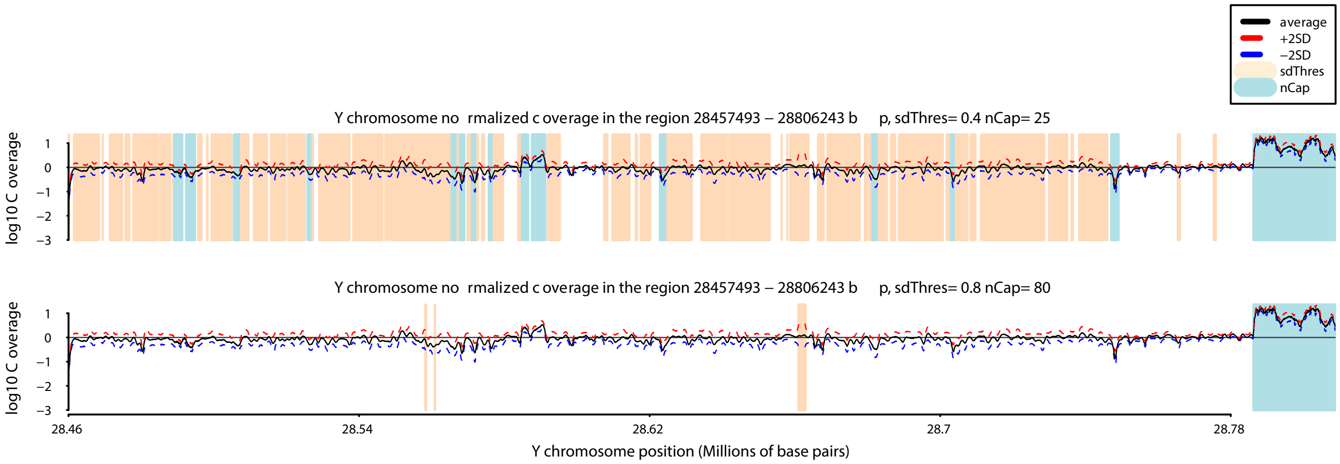


Figure S1. An example of a Y chromosome region filtering by local unique sequence coverage (UC). Combinations of filtering parameters (sdThres/nCap) applied to the same chrY region – upper panel 0.4/25; lower panel 0.8/80. In a sample of 307 Y chromosomes we normalized each individual's average UC in 1000 base pair (bp) windows (sliding by 50 bps) to that of the individual's average UC on chrX, then calculated the log10 average and standard deviation (SD) across all individuals. We then established two parameters for excluding sequence stretches of poor unique coverage: 1) sdThres – wide SD margins (exceeding a set value) indicate that coverage varies between individuals in the region; 2) nCap – very low or very high coverage (+2 SD < 0 or -2 SD > 0) in a set number of consecutive windows suggests a deletion/duplication in the region.

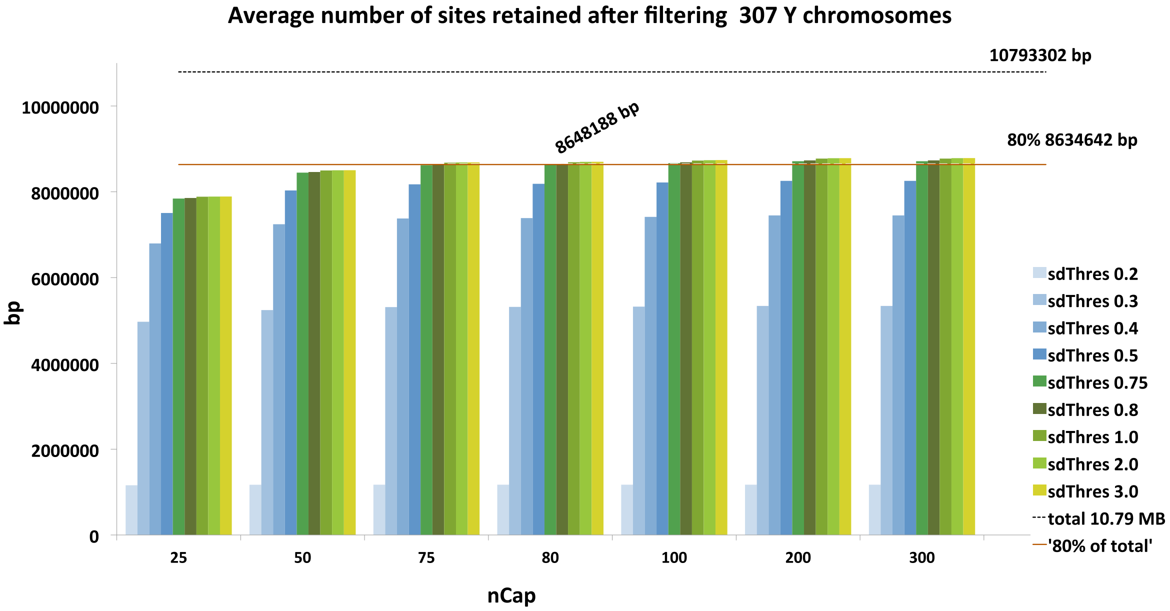


Figure S2. The effect of different combinations of nCap/sdThres on the average number of sites retained. The effect of different combinations of nCap/sdThres (filter b) on the sequence length (average number of sites) retained. Within 10.79 MB of chrY regions we applied a re-mapping filter, removed individual bases with less than 5x unique coverage and then tested the effect of nCap/sdThres combinations on the average number of nucleotides in the sample of 307 Ychromosomes for which coverage data was available. The combination of these two parameters reaches a plateau approximately at nCap=75/sdThres=0.75 where the gains in retaining the sequence length from further relaxing the parameters is negligible – the difference between nCap100/sdThres1.0 and nCap300/sdThres 3.0 is only 0.01%. To retain at least 80% of the original data we picked a point between nCap75/sdThres0.75 and nCap100/sdThres1.0 – and made final analyses with the parameters nCap80/sdThres0.8. Dashed line marks the 10.79MB, orange line marks the point at which 80% of the original sequence length was retained.

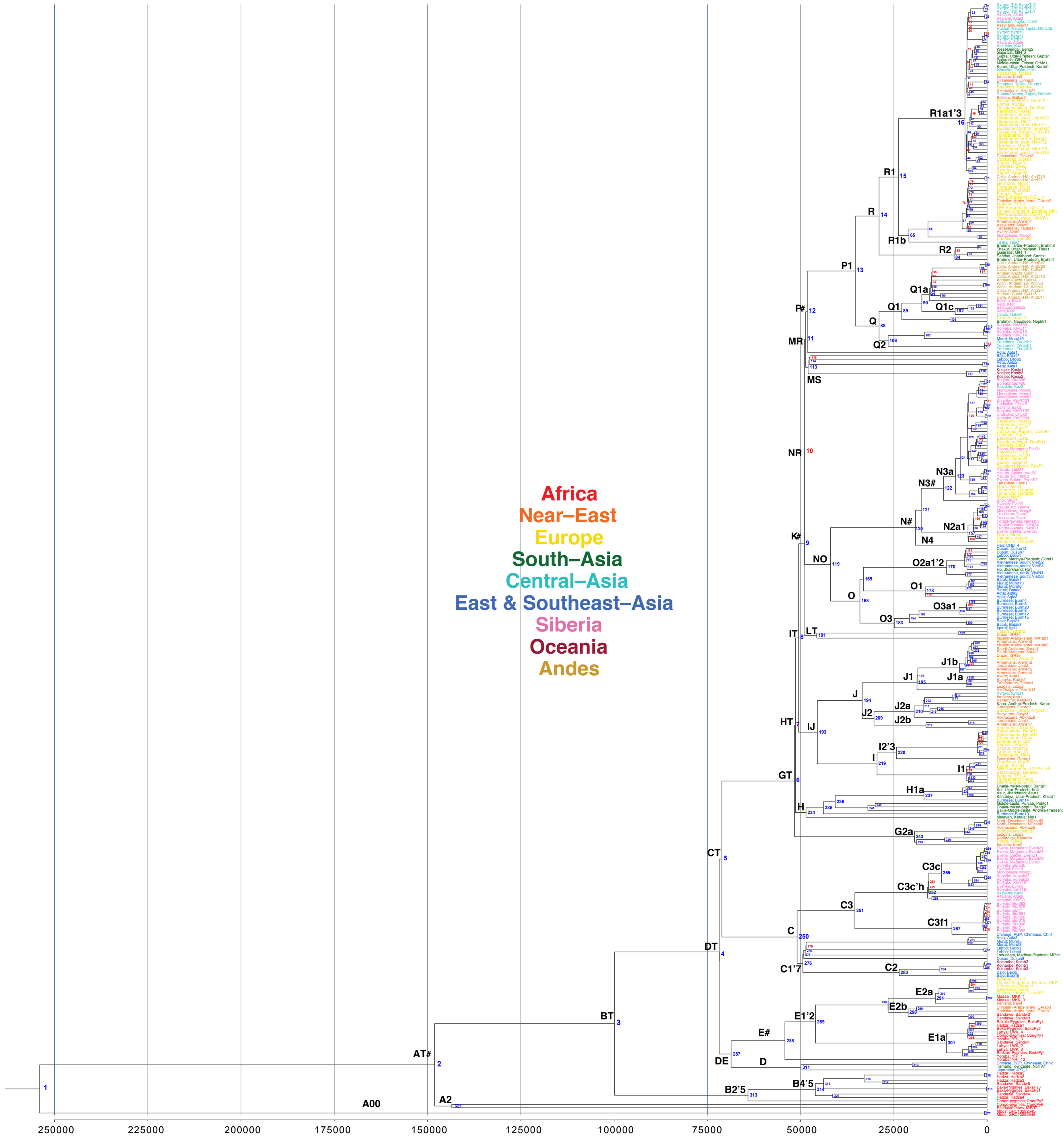


Figure S3. Y chromosome phylogeny based on BEAST analyses. Sample names correspond to those reported in Table S6. Age and posterior support estimates for each numbered clade are reported in Table S7. Six clades that showed significant (p_Delta1 <0.05) phylogenetic asymmetry in SymmeTree analyses are highlighted by # after the haplogroup label.

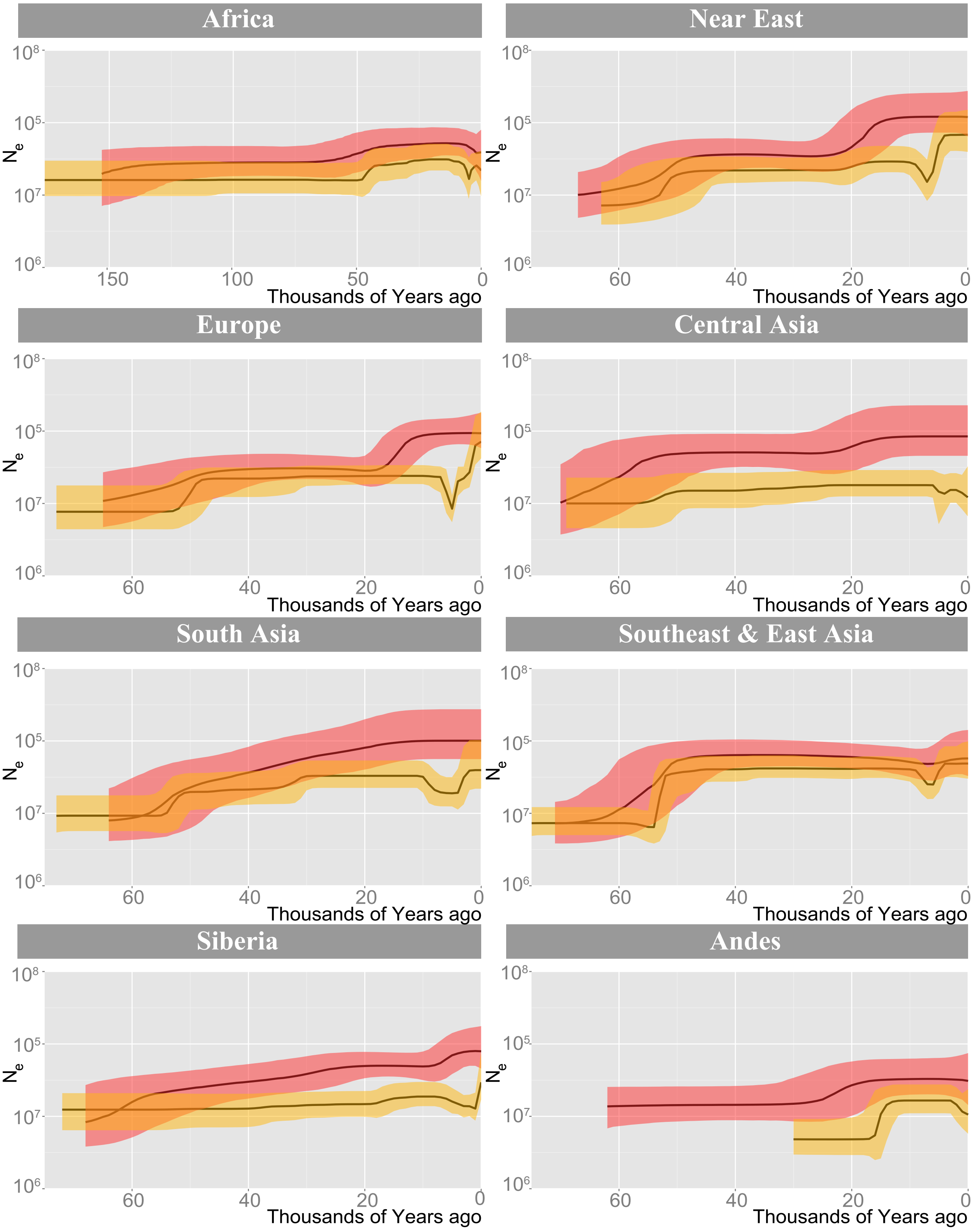


Figure S4A. NRY and mtDNA Bayesian Skyline Plots for each geographical region.

The 95% CI for mtDNA BSPs are represented in red and for NRY in yellow.
The BSPs were created using a piecewise-linear coalescence model.

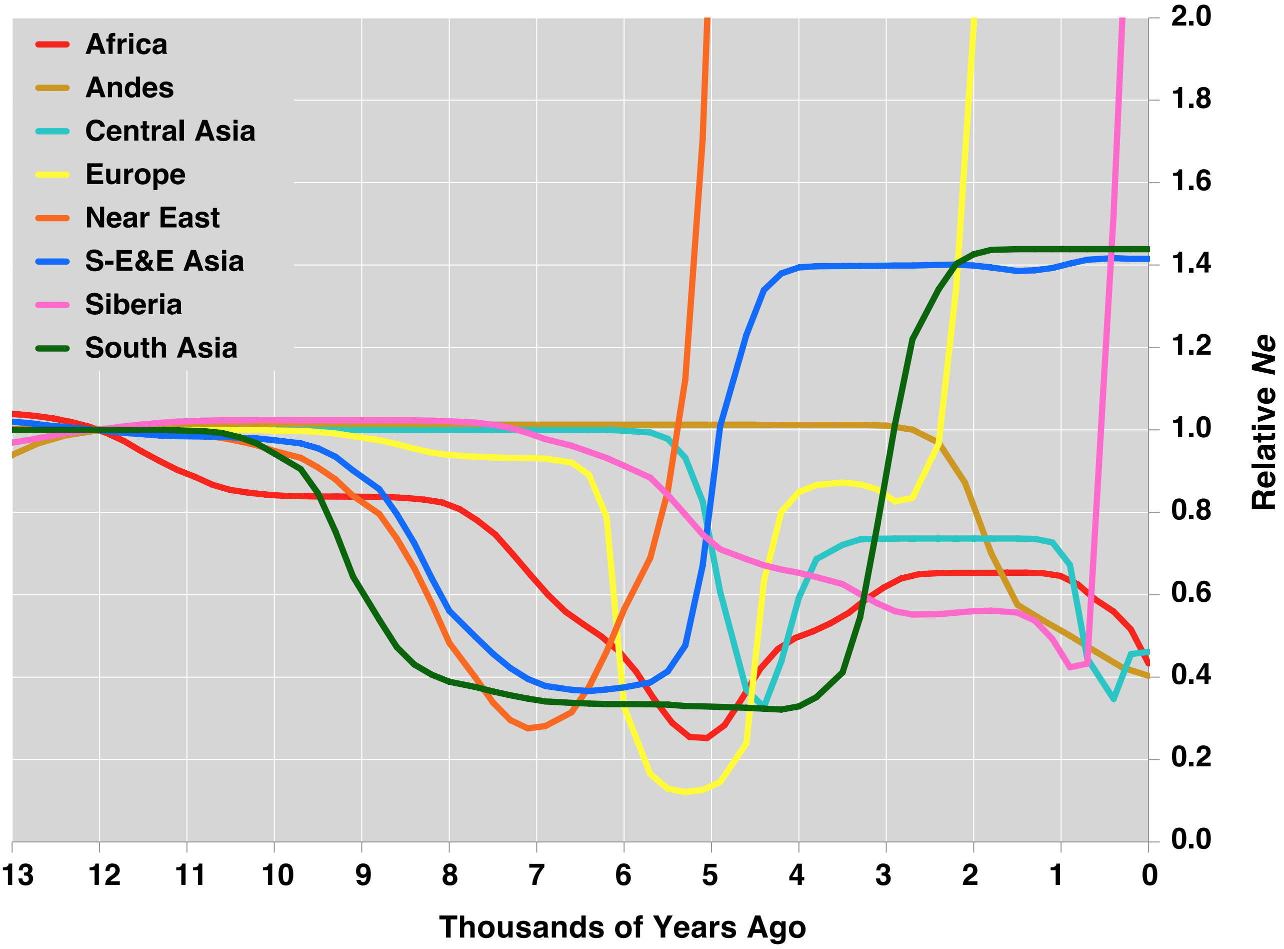


Figure S4B. Variation of Y chromosome effective population size over the past 13 kya. The regional NRY BSPs presented in Figure S4A are shown relative to their values at 12 kya.

