

ORIGINAL ARTICLE

Different contributions of ancient mitochondrial and Y-chromosomal lineages in ‘Karretjie people’ of the Great Karoo in South Africa

Carina M Schlebusch¹, Michael de Jongh² and Himla Soodyall¹

The Karretjie people of the South African Great Karoo are itinerants who subsist by sheep shearing. Although officially classified ‘Coloured’, they are aware of their Khoen and San roots. To investigate the maternal and paternal ancestries of the Karretjie people we analyzed their mitochondrial and Y-chromosome DNA variation. Their genetic ancestry was compared with a neighboring group of ‘Coloured’ individuals. We found that the mitochondrial DNA (mtDNA) haplogroup L0d was present in all the Karretjie people examined, suggesting a maternal contribution, exclusively from the Khoen and San, whereas the paternal ancestry in males was more heterogeneous. The Coloured sample, on the other hand, were found to have a lower frequency of L0d (64.5%), but did harbor other African (27.6%) and non-African (7.9%) mtDNA haplogroups. Similar to the Karretjie people, the Y-chromosome lineages identified in the Coloured group had heterogeneous origins. This study also enabled an assessment of mtDNA variation within L0d sub-haplogroups. All seven of the L0d sub-clades were identified in the combined sample and were used to construct an L0d network.

Journal of Human Genetics (2011) 56, 623–630; doi:10.1038/jhg.2011.71; published online 21 July 2011

Keywords: Karretjie people; Khoen; Khoen–San; L0d; mitochondrial DNA haplogroup; San; Y-chromosome haplogroup

INTRODUCTION

By the time early explorers and subsequently eighteenth-century colonial hunters and farmers advanced into the Great Karoo of South Africa, the pastoralist Khoen and foraging /Xam (San) had been roaming and using the region for centuries. Competition for resources ensued, and in the resulting conflict, the /Xam particularly, were unable to resist the better-armed and mounted colonists. The males were hunted down or they fled into the hills. Many of the captured /Xam, especially women and children, were taken back to the farms of members of the colonists’ punitive commandos. By the late 1700s, most colonist farms in the area had a number of so-called ‘tame Bushmen’.¹

Although these farm /Xam tended to come and go, they proved to be excellent herders and also acquired the various skills of a sheep-farming enterprise—including shearing. The advent of wire fencing, the growth in demand for wool and hence more intensive farming, eventually saw colonist farmers developing larger flocks of wool-bearing sheep and this necessitated laborers—as shearers—to be available in numbers at a particular time in the agricultural cycle—the shearing season. The farm /Xam and their descendants adapted to this need of a floating pool of sheep shearers by moving from farm to farm, first on foot with their possessions on pack animals, but eventually by means of donkey-drawn carts fashioned from the

parts of derelict horse carriages and motorized vehicles.² Thus, came into use the term Karretjiemense or Karretjie people—a self-appellation with the Afrikaans word Karretjie, meaning donkey cart.³ The Karretjie people with their mobile lifestyle became a common sight in the Great Karoo region (an arid scrubland with dispersed hills that stretch over an area of 400 000 km² of the Northern, Eastern and Western Cape Provinces of South Africa). Typically, the Karretjie people have a home base or as they call it ‘uitspanning’ or ‘stanning’ where they erect their temporary overnight shelters in between shearing assignments. These ‘stannings’ are usually on a neutral piece of land such as the section of land between a road and a farm fence. They will then stay there until a farmer needs a shearing team or a temporary job to be done. The donkey cart will then be packed and the whole family and living unit will move to the farm until the work is done, whereupon they move back to the original ‘stanning’ again.

Changes in the agricultural economy of the region, and competition from unionized shearing teams, have increasingly contributed to the trend toward sedentarism among the Karretjie people. They were arbitrarily and incorrectly categorized as ‘Coloured’ under the previous political dispensation and were sociopolitically invisible to the government. The present South African democracy has only confirmed the realities of their disempowerment and they are still restricted to their temporary shelters on the verges of roads or in informal settlements

¹Human Genomic Diversity and Disease Research Unit, Division of Human Genetics, School of Pathology, University of the Witwatersrand/National Health Laboratory Services, Johannesburg, South Africa and ²Department of Anthropology and Archaeology, University of South Africa, Pretoria, South Africa
Correspondence: Dr CM Schlebusch, Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.
E-mail: cschlebu@gmail.com

Received 21 January 2011; revised 15 May 2011; accepted 17 May 2011; published online 21 July 2011

on the fringes of Great Karoo towns. Recognition of them locally often only comes in pejorative terms.¹

In trying to understand the genetic ancestry of the Karretjie people, we have taken advantage of the properties of mitochondrial DNA (mtDNA) and Y-chromosome markers to reconstruct their maternal and paternal ancestries, respectively. As mtDNA is transmitted maternally in the absence of recombination, the mtDNA lineages found in the Karretjie people should elucidate the maternal ancestors of the group. Similarly, the non-recombining region of the Y chromosome ought to highlight the paternal ancestry of the Karretjie people. We have included a group of Coloured individuals who live in close proximity to the Karretjie people for comparative analysis.

Previous studies on San populations have shown that they harbor some of the most ancient mtDNA and Y-chromosome lineages found globally. The mtDNA haplogroup L0d, dating to between 60 000 and 100 000 years^{4,5} has been found at varying frequencies among different San groups; 51% in the !Xun,⁶ 92% in the Ju'hoansi,⁷ and 16% in the Khwe.⁶ Recently, L0d was found at frequencies of 76% and 71% in a combined group of Khoe and San individuals as well as in a Coloured group, respectively.⁸ In addition, L0d is found at a frequency of 6% in the click-speaking Sandawe but not in the Hadza population from Tanzania.⁹ Haplogroup L0d is found at low frequencies in Southern-African Bantu speakers, and is a useful marker to estimate the extent of maternal gene flow from the Khoe and San.^{8,10,11} Using whole mitochondrial genome sequencing, Behar *et al.*⁵ published an updated phylogeny and nomenclature of African mtDNA haplogroups in which the L0d haplogroup was further resolved into L0d1, L0d2, L0d3 and their associated sub-haplogroups.

