American Journal of Human Genetics, Volume *92*

Supplemental Data

Ancient Substructure in Early mtDNA

Lineages of Southern Africa

Chiara Barbieri, Mário Vicente, Jorge Rocha, Sununguko W. Mpoloka, Mark Stoneking, and Brigitte Pakendorf

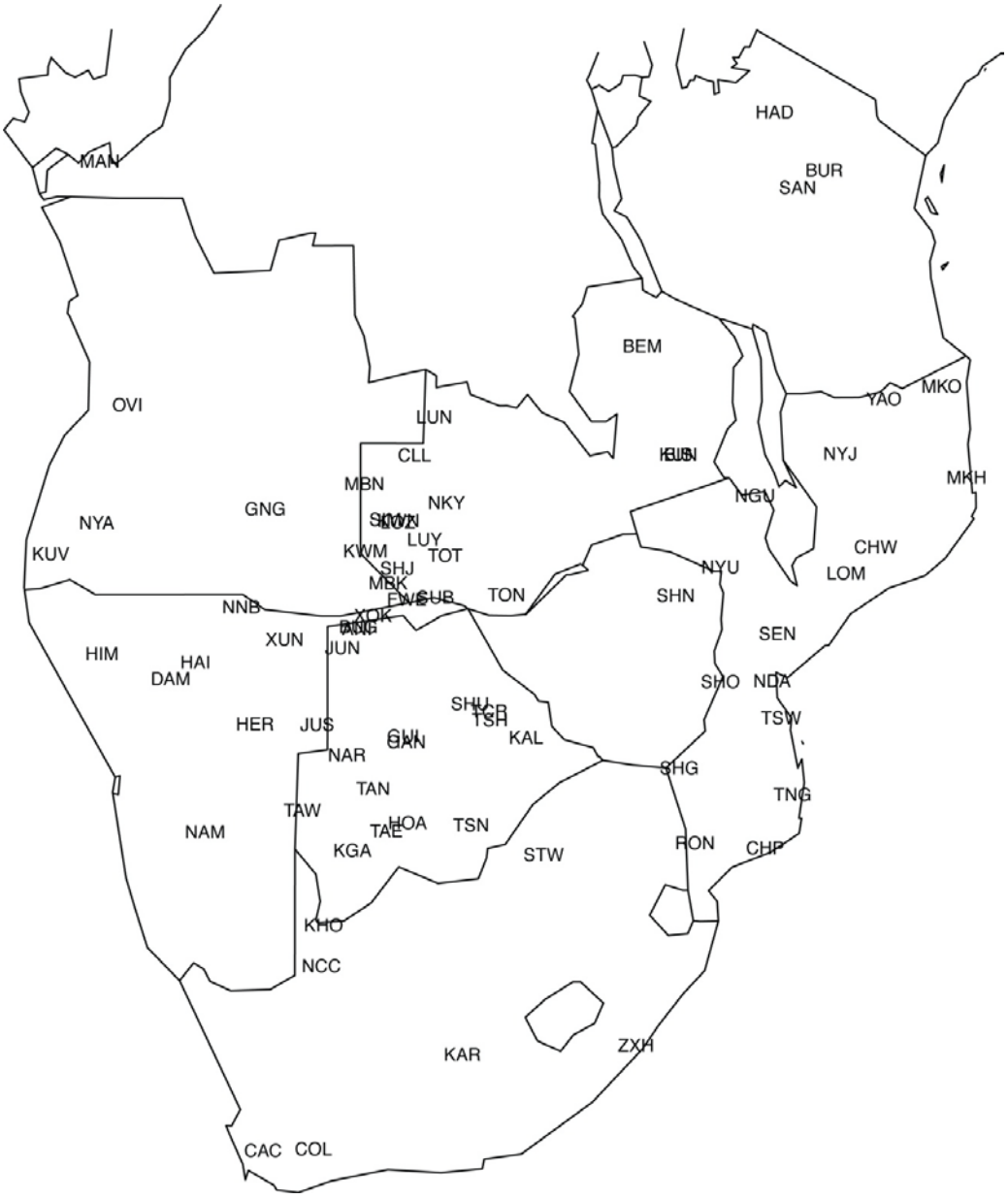

Figure S1: Map of approximate locations of the populations included in the surfer map (Figure 1 in the main text).

