Nature LETTER - CONFIDENTIAL

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

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], Hannes Rautenbach[6,7], Hagen E.A. Förtsch[8], M.S. Riana Bornman[7], & Vanessa M. Hayes[1,2,7,9,10]

[1]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia.

[2]St Vincent's Clinical School, University of New South Wales, Randwick, NSW, 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, South Africa.

[7]School of Health Systems and Public Health, University of Pretoria, South Africa.

[8]Windhoek Central Hospital, Windhoek Khomas, Namibia.

[9]Faculty of Health Sciences, University of Limpopo, Turfloop Campus, South Africa.

[10]Central Clinical School, University of Sydney, Camperdown, NSW, Australia.

Anatomically modern humans (AMHs) arose in Africa some 200 thousand years ago (ka)[1–4]. While some of the oldest skeletal remains suggest an eastern African origin[2], southern Africa is home to contemporary populations representing the earliest branch of human genetic phylogeny[5,6]. Generating the largest resource for the poorly represented and deepest-rooting maternal L0 mitochondrial DNA branch (198 new for a total of 1,217 mitogenomes), we show geographic isolation of L0d1'2, L0k and L0g KhoeSan descendants south of the Zambezi River. Establishing mitogenomic timelines, frequencies and dispersals, we propose L0 emerged within the residual Makgadikgadi-Okavango palaeo-wetland of southern Africa[7], some 200 ka (95% CI: 165-240 ka). Genetic divergence points to sustained 70,000 year-long existence before an out-of-homeland northeast-southwest dispersal, between 130 and 110 ka. Palaeo-climate proxy and model data suggest increased humidity opened green corridors, first to the northeast then southwest. Subsequent homeland drying corresponds with sustained effective population size (L0k), while wet-dry cycles and likely adaptation to marine foraging, allowed the southwest migrants to achieve population growth (L0d1'2), as supported by extensive south-coastal archaeological evidence[8–10]. Pinpointing a southern African human origin, we describe a sustained homeland occupation prior to the first human migrations, providing evidence for climate-driven dispersals.

Southern Africa has been a long-held contender for the origin of AMHs. Home to contemporary populations representing our 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, including the out-of-Africa ancestral L3 sub-branch (or haplogroup), and the rare deep-rooting L0. L0 is predominated by southern African haplogroups: L0d, L0k, and recently described L0g[6]. In contrast, the rare L0f and common L0a are dispersed throughout sub-Saharan Africa[1,3,6].

Through L0 pre-screening, we identified 198 southern Africans with poorly represented haplogroups to undergo mitogenome sequencing (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 click-speaking foragers, or non-KhoeSan. Non-KhoeSan with KhoeSan-derived L0 mitogenomes, are referred as KhoeSan-ancestral, with further geographical classification (**Fig. 1b** and Extended Data Table 2). Contemporary KhoeSan include Kalahari (Kw'a, Tuu and central Khoe-Kwadi speakers) and West-coastal (Khoe-Kwadi Nama speakers) KhoeSan[11]. Southern Bantu speaking peoples who migrated down the east coast of Africa some 1,500 years ago, may have acquired an East-coastal KhoeSan heritage[12]. The arrival of European colonists to the Cape in mid-1600's gave rise to the South African Coloured and Namibian Baster populations, acquiring a Cape KhoeSan heritage[13]. Excluding the east African click-speaking Sandawe and Hadza, indigenous KhoeSan populations appear to be absent northeast of the Zambezi River, supported by lack of skeletal remains representing KhoeSan-like hunter-forager morphology[14]. We classified the 198 new mitogenomes as: Kalahari (n=18), West-coastal (n=21), Cape (n=109), and East-coastal (n=29) KhoeSan, or non-KhoeSan (Bantu, n=19), while two were unknown. Using these identifiers, we have provided a best-fit classification for all 1,217 L0 mitogenomes (Supplementary Table 2).

Phylogenetic analysis confirms major L0 haplogroups, with the exclusion of L0b (Extended Data Fig. 1). Using a subset of 461 mitogenomes, including all rare lineages, we establish within L0 coalescence times (**Fig. 2a** and Supplementary Table 3) and use the complete dataset to reconstruct geographic dispersals (**Fig. 2b**). We redefine L0 emergence ~50-25 thousand years (kyr) prior to previous estimates[1,6], at ~200 ka (95% CI: 165-240 ka). L0d'k (n=309 ~187 ka) is largely KhoeSan-specific, emerging ~20 kyr prior to the widely dispersed L0a'b'f'g sister-branch (n=152 ~164 ka). Although exact branch resolution for L0k

remains undetermined, we observe a preference for L0d'k (posterior probability ~0.6) over L0a'b'f'g'k (~0.4). Irrespective, L0k (n=113) appears to remain stable for ~130 kyr before diverging into the Kalahari-specific L0k1, predominated by L0k1a (85 of 94), and rarer L0k1b and L0k2 lineages distributed around the Zambezi (Extended Data Fig. 2a). L0d remains stable for ~58 kyr before splitting into the KhoeSan-specific L0d1'2 and rarer L0d3.

