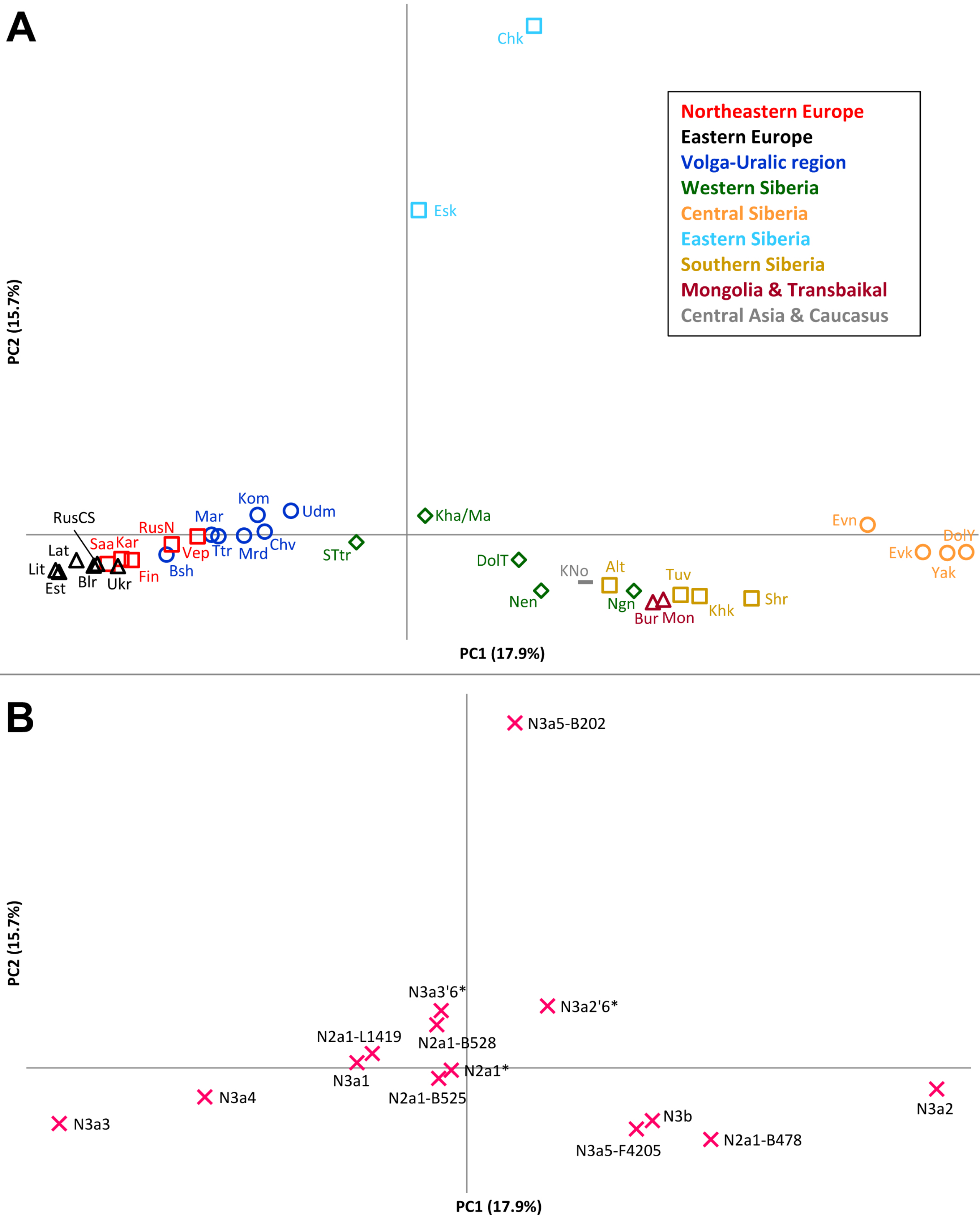

A
Chk
Esk
Northeastern Europe
Eastern Europe
Volga-Uralic region
Western Siberia
Central Siberia
Eastern Siberia
Southern Siberia
Mongolia & Transbaikal
Central Asia & Caucasus
PC2 (15.7%)
Kom
Udm
Mar
RusCS
RusN
Saa
Kar
Lat
Vep
Ttr
Mrd
Chv
Lit
Est
Blr
Ukr
Fin
Bsh
STtr
Kha/Ma
DolT
Nen
KNo
Alt
Ngn
Tuv
Bur
Mon
Khk
Shr
Evn
Evk
DolY
Yak
PC1 (17.9%)
B
N3a5-B202
PC2 (15.7%)
N3a3'6*
N3a2'6*
N2a1-L1419
N2a1-B528
N3a1
N2a1*
N2a1-B525
N3a4
N3a3
N3b
N3a5-F4205
N2a1-B478
N3a2
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.