The oldest Y-chromosome haplogroup, haplogroup A, is found at high frequencies among San populations (specifically A-M14 and A-M51), and ranges in frequencies from 36 to 44% in the !Xun and Ju'hoansi to around 12% in the Khwe.^{12–18} These findings were reaffirmed by Naidoo *et al.*,¹⁹ who examined a larger group of Khoe and San individuals from various sub-groups and found haplogroup A to be present at a frequency of 44% in the combined sample. The other Y haplogroups associated with the Khoe and San are haplogroup B-M112 (ref. 19) and E-M35 (specifically the E-M293 marker within E-M35).^{19,20}

MATERIALS AND METHODS

Subjects

Sampling was conducted in and around the Colesberg region (longitude, latitude: 25.101013, –30.712638) in the Northern Cape Province of South Africa (Figure 1). The sample group included Karretjie people, who resided at three locations (stanings) within a 40 km² radius of Colesberg, and people of mixed ancestry, more commonly referred to as Coloureds in South Africa.²¹ After excluding related individuals from the sample, 107 individuals (31 Karretjie people and 76 Coloureds) were examined for mtDNA variation and 54 males for Y-chromosome analysis (19 Karretjie people and 35 Coloureds). Kinship information was verified by personal communication with Riana Steyn (Department of Anthropology and Archaeology, UNISA).

This research was approved by the Human Research Ethics Committee at the University of the Witwatersrand (Protocol numbers M050902 (CMS) and M980553 (HS)). All samples were collected with the participants' informed and written consent. When individuals could not sign their name, they marked the signature section with a cross and provided a thumbprint in the presence of a witness.

DNA extraction

DNA was extracted from 10 ml EDTA–blood using the salting-out method.²² DNA samples were quantified using a NanoDrop ND-1000 Spectrophotometer (Coleman Technologies, Newtown Square, PA, USA, LabVIEW) and diluted to 10 ng ml⁻¹ using double distilled water.

MtDNA analysis

MtDNA types were assigned to macro-haplogroups L0–L6, M, N and R using the minisequencing technique.⁸ The mtDNA control region was amplified and sequenced according to previously published methods²³ with some modification.⁸ Sequence data were obtained for hypervariable segments I (HVS I; nucleotide positions 16 024–16 400) and HVS II (nucleotide positions 57–302). The base calls at positions 16 184–16 193 were ascertained by manual inspection of electropherograms and where there was any uncertainty about a mutation, the sequence was repeated with the reverse primer. Insertions or deletions were treated as a fifth state. When there was uncertainty whether a mutation in the 16 184–16 193 region was the result of a substitution or an indel, the mutation was always treated as a substitution. All sequences were compared with the revised Cambridge reference sequence²⁴ to establish the variant positions, and were assigned to haplogroups according to the proposed nomenclature.⁵ Haplotypes generated for the combined HVS I and II sequences were used in further analyses. The sequences obtained from the present study were

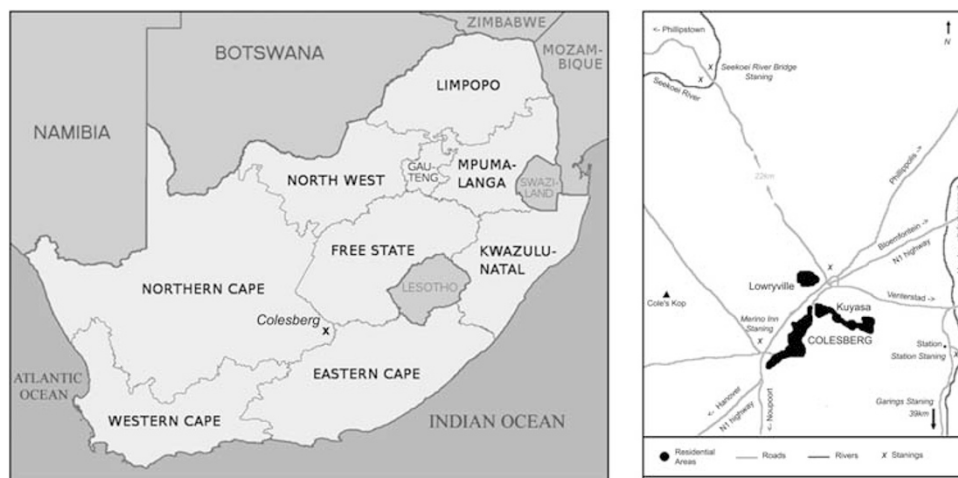


Figure 1 Maps illustrating the location of Colesberg in South Africa as well as the location of the different 'stanings' where sample collection were done in relation to the town. Samples were collected at the Merino Inn staning, the Garings staning and the Lowryville staning (some individuals from the Station staning and the Seekoei River Bridge staning were previously relocated to the Lowryville staning).

submitted to Genbank (accession numbers HQ173994—HQ174100 for HVS I and HQ174101—HQ174207 for HVS II sequences).

Y-chromosome analysis

A total of 23 single nucleotide polymorphisms were typed using restriction fragment length polymorphisms in a hierarchical manner to resolve the Y-chromosome haplogroups in the sample of 54 males (Supplement 1). In addition, 11 Y-chromosome short tandem repeats were typed using the PowerPlex Y System (Promega, Madison, WI, USA) according to kit instructions.

Computational analyses

Mitochondrial HVS sequences were aligned using the Clustal W algorithm²⁵ implemented in BioEdit v.7.0.5.3.²⁶ Gene diversity,²⁷ nucleotide diversity²⁷ and average number of nucleotide pairwise differences²⁸ were calculated using DnaSP v4.10.²⁹ Selective neutrality tests of Tajima (D)³⁰ and Fu (Fs)³¹ were also determined using the algorithms implemented in DnaSP. The genetic differentiation between the Karretjie people and Coloured group was evaluated by calculating Φ_{st} ,³² F_{st} based on pairwise differences and an exact test of population differentiation³³ using Arlequin v3.0b.³⁴

A network showing the relatedness between HVS sequences was constructed using Network v4.5.0.0,³⁵ based on the median-joining algorithm.³⁶ Networks were subjected to maximum parsimony post-analysis using the Steiner maximum parsimony algorithm³⁷ within Network v4.5.0.0. A phylogenetic tree was constructed using Bayesian likelihood analysis with the assistance of MrBayes v3.1.³⁸ The general time-reversible substitution model with Gamma-distributed rates and Invariable sites was used as it produced the best likelihood prediction using MrModeltest 2.2.^{39,40}