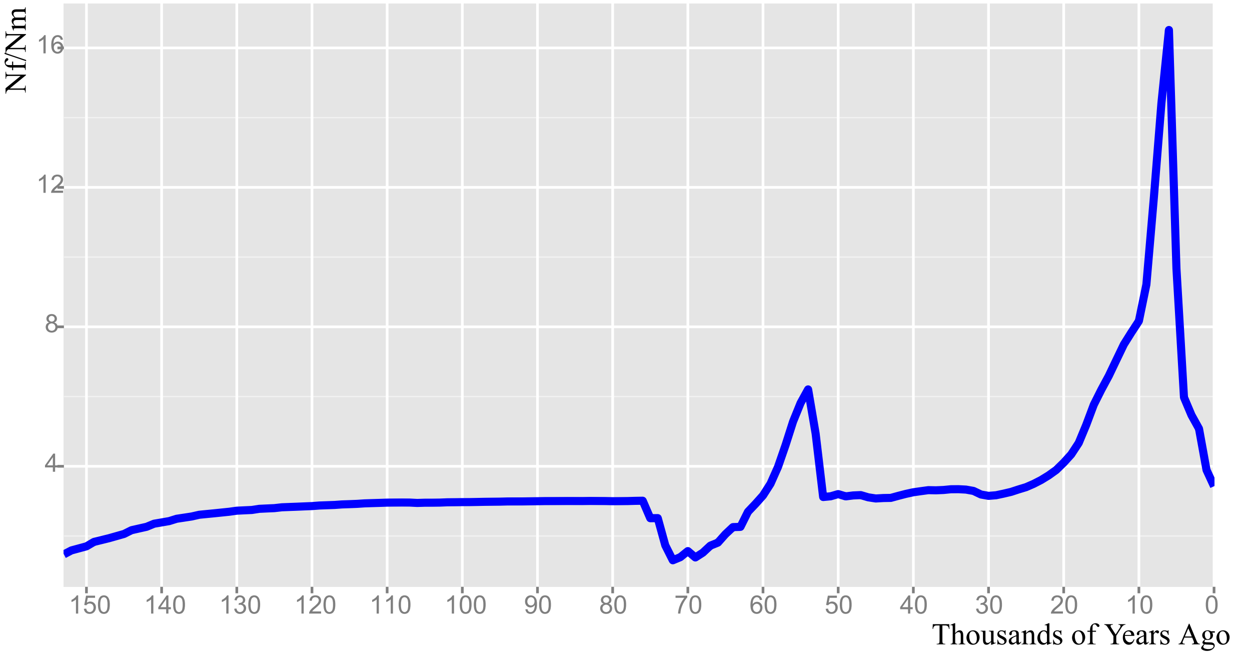


Figure S5 – The temporal dynamics of the ratio of female (Nf) and male (Nm) effective population size in the last 140KY. The ratios of the global accumulative Ne estimates of mtDNA (Nf) and Y chromosome (Nm), as presented in Figure 2, are plotted against the time (in thousands of years) back from the present (0). The BSPs estimates of Ne were obtained in BEAST using a piecewise-linear coalescence model.

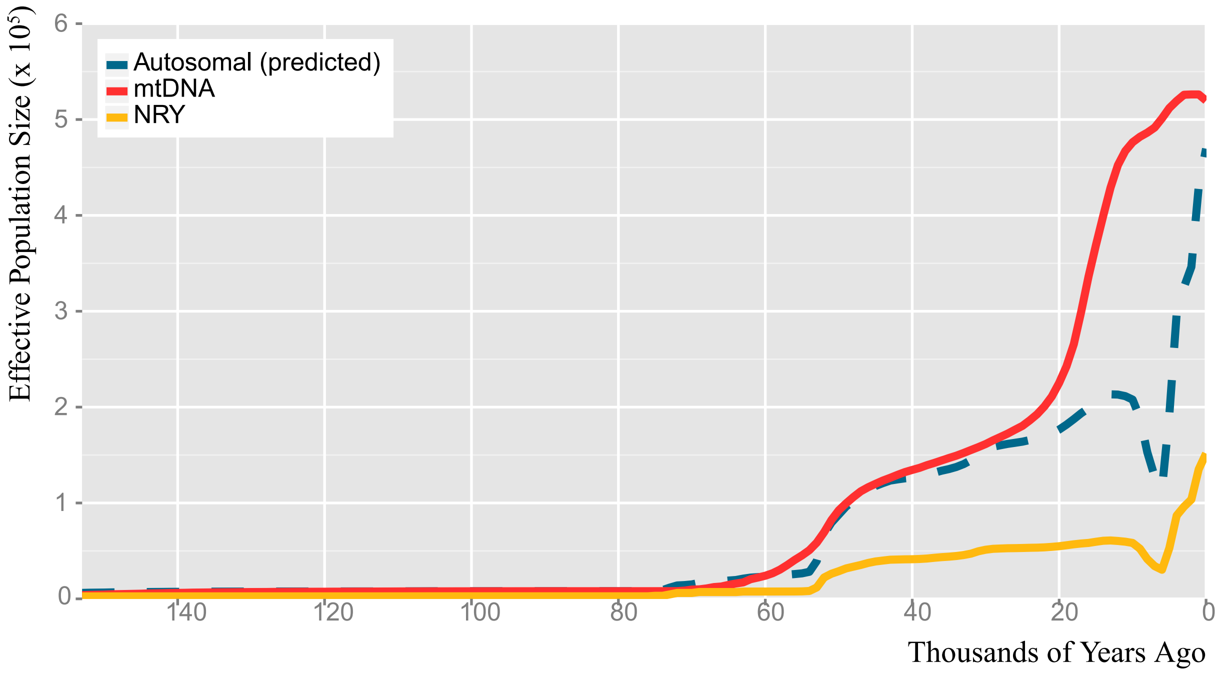


Figure S6 – The temporal dynamics of the autosomal effective population size in the last 140KY as predicted from female (Nf) and male (Nm) effective population sizes. The autosomal estimates were derived applying the 4NfNm/(Nf+Nm) formula on the mtDNA and Y chromosome based estimates.

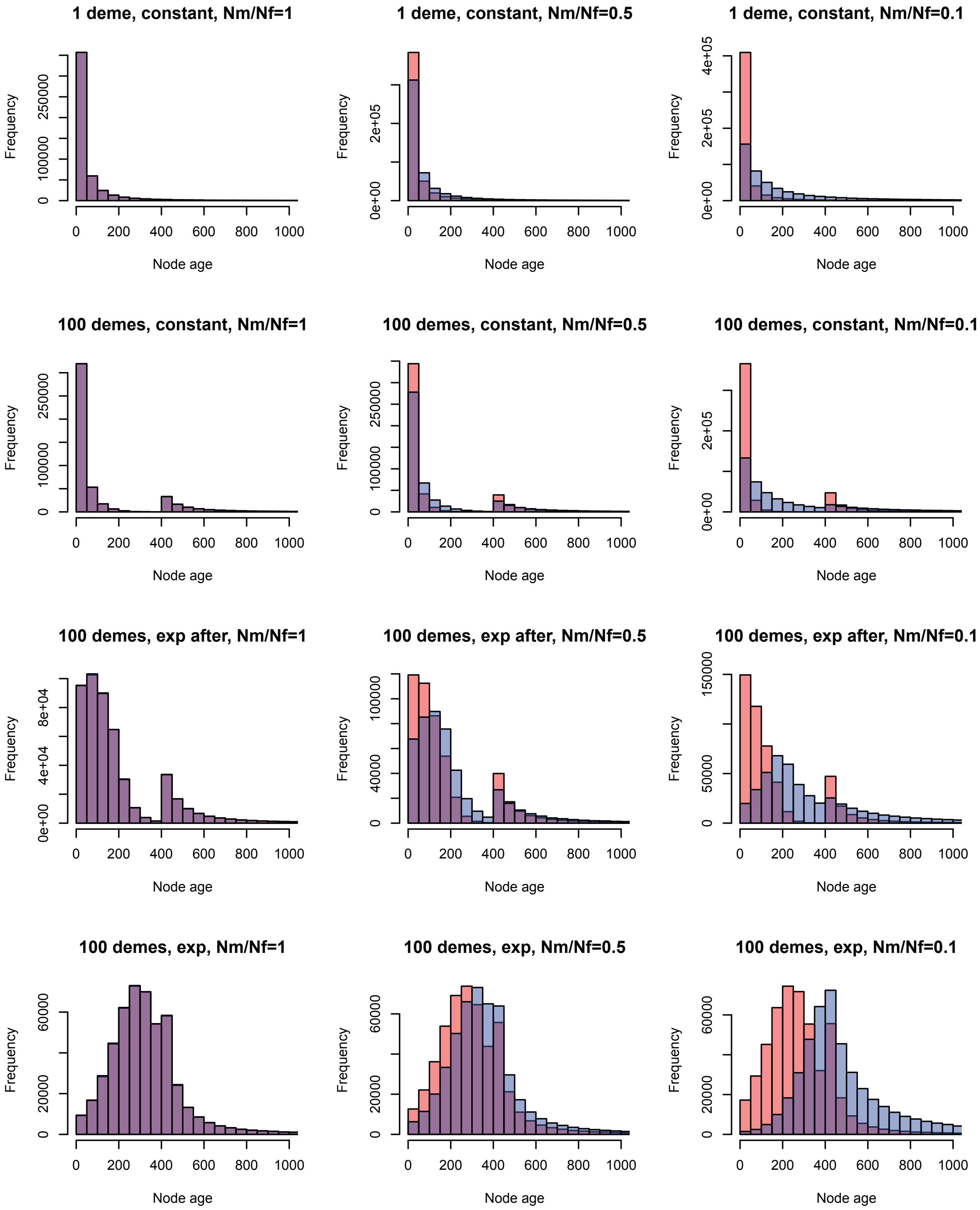


Figure S7 - Distribution of node ages for 1000 coalescent simulation replicates of Y chromosome (red) and mtDNA (blue) lineages under varying scenarios of population growth, deme formation, and effective number of males (Nm) and females (Nf).

Plots show histograms of node ages, measured in generations, from fastsimcoal2 coalescent simulations run for four scenarios: Row 1 - one deme with a constant population size; Row 2 - 1 deme divided into 100 equally-sized demes formed 400 generations ago with a constant population size; Row 3 - formation of 100 demes 400 generations ago, with exponential growth starting 200 generations ago; Row 4 - formation of 100 demes 400 generations ago, with exponential growth starting 400 generations ago. The first column shows simulations run, assuming the effective number of males relative to the effective number of females (Nm/Nf) is equal. The second column shows simulations where Nm/Nf = 0.5, and the third column shows simulations with an extreme Nm/Nf=0.1. Nm and Nf are computed assuming a constant autosomal effective size of 10,000. All simulations sample 500 individuals, distributed evenly between demes.

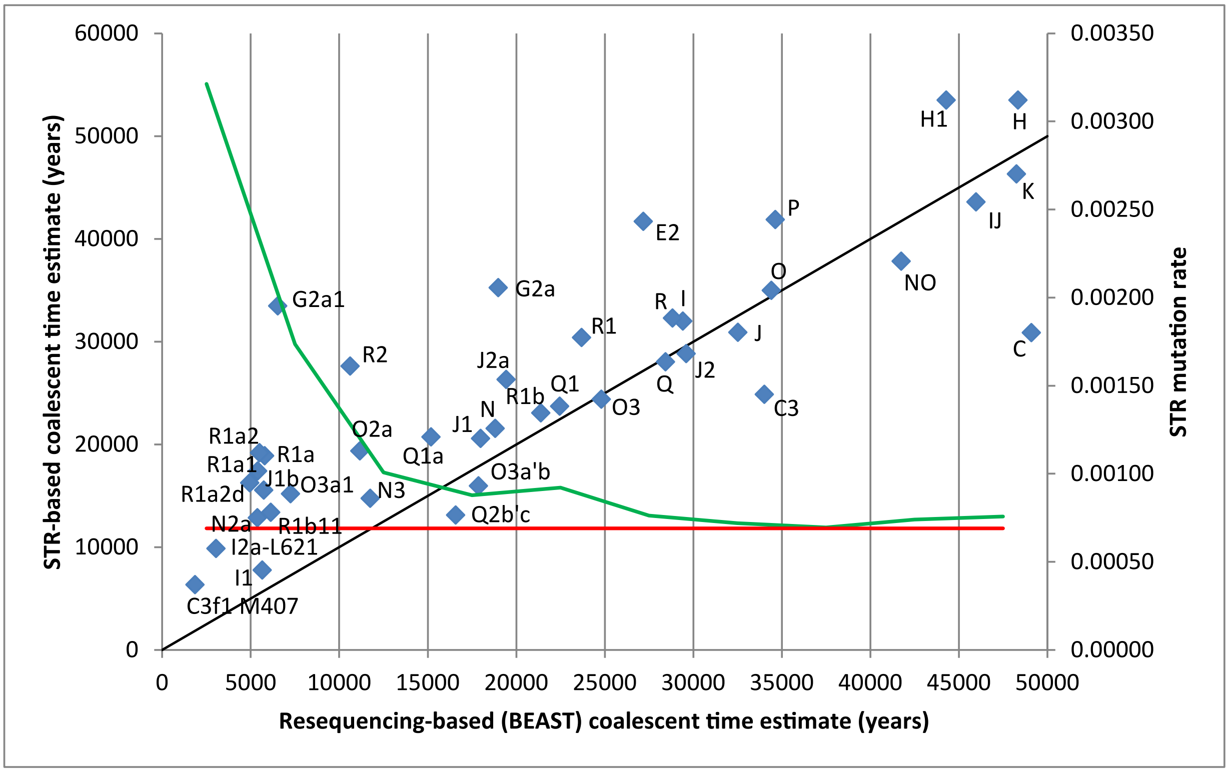
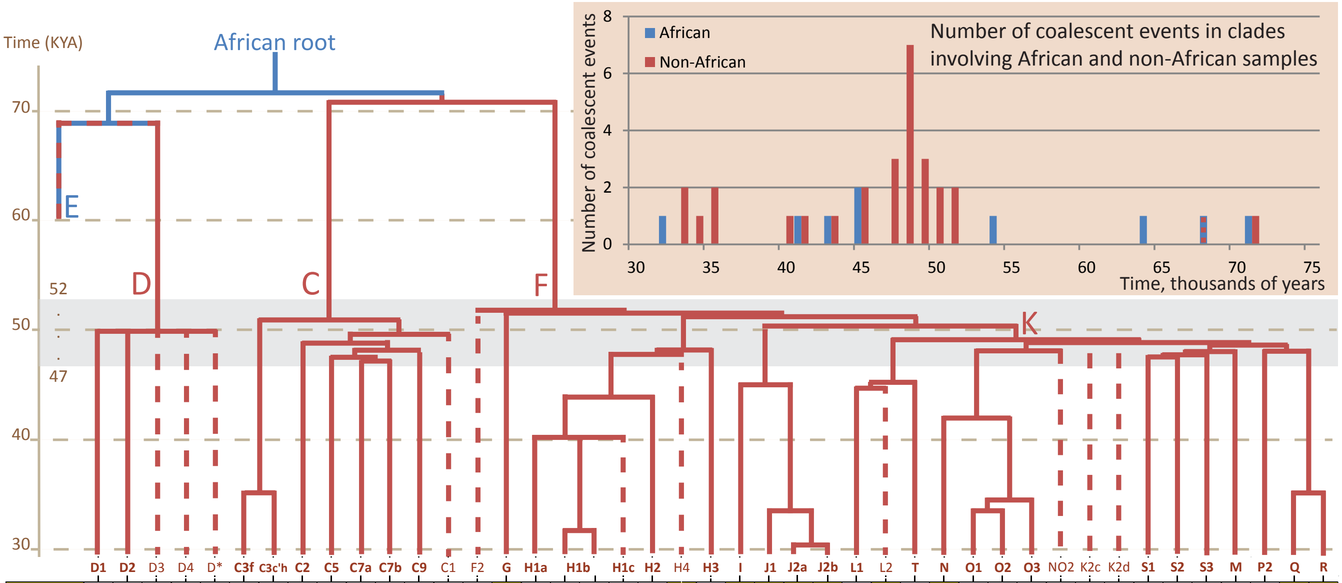


Figure S8. Comparison of STR and sequence data based coalescent time estimates.
STR-based estimates: 21 STR markers, "evolutionary" mutation rate 6.9x10^{-4} per 25 years (Zhivotovsky et al. 2004). Sequence based estimates: BEAST v.1.8.0, GTR substitution model, lognormal relaxed clock, mutation rate 7.4x10^{-10} mut/bp/year (SI3). The majority of the STR-based age estimates of "young" haplogroups deviate from the expected linear relationship (black line) with age estimates derived from sequence data. Adjusted STR mutation rates (green line, and the right y-axis) were calculated by 5000 year bins of haplogroup ages assuming linear relationship with sequence data based haplogroup age estimates. The "evolutionary" STR mutation rate that was used to calculate the raw STR-based age estimates is shown as a red line.



	D1	D2	D3	D4	D*	C3f	C3c'h	C2	C5	C7a	C7b	C9	C1	F2	G	H1a	H1b		H1c	H2	H4	H3	I	J1	J2a	J2b	L1	L2	T	N	O1	O2	O3	NO2	K2c	K2d	S1	S2	S3	M	P2	Q	R
Europe															■						■		■	■	■	■		■		■												■	■
West Asia Caucasus															■									■	■	■			■														■
South Asia					■				■								■	■	■			■				■	■					■											■
Northeast Asia	■	■	■			■	■						■																	■												■	
SE Asia Oceania				■				■		■	■	■		■						■											■	■	■	■	■	■	■	■	■	■	■	■	
Americas							■																																			■	

Figure S9. Coalescent times of the oldest Y chromosome haplogroups outside Africa. The tree was generated and coalescent times (presented in thousands of years, KY) calculated in BEAST. Shown are only those 32 branches that had coalescent date older than 30 KY. These branches are shown in solid lines. The plausible branching structure of 11 additional branches that have been reported in the literature is shown in dotted lines. The distribution of relevant clades by geographic regions is presented on the basis of our data and inferences made from literature ignoring cases of the presence of the haplogroup in a region that can be explained by historic gene flow. The distribution of coalescent events in time for African and non-African populations was obtained using a sliding window of 2000 and step size 1000 years. The red and blue line connecting D and E reflects the proposed model of an early back to Africa event involving putative DE* lineages (Hammer et al. 1998).

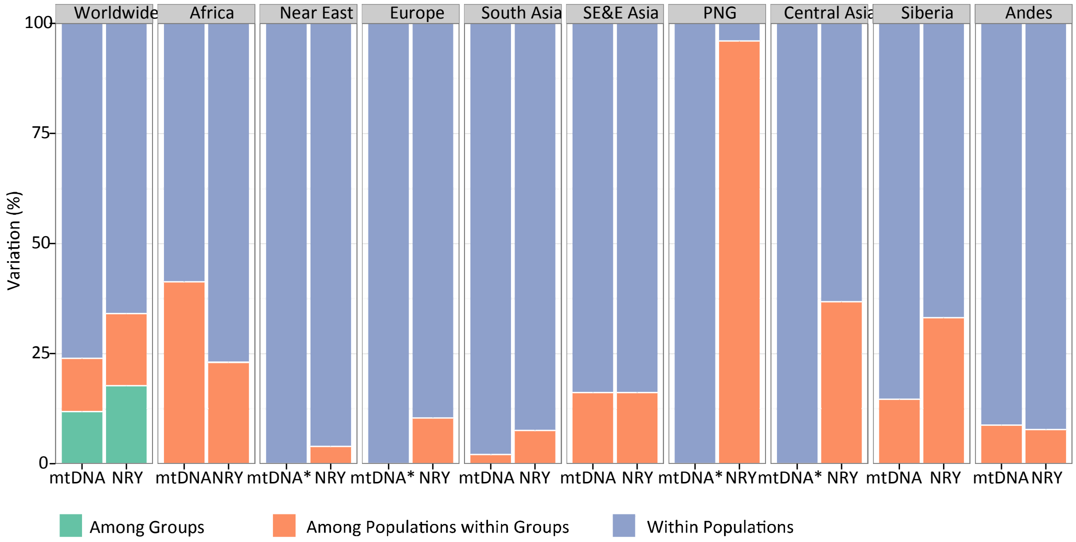


Figure S10. Distribution Y chromosome and mitochondrial diversity within and among populations.

