Article

Human origins in a southern African palaeo-wetland and first migrations

https://doi.org/10.1038/s41586-019-1714-1

Received: 30 October 2018

Accepted: 24 September 2019

Published online: 28 October 2019

Eva K. F. Chan[1,2], Axel Timmermann[3,4]*, Benedetta F. Baldi[1], Andy E. Moore[5], Ruth J. Lyons[1], Sun-Seon Lee[3,4], Anton M. F. Kalsbeek[1], Desiree C. Petersen[1,11], Hannes Rautenbach[6,7,12], Hagen E. A. Förtsch[8], M. S. Riana Bornman[7] & Vanessa M. Hayes[1,2,7,9,10]*

Anatomically modern humans originated in Africa around 200 thousand years ago (ka)[1–4]. Although some of the oldest skeletal remains suggest an eastern African origin[2], southern Africa is home to contemporary populations that represent the earliest branch of human genetic phylogeny[5,6]. Here we generate, to our knowledge, the largest resource for the poorly represented and deepest-rooting maternal L0 mitochondrial DNA branch (198 new mitogenomes for a total of 1,217 mitogenomes) from contemporary southern Africans and show the geographical isolation of L0d1′2, L0k and L0g KhoeSan descendants south of the Zambezi river in Africa. By establishing mitogenomic timelines, frequencies and dispersals, we show that the L0 lineage emerged within the residual Makgadikgadi–Okavango palaeo-wetland of southern Africa[7], approximately 200 ka (95% confidence interval, 240–165 ka). Genetic divergence points to a sustained 70,000-year-long existence of the L0 lineage before an out-of-homeland northeast–southwest dispersal between 130 and 110 ka. Palaeo-climate proxy and model data suggest that increased humidity opened green corridors, first to the northeast then to the southwest. Subsequent drying of the homeland corresponds to a sustained effective population size (L0k), whereas wet–dry cycles and probable adaptation to marine foraging allowed the southwestern migrants to achieve population growth (L0d1′2), as supported by extensive south-coastal archaeological evidence[8–10]. Taken together, we propose a southern African origin of anatomically modern humans with sustained homeland occupation before the first migrations of people that appear to have been driven by regional climate changes.

Southern Africa has long been considered to be one of the regions in which anatomically modern humans (AMHs) originated. Home to contemporary populations who represent the earliest human lineages, evolutionary time estimates have largely been based on mitochondrial DNA (mitogenomes)[1,6]. The maternal human phylogenetic tree consists of two major branches, the extensive L1′6—which includes the out-of-Africa ancestral L3 sub-branch (or haplogroup)—and the rare deep-rooting L0. The L0 lineage is predominated by southern African haplogroups: L0d, L0k and the recently described L0g[6]. By contrast, the rare L0f and common L0a lineages are dispersed throughout sub-Saharan Africa[1,3,6]. Through L0 pre-screening, we identified 198 southern Africans with poorly represented haplogroups for whom the mitogenome was sequenced (Supplementary Table 1), allowing for a combined analysis of 1,217 mitogenomes (Fig. 1a and Extended Data Table 1).

We ethno-linguistically classified study participants as KhoeSan—southern African populations who traditionally practiced foraging and spoke languages containing 'click' consonants—or non-KhoeSan individuals. Non-KhoeSan who have KhoeSan-derived L0 mitogenomes are referred to in this study as KhoeSan ancestral, with further geographical classification (Fig. 1b and Extended Data Table 2; terminology pertaining to southern African KhoeSan populations is complex and contentious, see Methods for further discussion). Contemporary KhoeSan include Kalahari KhoeSan (Kx'a, Tuu and central Khoe–Kwadi speakers) and west-coastal KhoeSan (Khoe–Kwadi Nama speakers)[11]. Peoples who speak Southern Bantu languages, who migrated down the east coast of Africa around 1,500 years ago, may have acquired an east-coastal KhoeSan heritage[12]. The arrival of European colonists to the Cape in mid-1600s gave rise to the South African Coloured and Namibian Baster populations (of Eurasian and indigenous descent), who acquired a Cape KhoeSan heritage[13]. Excluding the east African Sandawe and Hadza (whose languages also contain click consonants), indigenous KhoeSan populations appear to be absent northeast of the Zambezi river, supported by the lack of skeletal remains representing the KhoeSan-like hunter–forager morphology[14]. We classified the 198 new

[1]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. [2]St Vincent's Clinical School, University of New South Wales, Sydney, New South Wales, Australia. [3]Center for Climate Physics, Institute for Basic Science, Busan, South Korea. [4]Pusan National University, Busan, South Korea. [5]Department of Geology, Rhodes University, Grahamstown, South Africa. [6]Climate Change and Variability, South African Weather Service, Pretoria, South Africa. [7]School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa. [8]Windhoek Central Hospital, Windhoek, Namibia. [9]Faculty of Health Sciences, University of Limpopo, Sovenga, South Africa. [10]Central Clinical School, University of Sydney, Sydney, New South Wales, Australia. [11]Present address: The Centre for Proteomic and Genomic Research, Cape Town, South Africa. [12]Present address: Akademia, Johannesburg, South Africa. *e-mail: timmermann@pusan.ac.kr; v.hayes@garvan.org.au

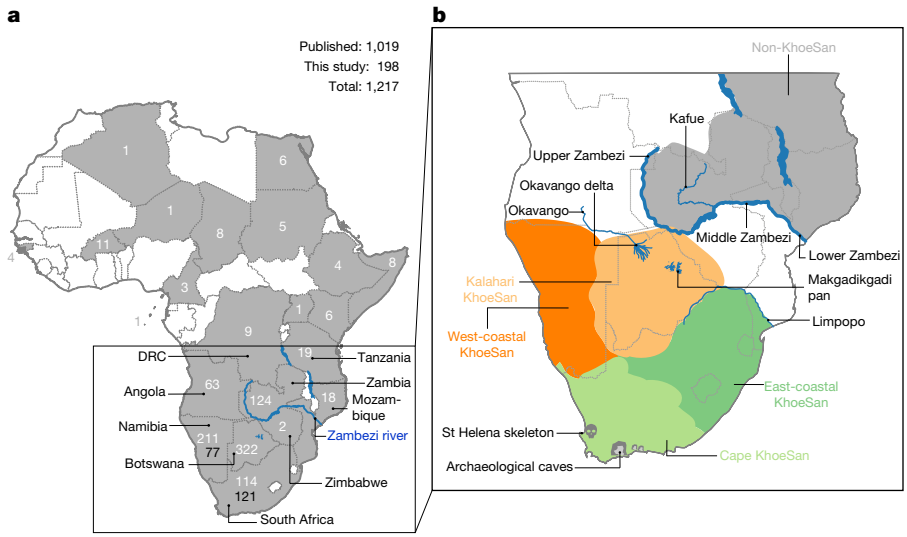


Fig. 1 | Geographical distribution of 1,217 L0 mitogenomes. a, Countries within (n = 1,139) or outside (n = 78) Africa from which L0 mitogenomes were sourced, including 198 new L0 mitogenomes (black numbers on the map). DRC, Democratic Republic of the Congo. **b**, Present-day southern Africa showing the geographical distribution of KhoeSan population identifiers defined as KhoeSan (orange), Kalahari and west-coastal; KhoeSan ancestral (green), Cape or east-coastal. The Zambezi river provides a geographical division between the KhoeSan and mostly non-KhoeSan population identifiers. Maps were generated in the R package 'maps' v.3.3.0[37].

mitogenomes as Kalahari (n = 18), west-coastal (n = 21), Cape (n = 109) and east-coastal (n = 29) KhoeSan, or non-KhoeSan (Bantu, n = 19), although two mitogenomes were classed as unknown. Using these identifiers, we provide a best-fit classification for all 1,217 L0 mitogenomes (Supplementary Table 2).

Phylogenetic analysis confirms the major L0 haplogroups, with the exclusion of L0b (Extended Data Fig. 1). Using a subset of 461 mitogenomes, including all of the rare lineages, we establish the coalescence times within the L0 lineage (Fig. 2a and Supplementary Table 3) and use the complete dataset to reconstruct geographical dispersals (Fig. 2b). We redefine the emergence of the L0 lineage to 50–25 thousand years (kyr) before previous estimates[1,6], around 200 ka (95% confidence interval, 240–165 ka). L0d'k (n = 309; coalesced around 187 ka (the number of mitogenomes and the coalescence time are provided for each lineage)) is largely KhoeSan-specific, emerging approximately 20 kyr before the widely dispersed L0a'b'f'g sister branch (n = 152; around 164 ka). Although the exact branch resolution for L0k remains undetermined, we observe a preference for L0d'k (posterior probability of approximately 0.6) over L0a'b'f'g'k (posterior probability of about 0.4). Irrespective of this, the L0k (n = 113) lineage appears to remain stable for around 130 kyr before diverging into the Kalahari-specific L0k1 lineage, which is predominated by L0k1a (85 out of 94), and rarer L0k1b and L0k2 lineages distributed around the Zambezi river (Extended Data Fig. 2a). The L0d lineage remains stable for almost 60 kyr before splitting into the KhoeSan-specific L0d1'2 and rarer L0d3 lineages.

Coalescing around 113 ka, L0d2 (n = 226) emerges approximately 15 kyr before L0d1 (n = 452). Within L0d2 (emerging about 91 ka), L0d2c diverged the earliest (n = 53; around 84 ka) with a broad and almost even KhoeSan-regional distribution (Extended Data Fig. 3 and Supplementary Table 4). In 2014, we derived an ancient L0d2c1c mitogenome from a sample of the skeleton of an approximately 2,330-year-old Cape-coastal marine forager (St Helena (StHe)/UCT606)[15]. Predating archaeological evidence for sheep herding in the region[12,16], we proposed that this L0d2c sub-clade represented a pre-pastoral indigenous southern African lineage. Recently, whole-genome sequencing confirmed a unique southern African heritage, whereas two younger (less than 2 kyr old) Cape skeletons showed a genetic link to eastern Africa and the associated pastoralist migration[17]. Previously, an overrepresentation of the L0d2b (28 out of 44; around 65 ka) and L0d2a (62 out of 118; around 60 ka) lineages within the Kalahari KhoeSan has been observed; however, by doubling the contribution of the L0d2d (6 out of 11) lineage, we show a broad southern African distribution (Extended Data Fig. 3 and Supplementary Tables 5, 6). While L0d1 is also spread throughout the KhoeSan-regional identifier, we show notable overrepresentation of the L0d1b (104 out of 174; about 69 ka) and L0d1c (151 out of 184; approximately 59 ka) lineages within the Kalahari and of the L0d1a (32 out of 91; around 44 ka) lineage within the Cape (Extended Data Fig. 4). We contribute two new KhoeSan-ancestral L0d1d mitogenomes to the single published mitogenome[6].

In contrast to L0d1'2, the L0d3 lineage is not specific to southern Africa. Although L0d3b (around 30 ka) appears to be KhoeSan-specific, the rarer L0d3a (about 42 ka) lineage is exclusively found north of the Zambezi river. Notably, three out of six L0d3a mitogenomes were derived from east African Sandawe individuals. Our data support previous studies that have suggested a genetic link between east Africa and the earliest southern Africans[17], who last shared a common ancestor around 59 ka. By adding a large number of mitogenomes (27 out of 40) to the L0d3 lineage, we observe overrepresentation of L0d3b in the Cape KhoeSan identifier (21 out of 34) (Extended Data Fig. 2b and Supplementary Table 7). Using a previously reported identifier that distinguishes maternal KhoeSan ancestry for the Coloured and Baster populations[13], we show that the L0d3b lineage is specific to the Coloured population, whereas the new L0d2b1a2a sub-clade is specific to the Baster population (Extended Data Fig. 3b).