Coalescing ~113 ka, L0d2 (n=226) emerges ~15 kyr prior to L0d1 (n=452). Within L0d2 (~91 ka), L0d2c is the earliest diverged (n=53 ~84 ka) with broad and almost even KhoeSan-regional distribution (Extended Data Fig. 3 and Supplementary Table 4). In 2014, we derived an ancient L0d2c1c mitogenome from the skeleton of a 2,330(±25)-year-old Cape-coastal marine forager (StHe/UCT606)[15]. Predating archaeological evidence for sheep-herding in the region[12,16], we proposed this L0d2c sub-clade represent a pre-pastoral indigenous southern African lineage. Recently, whole genome sequencing confirmed a unique southern African heritage, while two younger (<2k years) Cape skeletons showed a genetic link to east Africa and associated pastoralist migration[17]. Observing a bias towards Kalahari representation for L0d2b (28/44 ~65 ka) and L0d2a (62/118 ~60 ka), doubling the contribution for L0d2d (6/11) we show broad KhoeSan distribution (Extended Data Fig. 3 and Supplementary Tables 5,6). As with L0d2, L0d1 is spread throughout the KhoeSan-regional identifier, with notable over-representation including: L0d1b (104/174 ~69 ka) and L0d1c (151/184 ~59 ka) within the Kalahari, and L0d1a (32/91 ~44 ka) within the Cape (Extended Data Fig. 4). We contribute two new KhoeSan-ancestral L0d1d mitogenomes to the single published[6].

In contrast to L01'2, L0d3 is not specific to southern Africa. While L0d3b (~30 ka) appears to be KhoeSan-specific, the rarer L0d3a (~42 ka) is exclusively found north of the Zambezi. Notably, three of six L0d3a mitogenomes were derived from click-speaking east African Sandawe. Our data supports previous studies suggesting a genetic link between east Africa and the earliest southern Africans[17], last sharing a common ancestor ~59 ka. Providing

99 the largest mitogenome contribution to L0d3 (27/40), we observe over-representation of L0d3b
100 in the Cape KhoeSan identifier (21/34), with notable absence within the Namibian Baster
101 population (Extended Data Fig. 2 and Supplementary Table 7). Previously suggesting a
102 maternal KhoeSan-ancestral distinguishing identifier for the Coloured and Baster
103 populations[13], we show L0d3b to be specific to the Coloured and the new L0d2b1a2a sub-clade
104 to the Baster population (Extended Data Fig. 3b).

105 Within L0a'b'f'g, L0f is highly divergent (~125 ka 95% CI: 101-149 ka). Adding five
106 L0f mitogenomes, L0f1 (13/27 ~113 ka) predominates south and L0f2'3 (14/27 ~121 ka) north
107 of the Zambezi (Extended Data Fig. 2 and Supplementary Table 8). Within L0f1 we recognise
108 three new branches: the northeast sister-clades L0f1c (Zambian) and L0f1b (Tanzanian), and
109 the East-to-Cape-coastal southern clade L0f1a (n=8). Lack of L0f representation within
110 contemporary KhoeSan, suggests L0f1a presence within southern Africa is likely a result of
111 more recent east-coastal agropastoral back-migration. While L0a'g coalesce ~117 ka (95% CI:
112 94-145 ka), contributing 19 southern African to 347 L0a mitogenomes, we concur that L0a
113 likely diverged northeast of the Zambezi (~85 ka) and spread throughout Africa[3], with southern
114 L0a1b and L0a2a representation a result of Bantu back-migration (Extended Data Fig. 5). First
115 described within a Kw'a speaking hunter-gatherer[6], we contribute three new and reclassify five
116 published mitogenomes as L0g (Extended Data Fig. 2 and Supplementary Table 9). With broad
117 KhoeSan and KhoeSan-ancestral distribution, we hypothesize that L0g diverged southwest of
118 the Zambezi (~69 ka), akin to L0d1'2.