Population codes as indicated in Supplementary Table 1.

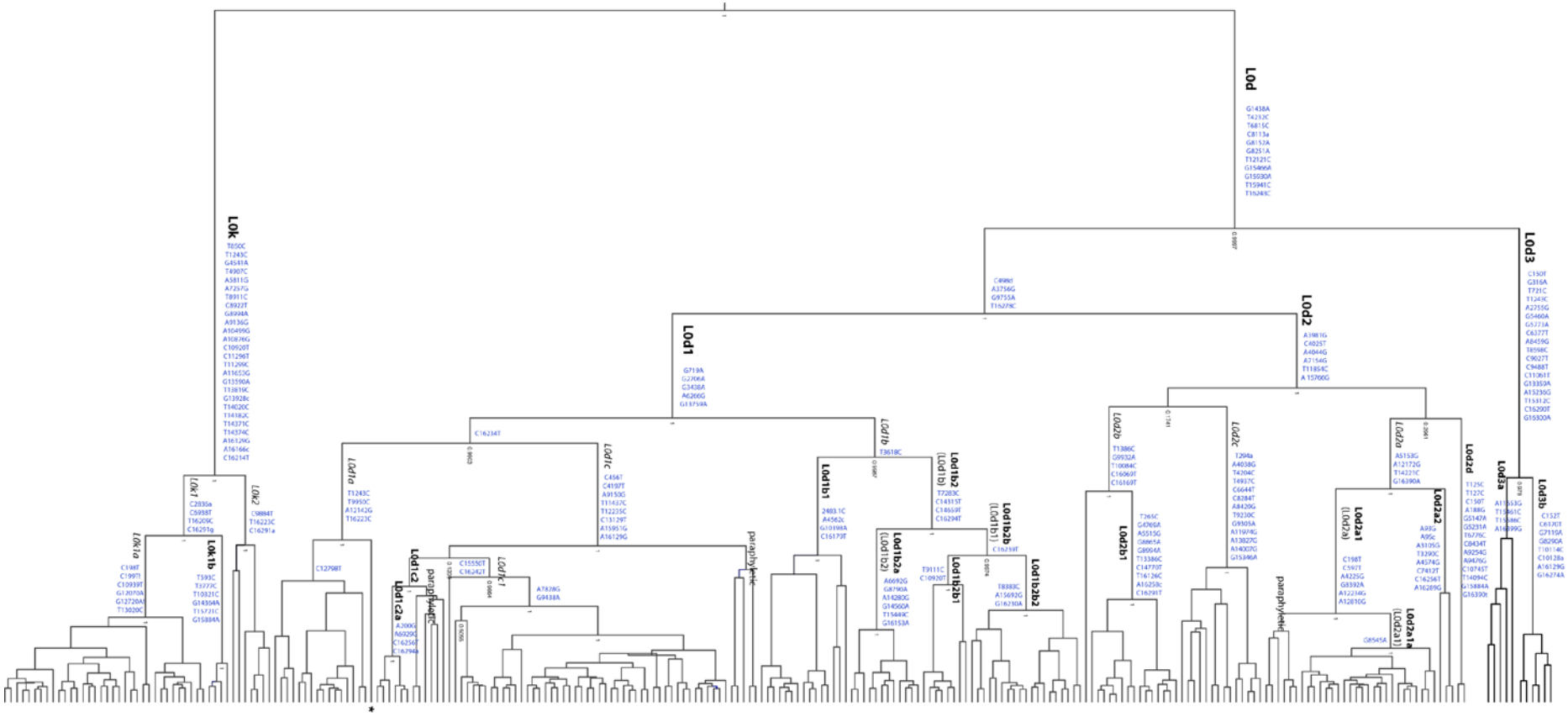

Figure S2: Tree of the 254 unique haplotypes of the dataset.