Within the L0a'b'f'g lineage, L0f is highly divergent (emerging around 125 ka; 95% confidence interval,149–101 ka). By including a further five L0f mitogenomes, we were able to show that L0f1 (13 out of 27; around 113 ka) predominates south and L0f2'3 (14 out of 27; about 121 ka) north of the Zambezi river (Extended Data Fig. 2c and Supplementary Table 8). Within L0f1, we recognize three new branches: the northeast sister clades L0f1c (Zambian) and L0f1b (Tanzanian), and the South African clade L0f1a (n = 8). Lack of L0f representation within contemporary KhoeSan suggests that the presence of L0f1a within South Africa is probably a result of more recent east-coastal agropastoral back-migration. While the L0a'g lineages coalesce around 117 ka (95% confidence interval, 145–94 ka), contributing 19 southern African to 347 L0a mitogenomes, we concur that the L0a lineage probably diverged northeast of the Zambezi river (around 85 ka) and spread throughout Africa[3]; the

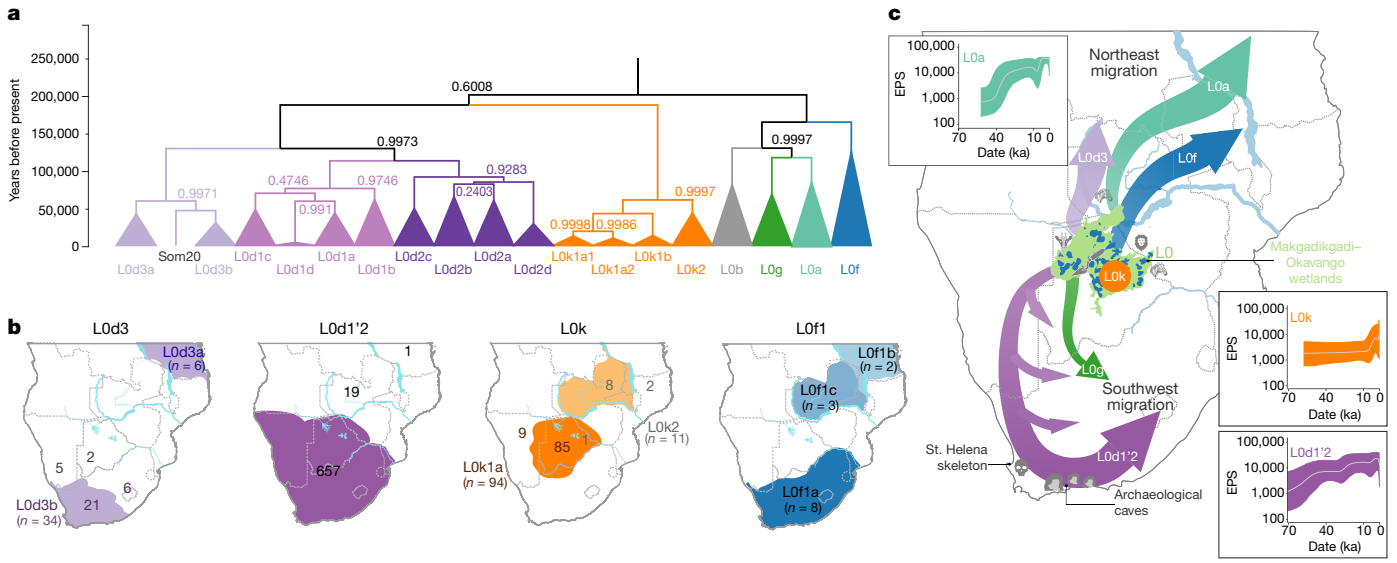


Fig. 2 | L0 phylogenetic tree, geographical distributions of the major southern African L0 haplogroup and out-of-homeland L0 dispersal routes. **a**, Phylogenetic branching and coalescence times derived from a focused subset of 461 L0 mitogenomes, including all rare branches, and anchored to Neanderthals (*Homo neanderthalensis*; $n = 7$). The Somalian-derived (Som20) L0d3 mitogenome[3] could not be assigned. **b**, Geographical distribution (identifiers described in Fig. 1b) for all KhoeSan-specific mitogenomes (out of 1,217): L0d3 ($n = 40$), L0d1'2 ($n = 677$, excluding one unknown), L0k ($n = 105$, excluding seven L0k1b and a single Yemen-derived L0k2), and L0f1 ($n = 13$). Predominant geographical representation (shaded regions), with region-specific overflow represented by the total number of mitogenomes, including the country-specific representation north of the Zambezi river. **c**, Schematic map of southern Africa representing the Makgadikgadi–Okavango palaeo-wetland sustained AMH homeland (200–130 ka), supported by archaeological data (represented by the trowel symbol)[7] and genetic wildlife data (represented by the lion, zebra and giraffe symbols)[23–25]. The out-of-homeland migration (130–110 ka), results in the split of L0d with L0a'g and L0f divergence. L0d3, L0a and L0f migrate in a northeast direction, L0d1'2 and L0g migrate southwest, while L0k remains in the homeland. Insets show BSP analyses of effective population sizes (EPS) of major L0 haplogroups over time, predicting the maintenance of the homeland L0k population (orange), population growth for the broadly dispersed southwest L0d1'2 migrants (purple), which is supported by archaeological evidence (100–60 ka)[8–10] and the StHe mitogenome[15], while population growth of the northeast L0a migrants coincides with the out-of-Africa migration (aqua). Maps were generated in the R package maps v.3.3.0[37].

southern representation of the L0a1b and L0a2a lineages are probably a result of a Bantu back-migration (Extended Data Fig. 5). First described in a Kx'a-speaking hunter-gatherer[6], we now contribute three additional and reclassify five published mitogenomes as L0g (Extended Data Fig. 2d and Supplementary Table 9). As the L0g lineage has a broad KhoeSan and KhoeSan-ancestral distribution, we hypothesize that this lineage diverged southwest of the Zambezi river (around 69 ka), similar to the L0d1'2 lineage.

Our results suggest that the greater Zambezi river basin, particularly the Kalahari region, had a critical role in shaping the emergence and prehistory of AMHs. Now a semi-desert, this region consists of salt pans within northern Botswana that represent desiccated vestiges of palaeo-lake Makgadikgadi, which at its peak in the early Pleistocene would have been the largest lake in Africa[7,18]. Contraction of the Makgadikgadi palaeo-lake during the Middle Pleistocene was accompanied by development of the Okavango delta as a result of neotectonic rifting, which—together with smaller lakes from the upper Zambezi to the Kafue rivers—would have created a vast residual wetland favourable for habitation by humans and mammals more broadly[19] (Fig. 2c). Today, the harsh Kalahari climate and oxygen-rich salt pans are not ideal for fossil and pollen preservation, respectively. However, period-relevant lithic artefacts are documented from the Makgadikgadi pans and surroundings[7,20,21], while palynology suggests that this region was once a grassland and forest biome[22]. Our data further suggest that the Makgadikgadi–Okavango palaeo-wetland sustained the existence of AMHs for around 70 kyr, supported by mitochondrial data of ancestral giraffe, lion and zebra[23–25], before out-of-homeland migrations split the founder homeland populations of the L0d, L0f and L0a'g lineages.

Southwest of their homeland, the L0d1'2 lineage experienced episodic splits and showed a broad south-coastal occupation of the emerged sub-populations, whereas the ancestors of the L0g lineage were less successful. Bayesian skyline plot (BSP) (Fig. 2c) analysis confirms effective population growth for the L0d1'2 lineage (BSP L0d1'2), whereas extensive archaeological evidence indicates cognitively modern human behaviour at the southern tip of Africa[8–10] between approximately 100 and 60 ka, together with an associated increase in the density of time-appropriate archaeological sites in coastal compared to inland regions[26]. Northeast of their homeland, the L0d3 and L0f lineages are less successful, whereas the L0a lineage underwent considerable diversification, which post-dates the out-of-Africa migration (BSP L0a; Fig. 2c). The northeast migration route is further supported by the appearance of data-appropriate archaeological sites[26]. Within their homeland, the population carrying the L0k lineage sustained a constant effective population size (BSP L0k), as did the Kalahari-predominant L0d2b, L0d2a and L0d1c lineages. Although the presence of L0k in Zambia has been suggested to represent contact with an ancient pre-Bantu population[27], we propose that these rare lineages represent an ancient out-of-homeland branch of the ancestral KhoeSan population.

Orbitally driven large-scale hydroclimate variations have been proposed as a contributor of early human migrations[28,29]. In some studies, wetter conditions and resulting 'green corridors' have been proposed to explain the out-of-Africa migration (a 'pull' scenario), whereas others have proposed that drier conditions and resulting food shortages forced dispersals (a 'push' scenario)[30]. To determine whether our predicted homeland isolation and major dispersals may have been driven by climate shifts, we analysed four key palaeo-hydroclimate datasets[29,31–33], along with a transient 784-kyr-long glacial–interglacial simulation conducted with the LOVECLIM Earth system model[28] (Fig. 3). Although limited by available palaeo-proxy records and a climate model of intermediate complexity, we observe a considerable degree of coherence on orbital timescales (Extended Data Fig. 6). During the homeland period (200–130 ka), palaeo-data link the 21-kyr-long precession cycle, which

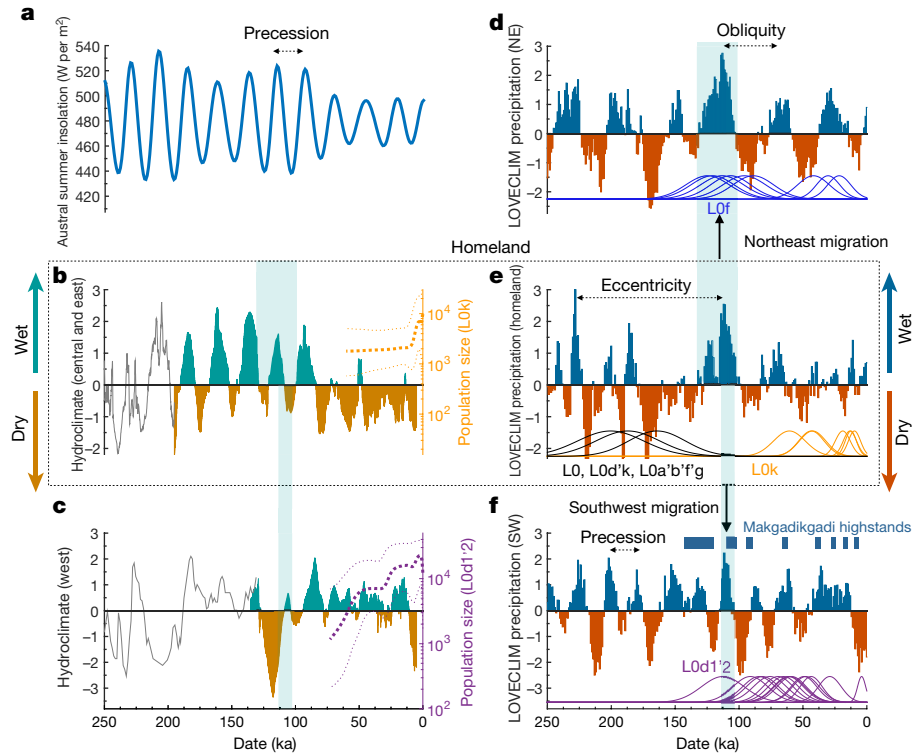


Fig. 3 | Reconstructed and simulated climatic conditions during the out-of-homeland migration. a, Austral summer insolation changes (blue) at 27° S. **b**, A hydroclimate composite of eastern and central southern Africa (shading) was obtained by averaging the Fe/K runoff record from core CD154-1006P[31] and the Pretoria Salt Pan rainfall reconstruction[29], extended from 250 to 190 ka (grey line). The plot shows the effective population size for homeland L0k as analysed by BSP (orange dashed lines). **c**, Southwestern hydroclimate reconstruction (shading) obtained by averaging normalized leaf wax data (MD08-3167)[33] and the aridity index from cores (MD96-2094)[32], for which the aridity record extended from 250 to 140 ka (grey line) and the effective population size of L0d1'2 as analysed by BSP is shown (purple dashed lines). **d**, Simulated LOVECLIM normalized precipitation changes (shading) northeast of the homeland (33° E, 13° S) and coalescence time probabilities for L0f haplogroup (blue bell curves). **e**, Same as for **d**, but for the homeland and coalescence probabilities for L0, L0d'k, L0a'b'f'g (black) and L0k haplogroups (orange). **f**, Same as for **d**, but for the area southwest of the homeland (17° E, 30° S) and L0d1'2 coalescence times (purple). Blue bars indicate predicted Makgadikgadi high stand phases[35]. NE, northeast; SW, southwest.

arises from a combination of Earth's axis wobble and a slow rotation of Earth's entire orbit around the Sun (Fig. 3a), with three wet–dry cycles (Fig. 3b). By contrast, the climate model simulates an extended drought, owing to a more pronounced eccentricity signal (Fig. 3e), suggestive of a wetland oasis in an otherwise vast harsh environment.