119 Our data suggests that the Greater Zambezi River Basin region, particularly the
120 southwest Kalahari, played a significant role in shaping AMH emergence and prehistory.
121 Today a semi-desert, salt pans within northern Botswana represent desiccated vestiges of
122 Palaeo-Lake Makgadikgadi, which at its peak would have been the largest lake in Africa,
123 flanked by smaller lakes from the Upper Zambezi to the Kafue Rivers[7,18]. Contraction of the

lake, accompanied by development of the Okavango Delta, as a result of neotectonic rifting, would have created a residual wetland favourable for human and broad mammal habitation[19] (**Fig. 2c**). Today the harsh Kalahari climate and oxygen-rich salt pans are not ideal for fossil and pollen preservation, respectively. However, lithic artefacts representing the Middle to Late Stone Age, are documented from the Makgadikgadi Pans and surroundings[7,20,21], while palynology suggests a once grassland and forest biome[22]. Our data further suggests the Makgadikgadi-Okavango palaeo-wetland sustained AMH existence for ~70 kyr, supported by mitochondrial data for ancestral giraffe, lion and zebra[23–25], before out-of-homeland migrations split founder homeland populations L0d, L0f and L0a'g.

Southwest of the homeland L0d1'2 experienced episodic splits, with broad south-coastal occupation, while L0g is less successful. Bayesian Skyline Plot (BSP, **Fig. 2c insets**) analysis confirms effective population growth for L0d1'2, while extensive archaeological evidence suggests cognitive human behaviour at the southern tip of Africa between ~100-60 ka[8–10] with associated increase in the density of coastal over inland time-appropriate archaeological sites[26]. Northeast of the homeland, L0d3 and L0f are less successful, while L0a underwent significant population trajectory post-dating the out-of-Africa migration (BSP L0a). The northeast migration route is supported by the appearance of data-appropriate archaeological sites[26]. Within the homeland, L0k sustained a constant effective population size (BSP L0k), as did Kalahari-predominant L0d2b, L0d2a and L0d1c. While the presence of L0k in Zambia has been suggested to represent contact with an ancient pre-Bantu population[27], we propose these rare lineages represent an ancient out-of-homeland KhoeSan offshoot.

Orbitally-driven large-scale hydroclimate variations have been proposed as a pacemaker for early human migrations[28,29], with wetter conditions and resulting 'green corridors' invoked to explain the out-of-Africa migration ("pull") or drier conditions and resulting food shortages forcing dispersals ("push")[30]. To determine whether our predicted

149 homeland isolation and major dispersals may have been driven by climate shifts, we analysed
150 four key palaeo-hydroclimate datasets[29,31–33], along with a transient 784 kyr-long
151 glacial/interglacial simulation conducted with the LOVECLIM earth system model[28] (**Fig. 3**).
152 Although limited by available palaeo-proxy records and a climate model of intermediate
153 complexity, we observed a considerable degree of coherence on orbital timescales (Extended
154 Data Fig. 6). During the homeland period (200-130 ka), palaeo-data links the 21,000 year-long
155 precession cycle (**Fig. 3a**), with three wet-dry cycles (**Fig. 3b**). Conversely, climate model
156 simulates an extended drought, owing to a more pronounced eccentricity signal (**Fig. 3e**),
157 proposing a wetland oasis in an otherwise vast harsh environment.

158 During the out-of-homeland period (130-110 ka), our model simulation supports humid
159 conditions to the northeast facilitated the first dispersals, concurring with L0f coalescence
160 (~125 ka) (**Fig. 3d**). In contrast, the region southwest of the homeland experienced a ~15 kyr-
161 long megadrought before an orbital shift created favourable humid conditions for L0d1'2
162 dispersal (~113 ka) (**Fig. 3f**), as supported by palaeo-data (**Fig. 3c**). This is also around the
163 time the northeast L0a and southwest L0g migrants last share a common ancestor (~117 ka).
164 During the last glacial period (~100-11 ka), we observe reduced amplitude of orbital-scale
165 hydroclimate changes and overall drying within the homeland (**Fig. 3b**), while the southwest
166 coastal hydroclimate was dominated by precessional variability and relatively agreeable
167 environmental conditions (**Fig. 3c,f**). Interestingly, periods of slowdown and acceleration in
168 effective population size estimates for L0d1'2, coincide with regional changes in hydroclimate,
169 further linking climate, population size and evolution.