The tree is based on full sequences, without a time scale, and includes the major branches L0k, L0d1, L0d2 and L0d3. Sub-branches that have not changed are labeled in italic font; new branches defined here are labeled in bold font. When previously defined branches have to be renamed, the older label is indicated in brackets. The posterior probabilities associated with major nodes are shown. Mutations defining branches are shown in blue font: transversions are indicated with lowercase, and back mutations to an ancestral state are indicated with an exclamation mark (!). The individual marked with an asterisk is mentioned in the note about positions 199 and 16266 in Supplementary Table 3.

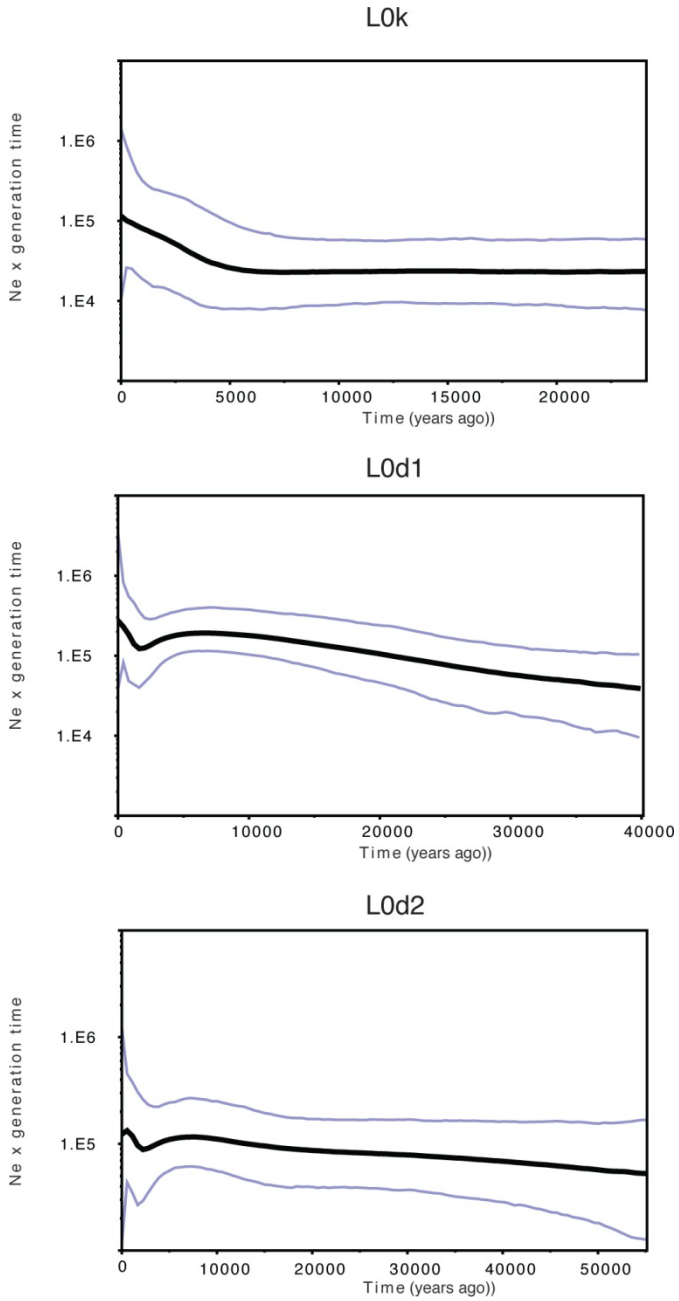

Figure S3: Bayesian Skyline Plots (BSP) of the L0k, L0d1 and L0d2 haplogroups.

The BSPs are based on the mtDNA coding region, estimated with 10 million iterations. The y axis for each plot is the product of the effective population size and the generation time and the x axis shows time using a linear relaxed clock with the substitution rate of 1.26×10^{-8} per site per year.

