

African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages

Alec Knight, Peter A. Underhill,
 Holly M. Mortensen, Lev A. Zhivotovsky,
 Alice A. Lin, Brenna M. Henn, Dorothy Louis,
 Merritt Ruhlen, and Joanna L. Mountain

Supplementary References

- S1. Tajima, F., and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* 1, 269–285.
- S2. Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., and Griffiths, R.C. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* 15, 427–441.
- S3. Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Grazioli, G., Heidorn, F., Herrmann, S., Herzog, B. Hidding, M., et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal. Med.* 110, 125–133, 141–149.
- S4. Rolf, B., Röhl, A., Forster, P., and Brinkmann, B. (1999). On the genetic origins of the Turks: study of six Y-chromosome short tandem repeats. In *Genomic Diversity: Applications in Human Population Genetics*, S.S. Papiha, R. Deka, and R. Chakraborty, eds. (New York: Plenum), pp. 75–82.
- S5. Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43–62.

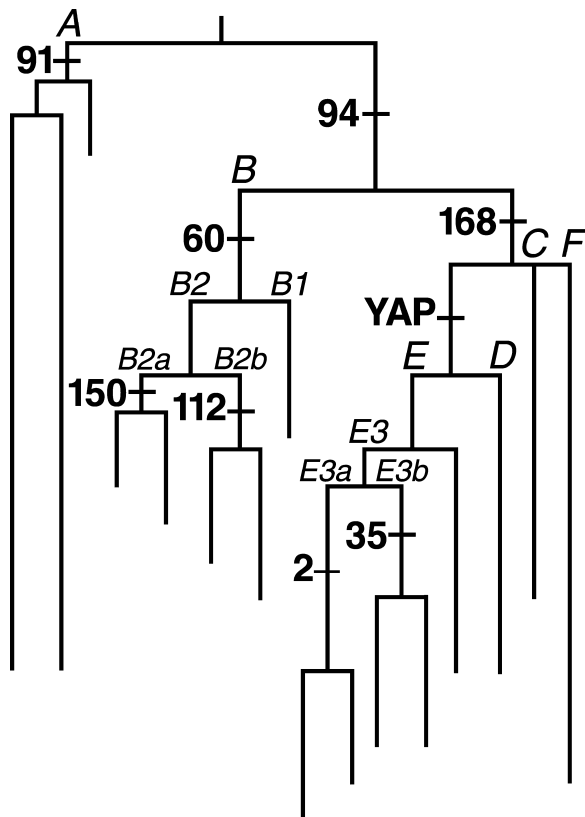


Figure S1. Relationships among Y Chromosome UEP Haplotypes Considered in This Study

Typed UEPs are indicated by the numbers at the horizontal bars. Haplogroup labels (YCC) are indicated in *italics*. The branch lengths correspond to the number of known mutations, a factor likely influenced by ascertainment bias. Note, for example, that the most basal branch of the M91 clade is older than the two more derived branches yet is much shorter. Branch length, therefore, does not correspond to time depth. For haplogroups studied less extensively, only representative branches are included. Haplogroups A, B, and E are most frequent among African populations. Haplogroups C, D, and F are highly diversified but are rare in African populations and are therefore represented only by the longest known branch within a clade.

Table S1. Genetic Distances between African Populations Estimated from mtDNA HV1 Sequence Variation