Time to most recent common ancestor of the L0d haplogroup and the L0d sub-haplogroups was calculated from the median-joining network using the Rho statistic.⁴¹ Time to most recent common ancestor was also estimated using a coalescence approach by means of the Markov Chain Monte Carlo sampling algorithm, as implemented in BEAST v.1.4.8.⁴² The general time-reversible substitution model with estimated base frequencies and a Gamma plus Invariant Sites heterogeneity model was used to infer the ancestral gene trees for each haplogroup. The mean substitution rate was fixed to the rate of Ward *et al.*,⁴³ and a relaxed molecular clock (uncorrelated lognormal) was used. The Bayesian Skyline coalescent was used as tree prior. The population size function can be implemented using either a piecewise constant or a piecewise linear function of population size change. In the present study, a piecewise linear model made up of 10 control points was used. Each Markov Chain Monte Carlo sampling was repeated for 40 000 000 generations, sampled every 4000, with the first 4 000 000 generations discarded as burn-in. All runs had an effective sample size of at least 1000 for the parameters of interest. Each independent run was repeated at least twice and results were combined using the LogCombiner v1.4.8 tool included in the BEAST package. Monophyletic clades were imposed in BEAST according to clades defined by network and Bayesian phylogenetic analysis. As L0d1c sequences were not monophyletic in the Bayesian-based phylogenetic tree but in the network analysis, these sequences were grouped according to the network analysis result.

A Neanderthal HVS sequence (Genbank accession numbers: AF011222 and AF142095) and the control region reference sequence²⁴ were included in phylogenetic and network analyses. Fifty additional published L0d HVS sequences were also included for comparative analysis (30 HVS sequences from Behar *et al.*,⁵ 20 HVS sequences from Gonder *et al.*⁴⁴ and Tishkoff *et al.*⁹ after removal of the common sequences in these studies). The two sequences classified as isolate TZSW084 (Sandawe) in the Genbank files (accession numbers: EF184586.1 in Gonder *et al.*⁴⁴ and EF999659.1 in Tishkoff *et al.*⁹) do not have the same L0d sequence when compared (in Gonder *et al.*,⁴⁴ it is an L0d1b sequence and in Tishkoff *et al.*,⁹ it is a L0d3 sequence). The Genbank file for isolate TZSW084 in Gonder *et al.*⁴⁴ seems to contain a duplicate sequence of isolate SanC5 (San) in Gonder *et al.*⁴⁴ as these two whole genome sequences are identical. Therefore, SanC5 (L0d1b) and TZSW084 (L0d3) from Tishkoff *et al.*⁹ were included in the present analysis and TZSW084 (L0d1b) from Gonder *et al.*⁴⁴ was excluded. This is an important note as a subsequent publication⁴⁵ used this sequence as a Tanzanian sequence, which lead to a tree in which this sequence represents the only Tanzanian individual outside the

L0d3 clade (it is shown to group with San individuals and not with other Tanzanian individuals in the L0d3 clade in Supplementary Figure 1A from Schuster *et al.*⁴⁵). Problems in the Gonder *et al.*⁴⁴ Genbank submissions were mentioned previously;⁴⁶ however, the misclassification of this particular sequence was not addressed.

The genetic affinities of the Karretjie people and the Coloured group from Colesberg were compared with other sub-Saharan African populations using some previously published HVS sequences.^{9,47} These included two Bantu-speaking groups from Zambia (Bisa and Kundu),⁴⁷ two Bantu-speaking groups from Tanzania, (Turu and Sukuma); a Pygmy group from Cameroon, (Bakola); an Afro-Asiatic (Burunge) and a Nilo-Saharan speaking group (Datoga) from Tanzania; and two San groups from Southern Africa (!Xun and Khwe).⁹ Although the !Xun and Khwe were sampled in South Africa, they originated in Angola⁴⁸ and belong to the northern San linguistic group (Ju) and central San linguistic group (Khoe–Kalahari branch), respectively.⁴⁹ Sequence data from HVS regions (16 032–16 391 for HVS I and 57–302 for HVS II) were in the comparative analysis. Seven sequences from the published data were excluded to minimize missing data (one San—EF999719; three Bisa—GQ131434, GQ131445, GQ131497; three Kundu GQ131484, GQ131502, GQ131503). A Φ_{st} ³² distance matrix of the pairwise comparisons between the different populations was created using Arlequin v3.0b.³⁴ and visualized as a UPGMA tree in MEGA v.4.⁵⁰

RESULTS

MtDNA and the maternal ancestry of the Karretjie people

We derived 56 unique mitochondrial HVS sequences in the combined sample of 107 individuals (Supplement 2). Haplogroup L0d was found in all the Karretjie people tested and was also the most common haplogroup (64.5%) found in the Coloured sample (Table 1). Of the remaining haplogroups found in the Coloured group, 27.6% were traced to African origins and 7.9% to non-African origins (Table 1). All seven of the L0d sub-groups (L0d1a, L0d1b, L0d1c, L0d2a, L0d2b, L0d2c and L0d3) were found in the combined sample (Supplement 2). The number of individuals belonging to the L0d haplogroup in the Karretjie people was significantly higher than in the Coloured group (Fishers exact $P < 0.0001$).

A complete list of the 77 haplotypes (56 from this study) used in phylogenetic and network analyses are given in Supplement 2. There was good concordance between the placement of the different mtDNA sequences in the network (Figure 2), and their haplogroup and sub-haplogroup assignment. MtDNA haplogroup L0d was most closely associated with the Neanderthal sequence in the network followed by L0a (Figure 2).

The L0d clade on the network consisted of 54 unique haplotypes; 36 were from the present study and the remaining 18 from published data (Supplement 2). With the exception of haplogroup L0d3 (which harbored mtDNA types found in the Sandawe from Tanzania^{9,44} and the sample from Kuwait⁵), the remaining L0d sub-haplogroups were from individuals who were sampled in Southern Africa. Haplogroup L0d2a harbored the most number of haplotypes (12) and displays a star phylogeny, which is consistent with a recent and rapid expansion of mtDNA types. This is supported by a significantly negative F_s statistic ($F_s = -4.670$, $P < 0.000$) for this sub-haplogroup.