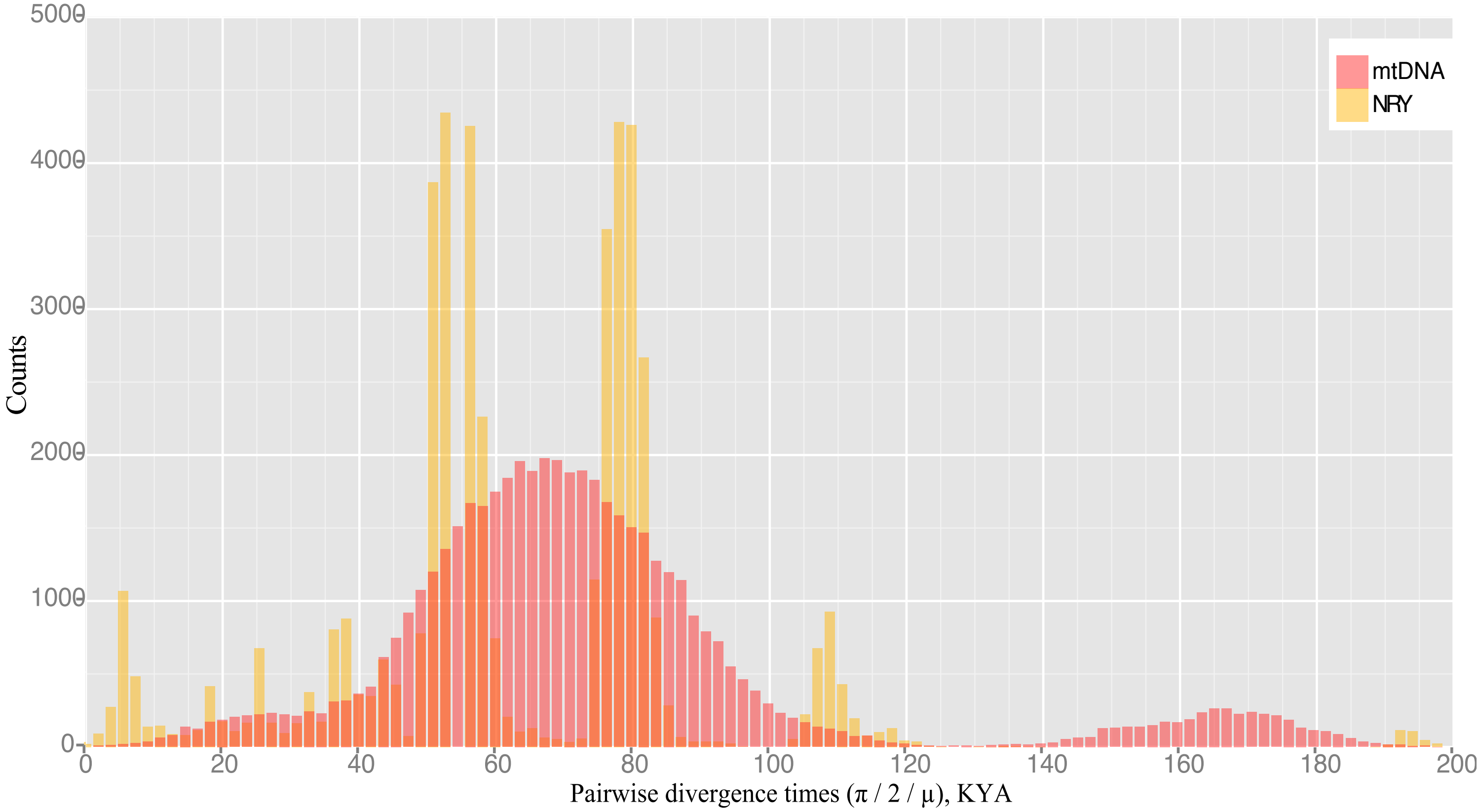


Figure S11. Distribution Y chromosome and mitochondrial DNA pairwise divergence times. The plot is based on the analysis of 320 mtDNA and Y chromosome sequences including African and non-African populations. The two highest peaks in the Y chromosome plot correspond to the two clusters of coalescent nodes in FigureS9, one at 43-48KYA involving non-African coalescent events within haplogroups C, D, and F, and the 62-64KYA cluster to the coalescent events among the non-African founders and between African E lineages and non-Africans.

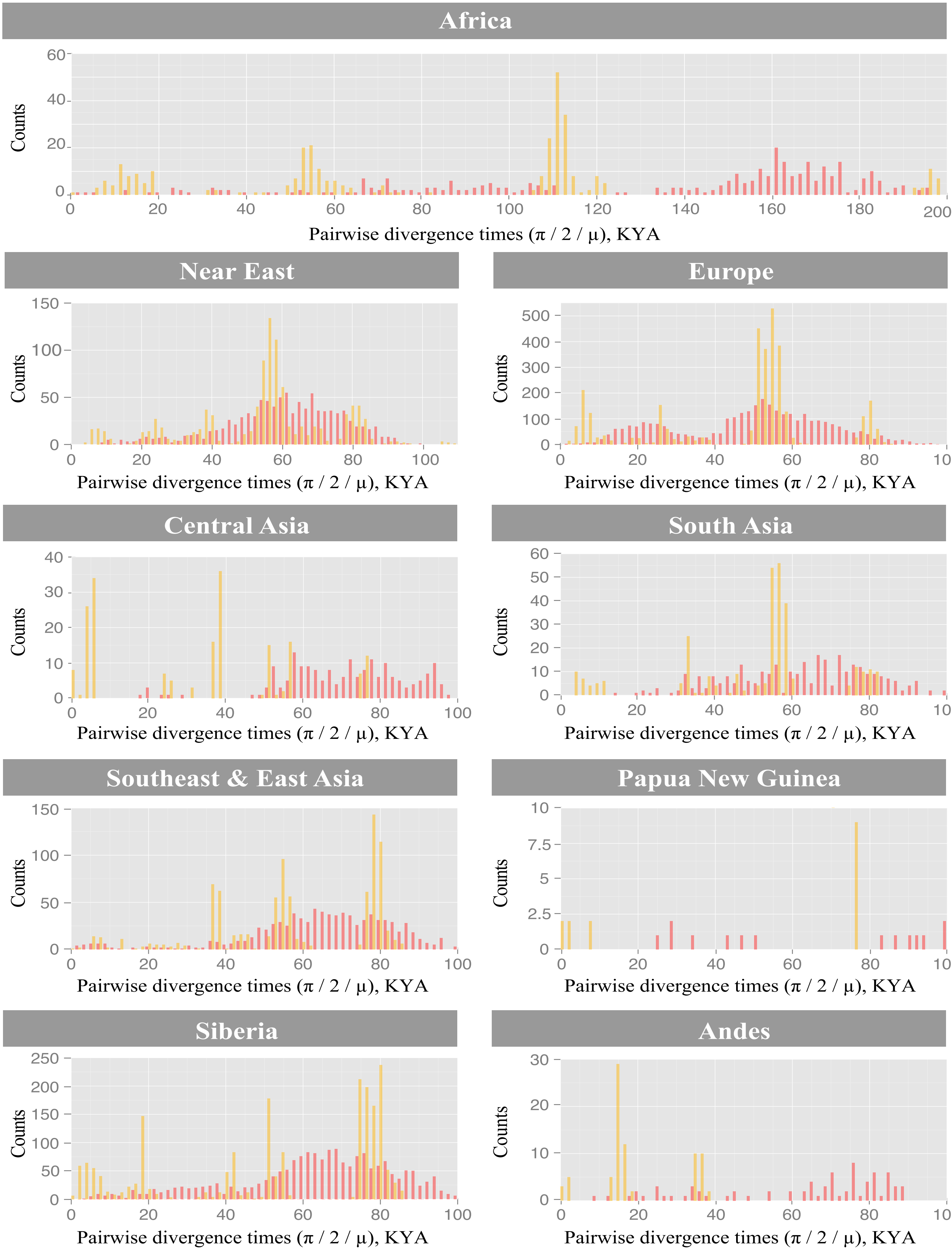


Figure S12. Regional plots of Y chromosome and mitochondrial DNA pairwise divergence times. The plot is based on the regional analysis of the 320 mtDNA and Y chromosome sequences as shown combined in Figure S11.

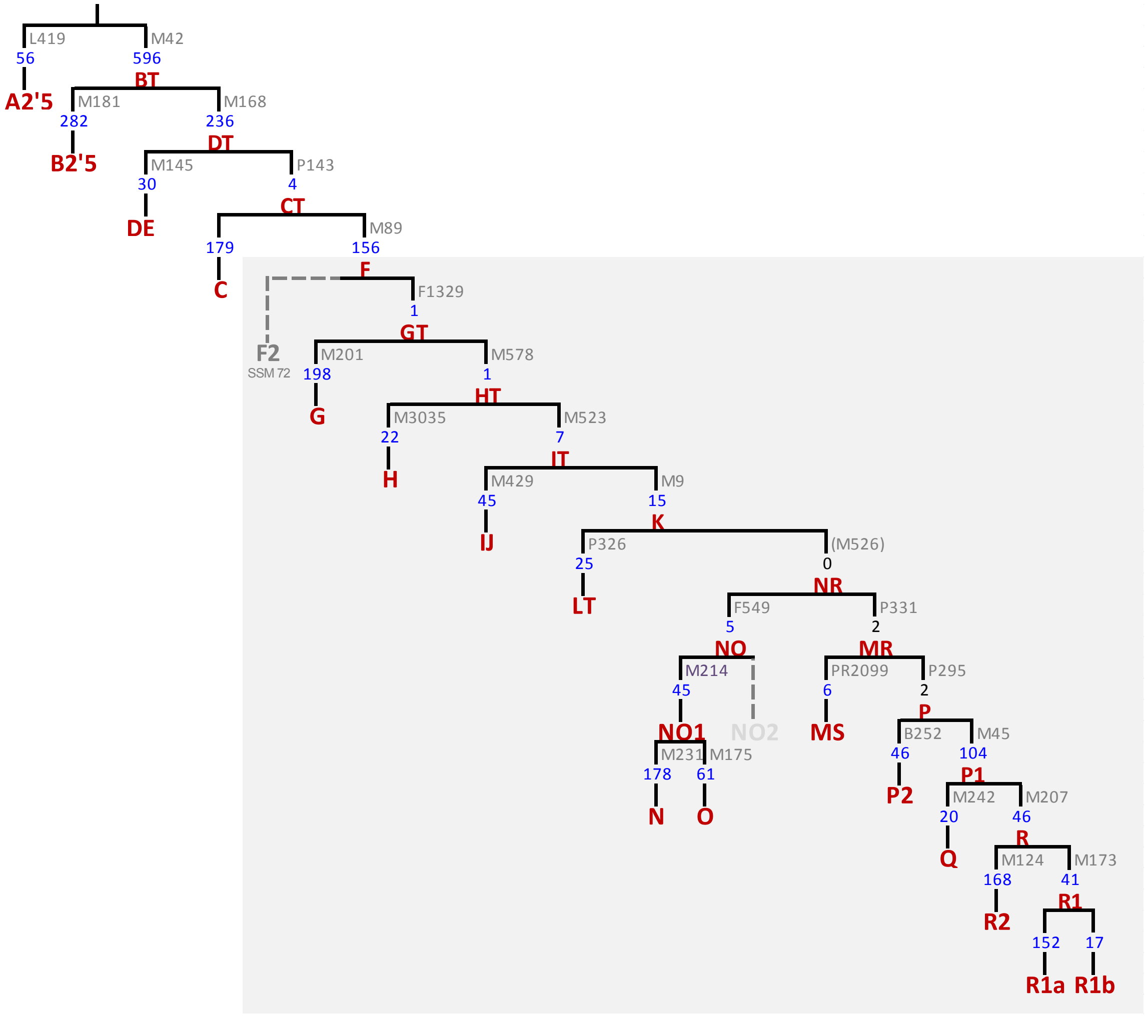
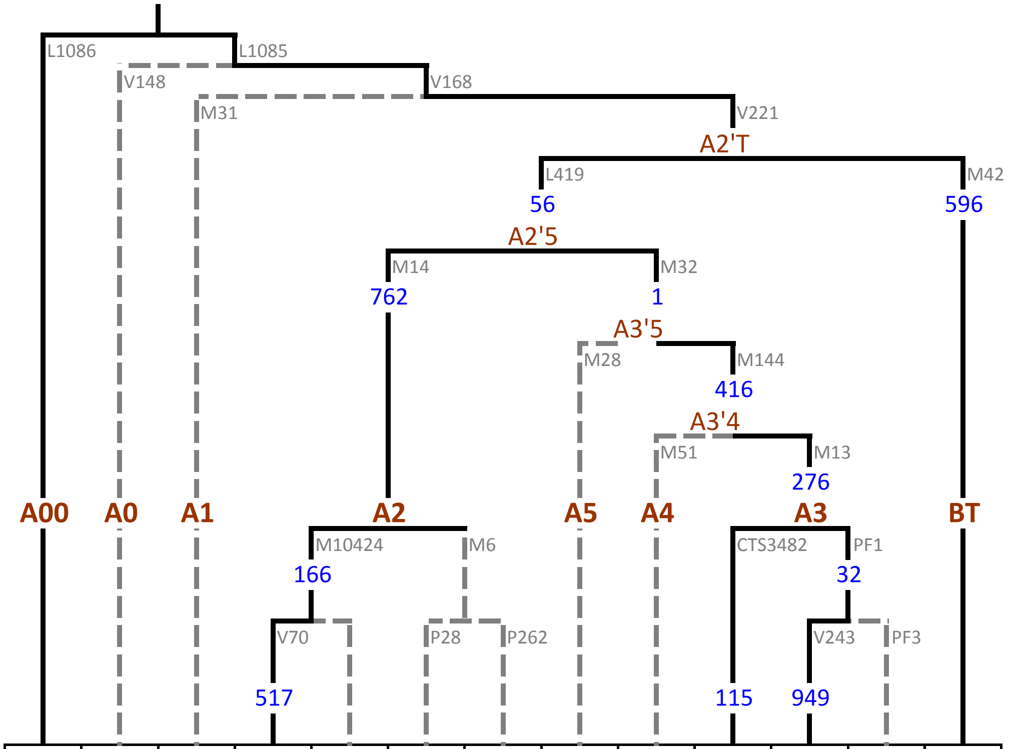
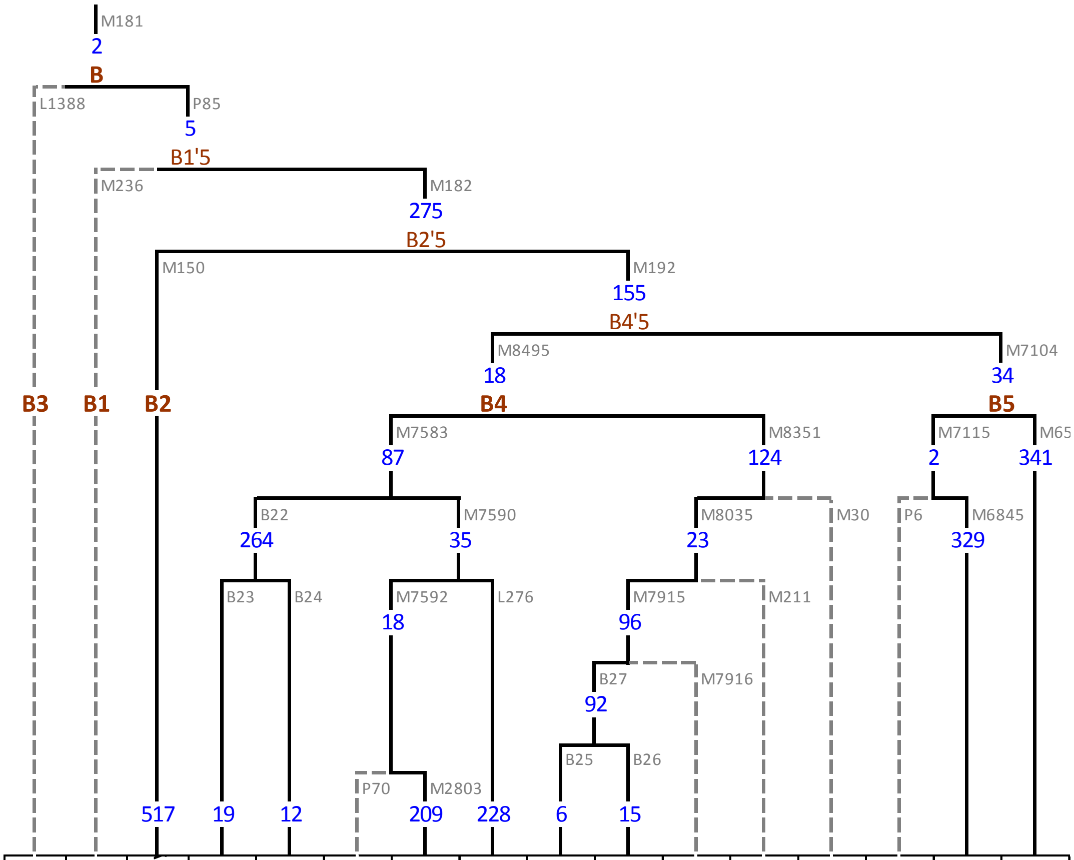


Figure S13. Refined topology of the basic branches of the Y-chromosome tree with an emphasis on the short branch lengths defining haplogroup F sub-clades. Mutation counts and branch defining markers (included in counts) are shown on the branches. If no previously named mutation existed on a given branch, a name from the B-series was given to one of the mutations discovered, and the rest were denoted with _eq in the annotations table. Branch lengths include recurrent mutations if such are present in no more than two (exceptionally three) locations in the Y-chromosome phylogeny. Given comparative data from Malaysian high coverage sequences (Wong et al. 2013), we map the position of the F2 clade represented by Malay sample SSM072 and redefine haplogroup NO by F549 and 4 equivalent mutations. NO splits further into NO1-M214 and the novel NO2-F2755, as defined by the SSM016 sequence.



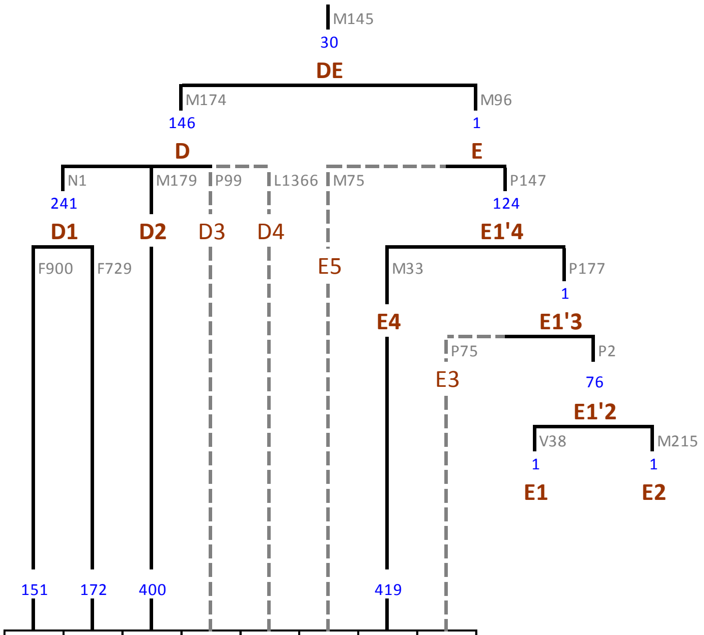
Sample identifier	2 Mbos			Congo Pygmy	Baka	San	San		San	MKK	Ethiopian Jew ISR07	7 Sardinians	
	GRC13292545 GRC13292546			35247	919	HGDP01029	HGDP00987		HGDP01036	NA21313	16204		
Karafet 2008		A1b	A1a	A2*		A2b	A2c	A3a	A3b1	A3b2			BT
Van Oven 2013	A00	A0	A1	A2		A2-P28	A2-P262	A3-M28	A3-M51	A3-M13			BT
ISOGG	A00	A0	A1a	A1b1a1		A1b1a1a1	A1b1a1a2b	A1b1b1	A1b1b2a	A1b1b2b			BT

Figure S14. Refined topology of the Y-chromosome haplogroup A. Grey dashed lines denote branches defined by published sequences generated with Illumina technology.