	Had	Ira	Dat	Suk	Jul'	Som	Ful	Tur	Nub	Din	Bia	Mbu	Tua	Kik	Hau	Kan	Yor	Man	Egy	Nue	Nub	Ber
Hadzabe	5.6	11.1	10.9	9.0	9.9	7.9	7.6	9.7	8.7	8.2	10.8	9.7	7.2	8.4	6.7	7.3	7.7	7.4	8.6	7.5	8.1	8.1
Iraqw	3.0	<i>7.7</i>	11.0	11.2	10.1	11.4	11.3	11.2	11.1	11.0	12.2	10.4	11.1	10.5	10.9	11.2	11.1	11.2	11.8	11.7	11.1	11.5
Datog	2.9	0.3	<i>7.6</i>	10.9	10.4	11.0	11.0	11.0	10.7	10.7	12.3	10.0	10.9	10.4	10.6	10.9	10.9	11.1	11.4	11.3	10.7	11.2
Sukuma	1.2	0.8	0.6	<i>7.0</i>	10.1	9.4	9.4	10.4	9.6	9.5	11.6	10.4	9.3	9.5	8.8	9.4	9.3	9.5	10.0	9.5	9.7	9.7
Jul'hoansi	5.6	3.2	3.6	3.6	3.0	10.0	10.3	10.1	9.6	9.3	8.9	9.4	9.8	9.6	9.7	10.3	9.6	10.2	10.2	10.3	9.6	10.5
Somali	1.3	2.1	1.8	0.5	4.6	7.8	7.8	9.9	8.3	8.5	11.7	10.5	7.5	8.4	7.1	7.7	8.0	7.8	8.1	7.8	8.5	7.2
Fulbe	1.2	2.4	2.2	0.8	5.2	0.3	7.1	9.9	8.2	8.4	11.6	10.5	7.2	8.3	6.5	7.2	7.4	7.0	8.1	7.6	8.6	7.1
Turkana	1.5	0.4	0.4	0.0	3.2	0.7	1.0	<i>7.7</i>	10.0	9.9	11.8	10.4	9.7	9.7	9.3	9.8	9.8	9.8	10.5	10.0	10.1	10.1
Nubia	1.7	1.5	1.2	0.4	3.9	0.2	0.4	0.5	8.4	8.8	11.5	10.5	8.0	8.7	7.7	8.2	8.3	8.3	8.2	8.3	9.0	7.3
Dinka	1.1	1.2	1.0	0.1	3.4	0.2	0.5	0.1	0.3	8.7	10.8	9.8	8.2	8.7	7.7	8.4	8.4	8.3	9.0	8.4	8.7	8.5
Biaka	3.9	2.7	2.9	2.6	3.4	3.8	3.9	2.4	3.3	2.4	8.1	11.2	11.1	11.3	10.9	11.7	11.0	11.3	12.1	11.8	11.2	12.3
Mbuti	3.2	1.3	1.0	1.7	4.2	2.9	3.2	1.3	2.6	1.8	3.4	7.4	10.2	9.8	9.9	10.2	10.3	10.5	11.2	10.7	9.9	11.0
Tuareg	0.9	2.2	2.1	0.7	4.8	0.1	0.1	0.8	0.3	0.2	3.5	3.0	7.1	8.1	6.5	7.1	7.4	7.1	7.9	7.4	8.2	7.0
Kikuyu	1.3	0.8	0.7	0.2	3.7	0.2	0.4	0.0	0.1	0.0	2.9	1.7	0.2	8.7	7.7	8.2	8.3	8.2	8.9	8.6	8.8	8.1
Hausa	0.9	2.6	2.3	0.8	5.2	0.2	0.0	0.9	0.5	0.4	3.8	3.3	-0.1	0.4	5.9	6.5	6.8	6.4	7.6	6.9	7.8	6.7
Kanuri	0.8	2.2	1.9	0.6	5.1	0.1	-0.1	0.7	0.3	0.3	3.9	2.8	-0.1	0.2	-0.2	7.5	7.5	7.1	8.1	7.6	8.5	7.1
Yoruba	1.0	1.9	1.7	0.5	4.3	0.2	0.0	0.6	0.3	0.2	3.1	2.7	0.0	0.2	0.0	-0.1	7.7	7.2	8.3	7.9	8.6	7.6
Mandenka	1.3	2.6	2.5	1.2	5.5	0.7	0.2	1.2	0.8	0.7	4.0	3.5	0.3	0.6	0.2	0.1	0.2	6.5	8.1	7.6	8.5	7.2
Egyptian	2.0	2.6	2.3	1.2	4.9	0.4	0.7	1.3	0.2	0.8	4.2	3.6	0.5	0.7	0.8	0.5	0.7	1.1	7.7	8.0	9.2	6.4
Nuer	0.9	2.5	2.2	0.7	5.0	0.1	0.2	0.8	0.3	0.2	3.9	3.2	0.0	0.4	0.1	0.1	0.3	0.6	0.3	7.6	8.5	7.0
Nuba	0.8	1.3	0.9	0.2	3.6	0.1	0.5	0.2	0.3	-0.1	2.6	1.7	0.2	0.0	0.3	0.3	0.2	0.8	0.8	0.2	9.0	8.7
Berber	3.1	3.9	3.6	2.4	6.8	1.1	1.3	2.5	0.9	1.9	6.0	5.1	1.2	1.6	1.5	1.2	1.5	1.7	0.3	0.9	2.0	4.4