A Bayesian-based phylogenetic tree (rooted with Neanderthal) showed high clade support (> 0.80) for five of the seven L0d sub-haplogroups (Supplement 3). L0d1c did not form a monophyletic clade; this was probably due to the hypermutable 16 129 mutation also observed in the network analysis (Figure 2).

The ages of the mtDNA L0d sub-haplogroups (Table 2) were estimated using the Rho statistic,⁴¹ with the associated s.d., σ^{51} and a mutation rate of 2.5×10^{-6} per nucleotide per generation.⁴³ L0d was dated to 88 000 ($\pm 18 000$) years before present; this value was older than previous estimates (49 600–62 000 years before present),^{4,11} but

Table 1 Haplogroup frequency distributions and summary statistics in the Karretjie people and their Coloured neighbors

Haplogroup	Coloured (N=76)	Karretjie people (N=31)
<i>L0d</i>	49 (64.5%)	31 (100%)
L0d1a	10 (13.2%)	2 (6.5%)
L0d1b	11 (14.5%)	6 (19.4%)
L0d1c	1 (1.3%)	—
L0d2a	16 (21.1%)	18 (58.1%)
L0d2b	2 (2.6%)	—
L0d2c	2 (2.6%)	—
L0d3	7 (9.2%)	5 (16.1%)
<i>L0a</i>	6 (7.9%)	—
L0a1b	1 (1.3%)	—
L0a1b1	2 (2.6%)	—
L0a2	2 (2.6%)	—
L0a2a2	1 (1.3%)	—
<i>L1b</i>	1 (1.3%)	—
L1b	1 (1.3%)	—
<i>L2a</i>	10 (13.2%)	—
L2a1	10 (13.2%)	—
<i>L2b</i>	2 (2.6%)	—
L2b	1 (1.3%)	—
L2b2	1 (1.3%)	—
<i>L3d</i>	1 (1.3%)	—
L3d1a	1 (1.3%)	—
<i>L3e</i>	1 (1.3%)	—
L3e1g	1 (1.3%)	—
M-derived	2 (2.6%)	—
N-derived	4 (5.3%)	—
Gene diversity	0.966	0.757
Nucleotide diversity π	0.021	0.010
MPSD	12.993	5.963
Tajima's D (<i>P</i> -value)	-1.090 (0.125)	-0.740 (0.255)
Fu's F_s (<i>P</i> -value)	-15.867 (0.002)	-0.157 (0.514)

comparable to other studies (90 400–100 800 years before present^{5,44}). The coalescent approach to estimate time to most recent common ancestor yielded older dates for all haplogroups (Table 2).

Nucleotide diversity and average number of nucleotide pairwise differences in the Coloured population was more than twice that of the Karretjie people (Table 1), in keeping with a more heterogeneous origin of mtDNA types in the Coloured group. About 65% of the maternal ancestors of the Coloured group (placed in the different L0d sub-haplogroups in the network) can be traced back to the Khoe and San, and six haplotypes are common to both the Karretjie people and the Coloured group (Figure 2). The more admixed nature of the Coloured group compared with the Karretjie people is apparent in the summary statistics and the haplogroup frequency distributions (Table 1). Interpopulation differences, as suggested by the low Φ_{st} (0.069) and F_{st} (0.062) measures, however, do not suggest significant difference in the maternal contribution to the two populations.

Also, the exact test for population differentiation between the Karretjie people and Coloured group was consistent with little population structure (Exact *P*-value=0.596, s.d.=0.063).

The Φ_{st} distance matrix between sub-Saharan African groups (Supplement 4) shows that the Coloured group and the Karretjie people are closely related to !Xun and Khwe San groups. This relationship is supported further in the UPGMA tree (Figure 3), in which the Coloured, Karretjie people and !Xun/Khwe groups form a monophyletic group. In this tree, the other click-speaking groups from Tanzania (Sandawe and Hadza) are found on different branches; this can be attributed to the differences in frequencies of haplogroup L0d and its associated sub-haplogroups in the Tanzanian and Southern-African click speakers.

Tajima's (*D*) and Fu's (F_s) tests for neutrality yielded negative values for both the Coloured group as well as in the Karretjie people (Table 1), but the F_s statistic was only significant in the Coloured population. The F_s statistic, a good predictor of population expansions,^{52,53} was lower in the Coloured population when compared with the Karretjie people, suggesting a stronger signal for growth and/or positive selection in the Coloured population (Table 1). To overcome the effect of sample size biases between the Karretjie people (*N*=31) and the Coloured group (*N*=76), we performed random sub-sampling with replacement in the Coloured group (100 times for 31 individuals and tested the F_s statistic for the 100 data sets of 31 Coloured individuals). In this re-analysis, F_s had an average value of -2.360 (s.d.=1.724), which is significantly lower than the -0.157 we previously obtained for the Karretjie people (95% confidence interval=0.342).

Y-chromosome DNA variation

The sample sizes for the Karretjie people (*N*=19) and the Coloured group (*N*=35) were small, but the combined group of 54 males harbored Y chromosomes that were assigned to 12 different Y-chromosome haplogroups delineated into 48 Y-short tandem repeat haplotypes (Figure 4, Supplement 5). Gene diversity estimated from short tandem repeat data was higher in the Coloured group compared with the Karretjie people (Figure 4). In contrast to the mtDNA data, the Y-chromosome haplogroups in the Karretjie people were derived from a heterogeneous background (Figure 4; Supplement 5). Haplogroup E, the commonest Y-chromosome haplogroup found in sub-Saharan Africa, was found at a frequency of 63.2% in the Karretjie people and 60% in the Coloured group (Figure 4). Haplogroups E-M75, together with E-M191 and E-M2 (58% in the Karretjie people; 54.2% in the Coloured group), has a wide geographic distribution in Africa, and is suggested to have spread as a result of migrations of people speaking Bantu languages.^{54–56} Haplogroup E-M35 (present in ~5% of the Karretjie people and 5.7% in the Coloured group) is a widely distributed haplogroup that occurs throughout Africa but also has a circum Mediterranean distribution. However, a recent study linked a sub-marker in the E-M35 clade (E-M293) with the introduction of pastoralism to the southern parts of Africa.²⁰ E-M293 was not used in the present study and individuals who possessed the M35 mutation were as assigned to haplogroup E-M35*. Henn *et al.*²⁰ found that the DYS389I-10 short tandem repeat allele was associated with the E-M293 marker in the San populations they examined. In the present study, two individuals from the Coloured population had E-M35* DYS389I-10 Y chromosomes (Supplement 5), suggesting San ancestry, but the individual from the Karretjie people group who had the E-M35* Y chromosome harbored the DYS389I-13 allele and had identical matches to nine east Europeans in the YHRD database (<http://www.yhrd.org>). It seems more likely that the latter individual has a E-M35 Y chromosome of Eurasian origin.