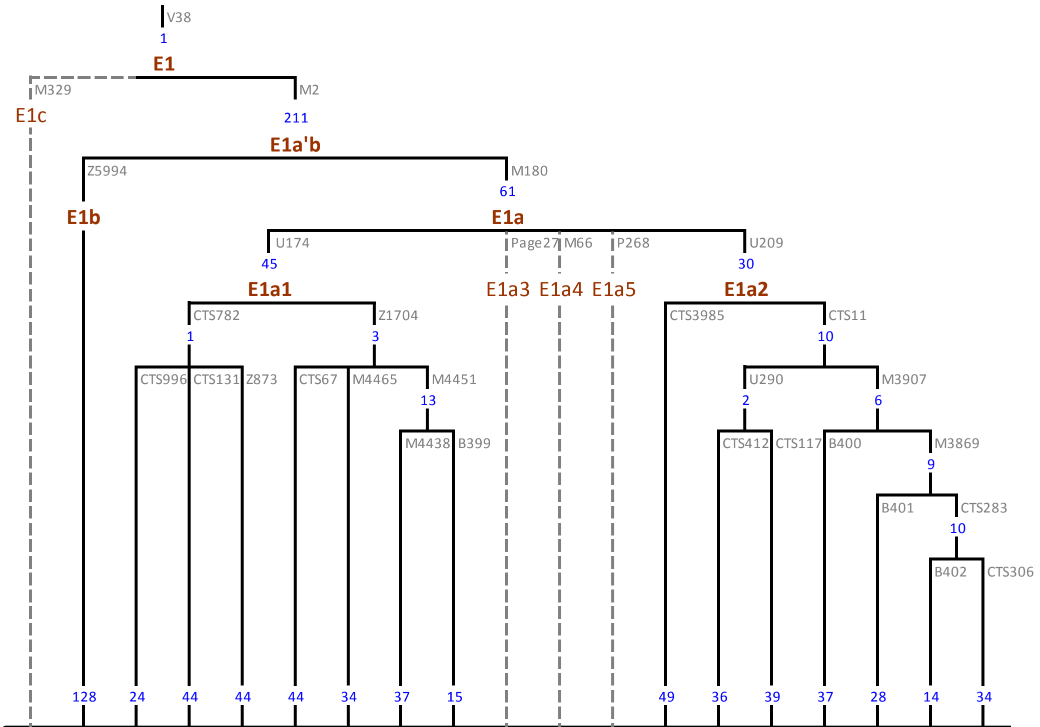
Sample identifier			**CongoPygm**	**Hadza2**	**Hadza5**		**Hadza3**	**Sandawe5**	**PygmyBaka**	**PygmyBaka**					**Hadza4**	**Sandawe4**
			35246	3124	1807	HGDP0099	319- DNA_G01 1100_37	319- DNA_G02 1100_37	319- DNA_A01 1100_37	319- DNA_C01 1100_37	Baka927	Baka908, 937	Baka 918, 920, 909	HGDP009 91 HGDP010	3125	1577
Karafet 2008		B1	B2a	B2b		B2b4a	B2b					B2b4b	B2b3	B2b1	B2b	
Van Oven 2013	B3	B1	B2a	B-M7583					B-M8035				B-M8351	B-P6	B-M7104	
ISOGG		B1	B2a	B2b1		B2b1a2	B2b1					B2b1a1c	B2b1a1b	B2b1b	B2b1	

Figure S15. Refined topology of the Y-chromosome haplogroup B. Grey dashed lines denote branches defined by published sequences generated with Illumina technology.



	Nepalese Tamang	Chinese	Japanese				Yoruba	
Sample identifier	GS000035037	GS000010424	NA18940-200-37				NA19239-L2-200-37	
Karafet 2008	D1a*		D2	D3	D*	E2	E1a	E1b2
Van Oven 2013	D-N1		D2	D3	D*	E2	E-M33*	E-P177*
ISOGG	D1a		D2	D3	D4	E2	E1a	E1b2

Figure S16. Refined topology of the Y-chromosome haplogroup DE. Grey dashed lines denote branches deriving from the respective nodes in previous phylogenetic reports but not found in our sample (Karafet et al 2008; van Oven et al 2013). Branches E1, E2, E1'3 are shown with length 1 because only one of the sub-branches was present in our sample and the rest of the identified mutations were counted to be equal with the sub-branch definition although some of them are actually defining the parent-clade.



Sample identifier		NA	Congo-pygmies	Yoruba	African-Americans	Luhya	Baka-Pygmies	Bakola-Pygmies	Hadza				Yoruba	African-Americans	African-Americans	Bedzan-Pygmies	Luhya	Sandawe	Luhya
		GS00253-DNA_B02_200_37	GS000035245	NA18504-200-37	NA19834-200-37	NA19026-200-37	GS00319-DNA_B01_1100_37	GS00319-DNA_D01_1100_37	GS00319-DNA_F01_1100_37				NA18501-200-37	NA19700-200-37	NA19703-200-37	GS00319-DNA_E01_1100_37	NA19025-200-37	GS00319-DNA_C02_1100_37	NA19020-200-37
Karafet 2008	E1b1c	E1b1a*	E1b1a7a*					E1b1a7a3		E1b1a1	E1b1a6	E1b1a9	E1b1a8a*	E1b1a8a1*	E1b1a8a1a	E1b1a8a*			
Van Oven 2013	E-V38*	E-Z5994	E-U174							E-M180*			E-U209*	E-U290*	E-U181	E-U209*			
ISOGG	E1b1a2	E1b1a1a2	E1b1a1a1f1a1*			E1b1a1a1f1a1d		E1b1a1a1f1a1c		E1b1a1a1a	E1b1a1a1e	E1b1a1a1h	E1b1a1a1g1*	E1b1a1a1g1a*	E1b1a1a1g1a1	E1b1a1a1g1*			

Figure S17. Refined topology of the Y-chromosome haplogroup E1. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

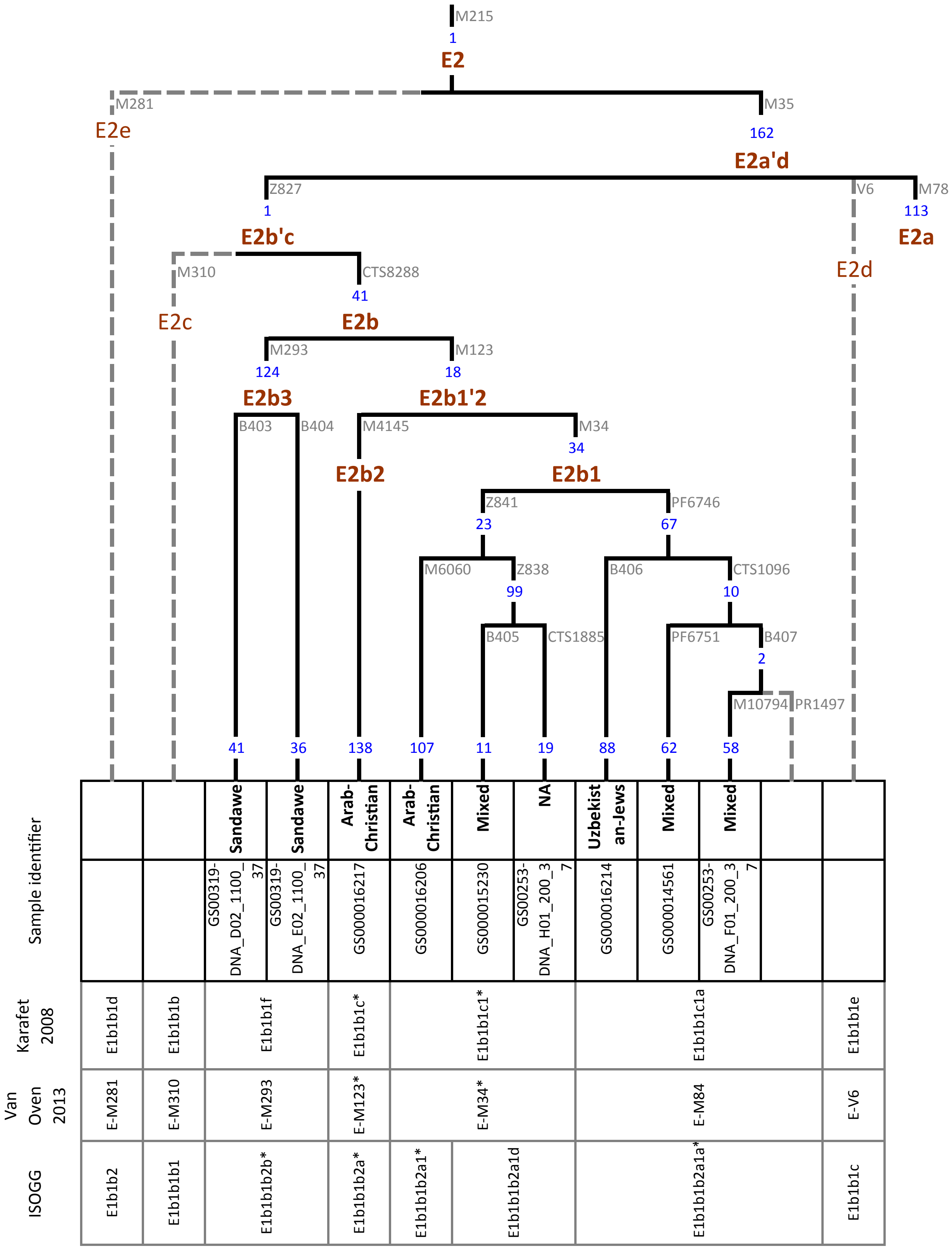
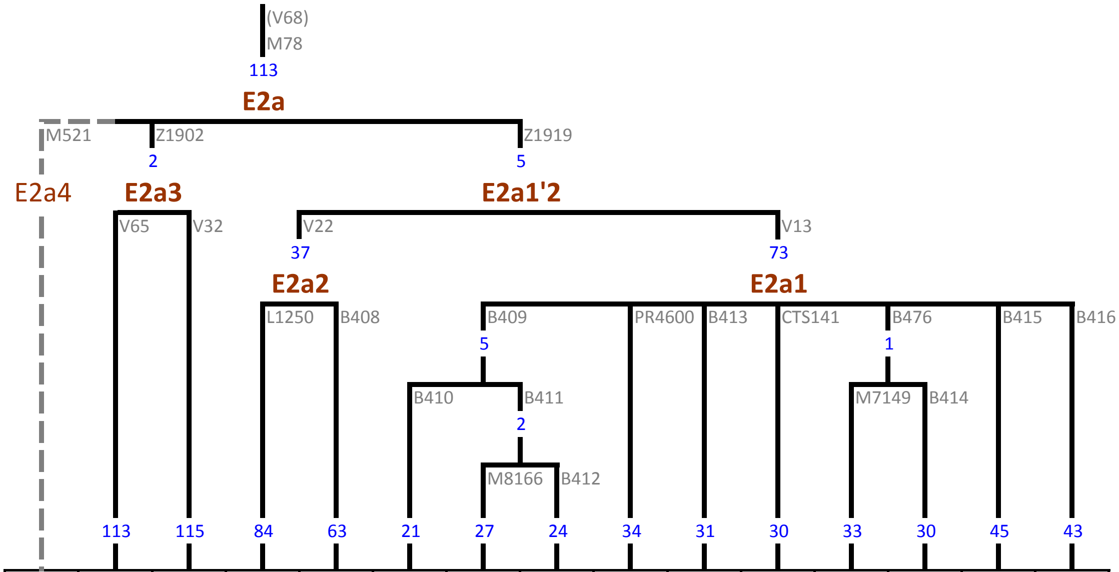
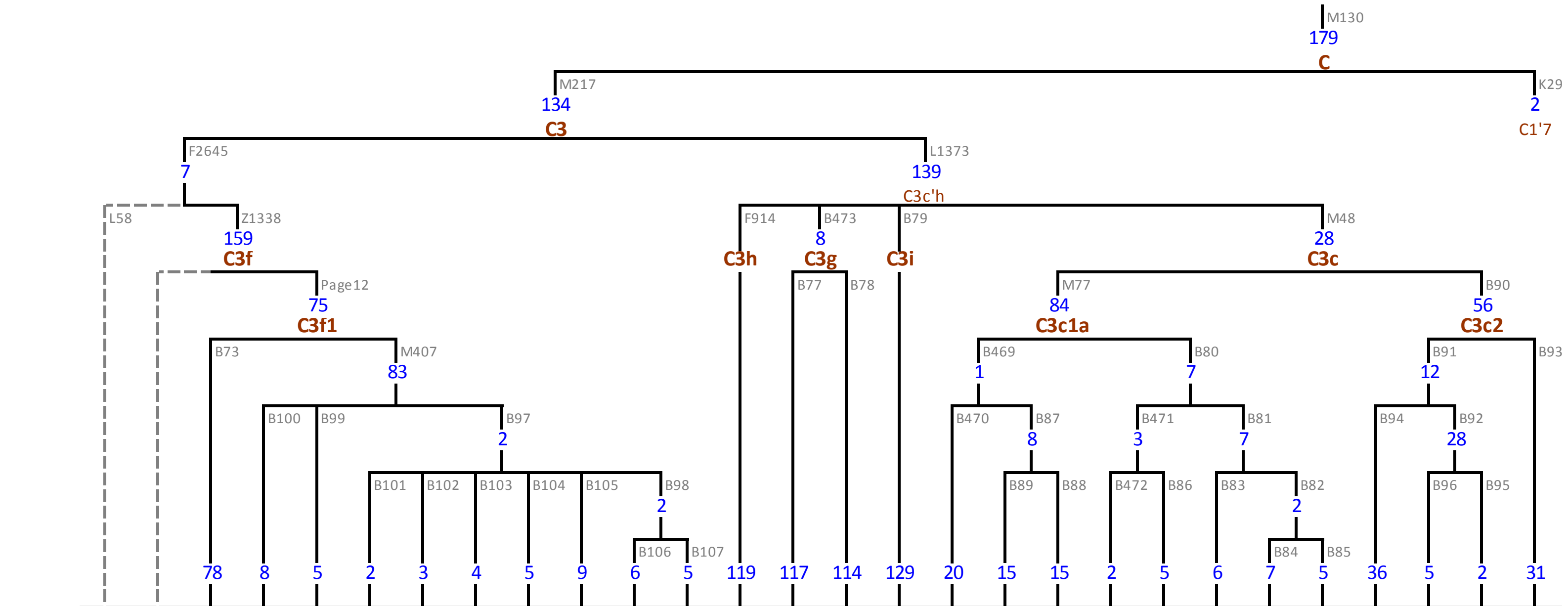


Figure S18. Refined topology of the Y-chromosome haplogroup E2. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



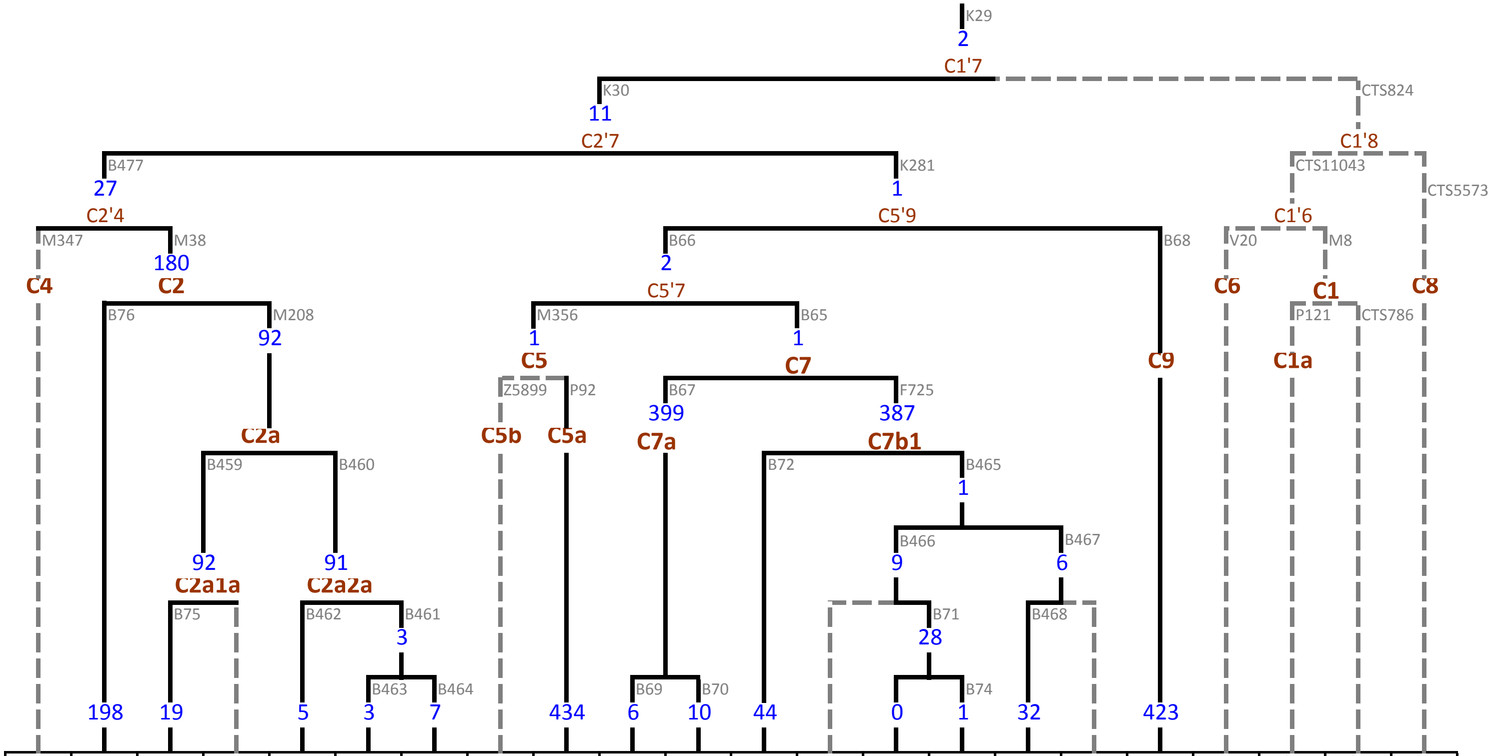
Sample identifier		US	Iranians	Mixed	Maasai family(2)	Estonians	US	Albanians	British	Mixed	Italians	US	Australian	Mishar-Tatars	Mixed
		GS000011811	GS000016179	GS000015708	NA21732-200-37; NA21737-200-37	GS000017206	GS000015892	GS000035021	GS000016441	GS000014562	NA20510-200-37	GS000015153	GS000035227	GS000013763	GS000015172
Karafet 2008	E1b1b1a*	E1b1b1a4	E1b1b1a1b	E1b1b1a3*		E1b1b1a2*									
Van Oven 2013	E-M521	E-V65	E-V32	E-V22		E-V13									
ISOGG	E1b1b1a1c	E1b1b1a1a2	E1b1b1a1a1b	E1b1b1a1b2*		E1b1b1a1b1a*									

Figure S19. Refined topology of the Y-chromosome haplogroup E2a. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