170 We propose the Makgadikgadi-Okavango palaeo-wetland as the possible homeland for
171 AMHs. Although one cannot exclude the possibility of a polycentric origin[34], this deltaic-
172 lacustrine ecosystem would have provided an ideal geographic locality for the evolution and 70
173 kyr sustained existence for AMHs deepest-branching maternal founder population. 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 model simulation (100-80 ka), would have created a "push' scenario, where reduced land carrying capacity would have increased pressure to seek out climatically more favourable regions. We speculate, the southwest migrants maintain a successful coastal forager existence, while the northeast migrants, as with the later branching L1'6, gave rise to ancestral pastoral and farming populations. A recent publication, provides further mitochondrial evidence to support the northeast out-of-homeland migration route and expansion into east Africa around 70-60 ka[36]. Revealing a southern African homeland for L0 emergence and extended subsistence, we propose an out-of-homeland migration event, likely driven by astronomically-induced regional shifts in hydroclimate, shaped present-day ethnic/genetic diversity.

REFERENCES

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. Palaeoanthropology: 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.* **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. *South 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. 'Khoisan' linguistic classification today. in *Current Issues in Linguistic Theory* (eds. Güldemann, T. & Fehn, A.-M.) **330**, 1–40 (John Benjamins Publishing Company, 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.e21 (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. The Delta hypothesis: Hominoid ecology and Hominin origins. in *Interpreting the past: Essays on Humans, Primates and Mammal Evolution* (eds. Pilbeam, D. R., Lieberman, D., Smith, R. J. & Kelley, J.) 231–242 (Brill Academic Publishers, 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**, (2014).
24. Pedersen, C.-E. 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**, (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 Pleistoscene record of the Palaeolake Makgadikgadi system. *Quaternary Sci. Rev*. **28**, 1392-1411 (2019).

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**, (2019).

Supplementary Information is available in the online version of this paper.

Acknowledgements We thank all 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 (www.evolvingpictures.com), R. Wilkinson and J. Sinvula (Namibian Blood Transfusion Services), H. Money (Western Cape Blood Transfusion Services, South Africa), the late C.F. Heyns, R.H. Glashoff, D. de Swart and P. Fernandez (University of Stellenbosch, South Africa), P.A. Venter (University of Limpopo, South Africa), S.C. Schuster (Penn State University, U.S.A.), M.P. Marx (Unistel Medical Laboratories, South Africa), and local Namibians the late Chief S.M. Kooitjie (39th leader of the #Aonin clan and chairperson of the Nama Traditional Leaders Association (NTLA)), 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 major contribution to Kalahari geomorphology made by the late Dr. M. McFarlane, which included identification of the Deception Ridge and its significance for the evolution of the Makgadikgadi Palaeo-lake. This work was supported by an Australian Research Council (ARC) Discovery Project grant awarded to V.M.H. (DP170103071) and sampling contributed by the Cancer Association of South Africa (CANSA) 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. The high-resolution CESM simulations were conducted on the Aleph supercomputer at the ICCP/IBS supercomputing facility.

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 geographic 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 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 the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Current affiliation for D.C.P. is The Centre for Proteomic and Genomic Research, Cape Town, South Africa and for H.R. is Akademia, Johannesburg, South Africa. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to V.M.H.

(v.hayes@garvan.org.au) or A.T. (timmermann@pusan.ac.kr) and requests for materials to V.M.H.