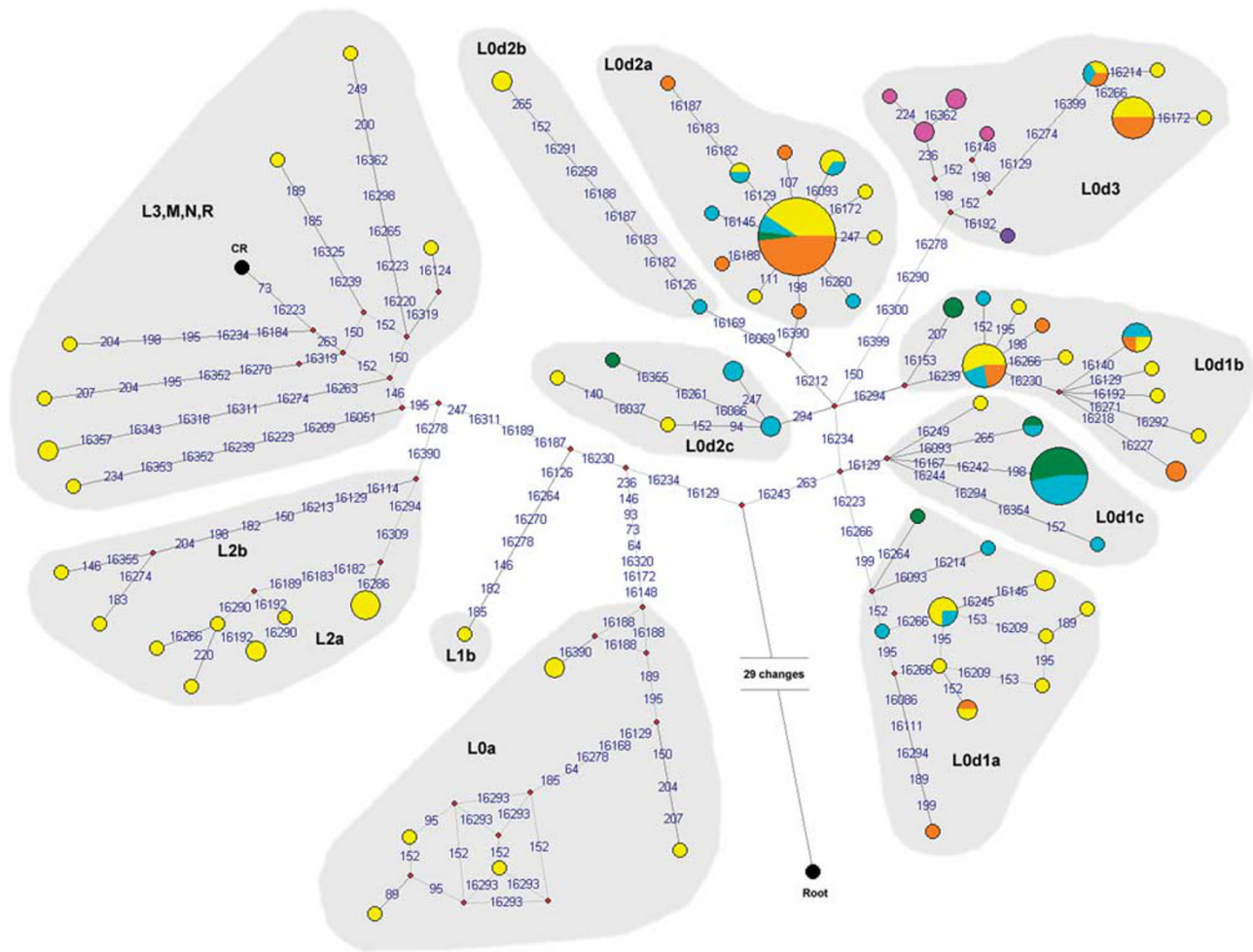


Figure 2 Median-joining network showing Karretjie people (orange) and Coloured group (yellow). Comparative published LOd sequences are included and indicated by different colours: Pink—Tanzanian Sandawe,^{9,44} Green—Southern-African San (IXun and Khwe),^{9,44} Turquoise—Various southern African LOd sequences from Bantu speakers, Khoe and San;⁵ Purple—One LOd sequence from Kuwait.⁵ Network is rooted with Neanderthal. CR indicates the control region reference sequence.

Table 2 Estimated haplogroup coalescent times

Haplogroup	Number of haplotypes				TMRCA using ρ^c		
	Total ^a	Coloured	Karretjie people	Other publications ^b	$\rho \pm \sigma^c$	Age in years ^d	TMRCA using coalescence ^e
LO	59	33	13	25	7.85 ± 1.7	126 000 ± 28 000	
LOd	54	28	13	25	5.48 ± 1.1	88 000 ± 18 000	130 000 ± 30 000
LOd1a	11	7	2	4	3.25 ± 1.2	52 000 ± 20 000	58 000 ± 18 000
LOd1b	11	7	4	4	2.13 ± 1.0	34 000 ± 16 000	54 000 ± 19 000
LOd1c	4	1	—	3	2.84 ± 1.4	46 000 ± 22 000	63 000 ± 22 000
LOd2a	12	6	5	5	1.37 ± 1.0	22 000 ± 16 000	37 000 ± 14 000
LOd2c	5	2	—	3	1.57 ± 0.6	25 000 ± 9 000	41 000 ± 18 000
LOd3	9	4	2	6	4.00 ± 1.4	64 000 ± 23 000	61 000 ± 19 000
All	76	48	13	25	8.80 ± 1.5	141 000 ± 25 000	

Abbreviation: TMRCA, time to most recent common ancestor.