Please see [S1] for information on genetic distances. The numbers on the diagonal (italicized) represent the average number of uncorrected pairwise differences within the population (π_X). The numbers above the diagonal represent the average number of uncorrected pairwise differences between populations (π_{XY}). The numbers below the diagonal represent corrected average pairwise difference ($(\pi_X + \pi_Y)/2$).

Table S2. Y Chromosome STR Haplotypes within UEP Marker-Defined Haplogroups

UEP (hg)	Pop. ^a	n ^b	19	389m	389n	389q	390m	390n	390p	392 ^c
M112 (B2b)	H	1	15	5	10	9	8	12	A	13
	H	7	15	5	10	10	9	13	A	13
	H	2	15	5	10	9	8	13	A	13
	H	1	17	5	11	9	8	16	G	13
	H	1	16	5	10	10	8	16	G	13
	S	1	16	5	10	10	8	16	G	13
	S	1	15	5	10	10	9	11	A	13
	K	1	16	5	7	9	8	16	G	13
	K	1	16	5	11	11	8	12	A	13
	K	1	16	5	7	10	8	16	G	13
	K	1	17	5	11	11	8	12	A	13
	K	1	17	5	7	11	8	16	G	13
	K	1	16	5	7	11	8	16	G	13
	M	1	16	5	11	11	8	15	G	13
	M	1	14	5	11	10	8	17	G	14
	M	1	15	5	11	11	8	15	G	13
	L	1	16	5	11	9	8	16	G	13
	B	1	18	5	11	11	8	15	G	13
	B	2	16	5	11	8	8	16	G	13
	B	1	16	5	11	10	8	17	G	13
B	1	17	5	11	8	8	16	G	13	
M150 (B2a)	S	3	16	7	11	10	8	16	G	13
M35 (E3b)	H	2	14	6	11	11	8	16	G	13
	D	3	14	6	11	11	8	16	G	13
	H	1	14	6	11	7	8	16	G	13
	I	1	14	6	11	7	8	16	G	13
	S	1	14	6	11	7	8	16	G	13
	D	1	15	6	12	10	8	16	G	13
	D	1	14	6	12	10	8	16	G	14
	I	1	15	6	11	11	8	15	A	13
	S	1	12	6	12	10	8	16	G	14
	M2 (E3a)	H	1	16	6	11	10	8	13	G
S		4	16	6	11	10	8	13	G	13
H		3	18	6	11	11	8	13	G	13
S		4	16	6	12	10	8	13	G	13
H		1	18	6	10	10	8	13	G	13
H		1	18	6	11	10	8	13	G	13
S		4	18	6	11	10	8	13	G	13
D		1	15	?	?	?	8	13	G	13
S		3	17	6	11	10	8	13	G	13
H		1	17	6	11	11	8	13	G	13
S		2	16	6	11	11	8	13	G	13
S		2	17	6	11	10	8	14	G	13
S		1	16	6	11	10	?	?	?	13
Other	I	1	15	6	11	11	8	16	G	13
YAP+ (E1, E2, E*, D)	S	1	16	6	11	10	8	13	G	11
	S	1	15	6	11	9	8	17	G	13
	S	1	15	6	10	10	8	15	G	13
	S	1	15	6	10	9	8	17	G	13
YAP- (C, F)	S	1	16	6	11	9	8	17	G	13
	H	1	?	?	?	?	8	16	G	?
	D	1	15	4	11	11	8	15	G	13
	D	1	15	5	13	11	?	?	?	12
I	1	15	4	12	10	8	16	G	13	
I	1	16	4	11	11	8	15	G	15	

^aPop. (populations or ethnic groups) are: H, Hadzabe; S, Sukuma; K, San; M, Mbuti; L, Lisongo; B, Biaka.

^bIndicates the number of individuals.