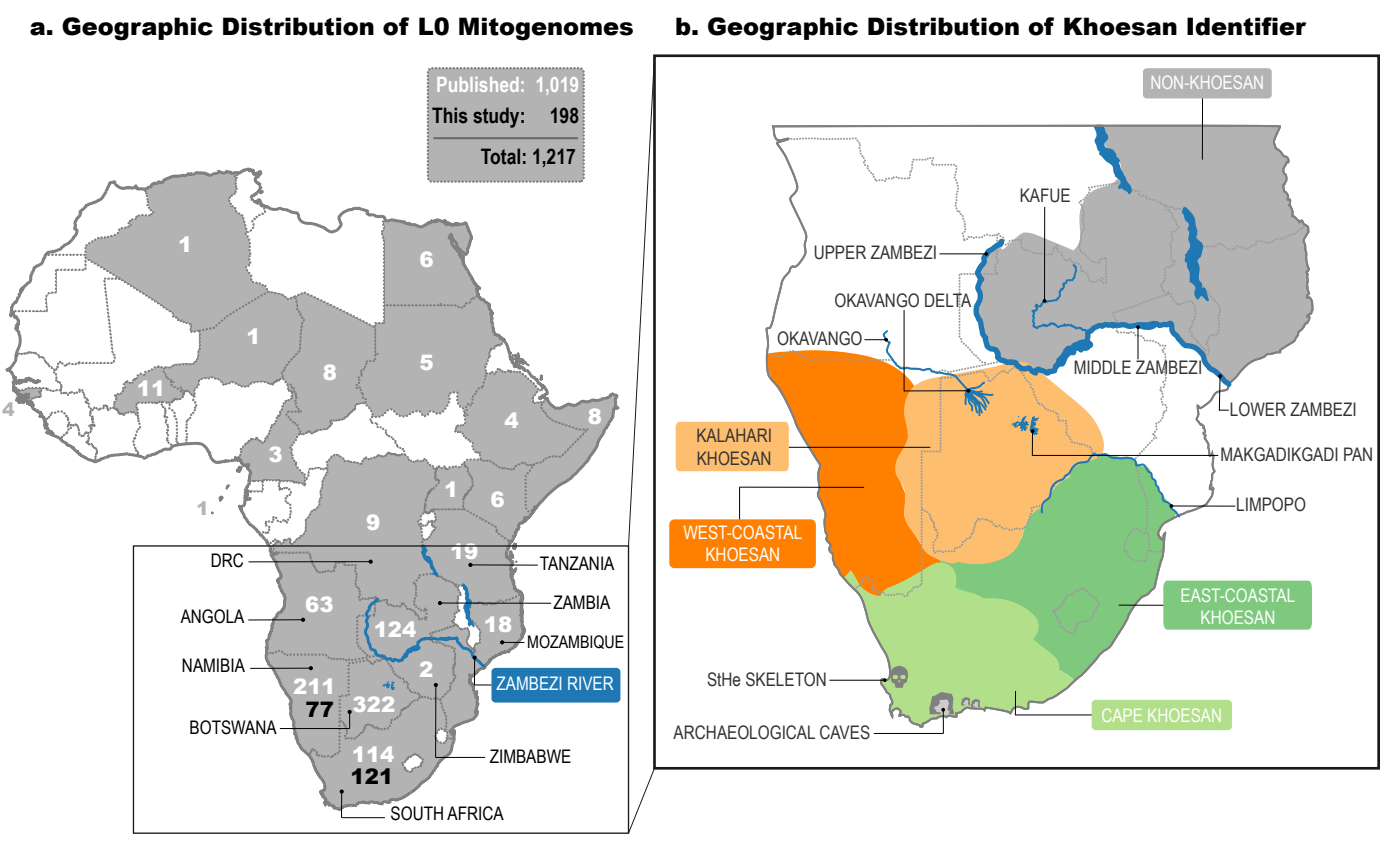
Figure Legends

Fig. 1. Geographical distribution of 1,217 L0-mitogenomes. a, Countries within (n=1,139) or outside (n=78) Africa from where L0-mitogenomes were sourced, including 198 new L0-mitogenomes (black numbers). **b,** Present day southern Africa with geographical distribution of KhoeSan population identifiers defined as; KhoeSan (orange): Kalahari and West-coastal, or KhoeSan-ancestral (green): Cape or East-coastal. The Zambezi River provides a geographic divide between the KhoeSan and largely non-KhoeSan population identifier.

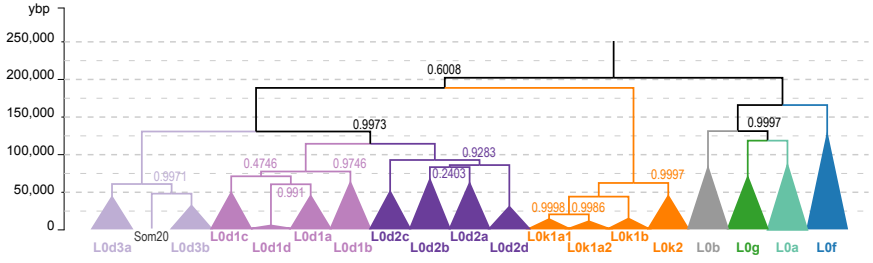
Fig. 2. L0 Phylogenetic tree, major southern African L0-haplogroup geographic distributions and out-of-homeland L0 dispersal routes. a, Phylogenetic branching and coalescence times derived from a focused subset of 461 L0-mitogenomes, including all rare branches, and anchored to Neanderthal (n=7). The Somalian-derived (Som20) L0d3 mitogenome[3] could not be assigned. **b,** Geographic distribution (identifiers described in Fig. 1b) for all KhoeSan-specific mitogenomes (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 geographic representation (shaded regions), with region specific overflow represented by total number of mitogenomes, including country specific representation north of the Zambezi. **c,** Schematic map of southern Africa representing the Makgadikgadi-Okavango palaeo-wetland sustained AMH homeland (200-130 ka), supported by archeological[7] and genetic wildlife data[23-25]. The out-of-homeland migration (130-110 ka), results in L0d, L0a'g and L0f splitting. L0d3, L0a and L0f migrate in a northeast direction, L0d1'2 and L0g migrate southwest, while L0k remains in the homeland. Bayesian Skyline Plot