^aTotal haplotypes according to Supplement 2. Haplotypes in three columns (Coloured, Karretjie and other) do not add up to 'Total' because of haplotypes that the three groups have in common.

^bOther publications.^{5,9,44}

^c ρ and σ as defined earlier.^{41,51}

^dUsing mutation rate of 2.5×10^{-6} per nucleotide per generation⁴³ (25 years per generation; 623 nucleotides).

^eEstimated using BEAST.⁴²

Haplogroup A, which was found in Southern-African Khoe and San groups, as well as in other populations with recent hunter-gatherer ancestry,^{12,16,19} occurred at a frequency of 10.5% in the

Karretjie people and 2.9% in the Coloured group. Haplogroup B-M150, more commonly found in Bantu speakers, was found at a frequency of 5% in the Karretjie People and 5.7% in the Coloured

group. Haplogroup B-M112 was only found in the Coloured group (Figure 4).

The Eurasian haplogroups I, J and R were found in the Karretjie people and Coloured group at frequencies of 21.1% and 28.6%, respectively, (Figure 4). This suggests that at least one-fifth of the Y chromosomes in these groups are derived from non-African sources.

DISCUSSION

MtDNA and Y-chromosome DNA analyses used in the present study were quite robust, despite the small sample size, in revealing the maternal and paternal ancestries, respectively, of the Karretjie people. The exclusive presence of mtDNA haplogroup L0d (and its associated sub-haplogroups) in the Karretjie people is highly suggestive of Khoe and San maternal ancestry. The Y-chromosome pool in the Karretjie people was more heterogeneous with contributions from Bantu-speaking (63.2%), Eurasian (26.3%) and only small number of

San (10.5%) lineages (Figure 4). The colonial period had an adverse effect on the San; men were either persecuted or dispersed by European colonists, whereas women were forcibly relocated to farms thereby disrupting their family units. The resulting unions between displaced San women and white male colonists as well as the incoming Xhosa-speakers have left an indelible record in the paternal contribution of the extant Karretjie people. Moreover, their San identity and hunter-gatherer lifestyle were lost, and as the Karretjie people, they acquired a new sociocultural and sociopolitical identity as wandering sheep shearers. With the first democratic elections in South Africa held in 1994, Karretjie people had, for the first time, a right to be classified as citizens of South Africa. Unfortunately, following the biases of the previous system they were still arbitrarily classified as Coloured. Although they distinguish themselves from the Coloured community and refer to themselves as the 'yellow' or 'Bushmen' (i.e., San) people, as opposed to the 'brown' or Coloured people, the system is still insensitive to such self-perception.

Using mtDNA haplogroup L0d as a marker to estimate Khoe and San contribution to maternal ancestry, the Karretjie people have a more exclusive Khoe and San ancestry (100%) when compared with other San groups such as !Xun and Ju'hoansi (51–92%).^{6,7} The Coloured group also had a high L0d mtDNA pool (65%) comparable to the !Xun group (51%), sampled at either Schmidtsdrift or Platfontein in the Northern Cape Province.⁶ The relatedness between the Karretjie people, Coloured and !Xun/Khwe San was also apparent from their association in the UPGMA tree when compared with other sub-Saharan African populations (Figure 3). It is very difficult to distinguish the Khoe and San on the basis of their mtDNA haplogroups; both groups have high frequencies of L0d and varying contributions from other African and non-African sources that have contributed to their mtDNA pool.⁵⁷ It might be suggested that Coloured communities, who have a higher social ranking, are descendants of Khoe pastoralists, whereas the Karretjie people are descendants of San hunter-gatherers. However, in the network featured in Figure 2, mtDNA types present in Karretjie people and the Coloured group were distributed throughout the L0d clade and could not be resolved to San or Khoe origins. The *F_s* statistic (Table 1) is also consistent with different demographic histories of the Karretjie people and the Coloured group, with a stronger signal for demographic expansion in the Coloured group (which could be attributed to the mtDNA haplogroups other than L0d present in the latter).

The data obtained for the Coloured population in the present study is similar to that published previously on Coloured groups from the

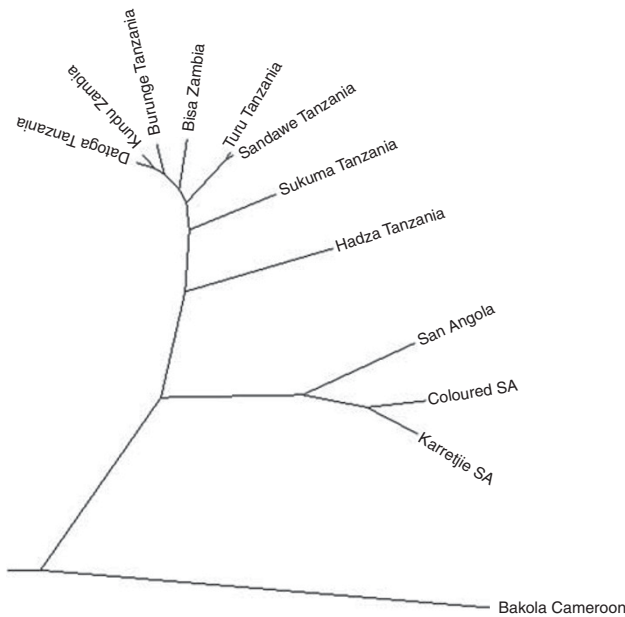


Figure 3 UPGMA tree illustrating mtDNA Φ_{st} distances between the Karretjie people and Coloured groups and published sub-Saharan African populations.^{9,47}