During the out-of-homeland period (130–110 ka), our model simulation supports humid conditions to the northeast that facilitated the first dispersals, concurring with L0f coalescence (around 125 ka) (Fig. 3d). By contrast, the region southwest of the homeland experienced an approximately 15-kyr-long megadrought before an orbital shift created the favourable humid conditions that led to the dispersal of the L0d1'2 lineage (around 113 ka) (Fig. 3f), which is also supported by palaeo-data (Fig. 3c). This is also around the time the northeast L0a and southwest L0g migrants last share a common ancestor (around 117 ka). During the last glacial period (approximately 100–11 ka), we observe a reduction in the amplitude of the changes in orbital-scale hydroclimate and overall drying within the homeland (Fig. 3b), whereas the southwest coastal hydroclimate was dominated by precessional variability and showed relatively agreeable environmental conditions (Fig. 3c, f). Notably, periods of deceleration and acceleration in the estimates of the effective population size of the L0d1'2 lineage coincide with regional changes in hydroclimate, further linking climate, population size and evolution.

We propose that the Makgadikgadi–Okavango palaeo-wetland was the possible homeland of AMHs. Although one cannot exclude the possibility of a polycentric origin[34], this deltaic–lacustrine ecosystem would have provided an ideal geographical locality for the evolution and 70-kyr-long sustained existence of the deepest-branching maternal founder population of AMHs. Increased humid conditions, supported by palaeo-lake system reconstructions[35], between 130 and 110 ka would have opened green corridors for successful northeast–southwest migrations, supporting a pull scenario. Drying within the homeland following the out-of-homeland period, supported by hydroclimate data (110–100 ka) and a model simulation (100–80 ka), would have created a push scenario, in which a reduced carrying capacity of the land would have increased pressure to seek out climatically more favourable regions. We propose that the southwest migrants maintained a successful coastal forager existence, while the northeast migrants—similar to the later-branching population of L1'6—gave rise to ancestral pastoral and farming populations. A recent publication[36] provides further mitochondrial evidence to support the northeast out-of-homeland migration route and expansion into eastern Africa around 70–60 ka. Revealing a southern African homeland for the emergence and extended subsistence of the L0 lineage, we propose that an out-of-homeland migration event, which was probably driven by astronomically induced regional shifts in hydroclimate, shaped the present-day ethnic and genetic diversity of modern humans.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1714-1.

1. Behar, D. M. et al. The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140 (2008).

2. Brown, F. H., McDougall, I. & Fleagle, J. G. Correlation of the KHS tuff of the Kibish Formation to volcanic ash layers at other sites, and the age of early *Homo sapiens* (Omo I and Omo II). *J. Hum. Evol.* **63**, 577–585 (2012).
3. Rito, T. et al. The first modern human dispersals across Africa. *PLoS ONE* **8**, e80031 (2013).
4. Stringer, C. & Galway-Witham, J. On the origin of our species. *Nature* **546**, 212–214 (2017).
5. Henn, B. M. et al. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl Acad. Sci. USA* **108**, 5154–5162 (2011).
6. Chan, E. K. F. et al. Revised timeline and distribution of the earliest diverged human maternal lineages in southern Africa. *PLoS ONE* **10**, e0121223 (2015).
7. Moore, A. E., Cotterill, F. P. D. & Eckardt, F. D. The evolution and ages of Makgadikgadi palaeo-lakes: consilient evidence from Kalahari drainage evolution south-central Africa. *S. Afr. J. Geol.* **115**, 385–413 (2012).
8. Henshilwood, C. S. et al. A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa. *Science* **334**, 219–222 (2011).
9. Douze, K., Wurz, S. & Henshilwood, C. S. Techno-cultural characterization of the MIS 5 (c. 105–90 ka) lithic industries at Blombos cave, Southern Cape, South Africa. *PLoS ONE* **10**, e0142151 (2015).
10. Henshilwood, C. S. et al. An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature* **562**, 115–118 (2018).
11. Güldemann, T. in *Beyond 'Khoisan': historical relations in the Kalahari Basin (Current Issues in Linguistic Theory 330)* (eds Güldemann, T. & Fehn, A.-M.) **330**, 1–40 (John Benjamins, 2014).
12. Lander, F. & Russell, T. The archaeological evidence for the appearance of pastoralism and farming in southern Africa. *PLoS ONE* **13**, e0198941 (2018).
13. Petersen, D. C. et al. Complex patterns of genomic admixture within southern Africa. *PLoS Genet.* **9**, e1003309 (2013).
14. Morris, A. G. Isolation and the origin of the Khoisan: late Pleistocene and early Holocene human evolution at the southern end of Africa. *Hum. Evol.* **17**, 231–240 (2002).
15. Morris, A. G., Heinze, A., Chan, E. K. F., Smith, A. B. & Hayes, V. M. First ancient mitochondrial human genome from a prepastoralist southern African. *Genome Biol. Evol.* **6**, 2647–2653 (2014).
16. Pleurdeau, D. et al. "Of sheep and men": earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia). *PLoS ONE* **7**, e40340 (2012).
17. Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71 (2017).
18. Eckardt, F. D. et al. Mapping the surface geomorphology of the Makgadikgadi Rift Zone (MRZ). *Quat. Int.* **404**, 115–120 (2016).
19. Wrangham, R. W. in *Interpreting the Past: Essays on Humans, Primates and Mammal Evolution* (eds Pilbeam, D. R. et al.) 231–242 (Brill Academic, 2005).
20. Robbins, L. H. et al. The advent of herding in southern Africa: early AMS dates on domestic livestock from the Kalahari Desert. *Curr. Anthropol.* **46**, 671–677 (2005).
21. Mackay, A., Stewart, B. A. & Chase, B. M. Coalescence and fragmentation in the late Pleistocene archaeology of southernmost Africa. *J. Hum. Evol.* **72**, 26–51 (2014).
22. Scott, L. & Neumann, F. H. Pollen-interpreted palaeoenvironments associated with the Middle and Late Pleistocene peopling of Southern Africa. *Quat. Int.* **495**, 169–184 (2018).
23. Bock, F. et al. Mitochondrial sequences reveal a clear separation between Angolan and South African giraffe along a cryptic rift valley. *BMC Evol. Biol.* **14**, 219 (2014).
24. Pedersen, C. T. et al. A southern African origin and cryptic structure in the highly mobile Plains zebra. *Nat. Ecol. Evol.* **2**, 491–498 (2018).
25. Moore, A. E. et al. Genetic evidence for contrasting wetland and savannah habitat specializations in different populations of lions (*Panthera leo*). *J. Hered.* **107**, 101–103 (2016).
26. Blome, M. W., Cohen, A. S., Tryon, C. A., Brooks, A. S. & Russell, J. The environmental context for the origins of modern human diversity: a synthesis of regional variability in African climate 150,000–30,000 years ago. *J. Hum. Evol.* **62**, 563–592 (2012).
27. Barbieri, C. et al. Ancient substructure in early mtDNA lineages of southern Africa. *Am. J. Hum. Genet.* **92**, 285–292 (2013).
28. Timmermann, A. & Friedrich, T. Late Pleistocene climate drivers of early human migration. *Nature* **538**, 92–95 (2016).
29. Partridge, T. C., Demenocal, P. B., Lorentz, S. A., Paiker, M. J. & Vogel, J. C. Orbital forcing of climate over South Africa: a 200,000-year rainfall record from the Pretoria saltpan. *Quat. Sci. Rev.* **16**, 1125–1133 (1997).
30. Tierney, J. E., deMenocal, P. B. & Zander, P. D. A climatic context for the out-of-Africa migration. *Geology* **45**, 1023–1026 (2017).
31. Simon, M. H. et al. Eastern South African hydroclimate over the past 270,000 years. *Sci. Rep.* **5**, 18153 (2015).
32. Stuut, J.-B. W. et al. A 300-kyr record of aridity and wind strength in southwestern Africa: inferences from grain-size distributions of sediments on Walvis Ridge, SE Atlantic. *Mar. Geol.* **180**, 221–233 (2002).
33. Collins, J. A., Schefuß, E., Govin, A., Mulitza, S. & Tiedemann, R. Insolation and glacial–interglacial control on southwestern African hydroclimate over the past 140 000 years. *Earth Planet. Sci. Lett.* **398**, 1–10 (2014).
34. Scerri, E. M. L. et al. Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* **33**, 582–594 (2018).
35. Burrough, S. L., Thomas, D. S. G. & Bailey, R. M. Mega-lake in the Kalahari: a late Pleistocene record of the palaeolake Makgadikgadi system. *Quat. Sci. Rev.* **28**, 1392–1411 (2009).
36. Rito, T. et al. A dispersal of *Homo sapiens* from southern to eastern Africa immediately preceded the out-of-Africa migration. *Sci. Rep.* **9**, 4728 (2019).
37. Becker, R. A. & Wilks, A. R. maps: Draw Geographical Maps. R package version 3.3.0 https://cran.r-project.org/web/packages/maps/index.html (2018).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Statement on population identifiers

The authors acknowledge that population identifiers (or ethnic labels) have different meanings to different peoples across different countries and between and within different ethnic groups. During the apartheid rule, South Africans were grouped according to ethnic identities, which resulted in discrimination based on population identifiers such as Bantu or Coloured. In turn, others view the very same population identifiers with cultural identity and pride. In 2013, we performed a study led by a Coloured co-author to assess the sensitivity in self-identification as Coloured. Of 521 participants, 91.2% self-identified as Coloured, Cape Coloured or South African Coloured, while 8.8% elected against the use of Coloured for self-identification[14]. In turn, using such population identifiers within the context of the United States would be seen as derogatory and highly offensive. We have previously genetically profiled the Baster population of Namibia[13] and again what could be to others a derogatory term, to the Baster community of Rehoboth in Namibia, the term is used with immense pride, who recognize themselves as a Republic with a national flag[38].

In this study, the authors have used linguistics, supported by ethnicity, to provide population identification, with further historical, geographical and genetic classification for deriving maternal contributions (described in the next section). KhoeSan (or KhoeSaan) languages are grouped together due to their use of click consonants as a unique language identifier. Once spread across the entire southern African region, KhoeSan languages are today restricted largely to populations residing in Namibia and Botswana (and southern Angola), although two Tanzanian isolates, Sandawe and Hadza, are believed to be linguistically related click languages (or east African KhoeSan)[39]. 'San' literally means 'forager' and Khoe means 'person'; culturally, the KhoiSan identifier refers to hunter–foragers (San) or herders (Khoi). At times linguistic and cultural identities clash. For example, Nama and Hai||om peoples both speak Nama (a Khoe–Kwadi language), while culturally and historically these two populations are quite different, representing a herder and hunter-gatherer ancestry, respectively. Additionally, autosomal genetic data have been used to provide further insights into KhoeSan admixture and substructures, highlighting at a genetic level the historical differences between the Nama and Hai||om[40]. We have attempted to capture both ethnic and linguistic identifiers that best reflect population ancestry. In contrast to KhoeSan languages, most Bantu languages do not contain click consonants; however, exceptions exist within Southern African Bantu languages (for example, isiXhosa and isiZulu languages, which have borrowed click consonants from their KhoeSan neighbours). Spoken across the entire sub-Saharan Africa (up to 500 groups), the Guthrie classification of languages further identifies the S-zone or Southern Bantu (South Africa, Zimbabwe, southern Mozambique and most of Botswana) and the R-zone or Southwest Bantu languages (northern Namibia, southern Angola and northwest Botswana)[41], which are of relevance to this study.