Table S1: List of African populations considered in the study, with frequencies of haplogroups L0d and L0k; frequencies of subhaplogroups are given only for populations from the present study.

Population	Code	Country (main)	Language family	lon	lat	N	L0d (%)	L0k (%)	L0d1	L0d2	L0d3	L0k1a	L0k1b	L0k2	Source
Kalanga	KAL	Botswana	Bantu	27.035522	-21.524627	17	**29.4**	**5.9**	29.4	0.0	0.0	0.0	0.0	5.9	PRESENT STUDY
Kgalagadi	KGA	Botswana	Bantu	21.751098	-24.796708	19	**52.6**	**0**	26.3	15.8	10.5	0.0	0.0	0.0	PRESENT STUDY
Tswana	TSN	Botswana	Bantu	25.3656	-24.066528	17	**29.4**	**0**	11.8	17.6	0.0	0.0	0.0	0.0	PRESENT STUDY
Herero	HER	Namibia	Bantu	18.7878333	-21.1344167	30	**16.7**	**0**	13.3	3.3	0.0	0.0	0.0	0.0	PRESENT STUDY
Himba	HIM	Namibia	Bantu	14.1235013	-19.100676	21	**9.5**	**0**	9.5	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
North Namibia Bantu	NNB	Namibia	Bantu	18.365478	-17.748687	10	**0**	**10**	0.0	0.0	0.0	0.0	10.0	0.0	PRESENT STUDY
Bemba	BEM	Zambia	Bantu	30.57312	-10.185187	12	**0**	**8.3**	0.0	0.0	0.0	0.0	0.0	8.3	PRESENT STUDY
Chokwe, Luchazi, Luvale	CLL	Zambia	Bantu	23.638916	-13.346865	33	**0**	**0**	0.0	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Kwangwa	KWN	Zambia	Bantu	23.152888	-15.2451259	36	**2.8**	**0**	2.8	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Lozi	LOZ	Zambia	Bantu	23.133544	-15.284185	118	**5.1**	**0.8**	4.2	0.8	0.0	0.0	0.8	0.0	PRESENT STUDY
Lunda	LUN	Zambia	Bantu	24.240417	-12.243392	9	**0**	**11.1**	0.0	0.0	0.0	0.0	0.0	11.1	PRESENT STUDY
Luyana	LUY	Zambia	Bantu	23.948364	-15.792254	8	**12.5**	**0**	12.5	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Mbunda	MBN	Zambia	Bantu	22.113647	-14.179186	67	**4.5**	**0**	4.5	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Nkoya	NKY	Zambia	Bantu	24.6118007	-14.723394	32	**0**	**0**	0.0	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Simaa	SIM	Zambia	Bantu	22.7506637	-15.2116789	44	**4.5**	**0**	4.5	0.0	0.0	0.0	0.0	0.0	Barbieri et a. 2012
Mbukushu	MBK	Zambia/ Namibia	Bantu	22.8387	-17.0567	20	**10**	**5**	5.0	5.0	0.0	0.0	5.0	0.0	Barbieri et a. 2012
Fwe	FWE	Zambia	Bantu	23.2241122	-17.5299131	33	**6.1**	**18.2**	6.1	0.0	0.0	0.0	6.1	12.1	Barbieri et a. 2012
Kwamashi	KWM	Zambia	Bantu	22.14452	-16.12411	32	**3.1**	**0**	3.1	0.0	0.0	0.0	0.0	0.0	Barbieri et a. 2012
Shanjo	SHJ	Zambia	Bantu	23.1145525	-16.6263951	24	**8.3**	**8.3**	4.2	4.2	0.0	0.0	0.0	8.3	Barbieri et a. 2012
Subiya	SUB	Zambia	Bantu	24.2796841	-17.4674193	17	**5.9**	**0**	0.0	5.9	0.0	0.0	0.0	0.0	Barbieri et a. 2012
Totela	TOT	Zambia	Bantu	24.596557	-16.256867	29	**0**	**0**	0.0	0.0	0.0	0.0	0.0	0.0	Barbieri et a. 2012
Tonga	TON	Zambia	Bantu	26.451416	-17.392579	35	**0**	**0**	0.0	0.0	0.0	0.0	0.0	0.0	Barbieri et a. 2012
Ganguela	GNG	Angola	Bantu	19.068603	-14.902322	20	**5**	**0**	5.0	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY
Kuvale	KUV	Angola	Bantu	12.564697	-16.214675	53	**24.5**	**0**	22.6	1.9	0.0	0.0	0.0	0.0	PRESENT STUDY
Nyaneka-Nkhumbi	NYA	Angola	Bantu	13.992919	-15.30538	59	**8.4**	**0**	8.4	0.0	0.0	0.0	0.0	0.0	PRESENT STUDY