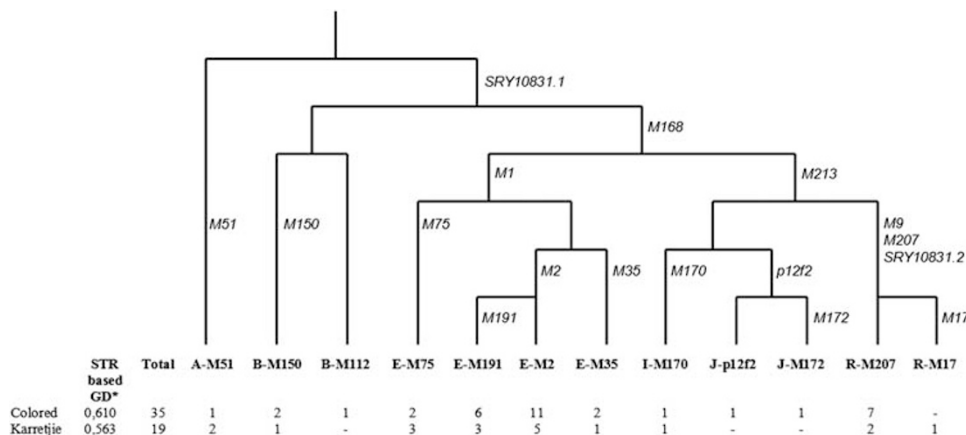


Figure 4 Y-chromosome haplogroup distribution in the Coloured and Karretjie people groups. *GD—gene diversity, based on short tandem repeat profiles.

Western Cape Province.⁵⁸ This group also had a high frequency of mtDNA haplogroup L0d (60%), which was comparable to that obtained in the sample we examined from Colesberg (64.5%). The paternal ancestry in the Western Cape sample, however, showed a much higher Eurasian contribution (55%) compared with the present study (29%). Y-chromosome data confirms that the male ancestors of Coloured populations were drawn from various sources (African, European, Indian, Malaysian and Middle Eastern) and at varying frequencies and the Y-chromosome pool among the different groups throughout the country have been shaped by history. Autosomal DNA studies in a Coloured population from the Western Cape Province reaffirmed the heterogeneity in the gene pool of this group.⁵⁹ However, although mtDNA shows a high maternal contribution from the Khoe and San, autosomal markers suggests higher genetic contributions from Bantu-speaking, Asian and European ancestry. Further genetic studies in Coloured groups of various geographic regions would be useful to establish the genetic ancestries of this diverse group of people.

This study permitted a judicious assessment of the sub-haplogroup structure within L0d. Behar *et al.*⁵ resolved L0d into seven sub-haplogroups (L0d1a, L0d1b, L0d1c, L0d2a, L0d2b, L0d2c and L0d3); our parsimony network analysis and Bayesian phylogenetic analysis also resolved the mtDNA types within L0d into seven sub-haplogroups. Of these, L0d2b, L0d2c and L0d1c contained a limited number of mtDNA types (Table 1) and were found only in the Coloured group, whereas the other four sub-haplogroup clades harbored mtDNA types found in both the Karretjie people and the Coloured group. The samples examined by Gonder *et al.*⁴⁴ and Tishkoff *et al.*⁹ were from northern San (!Xun and Khwe) and their mtDNA types were placed predominantly in the L0d1c clade with single individuals in other sub-clades. L0d2a seemed to have undergone a recent expansion in the Coloured and Karretjie people. The individuals, from the Behar *et al.*⁵ study, whose mtDNA types were placed in L0d2a included three Khoe and San, one Tswana from South Africa, one Mozambican and one Khwe.

L0d3 is the only sub-haplogroup, which has a geographic distribution outside of Southern Africa and occurs in Tanzania,^{9,44} and as far as in Kuwait.⁵ Previous studies suggested that the L0d clade could be subdivided into a Southern African and East-African subclade.^{9,44} However, the present study shows that this is only true for the L0d3 clade. Whole mtDNA genome sequence data places L0d3 as the oldest branch within the L0d clade.⁵ This is further supported by the deeper coalescence times estimated for L0d3 (64 000 ± 23 000 years before present) in the present study. In the absence of a wider geographic representation of sampling within Africa, several scenarios can be considered to explain the L0d geographic distribution pattern. One plausible scenario could be that L0d occurred over a wide geographic region that spanned the region between east Africa to the south. L0d3 had a clinal distribution across this region, whereas the other 'younger' branches of L0d, namely L0d1 and L0d2, only evolved in the ancestors of present day San. Intermediate branches of L0d3 were then lost because of drift and external factors such as the Bantu expansions, and only an eastern and southern branch of L0d3 remained. Another possibility is that there was an early split in people harboring L0d*, and that L0d3 evolved only in east Africa and L0d1 and L0d2 in the south. In this scenario, L0d3 would have been introduced into Southern Africa as a result of a subsequent migration of people from East Africa who harbored L0d3. A third possibility is that L0d originated in early inhabitants from Southern Africa, and, as some people migrated northward, L0d3 survived in their descendants while others were lost because of drift. This scenario is very plausible

given the convergence of data from other studies placing Southern Africa as the most likely geographic region for the origin of modern humans.⁶⁰

CONCLUSION

The Karretjie people are a good example of people who over the years have had to develop new identities. Despite this, their genetic ancestry, particularly their maternal ancestry, is consistent with an exclusive Khoe and/or San ancestry. The mtDNA haplogroup L0d and its sub-haplogroups is the only branch of the mtDNA tree that has survived in this group and there has been no maternal gene flow from Bantu speakers into the group. However, the Y-chromosome DNA findings indicate that both Bantu speakers and Europeans have contributed to the gene pool of the Karretjie people in the past. The Coloured group, who lives in close proximity to the Karretjie people in Colesberg, has a more heterogenous mtDNA and Y-chromosome DNA pool—an observation consistent with historical data. This study highlights the value of mtDNA and Y-chromosome DNA in gaining insights into the prehistory of populations in the absence of written records.

ACKNOWLEDGEMENTS

We are grateful to all subjects who participated in our research. This study was supported by grants awarded to HS from the South African Medical Research Council and the National Research Foundation; HS and CS, from the National Health Laboratory Service Research Trust, and CS were supported by the National Research Foundation. MDJ was supported by the KS Birch Trust, Research Committee, University of South Africa. We thank Thijessen Naidoo for providing comments on this manuscript.