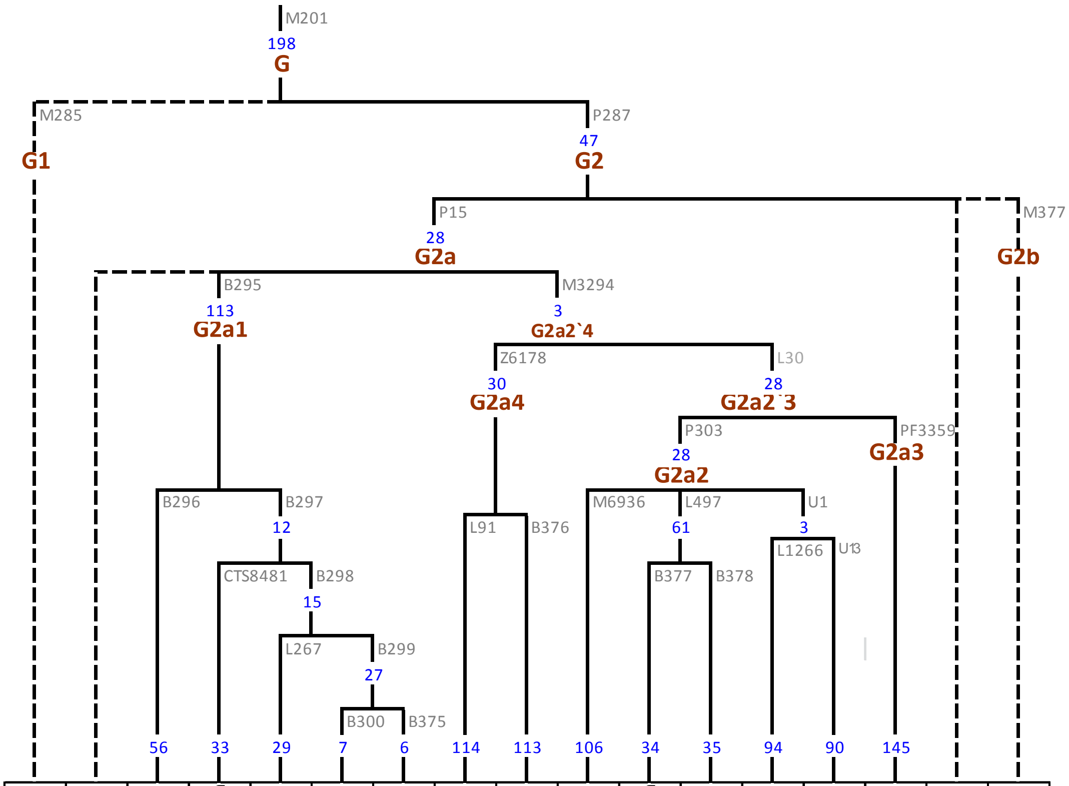
Sample identifier	CHS	CHB/JPT	Chinese	Buryat	Buryat	Buryat	Buryat	Buryat	Buryat	Buryat	Buryat	Buryat	Kazakh	Koryak	Kizhi	Koryak	Mongol-Zakhchin	Evenki	Buryat	Even	Even	Even	Even	Even	Koryak	Koryak	Koryak	Evenk
	HG00628	NA18749 NA18612 NA18620 NA19079	15179	22091	14418	22094	22090	13761	20113	22092	22093	22066	13752	22071	22440	19992	35117	20004	20126	34297	22293	16165	22292	16129	19993	16168	16169	14335
Karafet 2008	C*												C3*				C3c											
Van Oven 2013	C-M217*	C-Z1338											C-L1373*				C-M48											
ISOGG	C3*	C3e	C3e1a										C3*				C3b2a								C3b2			

Figure S20. Refined topology of the Y-chromosome haplogroup C3. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



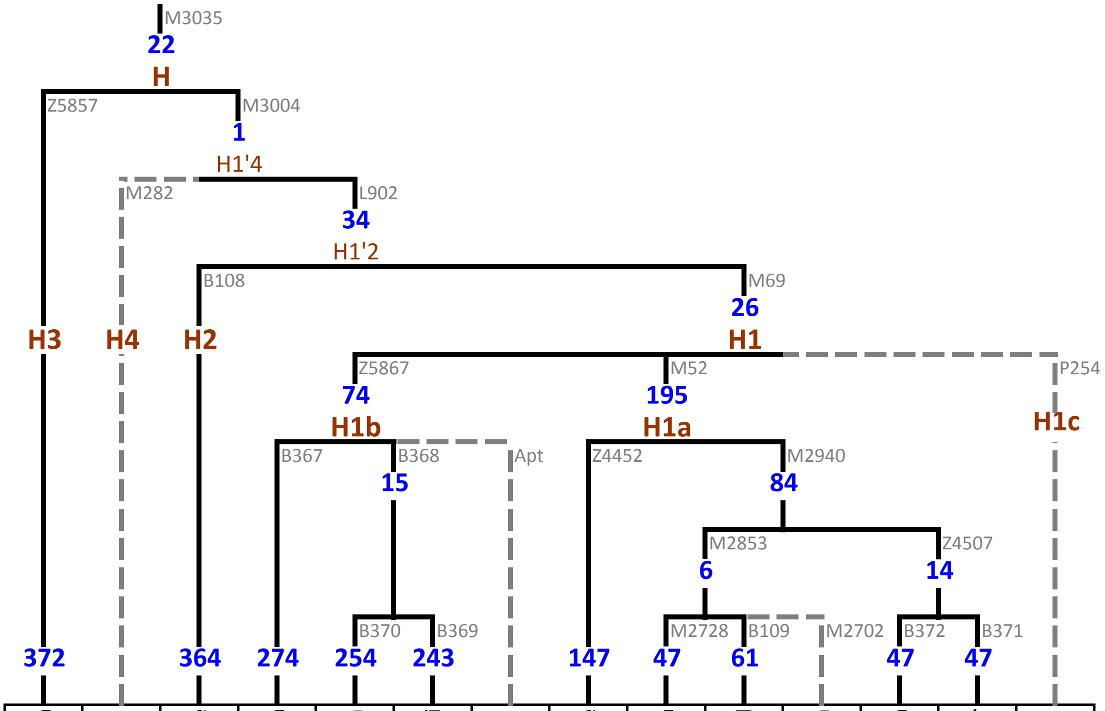
Sample identifier	Australian	Bajo	Bajo	Malay	Koinambe	Koinambe	Koinambe		Indian	Lebbo	Lebbo	Murut	Malay	Murut	Murut	Aeta	Malay	Dusun		JPT	JPT	JPT
	WON,M	17005	17004	ML99	35255	35258	35686		17146	16936	16937	19981	ML84	22099	20272	35257	ML18	20271		NA189	NA189	NA189
Karafet 2008	C4	C2*	C2a					C5*	C5a	C*										C1a	C1*	C*
Van Oven 2013	C4	C-M38*	C2a					C5b	C5	C-K29*									C6	C1		C-K29*
ISOGG	C4	C2*	C2a					C5*	C5a	C*										C1a	C1*	C*

Figure S21. Refined topology of the Y-chromosome haplogroup C1'7. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



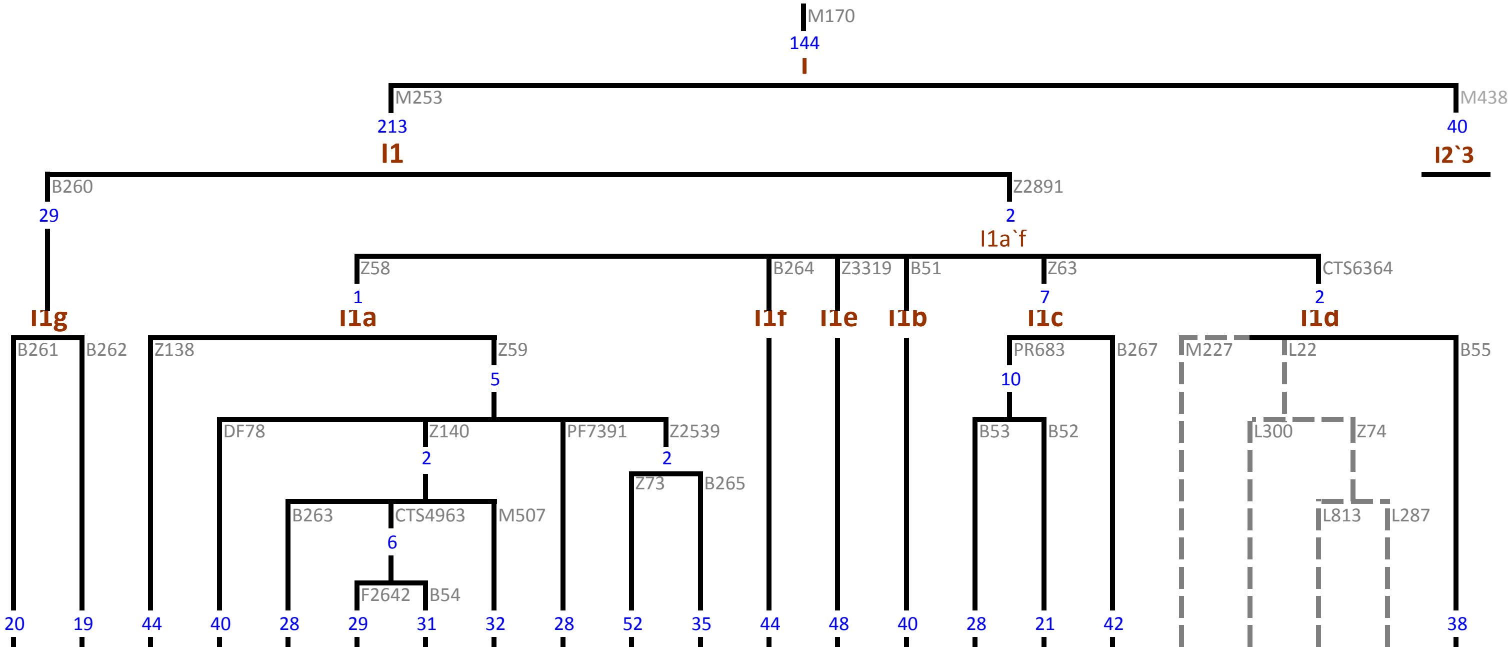
			Lezgin	Moldavian	Abkhazian	North Ossetian	North Ossetian	Iranian	Uzbekistan-Jew	Iraqi-Jew	Mexican	US	Kabardin	Pole	Dutch		
Sample ID	unsampled	unsampled	14366	16899	14409	13739	14414	14475	16215	16209	NA19670-20(	11806	14413	16372	F18	unsampled	unsampled
Karafet 2008	G1*	G2*	G2a1					G2a*									G2c
Van Oven 2013	G-M285	G-P287*	G-P16?					G-L91	G-PF3146*	G-P303*	G-L497*		G-U1*	G-U13	G-PF3359	G-P15*	G-M377
ISOGG	G1*	G2a						G2a2a1b	G2a2	G2a2b2a	G2a2b2a1b		G2a2b2a1a	G2a2b2a1a1	G2a2b2b	G2	G2b

Figure S22. Refined topology of the Y-chromosome haplogroup G. Grey dashed lines denote known branches with no sequence data in our dataset. In the case of haplogroup G all samples generated for this project belonged to clade G2a. Within this clade we distinguish four sub-clades, including G2a1-B295, which was earlier defined by a recurrent marker P16.



Sample identifier	Malayan	French	Burmese	Balija	Dhaka	Punjabi		Burmese	Dhaka	Kol	Cambodian	Kshatriya	Asur	
	16806	HGDP0 0528	19903	35123	35035	16965		20270	35034	16821	HGDP 00720	35039	16805	
Karafet 2008	F*			H*			H2	H1*	H1a					H*
Van Oven 2013	H3	H2	H-M3035*	H-Z5867			H-Apt	H-Z4452	H-M2853			H-Z4507		H-M69*
ISOGG	H3	H2	H*	H1b			H1b1	H1*	H1a					H1c

Figure S23. Refined topology of the Y-chromosome haplogroup H. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



Sample identifier	US	US	mixed	mixed	Australian	Dutch	Toscan	Hungarian	mixed	Utah,CEU	mixed	mixed	mixed	Belarus	Mordvin	Komi	NW European		Finnish		Finnish	Dutch
	15176	16442	11808	10320	35226	F19	NA20511-200-37	16901	15272	NA06994-200-37	10435	15888	GS00253-DNA_E01_200_37	16104	13698	14328	NA12891-200-37		HG00189		HG00325 HG00372 HG00186	F11
Karafet 2008	I1*							I1d	I1*									I1b	I1*			
Van Oven 2013	I-M253*		I-Z138	I-Z59		I-Z140				I-Z2539		I-M253*			I-Z63			I-M227	I-L22	I-L813	I-Z74*	I-CTS636
ISOGG	I1		I1a2b	I1a2a		I1a2a1a2	I1a2a1a	I1a2a		I1a2a1b		I1			I1a3			I1a1a	I1a1b4	I1a1b3b	I1a1b3a	I1a1*

Figure S24. Refined topology of the Y-chromosome haplogroup I1. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

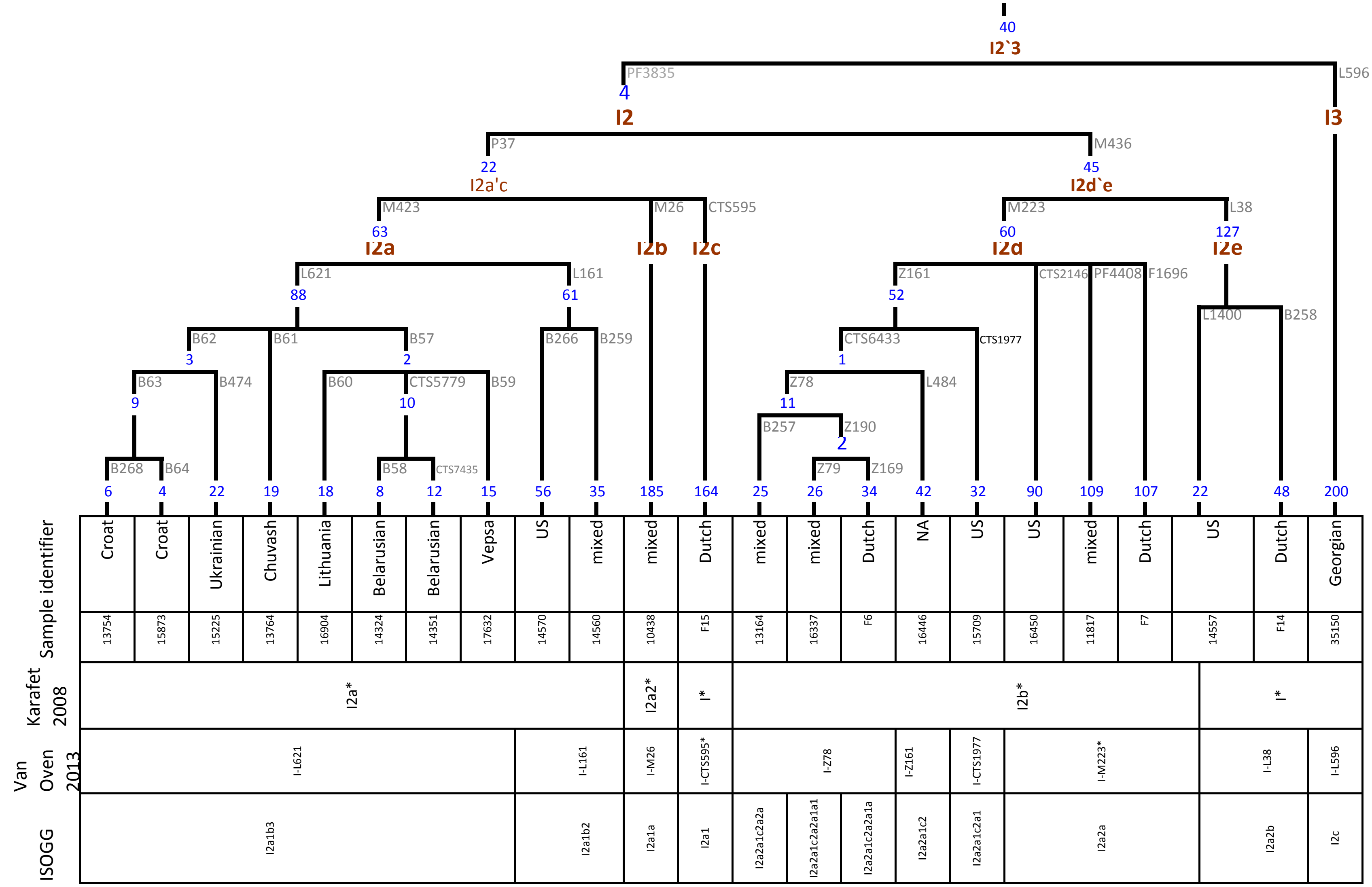
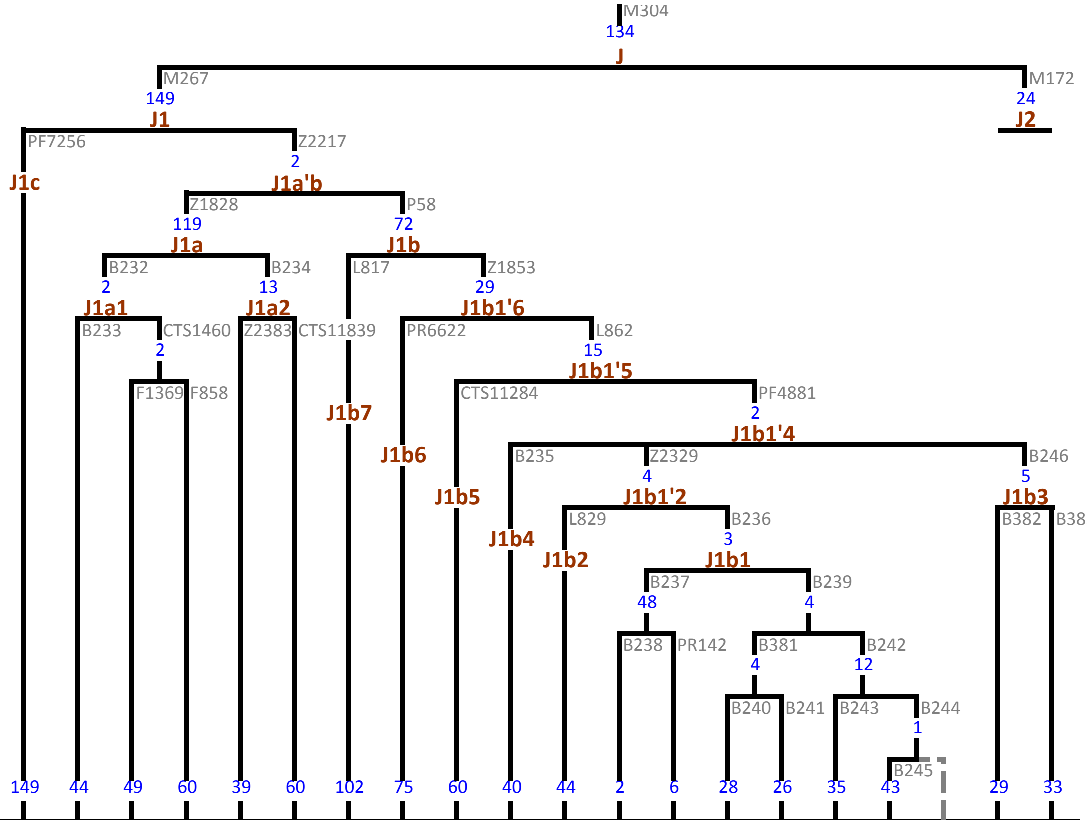
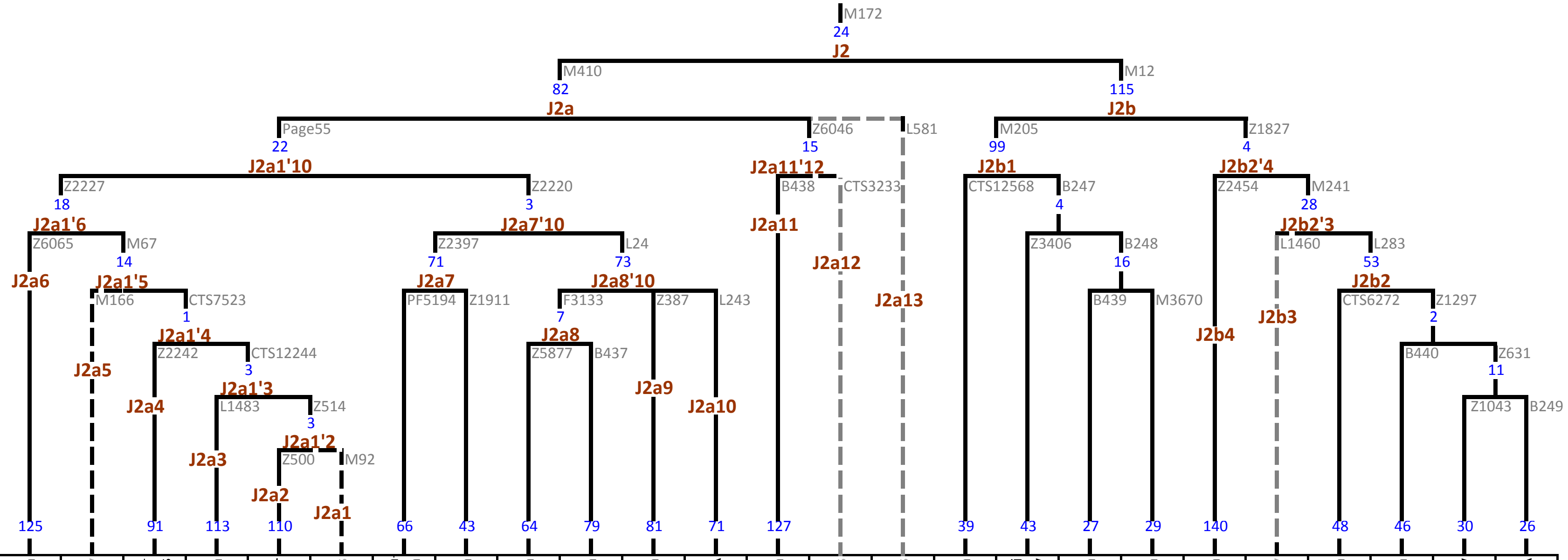