Ovimbundu	OVI	Angola	Bantu	14.915771	-11.888853	60	**3.3**	**1.6**	1.7	1.7	0.0	1.7	0.0	0.0	PRESENT STUDY
ǁAnikhoe	ANI	Botswana	Khoe	21.8850954	-18.3734521	18	**44.4**	**22.2**	44.4	0.0	0.0	22.2	0.0	0.0	PRESENT STUDY
ǁXokhwe	XOK	Botswana	Khoe	22.3761	-17.9957	17	**17.6**	**11.8**	17.6	0.0	0.0	11.8	0.0	0.0	PRESENT STUDY
Bugakhoe	BUG	Botswana	Khoe	21.9367	-18.3219	14	**42.9**	**28.6**	42.9	0.0	0.0	28.6	0.0	0.0	PRESENT STUDY
Naro	NAR	Botswana	Khoe	21.5840541	-22.0320817	35	**77.1**	**17.1**	51.4	25.7	0.0	17.1	0.0	0.0	PRESENT STUDY
Gǁana	GAN	Botswana	Khoe	23.3889	-21.6523	15	**93.3**	**6.7**	80.0	13.3	0.0	6.7	0.0	0.0	PRESENT STUDY
Gǀui	GUI	Botswana	Khoe	23.2946698	-21.486584	31	**93.5**	**3.2**	51.6	41.9	0.0	3.2	0.0	0.0	PRESENT STUDY
Haiǁom	HAI	Namibia	Khoe	16.9694944	-19.3450768	51	**68.6**	**13.7**	39.2	27.5	2.0	13.7	0.0	0.0	PRESENT STUDY
Nama	NAM	Namibia	Khoe	17.2608889	-24.2660935	29	**79.3**	**3.4**	37.9	34.5	6.9	0.0	3.4	0.0	PRESENT STUDY
Damara	DAM	Namibia	Khoe	16.2257392	-19.8301838	38	**13.2**	**0**	10.5	2.6	0.0	0.0	0.0	0.0	PRESENT STUDY
Shua	SHU	Botswana	Khoe	25.3321307	-20.5502369	42	**35.7**	**2.4**	35.7	0.0	0.0	2.4	0.0	0.0	PRESENT STUDY
TcireTcire	TCR	Botswana	Khoe	25.9166477	-20.7658488	12	**50**	**16.7**	41.7	8.3	0.0	16.7	0.0	0.0	PRESENT STUDY
Tshwa	TSH	Botswana	Khoe	25.9365757	-21.0249347	22	**54.5**	**0**	50.0	4.5	0.0	0.0	0.0	0.0	PRESENT STUDY
ǂHoan	HOA	Botswana	K'xa	23.4351167	-23.9989176	13	**100**	**0**	92.3	7.7	0.0	0.0	0.0	0.0	PRESENT STUDY
!Xuun	XUN	Botswana	K'xa	19.6826306	-18.6907202	27	**55.5**	**33.3**	44.4	11.1	0.0	33.3	0.0	0.0	PRESENT STUDY
Ju|'hoan North	JUN	Botswana	K'xa	21.4524476	-18.9372569	40	**72.5**	**22.5**	50.0	22.5	0.0	22.5	0.0	0.0	PRESENT STUDY
Ju|'hoan South	JUS	Botswana	K'xa	20.6815392	-21.151918	44	**70.5**	**25**	50.0	20.5	0.0	25.0	0.0	0.0	PRESENT STUDY
Taa East	TAE	Botswana	Tuu	22.8206545	-24.2365162	30	**100**	**0**	46.7	53.3	0.0	0.0	0.0	0.0	PRESENT STUDY
Taa North	TAN	Botswana	Tuu	22.4158579	-23.0145647	25	**84**	**16**	68.0	16.0	0.0	16.0	0.0	0.0	PRESENT STUDY
Taa West	TAW	Botswana	Tuu	20.2727412	-23.639938	31	**74.2**	**22.6**	51.6	22.6	0.0	22.6	0.0	0.0	PRESENT STUDY
Shona	SHN	Zimbawe	Bantu	31.593017	-17.413546	59	**1.7**	**1.7**							Castrì et al. 2009
Kunda	KUN	Zambia	Bantu	31.671753	-13.325485	36	**2.8**	**0**							De Filippo et al. 2010
Bisa	BIS	Zambia	Bantu	31.67175	-13.325483	46	**0**	**0**							De Filippo et al. 2010, present study
SA Coloured	COL	South Africa	Indoeuropean	20.562744	-33.449777	563	**60**	**0**							Quintana-Murci et al. 2010
Chopi	CHP	Mozambique	Bantu	34.317627	-24.726875	27	**0**	**0**							Salas et al. 2002
Chwabo	CHW	Mozambique	Bantu	37.679443	-16.003576	20	**0**	**0**							Salas et al. 2002
Lomwe	LOM	Mozambique	Bantu	36.778564	-16.762468	20	**0**	**0**							Salas et al. 2002
Makhwa	MKH	Mozambique	Bantu	40.447998	-13.987376	20	**0**	**0**							Salas et al. 2002
Makonde	MKO	Mozambique	Bantu	39.700927	-11.350797	19	**5.3**	**0**							Salas et al. 2002
Ndau	NDA	Mozambique	Bantu	34.537353	-19.890723	19	**30**	**0**							Salas et al. 2002