- 1 De Jongh, M. No fixed abode: the poorest of the poor and elusive identities in rural South Africa. *J South Afr Stud.* **28**, 441–460 (2002).
- 2 De Jongh, M. Human agency, asymmetrical relations and sociocultural systems, itinerancy and sedentarism in the Great Karoo of South Africa. *Int J Human.* **3**, 199–208 (2005/2006).
- 3 De Jongh, M. in *Customary Strangers. New Perspectives On Peripatetic Peoples in the Middle East, Africa and Asia*, (eds. Berland J. C., Rao A.), 155–178, (Praeger, London, 2004).
- 4 Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics.* **172**, 373–387 (2006).
- 5 Behar, D. M., Villemers, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E. *et al.* The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140 (2008).
- 6 Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383 (2000).
- 7 Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. African populations and the evolution of human mitochondrial DNA. *Science.* **253**, 1503–1507 (1991).
- 8 Schlebusch, C. M., Naidoo, T. & Soodyall, H. SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa. *Electrophoresis.* **30**, 3657–3664 (2009).
- 9 Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C. *et al.* History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* **24**, 2180–2195 (2007).
- 10 Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim, A. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* **65**, 439–458 (2001).
- 11 Salas, A., Richards, M., De la Fe, T., Lareu, M. V., Sobrino, B., Sanchez-Diz, P. *et al.* The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**, 1082–1111 (2002).
- 12 Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A. *et al.* A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* **70**, 1197–1214 (2002).
- 13 Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P. *et al.* Phylogeographic analysis of haplogroup E3b (E-M215). Y chromosomes reveals multiple migratory events within and out of Africa. *Am. J. Hum. Genet.* **74**, 1014–1022 (2004).
- 14 Knight, A., Underhill, P. A., Mortensen, H. M., Zhivotovskiy, L. A., Lin, A. A., Henn, B. M. *et al.* African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.* **13**, 464–473 (2003).

- 15 Jobling, M. A. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598–612 (2003).
- 16 Scozzari, R., Cruciani, F., Santolamazza, P., Malaspina, P., Torroni, A., Sellitto, D. *et al.* Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am. J. Hum. Genet.* **65**, 829–846 (1999).
- 17 Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
- 18 Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L. & Underhill, P. A. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**, 265–268 (2002).
- 19 Naidoo, T., Schlebusch, C. M., Makkan, H., Patel, P., Mahabeer, R., Erasmus, J. C. *et al.* Development of a single base extension method to resolve Y chromosome haplogroups in sub-Saharan African populations. *Investig. Genet.* **1**, 6 (2010).
- 20 Henn, B. M., Gignoux, C., Lin, A. A., Oefner, P. J., Shen, P., Scozzari, R. *et al.* Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proc. Natl Acad. Sci. USA.* **105**, 10693–10698 (2008).
- 21 Census 2001. Census in brief. Statistics South Africa, <http://www.statssa.gov.za/census01/HTML/CInBrief/CIB2001.pdf>.
- 22 Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
- 23 Behar, D. M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D. *et al.* The Genographic Project public participation mitochondrial DNA database. *PLoS Genet.* **3**, e104 (2007).
- 24 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
- 25 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- 26 Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95–98 (1999).
- 27 Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press: New York, USA, 1987).
- 28 Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics.* **105**, 437–460 (1983).
- 29 Rozas, J., SanchezDelBarrio, J. C., Messeguer, X. & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics.* **19**, 2496–2497 (2003).
- 30 Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**, 585–595 (1989).
- 31 Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* **147**, 915–925 (1997).
- 32 Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* **131**, 479–491 (1992).
- 33 Raymond, M. & Rousset, F. An exact test for population differentiation. *Evolution. Int. J. Org. Evolution.* **49**, 1280–1283 (1995).
- 34 Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online.* **1**, 47–50 (2005).
- 35 Network v.4.5.0.0. Fluxus engineering. <http://www.fluxus-engineering.com/>.
- 36 Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
- 37 Polzin, T. & Daneschmand, S. V. On Steiner trees and minimum spanning trees in hypergraphs. *Operations Res. Lett.* **31**, 12–20 (2003).
- 38 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* **19**, 1572–1574 (2003).
- 39 Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* **14**, 817–818 (1998).
- 40 Nylander, J. A. A. *MrModeltest v2. Program Distributed By The Author*. Evolutionary Biology Centre, Uppsala University (2004).
- 41 Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59**, 935–945 (1996).
- 42 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- 43 Ward, R. H., Frazier, B. L., Dew-Jager, K. & Paabo, S. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl Acad. Sci. USA.* **88**, 8720–8724 (1991).
- 44 Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* **24**, 757–768 (2007).
- 45 Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature.* **463**, 943–947 (2010).
- 46 Yao, Y. G., Salas, A., Logan, I. & Bandelt, H. J. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet.* **85**, 929–933 (2009); author reply 933.
- 47 de Filippo, C., Heyn, P., Barham, L., Stoneking, M. & Pakendorf, B. Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia. *Am J Phys Anthropol.* **141**, 382–94 (2009).
- 48 Sharp, J. & Douglas, S. in *Miscast: Negotiating the Presence of the Bushmen* (ed. Skotnes P.) 323–329 (UCT Press, Cape Town, 1996).
- 49 Güldemann, T. A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. *South Afr Human.* **20**, 93–132 (2008).
- 50 Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
- 51 Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J. & Norby, S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726 (2000).
- 52 Ramirez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F. & Navarro, A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics.* **179**, 555–567 (2008).
- 53 Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100 (2002).
- 54 de Filippo, C., Barbieri, C., Whitten, M., Mpoloka, S. W., Gunnarsdottir, E. D., Bostoen, K. *et al.* Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol. Biol. Evol.* **28**, 1255–1269 (2011).
- 55 Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
- 56 Wood, E. T., Stover, D. A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H. *et al.* Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* **13**, 867–876 (2005).
- 57 Soodyall, H., Makkan, H., Haycock, P. & Naidoo, T. The prehistory of the Khoe and San. *South Afr Human.* **20**, 37–48 (2008).
- 58 Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C. *et al.* Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am. J. Hum. Genet.* **86**, 611–620 (2010).
- 59 Patterson, N., Petersen, D. C., van der Ross, R. E., Sudoyo, H., Glashoff, R. H., Marzuki, S. *et al.* Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* **19**, 411–419 (2010).
- 60 Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J. M., Kidd, J. M. *et al.* From the cover: feature article: hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl Acad. Sci. USA.* **108**, 5154–5162 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)