Ethics statement and recruitment

The study was performed in accordance with the ethical standards of the overseeing human research ethics committees and local governance, as per the 1964 Helsinki Declaration. The study was reviewed and approved by the Ministry of Health and Social Services (MoHSS) in Namibia (17-3-3 2008, 2014 and 2019), with additional local approvals from participating community leaders, the University of Pretoria Human Research Ethics Committee (HREC 43/2010 and HREC 280/2017), including US Federal-wide assurance (FWA00002567 and IRB00002235 IORG0001762), as well as the South African National Blood Service (SANBS) HREC (HREC 2012/11). Participants were recruited within the borders of Namibia and South Africa and self-reported ethno-linguistic population identifiers were recorded. Blood samples were taken after receiving written and/or recorded informed consent. Isolated DNA was shipped under the Republic of South Africa Department of Health Export Permit (J1/2/4/2), in accordance with the National Health Act 2003, to the Garvan Institute of Medical Research in Australia. Mitogenome sequencing was performed in accordance with site-specific approval granted by St Vincent's Hospital HREC in Australia (SVH 15/227).

Participant population identifiers

Merging with published data for a total of 1,217 L0 mitogenomes, participants were broadly classified as KhoeSan, Bantu or Cape multi-ethnic heritage. Indigenous KhoeSan who inhabit the inland semi-desert Kalahari region of Botswana and Namibia include the Kx'a (Ju|'hoan or Hoan, and !Xun or !Xuun), Tuu (or Taa) and Khoe–Kwadi (Naro, ||Ani, Khwe, Buga, G||ana, G||ui, ||Xokhoe, Tshwa and Shua) speakers. Indigenous KhoeSan who inhabit the west-coastal region of Namibia speak a Khoe–Kwadi or Nama language and include the Nama, Damara, Topnaar (ǂAonin) and Hai||om speakers[42,43]. Novel mitogenomes were derived from 15 Kalahari KhoeSan, including Ju|'hoan ($n = 9$), !Xun ($n = 1$) and Naro ($n = 5$), and 21 west-coastal KhoeSan, including Nama ($n = 7$), Damara ($n = 8$) and Topnaar (ǂAonin, $n = 6$) from Namibia. Speakers of southwest Bantu (non-KhoeSan) languages (which do not contain click consonants) of Namibia, Botswana and southerly boarders of Angola, presenting with KhoeSan-predominant L0 maternal lineages, most likely carry a Kalahari or west-coastal KhoeSan mitogenome. As a result of refuge provided to the Herero by the Kalahari KhoeSan during the early 1900 German South West African genocide[44], we speculate in this study a probable Kalahari KhoeSan heritage for the three Herero mitogenomes.

Although indigenous KhoeSan are arguably absent from the coastal regions of South Africa, and while recognising and honouring the northwest inland (southern Kalahari) ǂKhomani San of South Africa (although not recruited within the context of this study), KhoeSan skeletal remains spread across the region[45]. Hunter-gatherer KhoeSan once inhabited a broad southwest to east-coastal region at the tip of Africa. These skeletal remains predate archaeological evidence supporting the arrival of sheep herders who appear to have crossed the Okavango river in northern Namibia around 2.2 ka, migrating along the southwest coast to the southern Cape[12,16,20,45] by around 2 ka. Recently, Cape KhoeSan skeletons younger than 2 ka have been genetically linked to east Africa and herder migration[17]. Migrating herders may have acquired indigenous KhoeSan maternal contributions. Along the east coast, southward-migrating Bantu farmers (Southern Bantu; who presumably did not speak languages containing click consonants) entered South Africa around 1,500 years ago, while a second wave of Bantu migrants (Southwest Bantu) crossed central Africa into Namibia around 800 years ago[12]. Maternal contributions to the South African Southern-Bantu-speaking populations ($n = 43$, this study) may therefore either be of Bantu origin (in this case, L0a lineages and therefore non-KhoeSan) or of east-coastal KhoeSan-ancestry. The arrival of European colonists and Dutch–East-Indian slaves to the Cape in the mid-1600s, gave rise to a multi-ethnic (European, Asian, KhoeSan and Bantu) Cape population, the ancestors of the South African Coloured ($n = 90$, this study) and Namibian Basters ($n = 24$, this study), who historically speak a Dutch-derived language known as Afrikaans[13,46]. Emerging from a common historical background to the Coloured, the Baster population have since the late 1800s distinguished themselves as independent from the Coloured, migrating to the Baster nation of Rehoboth in Namibia[47]. Although the vast majority of L0 mitogenomes represented in the Baster and Coloured populations are of Cape KhoeSan heritage (100% and 94.4%, respectively), we observe a percentage of non-KhoeSan (Bantu) L0a lineages within the Coloured population.

L0 haplogroup pre-screening

Subjects were selected for whole-mitogenome sequencing based on pre-screening for specific L0 markers using direct amplicon-specific Sanger sequencing. Specifically, a 2,673-bp region (Cambridge Reference Sequence (rCRS) position 3322–5995) was amplified and initially screened for the L0 variant T5442C. L0 samples were further screened to delineate L0d (T4232C), L0d1 (G3438A), L0d1b (T3618C), L0d1c (C4197T), L0d1'2 (A3756G), L0d2 (A3981G, C205T, A4044G), L0d2a (A5153G), L0d2d (G5147A, G5231A), L0d2C (A4038G, T4937C) and L0d3 (G5460A, G5773A) lineages. This identified 188 samples carrying a rare L0 haplogroup: L0d1b (n = 21), L0d1c (n = 13), L0d2a (n = 30), L0d2b (n = 7), L0d2c (n = 15), L0d2d (n = 6), L0d3 (n = 29), L0a1 (n = 6), L0a2 (n = 6), L0f (n = 5) and L0k (n = 5); as well as 55 samples that could not be unambiguously assigned to a major L0 sub-lineage: L0d1a'c (n = 2), L0a'b'f'k (n = 5), L0a'b (n = 2), L0d2 (n = 1) and L0d1 (n = 45, assumed L0d1a) (Supplementary Table 1).

Whole-mitogenome sequencing

Mitogenomes were isolated using two overlapping amplicons as previously described[6,48]. Specifically, two primer pairs were used to isolate and amplify fragments 12,250–3,005 (7.2 kb) and 2,583–12,337 (9.7 kb) of the circular mitogenome. This pair of primers has been demonstrated to effectively capture the mitogenome with high specificity while minimizing off-target capture of nuclear copies of mitochondrial-derived DNA. Following touchdown long-range amplification with the Platinum Taq DNA Polymerase High Fidelity (Invitrogen), the two amplicons were purified using AMPure XP beads (Agencourt) and combined in a 7:13 ratio of short:long fragments. Sequencing was performed on the Ion Torrent PGM platform. In brief, 200-bp single-end sequencing libraries were prepared using the Ion Xpress Plus Fragment Kit and Ion Xpress Barcode Adaptors (ThermoFisher), and 4–16 samples (barcodes) were pooled and sequenced on 314v2 Ion Chips. Using the Ion Torrent suite v.5.0.2.1, sequencing reads were quality trimmed and aligned to the human mitochondrial revised rCRS (accession NC_012920.1). Consensus mitogenome sequences were derived by first identifying variants relative to rCRS, using samtools (v.1.3.1) mpileup (with parameters -d 10000 -L 1000 -Q 7 -h 50 -o 10 -e 17 -m 4)[49] and bcftools (v.1.3.1) call (with parameters -c -M) (http://www.htslib.org/doc/bcftools.html), then converting to the FASTA format using the vcfutils.pl vcf2fq program in samtools.

Publicly available data

An exhaustive search for publicly available L0 mitogenomes was performed between 2015 and 2017, identifying 26 studies comprising a total of 6,334 mitogenomes. L0 status for all mitogenomes was deduced, either directly from the original publication or by downloading the nucleotide sequences from NCBI and evaluating their haplogroup using HaploGrep2 (v.2.1.13)[50] based on PhyloTree Build 17[51]. From this dataset, a subset of 1,019 L0 mitogenomes was identified and included in this study (Extended Data Table 1 and Supplementary Table 2). Publicly available genomes were broadly classified as KhoeSan, Bantu (KhoeSan ancestral), or non-KhoeSan based on the reported population and/or country of origin.

Whole-mitogenome haplotyping

HaploGrep2[50] was used to type all 1,217 sequences against PhyloTree Build 17[51]. This resulted in the refinement and reclassification of our 198 mitogenomes, resulting in L0d1 (n = 81, including 45 L0d1a, 21 L0d1b, 13 L0d1c and 2 L0d1d), L0d2 (n = 58, including 30 L0d2a, 8 L0d2b, 14 L0d2c and 6 L0d2d), L0d3 (n = 27), L0a (n = 19), L0f (n = 5), L0k (n = 5) and L0g (n = 3) mitogenomes (Supplementary Table 1). This refined, and in some cases reclassified, the haplogroups of the 1,019 publicly available mitogenomes (Supplementary Table 2).

Phylogenetic inference

Multiple sequence alignment was performed across all 1,217 mitogenomes along with 7 Neanderthal genomes (Supplementary Table 10), using MUSCLE v.3.8.31[52] with parameters -maxiters 3 -diags1. Phylogenetic inference was performed using FastTree v.2.1.7 (SSE3)[53] using the generalized time reversible (-gtr) and discrete gamma model with 20 rate categories (-gamma). A summary of the inferred phylogenetic tree is shown in Extended Data Fig. 1, with the tree rerooted to the 7 Neanderthal genomes.

Bayesian phylogenetic inferences and divergence times were calculated using BEAST2 v.2.4.2 with BEAGLE v.2.0[54]. Owing to the computational burden of this analysis, BEAST was performed on a subset of 461 mitogenomes, selected to include: (i) only complete mitogenomes (27 mitogenomes with only the coding region[55,56] were excluded); (ii) all 198 novel mitogenomes from this study; (iii) all 121 L0 mitogenomes from our previous studies, Chan et al.[6] (n = 77), Morris et al.[15] (StHe, defining the new haplogroup L0d2c1c), McCrow et al.[48] (n = 37) and Schuster et al.[57] (n = 6); (iv) all rare haplogroups, namely L0g (n = 9), L0f (n = 22), L0d3 (n = 30), L0d1d (n = 3), L0d2d (n = 11) and L0k2 (n = 12); (v) all mitogenomes that could not be unambiguously typed by HaploGrep2[50] (n = 14; none from this study); and (vi) a random subset of mitogenomes for all remaining sub-lineages not already represented.

Multiple sequence alignment of the subset of 461 AMH and 7 Neanderthal mitogenomes was converted to NEXUS format using the convert function of seqmagick v.0.6.1 (https://fhcrc.github.io/seqmagick) with parameter --alphabet dna-ambiguous. This provided the input to BEAST2. Specifically, BEAUTi v.2.4.2 was used to set up the phylogenetic model, assuming: (i) the gamma site model with six gamma categories and no invariant sites; (ii) the generalized time-reversible substitution model; (iii) a strict constant clock model with a normal prior with $\mu = 1.665 \times 10^{-8}$ and $\sigma = 1.479 \times 10^{-9}$ based on a previously published study[58]; and (iv) a coalescent constant population. Times were calibrated to the seven *H. neanderthalensis* mitogenomes with tip dates set to their reported approximate archaeological dating estimates: Feldhofer 1, 40 ka[59]; Vindija, 38 ka[59]; El Sidron, 39 ka[59]; Feldhofer 2, 40 ka[59]; Mezmaiskaya, 65 ka[59]; Croatia, 38.31 ka[60]; Altai, 50 ka[61] (Supplementary Table 10). No prior was set on the most recent common ancestor of this taxon set, and calibration was applied to the leaves instead of the most recent common ancestor. Further, a normal prior, $N(\mu = 200{,}000, \sigma = 50{,}000)$, was set on the coalescent time of the AMH genomes, and a tip date of 2,330 years before present was set for the StHe genome[15].