Figure S25. Refined topology of the Y-chromosome haplogroup I2'3.



Sample identifier	Azeri	Tabas-saran	Kumyk	Avar	Lezgin	USA	Belarussian	Armenian	Armenian	Jordanian	Druze	Yemeni Jew	Yemen	Armenian	Arab Muslim	Arabian	Arabian	Malay	Armenian	Albanian
	13741	14421	14368	14405	13724	9929	11735	35126	35125	14474	16136	16887	16973	13745	16213	16180	16277	ML21	35124	35434
Karafet 2008	J1*						J1e													
van Oven 2013	J1*	J1-Z1828					J1-P58*	J1-Z1865*			J1-L858*								J1-Z1865*	
ISOGG	J1a*	J1a3a					J1a2b*		J1a2b2*											

Figure S26. Refined topology of the Y-chromosome haplogroup J1.



Sample identifier	Assyrian	Malay	Georgian-Imeretins	Italian	Tatar	Sardinians	Kapu, Andhra Pradesh	unknown	Iranian	Kirgizian	Kabardinian	USA	Abkhazian	Sardinians	Sardinians	Jordanian	Bene Israeli Jew	Armenian	Dutch	Russian	Malay	German	Albanian	Cochin Jew	USA
	13751	ML79	35685	10431	15875	860-867	16802	16445	13747	16177	35115	15895	14410	841-847	831-840	16181	16882	13746	F16-S16	10434	ML82	16373	35147	16138	15174
Karafet 2008	J2a*	J2a2b	J2a2*			J2a2a	J2a*									J2b1				J2b*	J2b2*				
van Oven 2013	J2a-Z6065	J2a-M67*	J2a-M67*		J2a-PF7394	J2a-M92	J2a-PF5172		J2a-F3133		J2a-L25*		J2a-M410*		J2a-L581	J2b-M205				J2b*-M12	J2b-M241*	J2b-L283			
ISOGG	J2a1*	J2a1b2	J2a1b*			J2a1b1	J2a1*		J2a1h2*				J2a*		J2a2	J2b1				J2b*	J2b2*	J2b2a1*	J2b2a1a*	J2b2a1a1	

Figure S27. Refined topology of the Y-chromosome haplogroup J2. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

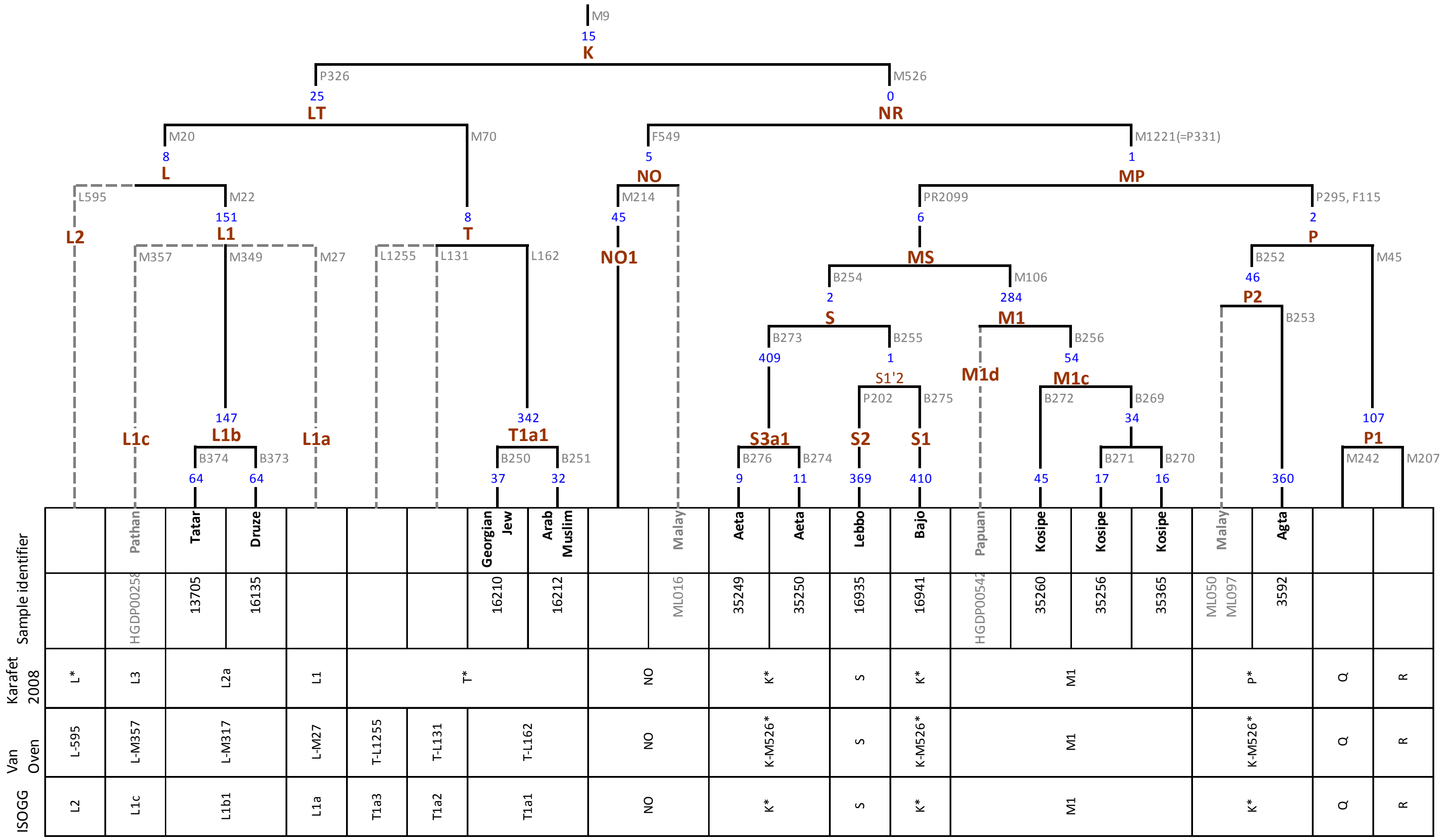
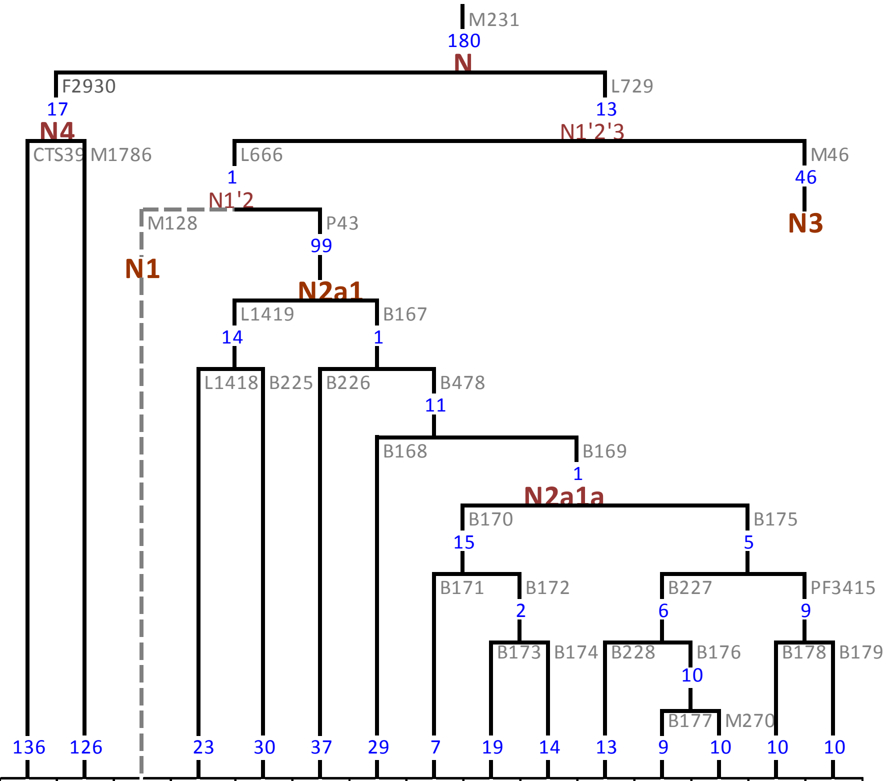
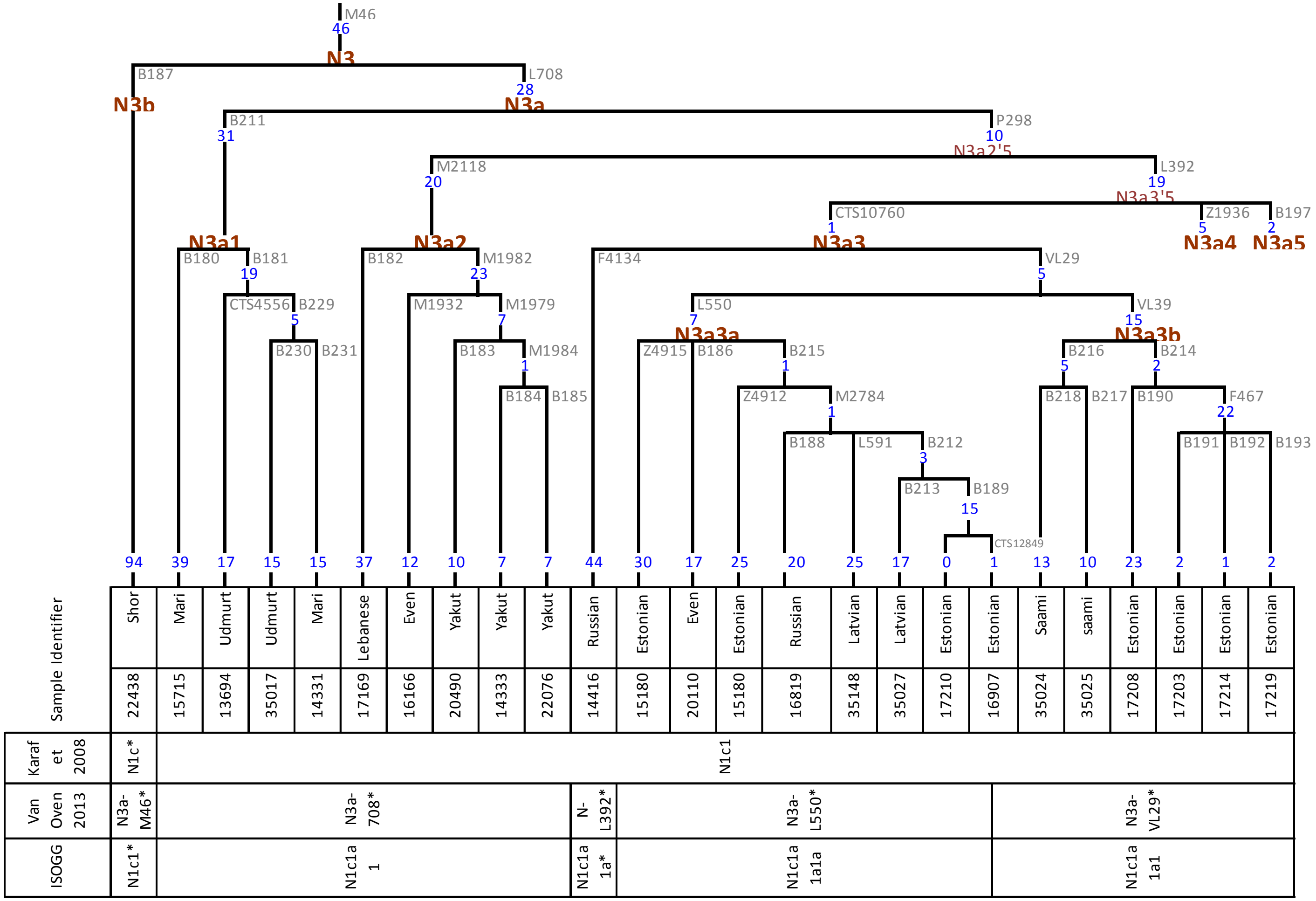


Figure S28. Refined topology of the Y-chromosome haplogroups L, T, S and M. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



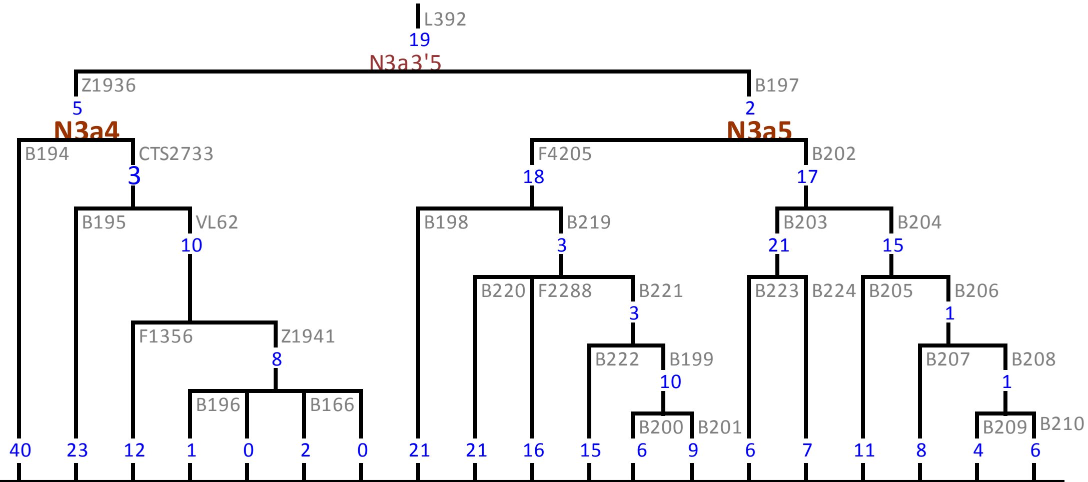
Sample Identifier	CHB	Cambodian	Unsampled	Vepsa	Mari	Udmurd	Even	Tundra Nenets	Tundra Nenets	Forest Nenets	Mongol	Evenki	Yakut	Tuvinian	Tuvinian
	20037	Poznik 2013		35244	14332	35432	16167	16163	16162	15879	35235	20108	22441	17247	16968
Karafet 2008	N*		N1a	N1b			N1b1								
Van Oven 2013	N4		N3b-M128	N3b-P43*			N3b-P63								
ISOGG	N2		N1c2a	N1c2b			N1c2b1								

Figure S29. Refined topology of the Y-chromosome haplogroup N. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing. Haplogroup N 1’3 is further split into N 1’2-L166 (former N3b, PhyloTree) and N3-M46 (former N3a, PhyloTree).



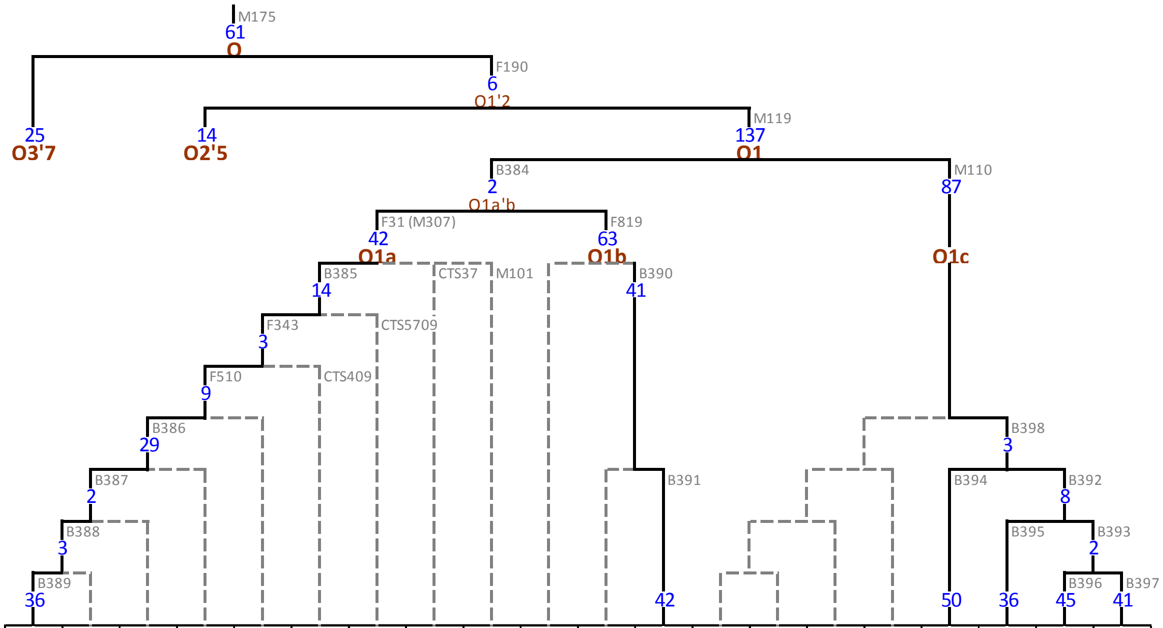
Sample Identifier	Shor	Mari	Udmurt	Udmurt	Mari	Lebanese	Even	Yakut	Yakut	Yakut	Russian	Estonian	Even	Estonian	Russian	Latvian	Latvian	Estonian	Estonian	Saami	saami	Estonian	Estonian	Estonian	Estonian
	22438	15715	13694	35017	14331	17169	16166	20490	14333	22076	14416	15180	20110	15180	16819	35148	35027	17210	16907	35024	35025	17208	17203	17214	17219
Karaf et 2008	N1c*	N1c1																							
Van Oven 2013	N3a-M46*	N3a-708*									N-L392*	N3a-L550*								N3a-VL29*					
ISOGG	N1c1*	N1c1a 1									N1c1a 1a*	N1c1a 1a1a								N1c1a 1a1					

Figure S30. Refined topology of the Y-chromosome haplogroup N. We re-define the internal structure of haplogroup N3a-L708 and report several novel subclades. Haplogroup N3a2'5 now has a novel geographically restricted sister clade N3a1, defined by the marker B211, and is further split into previously unreported subclade N3a2-M2118 and subclade N3a3'5 (former N3a-L392).