Nguni	NGU	Mozambique	Bantu	34.010009	-14.51978	11	**0**	**0**							Salas et al. 2002
Nyanja	NYJ	Mozambique	Bantu	36.602783	-13.304103	20	**0**	**0**							Salas et al. 2002
Nyungwe	NYU	Mozambique	Bantu	32.955322	-16.594081	20	**0**	**0**							Salas et al. 2002
Ronga	RON	Mozambique	Bantu	32.186279	-24.58709	21	**19**	**0**							Salas et al. 2002
Sena	SEN	Mozambique	Bantu	34.691162	-18.521283	21	**0**	**0**							Salas et al. 2002
Shangaan	SHG	Mozambique	Bantu	31.70288	-22.411029	22	**4.5**	**0**							Salas et al. 2002
Shona	SHO	Mozambique	Bantu	32.955322	-19.911384	18	**0**	**0**							Salas et al. 2002
Tonga	TNG	Mozambique	Bantu	35.152587	-23.180764	20	**5**	**0**							Salas et al. 2002
Tswa	TSW	Mozambique	Bantu	34.801025	-20.96144	19	**15.8**	**0**							Salas et al. 2002
Yao	YAO	Mozambique	Bantu	37.965087	-11.716788	10	**0**	**0**							Salas et al. 2002
Karretjie Mense	KAR	South Africa	Indoeuropean [a]	25.101013	-30.712638	30	**100**	**0**							Schlebusch et al. 2011
Cape Colured	CAC	South Africa	Indoeuropean	19.037475	-33.495598	20	**45**	**0**							Schlebusch 2010
Khomani	KHO	South Africa	Tuu	20.872192	-26.971038	57	**98.2**	**0**							Schlebusch 2010
Manyanga	MAN	DRC	Bantu	14.058837	-4.82826	14	**0**	**0**							Schlebusch2010
Northern Cape Coloured	NCC	South Africa	Indoeuropean	20.804443	-28.149503	40	**92.5**	**0**							Schlebusch 2010
Sotho Tswana	STW	South Africa	Bantu	27.572021	-24.926295	22	**22.7**	**0**							Schlebusch 2010
Zulu Xhosa	ZXH	South Africa	Bantu	30.384521	-30.448674	36	**44.4**	**2.8**							Schlebusch 2010
Burunge	BUR	Tanzania	Cushitic	36.119384	-5.090944	38	**3**	**0**							Tishkoff et al. 2007
Hadza	HAD	Tanzania	Khoisan (isolated)	34.603271	-3.403758	79	**0**	**0**							Tishkoff et al. 2007
Sandawe	SAN	Tanzania	Khoisan (isolated)	35.306396	-5.594118	82	**5**	**0**							Tishkoff et al. 2007