Five BEAST replicates were performed, each with 100 million Markov chain Monte Carlo iterations, sampling every 10,000. Tracer v.1.6 was used to evaluate BEAST trace files (Supplementary Table 11), ensuring all runs had converged. The five replicates were combined using LogCombiner v.2.4.2, discarding 10% of the samples as burn-in for each replicate and without resampling states at a lower frequency.

Sampled trees from BEAST were summarized into a single maximum clade credibility target tree using TreeAnnotator v.2.4.2 for each of the five replicates, discarding the first 10% as burn-in. To summarize across replicates, sampled trees from the five replicates were first combined using LogCombiner v.2.4.2, again discarding the first 10% as burn-in from each replicate, but resampling at a lower frequency of 50,000 (five replicates of 10,000 samples). The combined, resampled trees were then summarized with TreeAnnotator v.2.4.2 as for the individual replicate BEAST results.

FigTree v.1.4.2+ (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize all resulting trees.

BSP analysis

BSP analyses were performed to estimate the demographic history of each maternal haplogroup. Although maternal haplogroups do not necessarily equate to population groups, it has been suggested that the signal associated with a haplogroup can still provide insights into the demographic processes in the populations who carry the haplogroup[62,63].

For each haplogroup of interest (for example, L0a, L0d1'2 and L0k), a nexus file was derived using SeqMagic v.0.6.1 as described above.

BSP analyses were performed using BEAST2, using BEAUTi 2 for model setup as described in 'Phylogenetic inference', with the following key differences: (i) the gamma shape of the gamma site model was estimated with an exponential prior with mean = 1.0 and offset = 0.0; (ii) the molecular clock was fixed (not estimated) at 1.665×10^{-8} based on a previously published study[58]; and (iii) the phylogenetic tree prior was set to coalescent Bayesian skyline, assuming 20 intervals between the root of the tree and the present time.

Tracer v.1.6 was used to reconstruct the Bayesian skyline from the sampled trees for each analysis, using a stepwise constant variant and the lower 95% highest posterior density of the root height as the maximum time. Results of this analysis are summarized in Supplementary Table 12.

Geographical history of the palaeo-wetland Makgadigadi

Initiated around 2 million years ago, palaeo-lake Makgadikgadi[7] originally covered an area of around 170,000 km^2 at its highest lake stand, bounded by a shoreline of around 995 m. A degraded sand ridge (the Deception ridge), was associated with the 995-m shore in the southwest of the lake. This lake would have covered more than twice the area of modern Lake Victoria, and similar to the latter, would have caused a considerable climatic feedback, with locally enhanced rainfall. We previously proposed that this was, in turn, responsible for the initiation of the surrounding (now-fossil) drainages, creating a well-watered environment and very favourable habitat for mammals, including hominids[7]. Smaller lakes, now represented by residual wetlands, also formed on the upper Zambezi river and the modern Kafue Flats on the Kafue river, resulting in an archipelago of palaeo-lakes in south-central Africa during the Early and Middle Pleistocene epoch.

Palaeo-Makgadikgadi, bounded by the 995-m shoreline, was originally sustained by a major drainage line, which included the Chambeshi river as headwaters, connected to the upper Zambezi river via the upper Kafue river. Severance of the original links between the Chambeshi river and upper Kafue river, and the latter and the upper Zambezi river resulted in a sequential contraction of the Makgadikgadi to a much smaller water body. This is reflected in a series of fossil shorelines, associated with breaks in slope, at progressively lower levels (945 m, 936 m and 922 m). The Gidikwe ridge was associated with the 945-m shoreline. However, contraction of the lake was accompanied by the development of the modern Okavango delta. Timing of the contraction of the lake and initiation of the Okavango delta is not tightly constrained, but by the time that we propose that modern humans emerge within the region, around 200 ka, we speculate that the formerly extensive Makgadikgadi palaeo-lake had contracted to a much less extensive deltaic–lacustrine system. Together with the lakes that developed from the upper Zambezi and Kafue rivers to the north and the Okavango delta to the west, the region would have been a vast wetland, a favourable habitat for hominid occupation. It is this palaeo-wetland region that we propose as the homeland for the founder population of AMHs.

Climate model simulations and palaeo-climate data

To place the coalescence time estimates of the L0 branch into a climatic context and to test the robustness of simulated hydroclimate responses in South Africa to orbital-scale conditions, we use the LOVECLIM Earth system model of intermediate complexity[28]. It is based on a 3-layer atmosphere, a 20-level ocean general circulation model, a dynamic–thermodynamic sea–ice model and a terrestrial vegetation model. A transient simulation that covers several glacial–interglacial cycles was conducted using time-dependent boundary conditions. The experiment[28] (covering the past 784 kyr) uses time-varying boundary conditions for orbital parameters, CO_2 and other greenhouse gas concentrations obtained from Antarctic ice cores, and an estimate of Northern Hemispheric ice-sheet orography and albedo changes (data are used in Fig. 3 and Extended Data Fig. 6). The forcings are applied with an acceleration factor of five: one coupled model year corresponds to five orbital calendar years. Our analysis focuses on the past 250 kyr in both simulations. The climate sensitivity of this model to CO_2 variations was modified to capture the range of reconstructed global mean surface temperature changes in response to radiative forcing[64]. The transient LOVECLIM model simulations have previously been validated against other palaeo-climate records from around the world[28,64,65]. Our analysis here focuses on the simulated precipitation as well as changes in tree and grass fractions in central eastern Africa and western southern Africa (data used in Fig. 3 d–f and Extended Data Fig. 6b–d).

As a result of its coarse horizontal atmospheric resolution (5.6°) and the use of only parameterized ageostrophic wind components, LOVECLIM has several deficiencies. Of particular note are the lack of realistic El Niño–Southern Oscillation variability and the fact that annual mean freshwater flux corrections have been applied to mimic the atmospheric moisture transport from the Atlantic to the Pacific and to stabilize the Atlantic Meridional Overturning Circulation.

There exist only a few long-term hydroclimate datasets from southern Africa that cover the past >120 kyr. Here we compare the simulated LOVECLIM precipitation (normalized) in central southern Africa with a southern central African hydroclimate composite, obtained by averaging the normalized orbitally tuned rainfall reconstruction from the Pretoria salt pan[29] and the normalized Fe/K river run-off proxy obtained from marine sediment core CD154-1006P[31] (Fig. 3b). The composite index emphasizes the joint variability in both records. We find that some of the overall features in the observations—particularly the fact that rainfall is modulated by the precessional cycle of austral summer insolation[66] (Fig. 3a)—are well-captured by the LOVECLIM model experiment. However, we also find some discrepancies in the central part of southern Africa, such as in the phase of the precessional signal and the difference in overall wet and dry conditions during the homeland period from 200 to 120 ka. The overall glacial drying in the central part of southern Africa from 100 to 20 ka is, however, captured in both model simulation and palaeo-proxy reconstructions (Fig. 3b, e). Orbital-scale hydroclimate variations in southern Africa are clearly not spatially homogenous (Fig. 3b–f). To gain a better understanding of the spatial patterns of hydroclimate variability, we compared the model simulation with a composite index from southwestern Africa, obtained by averaging a normalized aridity index reconstructed from sediment core MD96-2094[33] and the normalized $\delta^{13}C$ isotope ratio data of leaf wax extracted from the South Atlantic sediment core MD08-3167[32] (Fig. 3c and Extended Data Fig. 6c, d). The results show a good correspondence between model and reconstructions on the western side of southern Africa, and in particular reproduce a major drought period that peaked around 120 ka and a subsequent increase in rainfall towards the last glacial period. This gradual increase in rainfall corresponds to an overall increase in lineage splitting of the L0d1'2 haplogroup (Fig. 3f) and growth of its population (Fig. 3c). This result further highlights the possibility that climate shifts may have played an important part in the southwestward migration of L0d1'2 descendants (Fig. 2).

To further test the fidelity of LOVECLIM in reproducing interhemispheric orbital rainfall shifts across Africa, we also compared the simulated vegetation changes with a leaf-wax index from stable hydrogen isotope data extracted from a sediment core in the Gulf of Aden[30], which is indicative of hydroclimate and vegetation changes in the northeastern Horn of Africa (Extended Data Fig. 6b). The comparison shows a good qualitative correspondence for the precessional-scale timing of rainfall and vegetation maxima and minima as well as of the eccentricity modulated amplitude of these changes, lending further support to the credibility of the simulated rainfall patterns across Africa. It should be noted that regional patterns of paleo-rainfall changes are in general difficult to simulate. In response to Last Glacial Maximum boundary conditions, different coupled general circulation models simulate widely varying responses in rainfall over Africa[28].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The consensus sequences for this set of 198 mitogenomes have been deposited in the NCBI GenBank with accession numbers MK248274–MK248471. Requests for materials should in the first instance be addressed to V.M.H.

38. Orizio, R. *Lost White Tribes: the End of Privilege and the Last Colonials in Sri Lanka, Jamaica, Brazil, Haiti, Namibia, and Guadeloupe* (Free, 2001).
39. Heine, B. & Nurse, D. (eds) *African Languages: an Introduction* (Cambridge Univ. Press, 2000).
40. Montinaro, F. et al. Complex ancient genetic structure and cultural transitions in southern African populations. *Genetics* **205**, 303–316 (2017).
41. Guthrie, M. *The Classification of the Bantu Languages* (Oxford Univ. Press, 1948).
42. Honken, H. & Heine, B. The Kx'a family: a new Khoisan genealogy. *J. Asian Afr. Stud.* **79**, 5–36 (2010).
43. Güldemann, T. & Elderkin, E. D. in *Khoisan Languages and Linguistics: Proc. 1st International Symposium January 4–8, 2003, Riezlern/Kleinwalsertal* (eds Brenzinger, M. & König, C.) 15–52 (Rüdiger Köppe, 2010).
44. Stockton, R. *The Herero genocide: Germany's first mass murder. All That's Interesting* https://allthatsinteresting.com/herero-genocide (2017).
45. Smith, A. B. *Excavations at Kasteelberg and the Origins of the Khoekhoen in the Western Cape, South Africa* (Archaeopress, 2006).
46. Patterson, N. et al. Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* **19**, 411–419 (2010).
47. van der Ross, R. E. *Up from Slavery: Slaves at the Cape: their Origins, Treatment and Contribution* (Ampersand, 2005).
48. McCrow, J. P. et al. Spectrum of mitochondrial genomic variation and associated clinical presentation of prostate cancer in South African men. *Prostate* **76**, 349–358 (2016).
49. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
50. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
51. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
52. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
53. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
54. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
55. Kivisild, T. et al. The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373–387 (2006).
56. Herrnstadt, C. et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* **70**, 1152–1171 (2002).
57. Schuster, S. C. et al. Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463**, 943–947 (2010).
58. Soares, P. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
59. Briggs, A. W. et al. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).
60. Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494–2502 (2009).
61. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
62. Gandini, F. et al. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).
63. Soares, P. et al. The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
64. Friedrich, T., Timmermann, A., Tigchelaar, M., Elison Timm, O. & Ganopolski, A. Nonlinear climate sensitivity and its implications for future greenhouse warming. *Sci. Adv.* **2**, e1501923 (2016).
65. Stockhecke, M. et al. Millennial to orbital-scale variations of drought intensity in the Eastern Mediterranean. *Quat. Sci. Rev.* **133**, 77–95 (2016).
66. Laskar, J. et al. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
67. Barbieri, C. et al. Unraveling the complex maternal history of Southern African Khoisan populations. *Am. J. Phys. Anthropol.* **153**, 435–448 (2014).
68. Barbieri, C., Butthof, A., Bostoen, K. & Pakendorf, B. Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *Eur. J. Hum. Genet.* **21**, 430–436 (2013).
69. Barbieri, C. et al. Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol. Biol. Evol.* **29**, 1213–1223 (2012).
70. Barbieri, C. et al. Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS ONE* **9**, e99117 (2014).
71. Batini, C. et al. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol. Biol. Evol.* **28**, 1099–1110 (2011).
72. Eaaswarkhanth, M. et al. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur. J. Hum. Genet.* **18**, 354–363 (2010).
73. 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).
74. Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl Acad. Sci. USA* **92**, 532–536 (1995).
75. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
76. Just, R. S., Diegoli, T. M., Saunier, J. L., Irwin, J. A. & Parsons, T. J. Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci. Int. Genet.* **2**, e45–e48 (2008).
77. Kujanová, M., Pereira, L., Fernandes, V., Pereira, J. B. & Cerný, V. Near eastern Neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am. J. Phys. Anthropol.* **140**, 336–346 (2009).
78. Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2**, 13 (2001).
79. Macaulay, V. et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
80. Margaryan, A. et al. Eight millennia of matrilineal genetic continuity in the South Caucasus. *Curr. Biol.* **27**, 2023–2028 (2017).
81. Olivieri, A. et al. Mitogenome diversity in Sardinians: a genetic window onto an island's past. *Mol. Biol. Evol.* **34**, 1230–1239 (2017).
82. van der Walt, E. M. et al. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. *Eur. J. Hum. Genet.* **20**, 650–656 (2012).
83. Vyas, D. N. et al. Bayesian analyses of Yemeni mitochondrial genomes suggest multiple migration events with Africa and Western Eurasia. *Am. J. Phys. Anthropol.* **159**, 382–393 (2016).