^cLocus 390p is a single nucleotide polymorphism within a tetranucleotide STR with four CTAT repeats adjacent downstream and the 390n CTAT STR adjacent upstream. As the A allele is likely the result of a deletion of the CTGT repeat rather than a base substitution, it was scored as a deletion. Therefore, the A allele was scored as four repeat units (including only the four CTAT downstream repeats), and the G allele was scored as five repeat units (four CTAT plus one CTGT).

Table S3. PCR and Sequencing Primers

Designation	Sequence (5'-3')
Mitochondrion	
L15988	TCTTTAACTCCACCATTAGCA
H598	ACATTTTCAGTGATTGCTTT
L15990	TTTAACTCCACCATTAGCACC
H587	TATTGCTTTGAGGAGGTAAG
H16405 ^a	GATATTGATTTACGGAGGAT
L3592	CTACCATCACCTCTACATC
H3650	CACCCTGATCAGAGGATTG
L10290	TCCTTTTACCCTACCATGAG
H10557	ATTATTCCTTCTAGGCATAGTAG
Y Chromosome	
YAPF	CAAGTTAGCTGTCCATACTG
YAPR	CGTTAATAAGGAGAGGACTAG
YAP.1 ^b	CAGGGGAAGATAAAGAAATA
YAP.2 ^b	ACTGCTAAAAGGGGATGGAT
M2F	AGGCACTGGTCAGAATGAAG
M2R	AATGGAAAATACAGCTCCCC
M2FN	GGTCAGAATGAAGTGAATGG
M2RN	AATACAGCTCCCCCTTTATC
M35F	TAAGCCTAAGAGCAGTCAGAG
M35R	AGAGGGAGCAATGAGGACA
M35FN	AGAGCAGTCAGAGTAGAATG
M35RN	GCAATGAGGACATTATTCTC
M60F	GCACTGGCGTTTCATCATCT
M60R	ATGTTTATTATGGTTCAGGAGG
M60FN	ATCTGGGAGCAGCTCAAAG
M60RN	TCAGGAGGAGCGACGTCTAC
M94F	CACATGGAGAACAGAGAAATGC
M94R	CTTGTGAAATGTTGTGAAAGTGG
M94FN	AACAGAGAAATGCAGTGCAG
M94RN	TGAAAGTGGGGTTCACAGAC
M112F	ACTTTTTCCAACAGTTATTTTTGA
M112R	TATATTTCTTGATGATGAGACCAAT
M112FN	CAAACCTACAGAAGTCTTG
M112RN	TTATTACACCTGATAATCAC
M150F	GCAGTGGAGATGAAGTGAGAC
M150R	CCTACTTTCCCCCTCTTCTG
M150FN	CTGGGCTTTGGAGAGGTGAG
M150RN	GGCAGATTGTGGCCTTCAAG
DYS19F	CTACTGAGTTTCTGTTATAGT
DYS19R ^b	ATGGCATGTAGTGAGGACA
DYS19FN	CTGAGTTTCTGTTATAGTGT
DYS19RN	GTGAGGACAAGGAGTCCATC
DYS389F ^b	CCAACTCTCATCTGTATTATCTAT
DYS389R ^b	TCTTATCTCCACCCACCAGA
DYS389FN	ATCTGTATTATCTATGTGTG
DYS389RN	TGTTTCCAGACATTGCCAAG
DYS390F ^b	TATATTTTACACATTTTGGGCC
DYS390R ^b	TGACAGTAAAATGAACACATTGC
DYS390FN	TTTTGGGCCTTGCAATTTGG
DYS390RN	GTAAAATGAACACATTGCTATG
DYS392F ^b	TCATTAATCTAGCTTTTAAAAACAA
DYS392R ^b	AGACCCAGTTGATGCAATGT
DYS392FN	AAAAACAATAATTTGATTTTC
DYS392RN	GCAATGTAAATTCCTACAGTC

^aInternal sequencing primer. In all other cases, PCR primers were used for sequencing. MtDNA HV1 primers are designated by the position of the 3' terminus in the Cambridge reference sequence. Y chromosome primers are designated by locus. The suffix "F" designates forward, "R" designates reverse, "FN" designates forward nested, and "RN" designates reverse nested.

^bPrimer sources (those not original to this study) can be found in [S2–S5].