[a] this population used to speak a Tuu language but has shifted to Afrikaans.

Table S2: List of positions (numbered in accordance with the RSRS/rCRS) with missing data that were excluded from the analysis. Polymorphic sites are underlined.

316
1243
3106
3492
3516
3981
4232
5515
5936
6716
6938
7412
8563
10550
10589
11854
13020
13198
13386
14770
15530
15930
15941
16069
16093
16169
16212
16215
16230
16242
16243
16474

Table S3: Notes on some of the haplogroup-defining mutations.	
Mutation	**Remarks**
C152T	This mutation defines L0d3b, but is also present in an individual belonging to L0d3a as well as being found sporadically in other branches of L0d and L0k.
A188G	In Supplementary Figure 2, this is shown only for L0d2d; however, this mutation also occurs in nearly all the individuals belonging to L0d1b1, with only 2 exceptions.
C198T	In Supplementary Figure 2, this is shown for L0d2a1 and L0k1a; however, this mutation also defines a minor subbranch of L0d1c1 (rather than defining L0d1c1 as a whole, as previously thought).
199	The evolutionary pathway involving L0k cannot be resolved, since L0k2 and L0k1b carry a C at this position, while L0k1a carries a T, which is the state reconstructed for the RSRS. In Supplementary Figure 2 we show the C199T back mutation as defining L0k1a; however, with our dataset it is equally likely that two independent T-C transitions occurred on the branches leading to L0k2 and L0k1b, with L0k1a retaining the ancestral T. In addition, L0d1a carries a C at this position with the exception of three lineages not forming a clade. One of these is a deeply divergent lineage represented by only one individual from Botswana (indicated by an asterisk in Supplementary Figure 2); thus, one could postulate either three back mutations from the mutation defining L0d1a as a whole, or consider T199C a defining mutation only for the subclade L0d1a1, with two back mutations having occurred subsequently. Since C16266a is also missing in the divergent lineage (see below), one should perhaps consider both T199C and C16266a as mutations defining the subclade L0d1a1, with subsequent back mutations (C199T) or novel mutations (A16266G) in some individuals.
294	In Supplementary Figure 2, we show the T-A transversion defining L0d2c; in addition, a T-C transition defines a subbranch of L0d1c1.
A7828G	Rather than defining branch L0d1c1 as a whole, as previously suggested, this is missing from one individual and thus defines only a subbranch, as shown in Supplementary Figure 2.
C8922T	This is found in L0k2, L0k1b, and several branches of L0k1a, but is missing from one subbranch of L0k1a. The most plausible reconstruction is that the transition occurred on the branch leading to L0k, as previously assumed,
G8994A	In Supplementary Figure 2, this is shown only for L0d2b1 and L0k; however, this mutation also defines a small subbranch of the paraphyletic branch of L0d1c.
A9136G	This mutation defining L0k mutates back to A in a subbranch of L0k2.
A9347G	This mutation is at the root of haplogroup L0, but almost all of the L0k2 individuals present a back mutation at this site, with the exception of the sample from Yemen.