Acknowledgements We thank all of the study participants, as well as the many people who provided assistance during participant recruitment and recording, or provided critical historical, cultural and linguistic insights including; C. P. Bennett (https://evolvingpicture.com/), R. Wilkinson, J. Sinvula, H. Money, the late C. F. Heyns, R. H. Glashoff, D. de Swart, P. Fernandez, P. A. Venter, S. C. Schuster, M. P. Marx, the late S. M. Kooitjie (39th leader of the ǂAonin clan and chairperson of the Nama Traditional Leaders Association), A. A. Collins, B. Kaesje, J. Kayimbi, H. Mische, F. Naque, D. Naque, H. Oosthuizen, E. Oosthuizen, A. Oosthuysen, E. Oosthuysen, D. Roux, C. Swau and T. Tsebe. We acknowledge the late M. McFarlane, who identified Deception ridge and its importance in the evolution of the Makgadikgadi palaeo-lake. This work was supported by an Australian Research Council Discovery Project grant awarded to V.M.H. (DP170103071) and sampling contributed by the Cancer Association of South Africa to M.S.R.B. and V.M.H. A.T. and S.-S.L. received funding from the Institute for Basic Science (IBS) under IBS-R028-D1. V.M.H. is supported by the University of Sydney Foundation in a Petre Foundation chair position. Computational resources were provided by the Australian Government through the National Computational Infrastructure, the Sydney Informatics Research Hub at the University of Sydney (Artemis HPC) and by the Garvan Institute of Medical Research Data Intensive Computer Engineering team.

Author contributions V.M.H. designed the study. M.S.R.B., H.E.A.F. and V.M.H. obtained and maintain study approvals and permits, as well as community leadership support. M.S.R.B., D.C.P. and V.M.H. performed recruitments, consenting, sampling and processing. R.J.L., A.M.F.K. and D.C.P. performed pre-screening and mitogenome data generation. E.K.F.C. performed the bioinformatics and phylogenetic analyses. A.E.M. performed geographical interpretation. S.-S.L. and A.T. performed climatological model analyses and interpretation, with additional local climatology interpretation provided by H.R. V.M.H. led the interpretation of the multiple-discipline analyses, with contributions from all of the authors. E.K.F.C., B.F.B., S.-S.L., A.T. and V.M.H. generated and interpreted the figures. V.M.H., E.K.F.C. and A.T. wrote the manuscript with contributions from all of the authors.

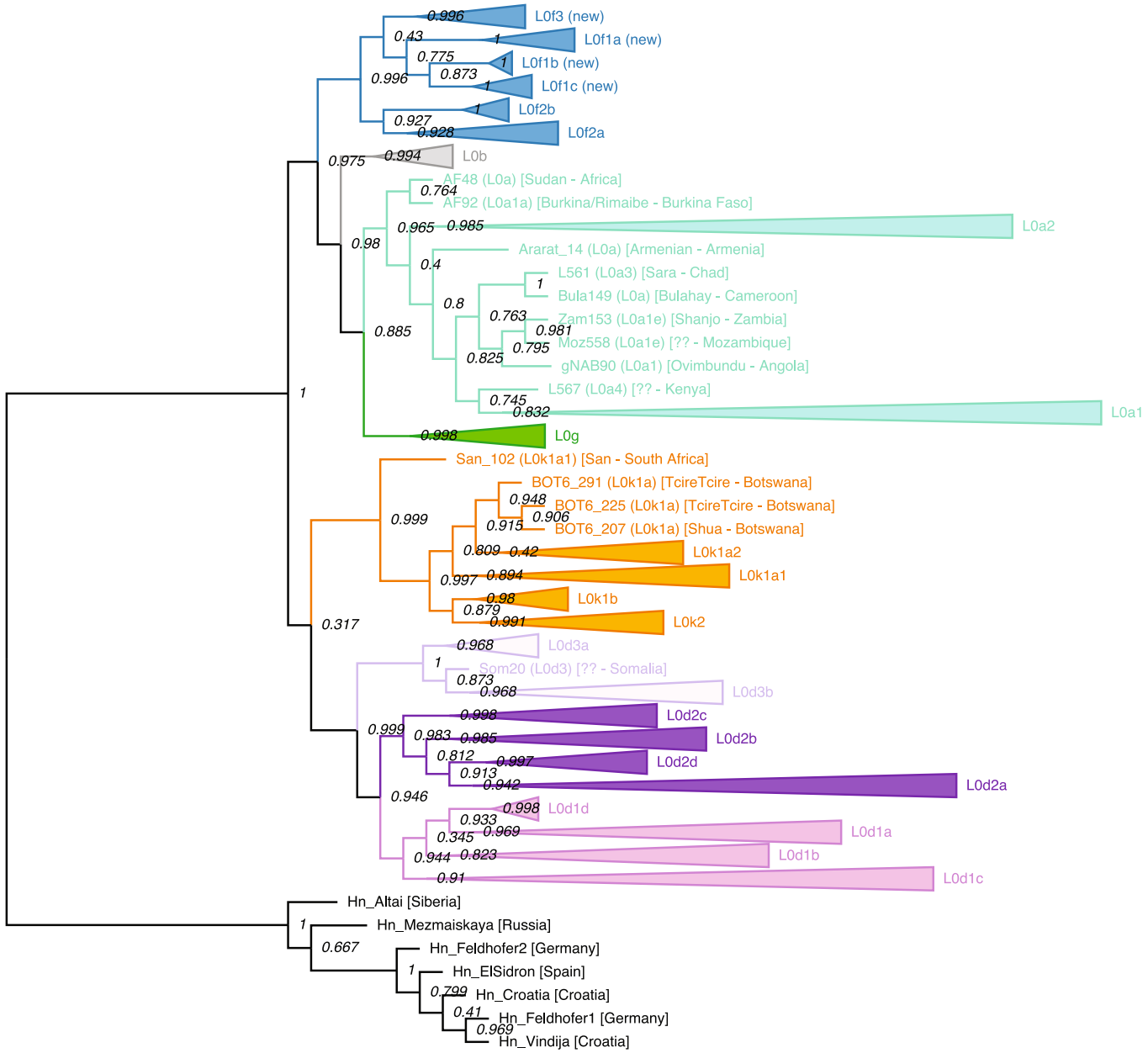
Competing interests The authors declare no competing interests.

Additional information

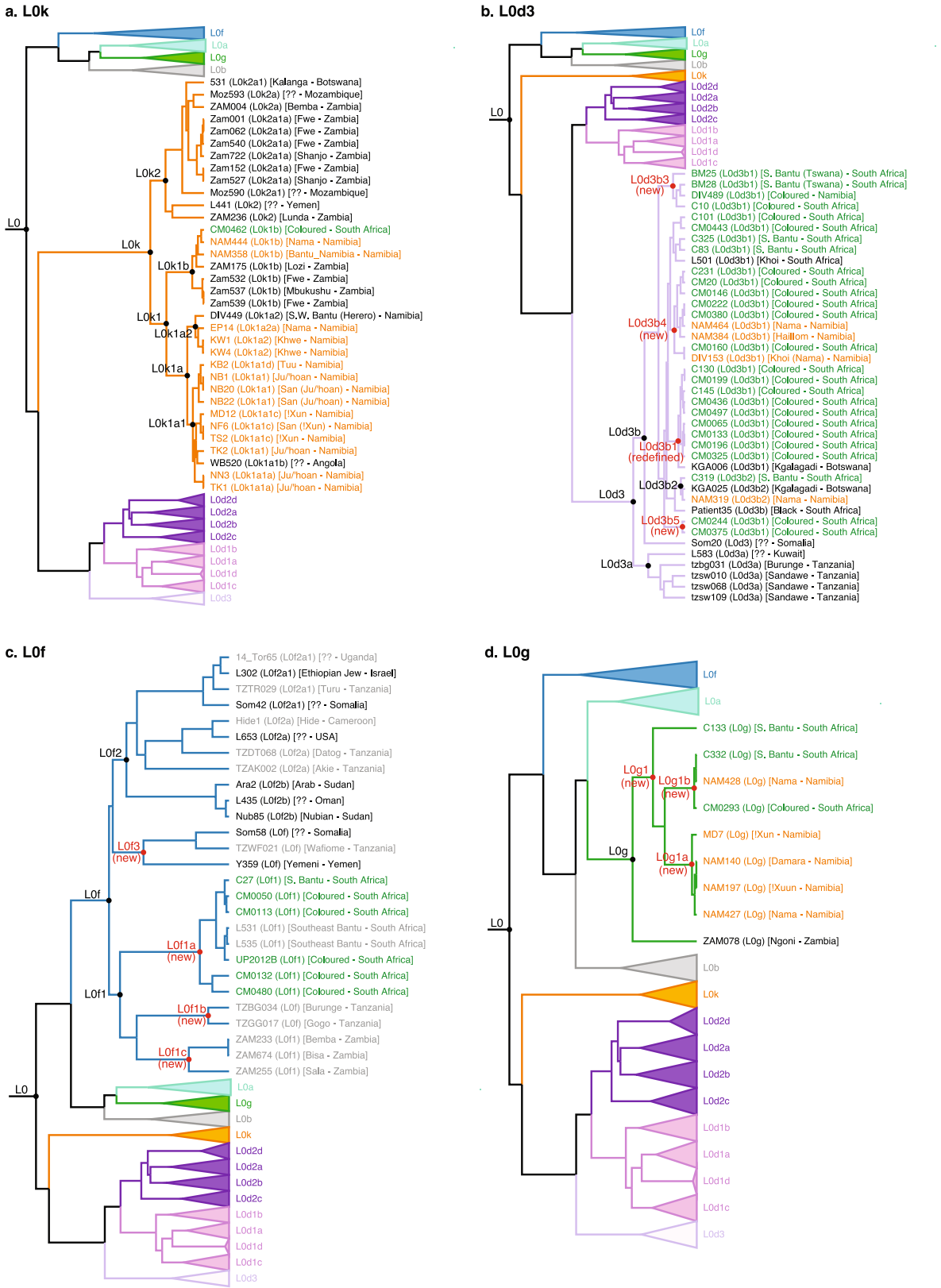
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1714-1.

Correspondence and requests for materials should be addressed to A.T. or V.M.H.