Sample Identifier	Cossac	Vepsa	Karelian	Estonian	Estonian	Estonian	Estonian	Mongol	Mongol	Mongol	Kazakh	Buryat	Buryat	Chukchi	Koryak	Koryak	Eskimo	Koryak	Chukchi
	16186	16971	35149	18757	17207	17380	17213	35116	35236	16190	35127	20491	22077	20002	33948	22068	19998	22069	20001
Karafet 2008																			
Van Oven 2013	N3a-Z1936*			N3a-Z1941				N-L392*											
ISOGG	N1c1a1a2			N1c1a1a2a1				N1c1a1a											

Figure S31. Refined topology of the Y-chromosome haplogroups N3a4 and N3a5.



	Agta	Malay	Malay	Malay	Malay	CHS	CHB	CHB		Malay	Malay	Agta	Malay	Malay	Malay	Malay	Batak	Murut	Murut	Batak
Sample identifier	35248	ML083	ML055	ML090	ML017	HG00595 HG00577	NA18536 NA18606	NA18632 NA18740		ML092	ML091	35435	ML074	ML100	ML081	ML031	35252	19983	19985	35251
Karafet 2008	O1a1								O1a1a	O1a1			O1a2							
Van Oven 2013	O-P203												O-M110							
ISOGG	O1a1								O1a1a	O1a1			O1a2							

Figure S32. Refined topology of the Y-chromosome haplogroup O1. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

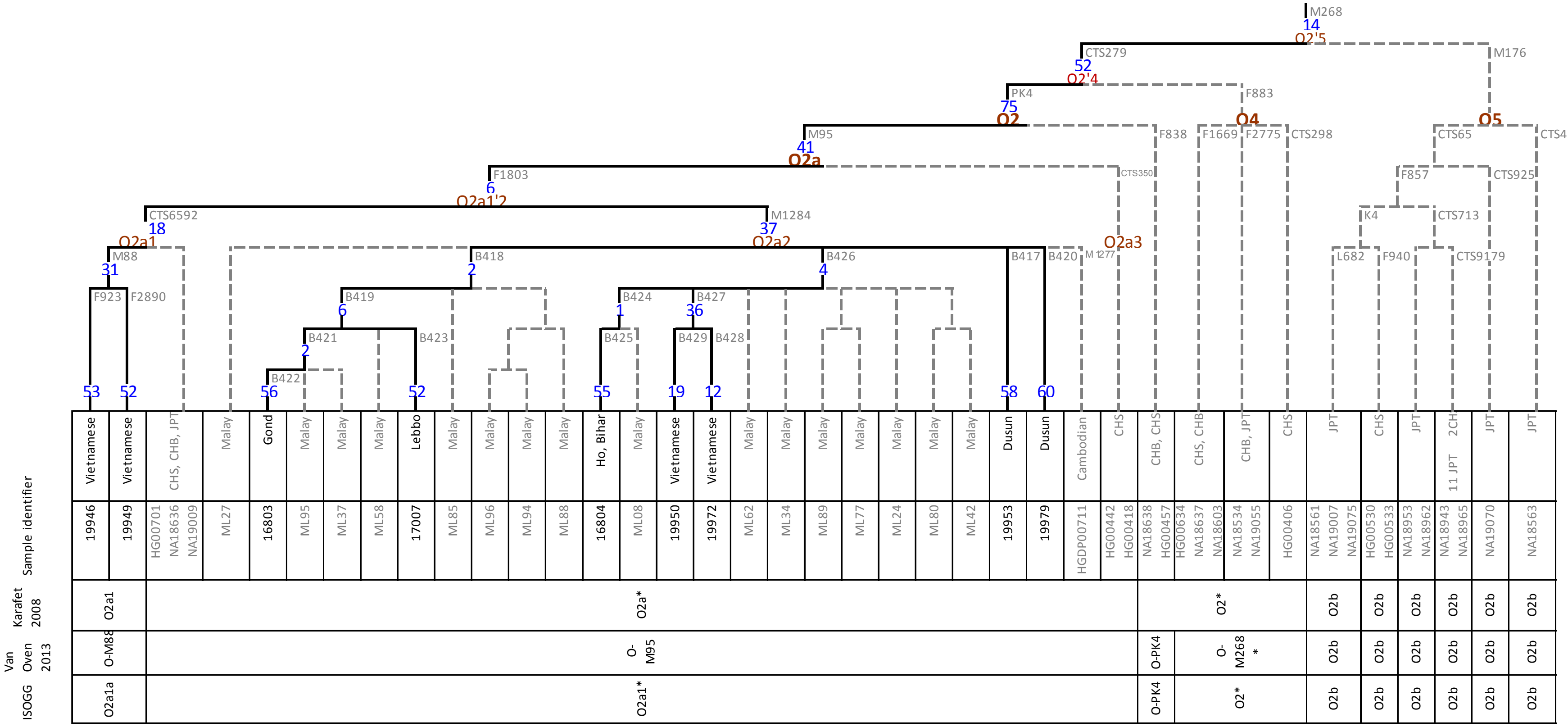


Figure S33. Refined topology of the Y-chromosome haplogroup O2. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

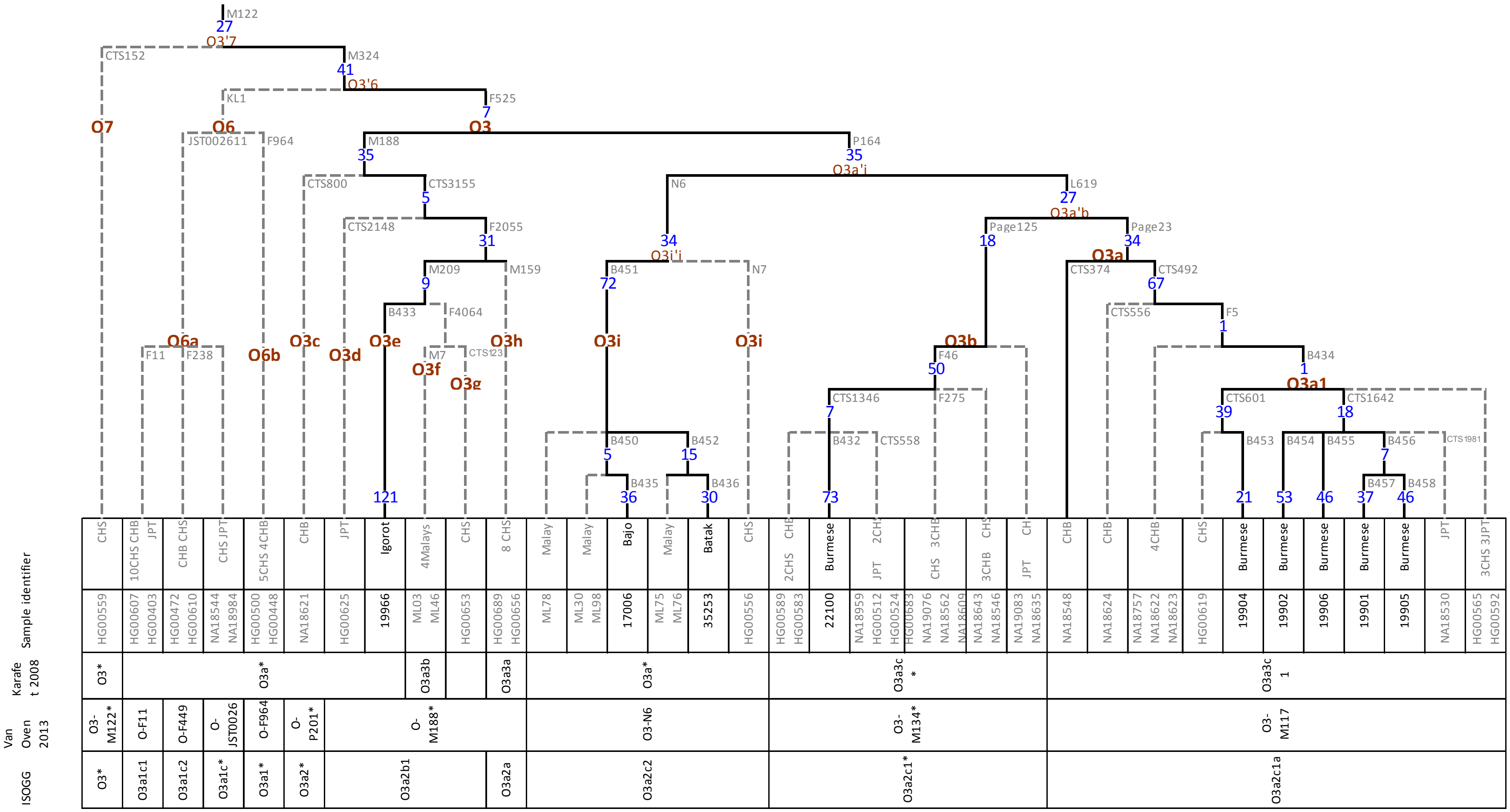


Figure S34. Refined topology of the Y-chromosome haplogroup O3. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

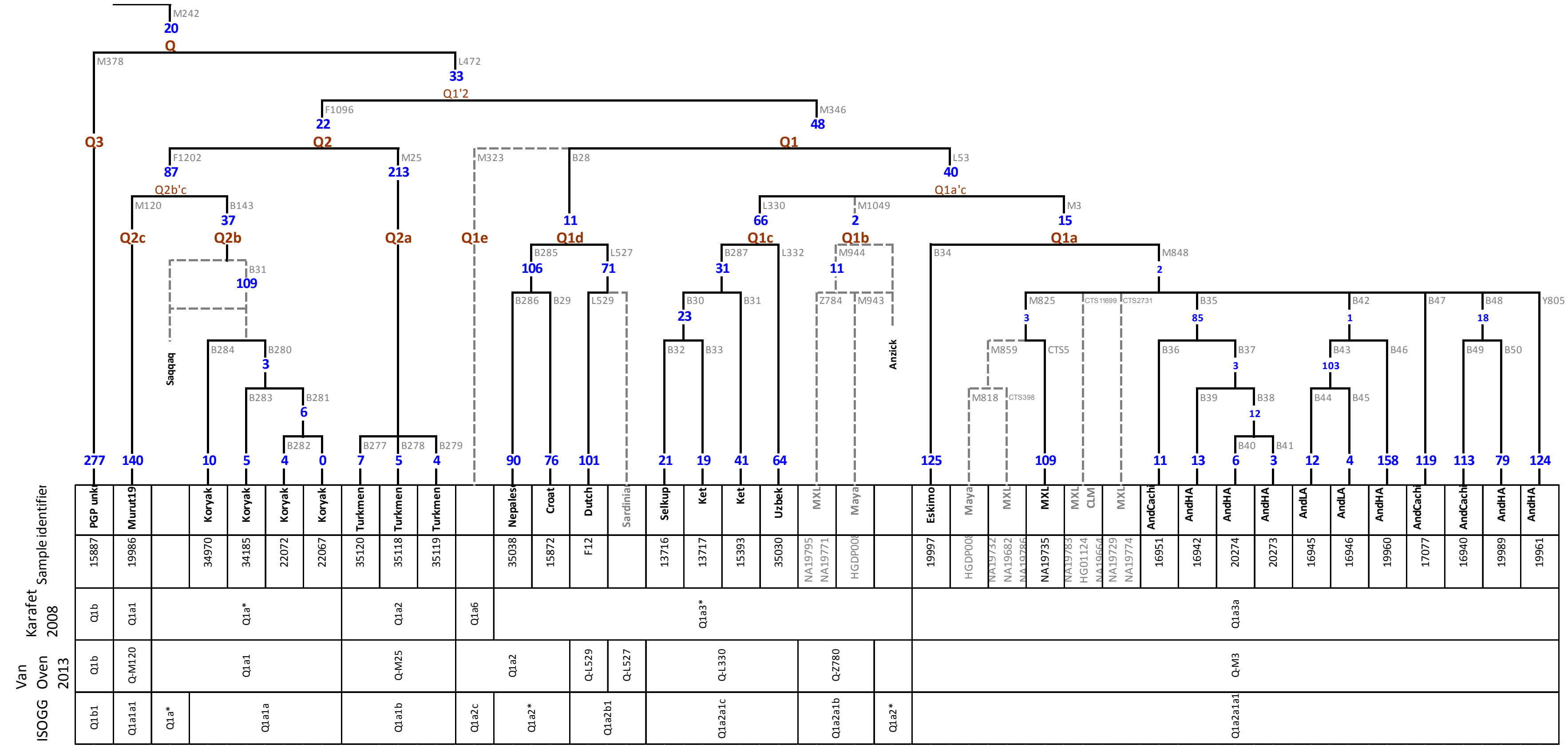
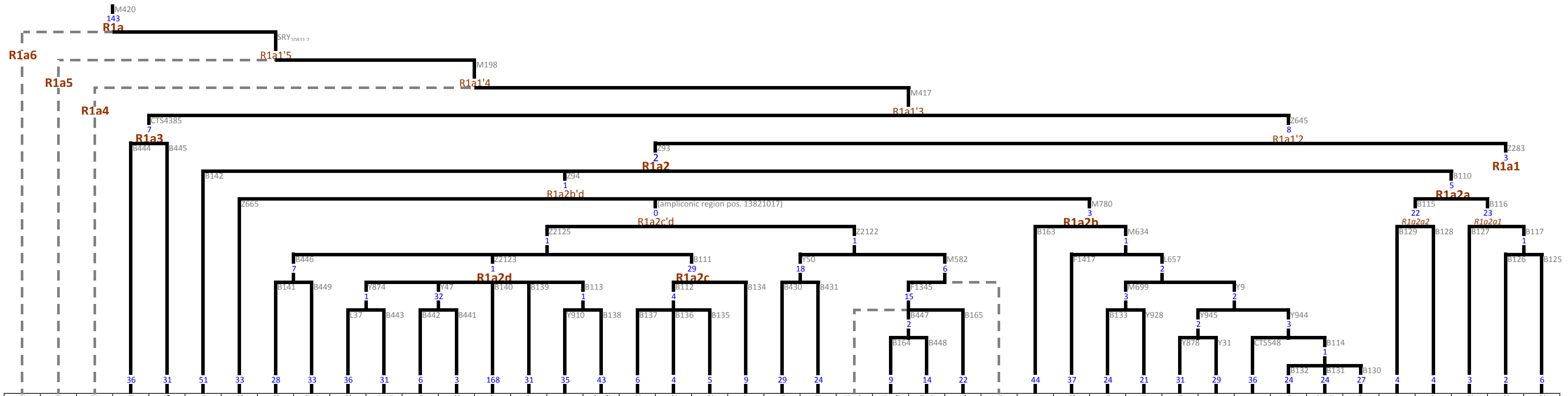
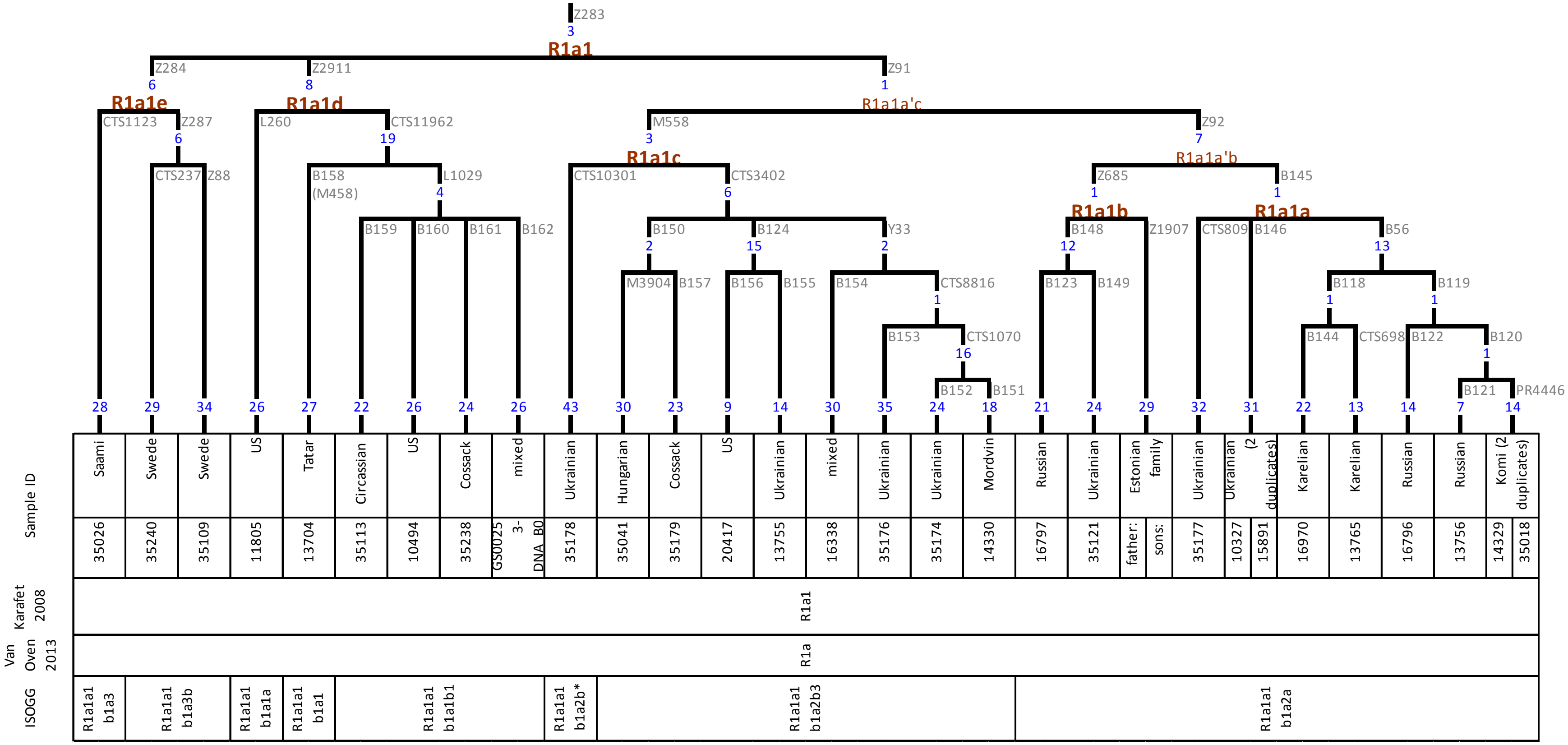