G9438A	Rather than defining branch L0d1c1 as a whole, as previously suggested, this is missing from one individual and thus defines only a subbranch, with a further back mutation to A9438G found in one sequence.
A11653G	This mutation defines L0k as well as L0d3a.
C15550T	Together with C16242T, this is the only mutation defining branch L0d1c1; A7828G and G9438A are missing from one divergent lineage and thus define only a subset of L0d1c1 (with a further back mutation to A9438G found in one sequence), while C198T defines an even smaller branch within L0d1c1.
T15586C	This mutation defining L0d3a mutates back to T in one individual of the same subbranch.
A16129G	This mutation defines L0d1c, L0k and L0d3b, as well as a subbranch of L0d1b2a. Given the hypervariability of this position, it is not surprising that several back mutations to A occur in the tree – the most notable being a back mutation in the individual from Yemen whose sequence up to now was the only lineage known for L0k2. Therefore, A16129G was previously considered a mutation defining only L0k1; with our extended dataset we show that it defines all of L0k.
T16209C	This mutation, which defines L0k1, also appears in a subbranch of L0d1a.
C16242T	Together with C15550T, this is the only mutation defining branch L0d1c1; A7828G and G9438A are missing from one divergent lineage and thus define only a subset of L0d1c1 (with a further back mutation to A9438G found in one sequence), while C198T defines an even smaller branch within L0d1c1.
16266	Like the T-C transition at 199, C16266a is not found in all the sequences belonging to L0d1a; rather, four sequences carry a G at this position. Since one of these is the divergent lineage represented by an asterisk in Supplementary Figure 2 (as mentioned for position 199 above), one should perhaps consider both T199C and C16266a as mutations defining the subclade L0d1a1, with subsequent back mutations (C199T) or novel mutations (A16266G) in some individuals.
16291	While a C-T transition defines branch L0d2b1, it also defines a subbranch of the paraphyletic sister clade of L0d1c2. Furthermore, L0k1 is defined by a G at this position, with a subsequent G to A transition on a subbranch of L0k1a; L0k2 carries an A at this position. While Phylotree (http://www.phylotree.org/tree/subtree_L.htm, Build 15) reconstructs a C-G transversion for L0k as a whole and a G-A transition for L0k2, from the data available to us it appears impossible to decide whether a C-G or C-A transversion took place on the branch leading to L0k. Therefore, in Supplementary Figure 2 the mutations defining L0k1 and L0k2 are both listed as transversions, even though the actual evolutionary path would have involved just one transversion (on the branch leading to L0k) and one transition (on either L0k1 or L0k2).
16294	While a C-T transition defines branch L0d1b2, and a C-A transversion defines branch L0d1c2a, the paraphyletic sister branch of L0d1c2a is defined by a G at this position, with the exception of one individual who carries an A.
A16300G	This mutation, which defines L0d3, mutates back to A in two individuals of branch L0d3b

The mutations are numbered in accordance with the RSRS/rCRS sequence.

Supplemental References:

Barbieri, C., Butthof, A., Bostoen, K., and Pakendorf, B. (2012). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. Eur J Hum Genet. doi: 10.1038/ejhg.2012.192. Aug 29. [Epub ahead of print]

Castrì, L., Tofanelli, S., Garagnani, P., Bini, C., Fosella, X., Pelotti, S., Paoli, G., Pettener, D., and Luiselli, D. (2009). mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. Am J Phys Anthropol 140, 302-311.

de Filippo, C., Heyn, P., Barham, L., Stoneking, M., and Pakendorf, B. (2010). Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. Am J Phys Anthropol 141, 382-394.

Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., van Helden, P.D., Hoal, E.G., and Behar, D.M. (2010). Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture. Am J Hum Genet 86, 611-620.

Schlebusch, C.M., de Jongh, M., and Soodyall, H. (2011). Different contributions of ancient mitochondrial and Y-chromosomal lineages in 'Karretjie people' of the Great Karoo in South Africa. J Hum Genet 56, 623-630.

Schlebusch, C.M. (2010). Genetic variation in Khoisan-speaking populations from southern Africa. PhD thesis, University of the Witwatersrand, Johannesburg.

Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., et al. (2007). History of click-speaking Populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24, 2180-2195