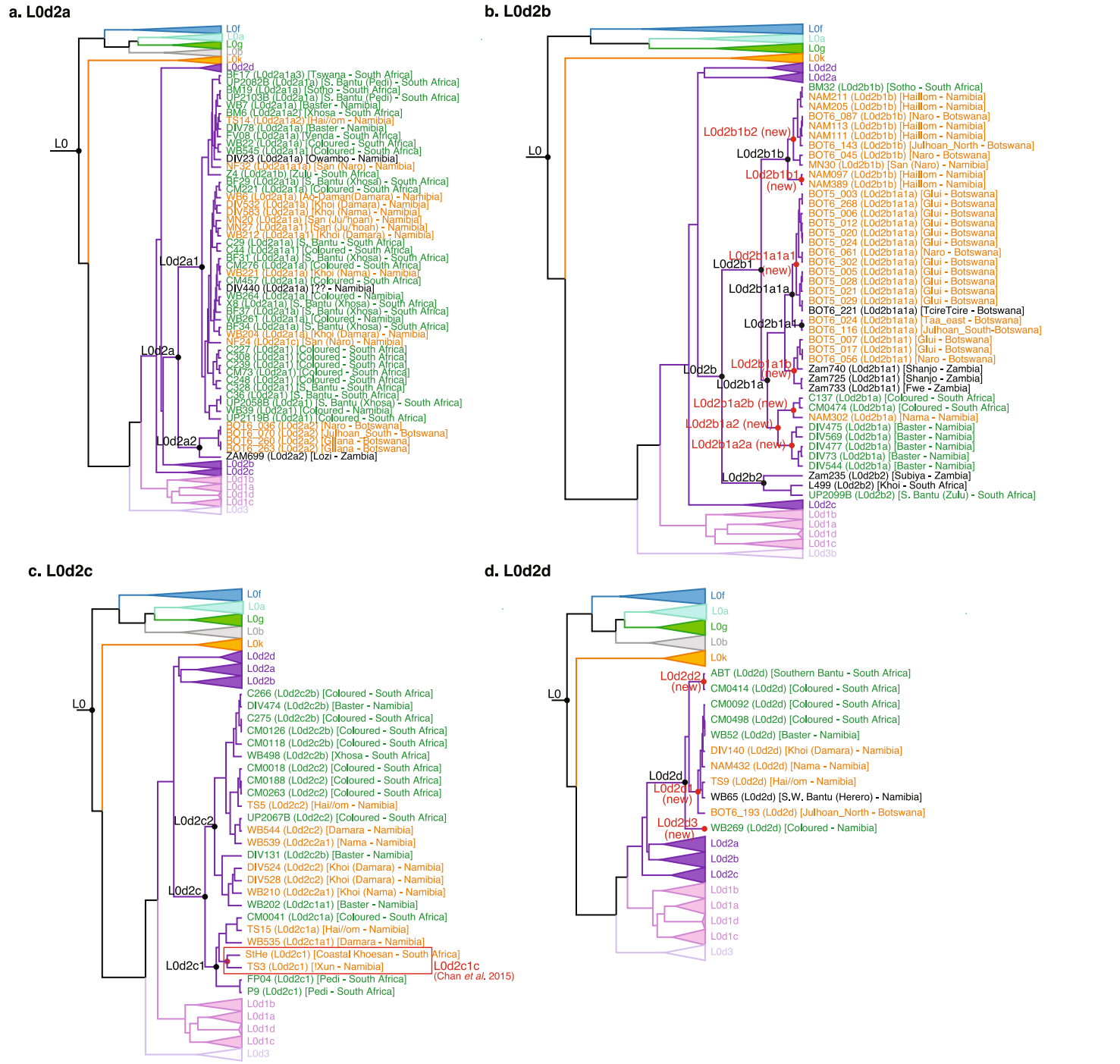
Peer review information *Nature* thanks Victor Brovkin, Rebecca Cann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Phylogenetic tree of all 1,217 L0 mitogenomes. Phylogeny was inferred using FastTree v.2.1.746, displayed using FigTree. Tips belonging to the same haplogroup are collapsed and coloured as in Fig. 2a. Local support values for each node are indicated and branch lengths are proportional to the number of substitutions per site. The tree is rooted to the seven Neanderthal mitogenomes as indicated.

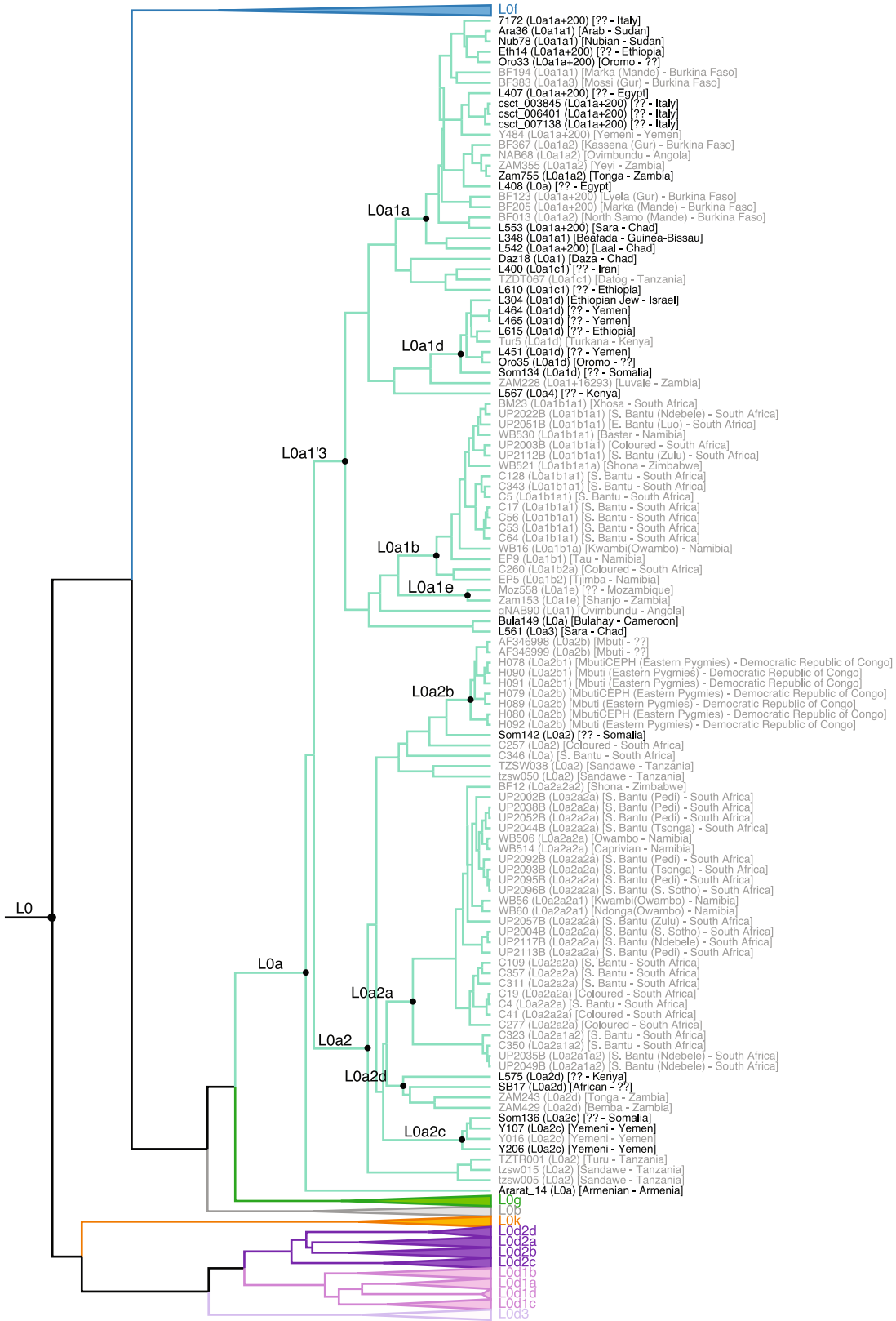


Extended Data Fig. 2 | Detailed phylogenetic branching of L0k, L0d3, L0f and L0g. a–**d**, Expanded sections of the phylogenetic tree depicted in Fig. 2a are shown, including 34 (out of a total of 113) L0k (**a**), all 40 L0d3 (**b**), all 27 L0f (**c**) and all 9 L0g (**d**) mitogenomes. Each mitogenome is represented as a tip and coloured based on their broad ethno-linguistic classification, if known. KhoeSan is shown in orange, non-KhoeSan in grey and Cape multi-ethnic (KhoeSan ancestral) in green. Publicly available mitogenomes for which we cannot be certain of their broad population identifier are labelled in black font. Proposed new sub-lineages for L0d3, L0f and L0g1 are indicated by red-coloured node labels and are further described in Supplementary Tables 7–9.

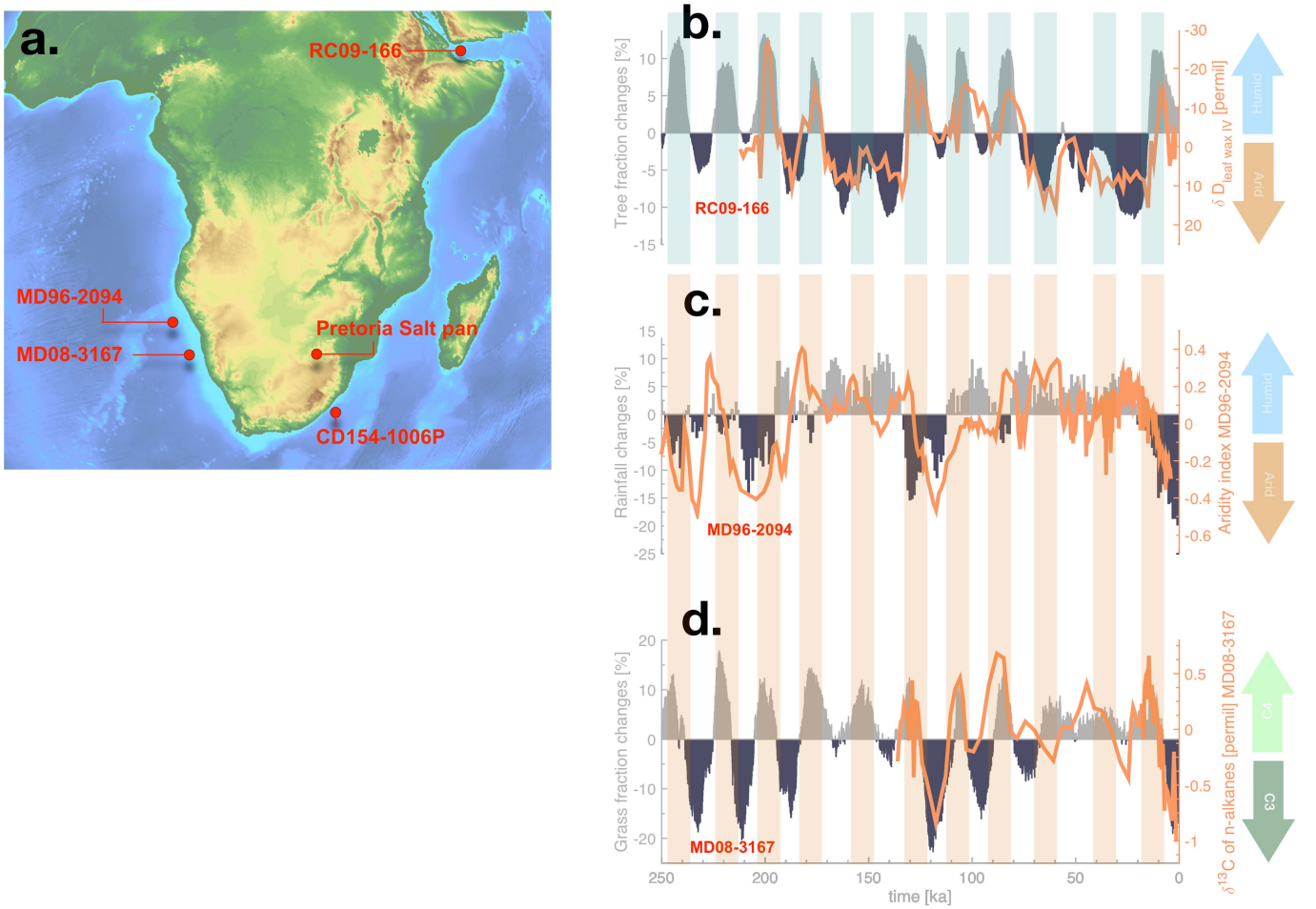


Extended Data Fig. 3 | Detailed phylogenetic branching of L0d2. **a**, **c**, **d**, Expanded branches of the phylogenetic tree depicted in Fig. 2a are shown, including 51 (out of a total of 118) L0d2a (**a**), 25 (out of 53) L0d2c (**c**) and all 11 L0d2d (**d**) mitogenomes. **b**, For L0d2b, an additional BEAST analysis was performed using an alternate subset of 441 mitogenomes that included all 43 L0d2b samples, as opposed to the $n = 461$ subset (Fig. 2a) that included only 13 L0d2b. The same model parameters were used for both data subsets. In all panels, each mitogenome is represented as a tip and coloured based on their broad ethno-linguistic classification, as in Extended Data Fig. 2. The previously defined L0d2c1c haplogroup, containing the coastal KhoeSan StHe skeleton[6] and other newly proposed sub-lineages are indicated by red node labels (Supplementary Tables 4–6).