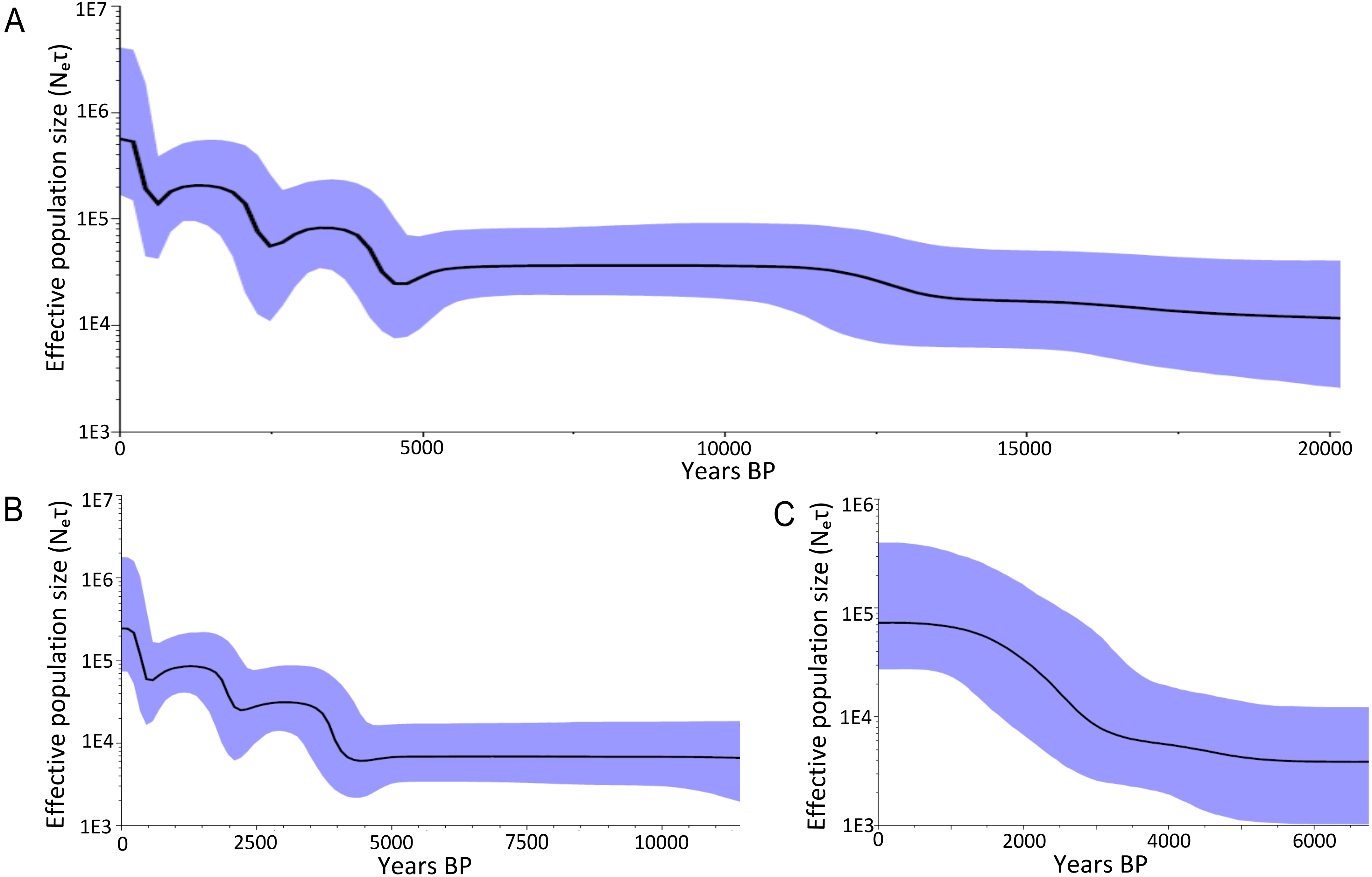


Figure S6. Bayesian Skyline Plots for Y-chromosome haplogroup N.
The BSPs show the variation of the effective population size through time for hg N (Panel A), N3 (Panel B) and N2a (Panel C). The BSPs were created using a piecewise-linear coalescence model and sequence data of chromosome Y. The y-axis is equal to $N_e\tau$ (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.