Figure S35. Refined topology of the Y-chromosome haplogroup Q. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



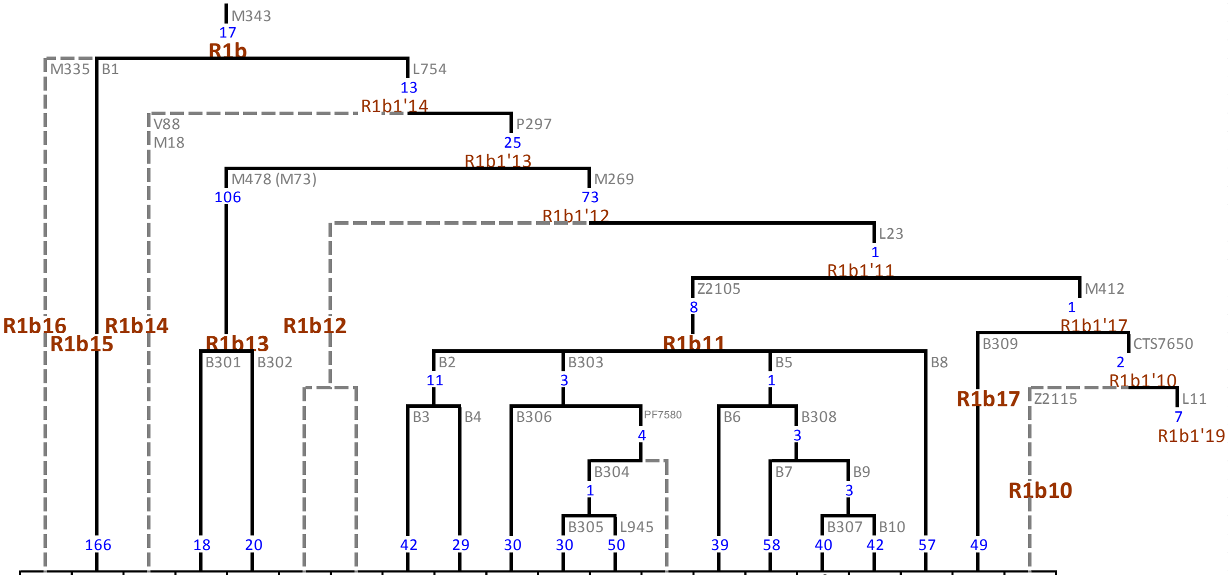
Sample ID	unsampled	unsampled	unsampled	mixed	US	Iranian	Tajik	Tajik	Dutch family	Tajik	Azerbaijani	Circassian	Tajik	Balkar	Iraqi Jew	Bashkir	Bene Israelite	Kyrgyz	Kyrgyz	Kyrgyz	Altay	Assyrian	mixed	Ashkenazi Jew	Ashkenazi Levite	Dutch family	N/A	Iberian	Cossack	Tajik	Orissa	Kurmi	Gupta	Gujarati	Gujarati	Kazakh	West Bengali	Iraqi Jew	Altay	Altay	Kyrgyz	Kyrgyz	Kyrgyz
				14566	15183	16875	16196	14417	father: F23 son: S23	16188	35237	35114	13758	16171	16198	16276	16974	16191	16197	13757	22439	13749	15893	P3 (Poznik et al. 2013_Illumina platform data)	16207	father: F3 son: S3	GS00253-DNA_D01	HG01617 (1000 Genomes Project data)	35239	16189	16808	16966	16799	20850	20846	14415	16801	16972	16967	16174	16187	16195	16193
Karafet et al 2008	-	R1a*	R1a1																																								
Van Oven 2013	R1a																																										
ISOGG	R1a*	R1a1*	R1a1a*	R1a1a1*		R1a1a1b2*	R1a1a1b2a*			R1a1a1b2a2a								R1a1a1b2a*				R1a1a1b2a2b*		R1a1a1b2a2b1				R1a1a1b2a2b*	R1a1a1b2a*		R1a1a1b2a1*		R1a1a1b2a1a						R1a1a1b2*				

Figure S36. Refined topology of the Y-chromosome haplogroup R1a2 and R1a3. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing. The topology of hg R1a has been covered extensively with our set of 70 samples (excluding 2 duplicates), three of which belong to an Estonian family and two each to two Dutch father-son pairs. In the absence of basal branches in our dataset, the 143 mutations at the root of hg R1a cannot be specifically distributed between the branches R1a1'6. Apart from M420, SRY10831.2, M198 and M417, known to define the branches R1a1'6 as they are drawn, the rest of the mutations (139) are marked in the annotations table as M420_eq, although they may or may not be equivalent to M420.



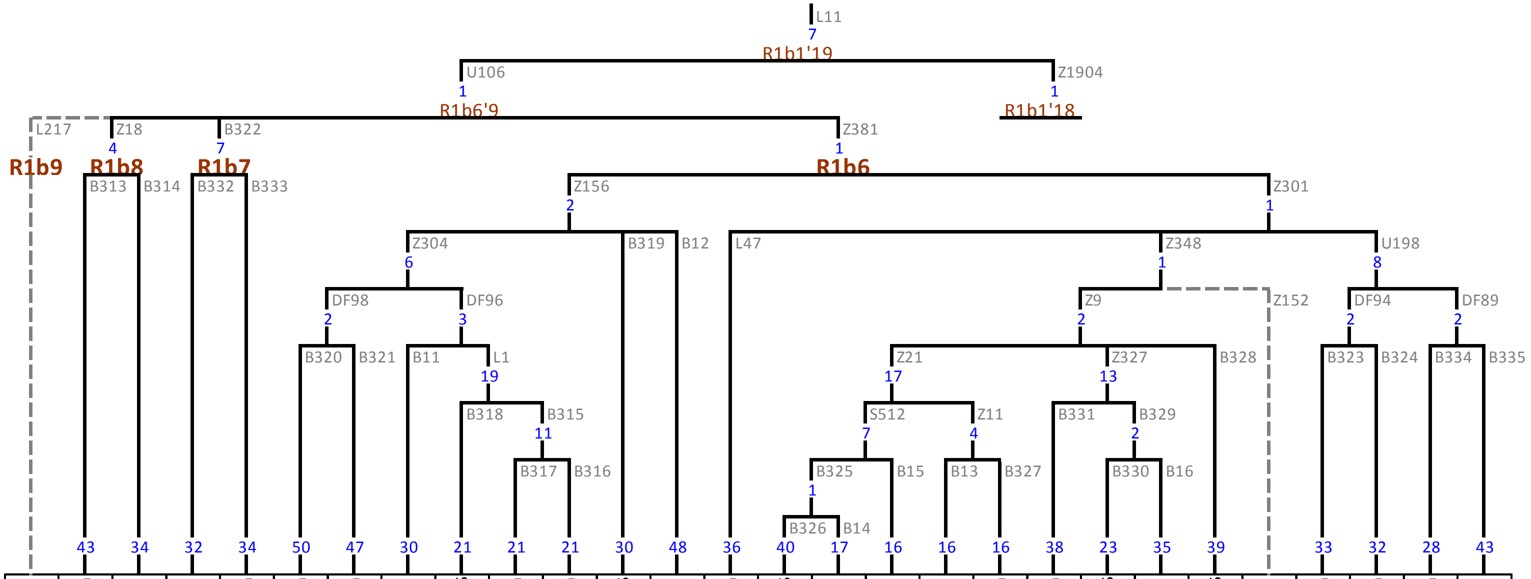
Sample ID	Saami	Swede	Swede	US	Tatar	Circassian	US	Cossack	mixed	Ukrainian	Hungarian	Cossack	US	Ukrainian	mixed	Ukrainian	Ukrainian	Mordvin	Russian	Ukrainian	Estonian family	Ukrainian	Ukrainian (2 duplicates)	Karelian	Karelian	Russian	Russian	Komi (2 duplicates)
	35026	35240	35109	11805	13704	35113	10494	35238	GS00253-DNA_B0	35178	35041	35179	20417	13755	16338	35176	35174	14330	16797	35121	father: sons:	35177	10327 15891	16970	13765	16796	13756	14329 35018
Karafet 2008	R1a1																											
Van Oven 2013	R1a																											
ISOGG	R1a1a1b1a3	R1a1a1b1a3b		R1a1a1b1a1a	R1a1a1b1a1	R1a1a1b1a1b1				R1a1a1b1a2b*	R1a1a1b1a2b3								R1a1a1b1a2a									

Figure S37. Refined topology of the Y-chromosome haplogroup (hg) R1a1. In one case, the relationship of the branches does not follow previous work: in the Y-chromosome minimal reference phylogeny (van Oven et al. 2013), R-L260 is nested within R-M458, whereas in our dataset, the branch defined by L260 constitutes an outgroup for the group containing the branch carrying M458.



Sample identifier		Tadjik	29 Sardinians	Bashkir	Mongol-Khalkh	4 Sardinians	6 Sardinians	TurkJew	MoroJew	Armenian	PGP unkn	Assyrian	2 Sardinians	Tabassaran	KurdiJew	PGP US	Dutch	Avar	PGP unkn	3 Sardinians
		16192		16134	35234			16133	16886	35173	14559	13750		14364	16881	16032	F2	14406	12713	
Karafet 2008	R1b1c	R1b1*		R1b1b1				R1b1b2												
Van Oven 2013	R1b*		R1b2	R1b1a																
ISOGG	R1b1b	R1b1*	R1b1c	R1b1a1				R1b1a2a2											R1b1a2a1*	

Figure S38. Refined topology of the Y-chromosome haplogroup R1b10-17. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



Sample identifier		PGP unkn	Ukrainian	Dutch	PGP unkn	PGP unkn	PGP unkn	Dutch	PGP US	PGP unkn	PGP unkn	PGP US	Dutch	PGP unkn	PGP US	Dutch	Dutch	Dutch	PGP unkn	PGP unkn	PGP US	Dutch	PGP US		PGP unkn	PGP unkn	PGP unkn	Dutch
		15227	35175	F8	9926	9932	16334	F1	15885	9920	16014	16027	F21	15155	9931	F4	F10	F9	15175	11740	20414	F24	6909		GS002 53	12712	16029	F22
Karafet 2008	R1b1b2g																							R1b1b2g 1	R1b1b2g			
Van Oven 2013																												
ISOGG	R1b1a2a1 a1a	R1b1a2a1 a1b		R1b1a2a1 a1*		R1b1a2a1 a1c1						R1b1a2a1 a1c1		R1b1a2a1 a1c2b1	R1b1a2a1 a1c2b2a1 a1b2			R1b1a2a1 a1c2b2a1 a1		R1b1a2a1 a1c2b2				R1b1a2a1 a1c2a	R1b1a2a1 a1c2*			

Figure S39. Refined topology of the Y-chromosome haplogroup R1b6-9. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

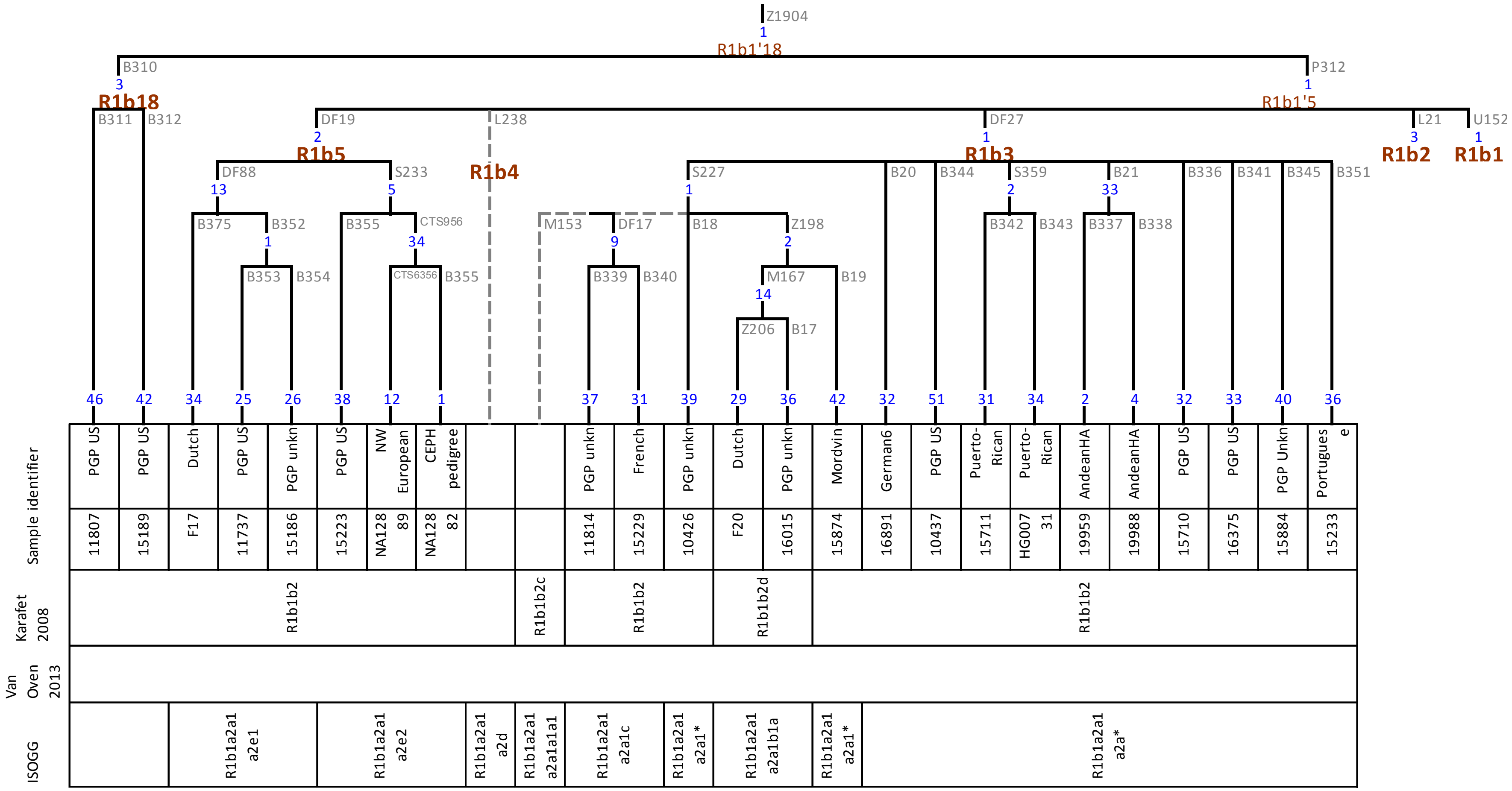


Figure S40. Refined topology of the Y-chromosome haplogroup R1b3-5. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.

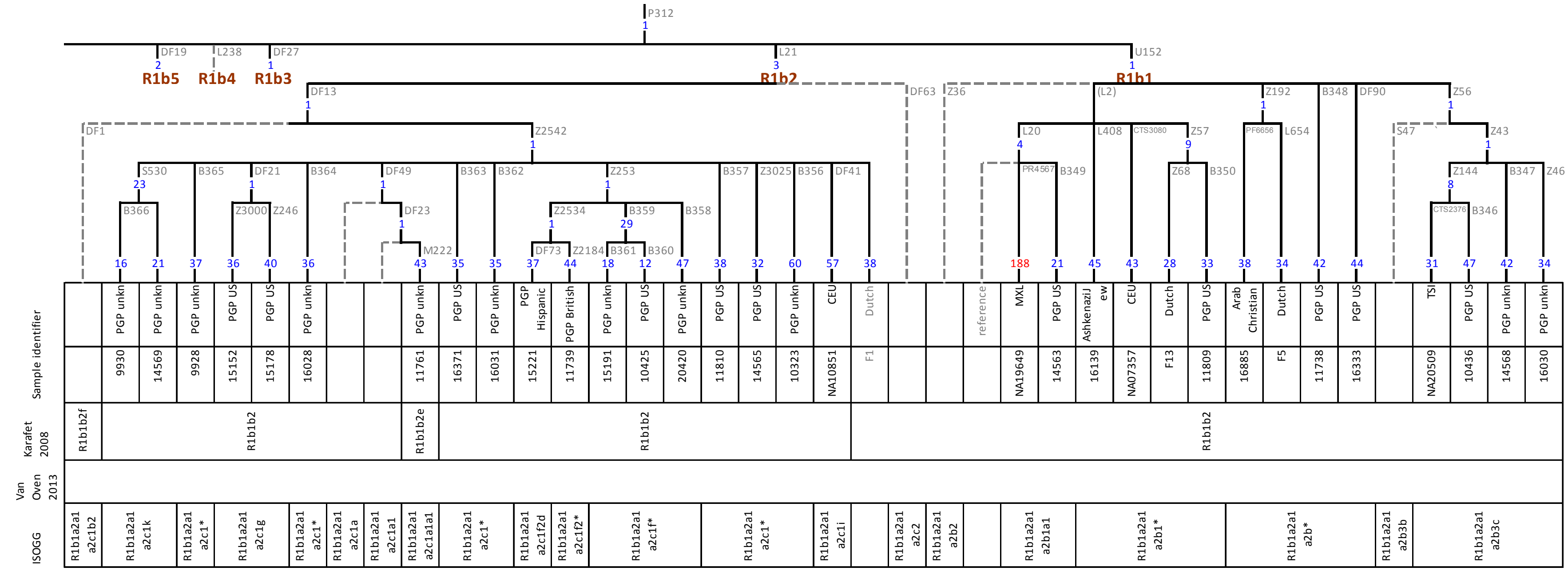
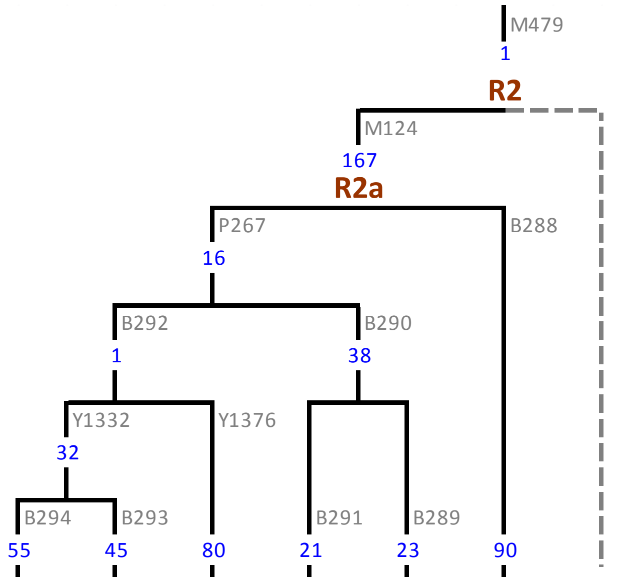


Figure S41. Refined topology of the Y-chromosome haplogroup R1b1-2. Grey dashed lines denote known branches with no sequence data and branches based on low coverage sequencing.



Sample identifier	Gujarati	Santhal	UP Brahmin	UP Brahmin	Thakur	Iran Jew	
	NA20845	16800	16809	16807	16798	16216	
Karafet 2008	R2						R*
Van Oven 2013	R2						
ISOGG	R2a1					R2a	R2*

Figure S42. Refined topology of the Y-chromosome haplogroup R2. Grey dashed lines denote known branches with no sequence data.