Extended Data Fig. 4 | Detailed phylogenetic branching of L0d1. **a**–**c**, Expanded branches of the phylogenetic tree depicted in Fig. 2a are shown, including 54 (out of a total of 91) L0d1a (**a**), 45 (out of 174) L0d1b (**b**) and 33 (out of 184) L0d1c (**c**) mitogenomes. Each mitogenome is represented as tips and coloured based on their broad ethno-linguistic classification as in Extended Data Fig. 2.



Extended Data Fig. 5 | Detailed phylogenetic branching of L0a. The L0a branch of the phylogenetic tree displayed in Fig. 2a is shown, which includes a subset of 114 (out of a total of 294) L0a mitogenomes. Each mitogenome is represented as tips and coloured based on their broad ethno-linguistic classification as in Extended Data Fig. 2.



Extended Data Fig. 6 | Comparison of the palaeo-data and palaeo-model. **a**, Locations of key sites that are used for the comparison of the palaeo-model and palaeo-data in this study are highlighted in red. The map was generated in Paraview v.5.6 (https://www.paraview.org/). **b**, Simulated tree fraction (%) at Horn of Africa (land grid points nearest to RC09-166) (grey, dark-blue bars) and stable hydrogen isotopic composition of leaf wax, corrected for ice volume contributions from the Gulf of Aden marine sediment core RC09-166[30] (orange), indicating changes in hydroclimate. **c**, Relative precipitation changes (%) simulated by LOVECLIM transient model (all forcings) for 11° E, 19° S (grey, dark-blue bars) and grain-size aridity index reconstructed from sediment core MD96-2094[32] (orange). **d**, Grass fraction changes simulated by LOVECLIM transient model (all forcings) at 11° E, 14–17° S (grey, dark-blue bars) and reconstructed $\delta^{13}C$ changes of *n*-alkanes (orange) (South Atlantic sediment core MD08-3167) indicative of abundance of C_3 and C_4 plants in the Namibian desert and further inland[33].

Extended Data Table 1 | L0 mitogenomes included in this study

Study Reference	Number of L0 Mitogenomes
Barbieri *et al.*, **AJHG** 2013[27]	485
Barbieri *et al.*, **AJPA** 2014[65]	26
Barbieri *et al.*, **EJHG** 2013[66]	33
Barbieri *et al.*, **MBE** 2012[67]	10
Barbieri *et al.*, **PlosOne** 2014[68]	115
Batini *et al.*, **MBE** 2011[69]	9
Behar *et al.*, **AJHG** 2008[1]	64
Chan *et al.*, **Plos One** 2015[6]	77 *
Eaaswarkhanth *et al.*, **EJHG** 2009[70]	3
Gonder *et al.*, **MBE** 2007[71]	27
Herrnstadt *et al.*, **AJHM** 2002[55]	2 †
Horai *et al.*, **PNAS** 1995[72]	1 ‡
Ingman *et al.*, **Nature** 2000[73]	5
Just *et al.*, **FSIG** 2008[74]	8
Kivisild *et al.*, **Genetics** 2006[54]	25 †
Kujanova *et al.*, **AJPA** 2009[75]	4 §
Maca-Meyer *et al.*, **BMC Genet** 2001[76]	1
Macaulay *et al.*, **Science** 2005[77]	1
Margaryan *et al.*, **Curr Biol** 2017[78]	1
McCrow *et al.*, **Prostate** 2016[47]	37 *
Morris *et al.*, **GBE** 2014[15]	1 *
Olivieri *et al.*, **MBE** 2017[79]	4
Rito *et al.*, **Plos One** 2013[3]	42
Schuster *et al.*, **Nature** 2010[56]	6 *
van der Walt *et al.*, **EJHG** 2012[80]	20
Vyas *et al.*, **AJPA** 2016[81]	12
Total Public	1,019
This Study	198
Total L0 Mitogenomes	**1,217**

Numbers of mitogenomes taken from previously published[67-83] studies.

*Previously published data by our group with verified population metadata.

†Mitochondrial DNA sequences of the coding-region only.

‡Sequence has non-canonical start position corresponding to position 577 of rCRS.

§Coriell cell lines.

Extended Data Table 2 | KhoeSan population identifiers used in this study

Broad identifier	Geographic identifier	Broad language group	Ethno-linguistic identifiers
KhoeSan	Kalahari KhoeSan	Kw'a	Ju/'hoansi; !Xun; !Xuun; Hoan;
		Tuu	/Hoa; Taa
		Khoe-Kwadi (central)	Naro; ǁAni; Buga; Gǁana; Gǀui; Khwe; Shua; Tshwa; ǁXokhoe
	West-coastal KhoeSan	Khoe-Kwadi (Nama)	Nama; Damara; #Anonin; Ao-Daman; Haiǁom; Orlams
KhoeSan ancestral	Cape KhoeSan	Afrikaans	Baster; Coloured
	East-Coastal KhoeSan	Southern Bantu	Ndebele; Pedi; Tswana; Tsonga; Venda; Xhosa; Zulu

nature research

Corresponding author(s): Vanessa M. Hayes

Last updated by author(s): 2019/09/04

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The nucleotide sequences of 1,019 L0 mitogenomes were downloaded from NCBI
Data analysis	All software used is described and referenced in the METHODS section. No new software algorithms or code was established. These include: Mitogenomic and phylogenetic analyses: Consensus mitogenome sequences were derived by first identifying variants relative to rCRS using samtools (v1.3.1) mpileup (with parameters -d 10000 -L 1000 -Q 7 -h 50 -o 10 -e 17 -m 4) and bcftools (v1.3.1). Variant data was converted to fasta format using samtools' vcfutils.pl vcf2fq. Mitochondrial haplogroups were evaluated (called) using HaploGrep2 (v2.1.13) based on PhyloTree Build 17. Multiple sequence alignment (for phylogenetic inference) was performed using FastTree v2.1.7 (SSE3) using the generalised time reversible (-gtr) and discrete gamma model with 20 rate categories (-gamma). Bayesian phylogenetic inferences and divergence times were calculated using BEAST2 v2.4.2 with BEAGLE 2.0. Multiple sequence alignment of the subset of 461 AMH and seven Neanderthal mitogenomes was converted to NEXUS format using the convert function of seqmagick v0.6.1 (https://fhcrc.github.io/seqmagick) with parameter --alphabet dna-ambiguous. This provided the input to BEAST2. Specifically, BEAUTi v2.4.2 was used to set up the phylogenetic model, assuming: (i) the Gamma Site Model with 6 gamma categories and no invariant sites; (ii) the generalized time reversible substitution model; (iii) a strict constant clock model with a normal prior of with μ = 1.665 x 10-8 and σ = 1.479 x 10-9 based on Soares et al. 200960; and (iv) a Coalescent Constant Population. Five BEAST replicates were performed, each with 100 million MCMC iterations, sampling every 10,000. Tracer v1.6 was used to evaluate BEAST trace files. Replicates were combined using LogCombiner v2.4.2. Sampled trees from BEAST were summarized into a single Maximum Clade Credibility target tree using TreeAnnotator v2.4.2 for each of the five replicates, discarding the first 10% as burn-ins. FigTree v1.4.2+ was used to visualize all resulting trees. Bayesian Skyline Plot (BSP) analysis: For each haplogroup of interest (e.g. L0a, L0d1'2, and L0k), a nexus file was derived using SeqMagic v0.6.1 as described above. BSP analyses were performed using BEAST2, using BEAUTi 2 for model setup as before, with the following key differences: (i) the gamma shape of the Gamma Site Model was estimated with an exponential prior with mean = 1.0 and offset = 0.0; (ii) the molecular clock was fixed (not estimated) at 1.665 x 10-8 based on Sores et al.60; and (iii) the phylogenetic tree prior was set to

Coalescent Bayesian Skyline, assuming 20 intervals between the root of the tree and the present time. Tracer v1.6 was used to reconstruct the Bayesian Skyline from the sampled trees for each analysis, using a stepwise constant variant and the lower 95% highest posterior density of the root height as the maximum time.

Climate model: LOVECLIM earth system model

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The consensus sequences for this set of 198 mitogenomes have been deposited to NCBI with Accession Numbers MK248274-MK248471.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed as all mitogenomes representing a rarer L0 haplogroup were included. Pre-screening was performed on roughly 500 population relevant samples, identifying 198 of interest to undergo complete mitogenome sequencing and inclusion in this study. All relevant published data was also downloaded for a total study of 1,217. Sample sizes are therefore ALL inclusive (no rare L0 lineage and mitogenome was excluded) and therefore sufficient.
Data exclusions	All 1,217 mitogenomes were used to determine haplogroup frequencies and reconstruct geographic dispersals. A focused subset of 461 mitogenomes, including all rare lineages, were used to establish within L0 coalescence times. The smaller subset was necessary due to the computational burden of the BEAST analysis.
Replication	L0 haplogroup pre-screening (2,673 bp region) using Sanger sequencing for the 198 mitogenomes that underwent Ion Torrent deep complete mitogenome sequencing, providing an internal experimental validation. All samples were validated.
Randomization	Not relevant to this study as grouping (haplogroups) were identified via mitogenomic data. KhoeSan geographic classification were based on self-identification (ethnic classifications) of participating subjects and country of origin.
Blinding	Investigators were blinded to the subject identifiers during analysis, as were the climate physicists blinded to the hypotheses and distributions associated with the mitogenomic data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	NA - no specimens were collected for the purpose of this study - only published data used.
Specimen deposition	NA - the study involved published data
Dating methods	No new specimens or dates are provided or used.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Southern Africans from Namibia and South Africa representing and self-identifying as ancestrally from a KhoeSan or KhoeSan ancestral population identifiers (as outlined within the METHODS) were recruited. There was no gender bias and all participants were greater than 18 years of age as per ethical requirements.
Recruitment	Participants were recruited based on their self-identified ethnicity. Recruitment was led by local researchers and coauthors (MSRB, DCP or VMH in South Africa) or in Namibia (HATF, as well as VMH). The study was explained to the participants and communities and in some communities, especially contemporary Namibian Kalahari KhoeSan populations, this took place over an extensive period, with regular engagement with the communities by VMH over a 10 year period. In these communities recruitment was a decision made by the entire community. All recruiters are familiar with local languages and cultures. There was no other biases that would impact this study. All possible populations that may carry a L0 mitogenome and live within the borders of Namibia and South Africa and were willing to participate, were included.
Ethics oversight	The study was reviewed and approved by the Ministry of Health and Social Services (MoHSS) in Namibia (#17-3-3 2008, 2014 and 2019), with additional local approvals from community leaders, the University of Pretoria Human Research Ethics Committee (HREC #43/2010 and HREC #280/2017), including US Federal-wide assurance (FWA00002567 and IRB00002235 IORG0001762), as well as the South African National Blood Service (SANBS) HREC (HREC #2012/11). Isolated DNA was shipped under the Republic of South Africa Department of Health Export Permit (#J1/2/4/2), in accordance with the National Health Act 2003, to the Garvan Institute of Medical Research in Australia. Mitogenome sequencing was performed in accordance with site-specific approval granted by St Vincent's Hospital HREC in Australia (SVH 15/227).

Note that full information on the approval of the study protocol must also be provided in the manuscript.