(BSP) analyses [insets] for major L0-haplogroups over time predicts: maintenance of the homeland L0k population (orange), population growth for the broadly dispersed southwest L0d1'2 migrants (purple), supported by archaeological evidence (100-60 ka)[8–10] and StHe mitogenome[15], while population growth for the northeast L0a migrants coincides with the out-of-Africa migration (aqua).

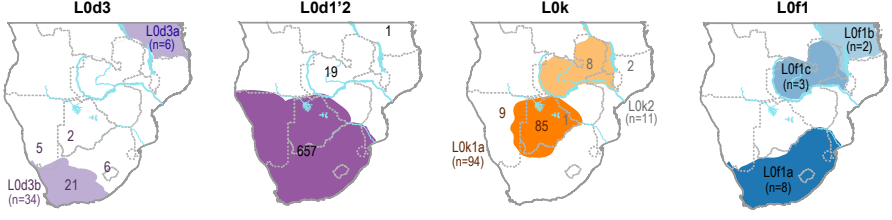
Fig. 3. Reconstructed (left panel) and simulated (right panel) climatic conditions during the out-of-homeland migration. a, Austral summer insolation changes (blue) at 27°S. **b**, Eastern and central southern Africa hydroclimate composite (shading) obtained by averaging Fe/K runoff record from core CD154-1006P[31] and the Pretoria Salt Pan rainfall reconstruction[29], extended from 250-190 ka (grey line), and depicting Bayesian Skyline Plot (BSP) effective population size for homeland L0k (orange dashed). **c,** Southwestern hydroclimate reconstruction (shading) obtained by averaging leaf wax data (MD08-3167[32]) and aridity index from cores (MD96-2094[33]), with aridity record extended from 250-140 ka (grey line) and L0d1'2 BSP effective population size (purple dashed). **d**, Simulated LOVECLIM normalized precipitation changes (shading) northeast of the homeland (33°E, 13°S) and coalescence time probabilities for L0f sub-branches (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 sub-branches (orange). **f**, Same as 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].


a. Geographic Distribution of L0 Mitogenomes
Published: 1,019
This study: 198
Total: 1,217
1
6
1
8
5
11
4
8
3
1
6
9
19
TANZANIA
DRC
63
ZAMBIA
ANGOLA
124
18
MOZAMBIQUE
NAMIBIA
2
211
77
322
ZAMBEZI RIVER
BOTSWANA
114
121
ZIMBABWE
SOUTH AFRICA
b. Geographic Distribution of Khoesan Identifier
NON-KHOESAN
KAFUE
UPPER ZAMBEZI
OKAVANGO DELTA
OKAVANGO
MIDDLE ZAMBEZI
LOWER ZAMBEZI
KALAHARI KHOESAN
MAKGADIKGADI PAN
LIMPOPO
WEST-COASTAL KHOESAN
EAST-COASTAL KHOESAN
StHe SKELETON
CAPE KHOESAN
ARCHAEOLOGICAL CAVES

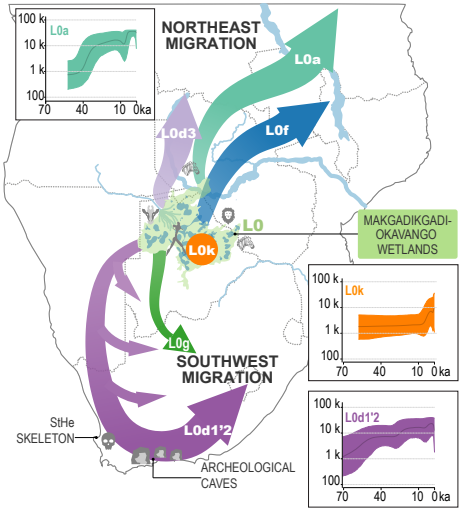
a. Phylogenetic Tree For 461 L0 Mitogenomes

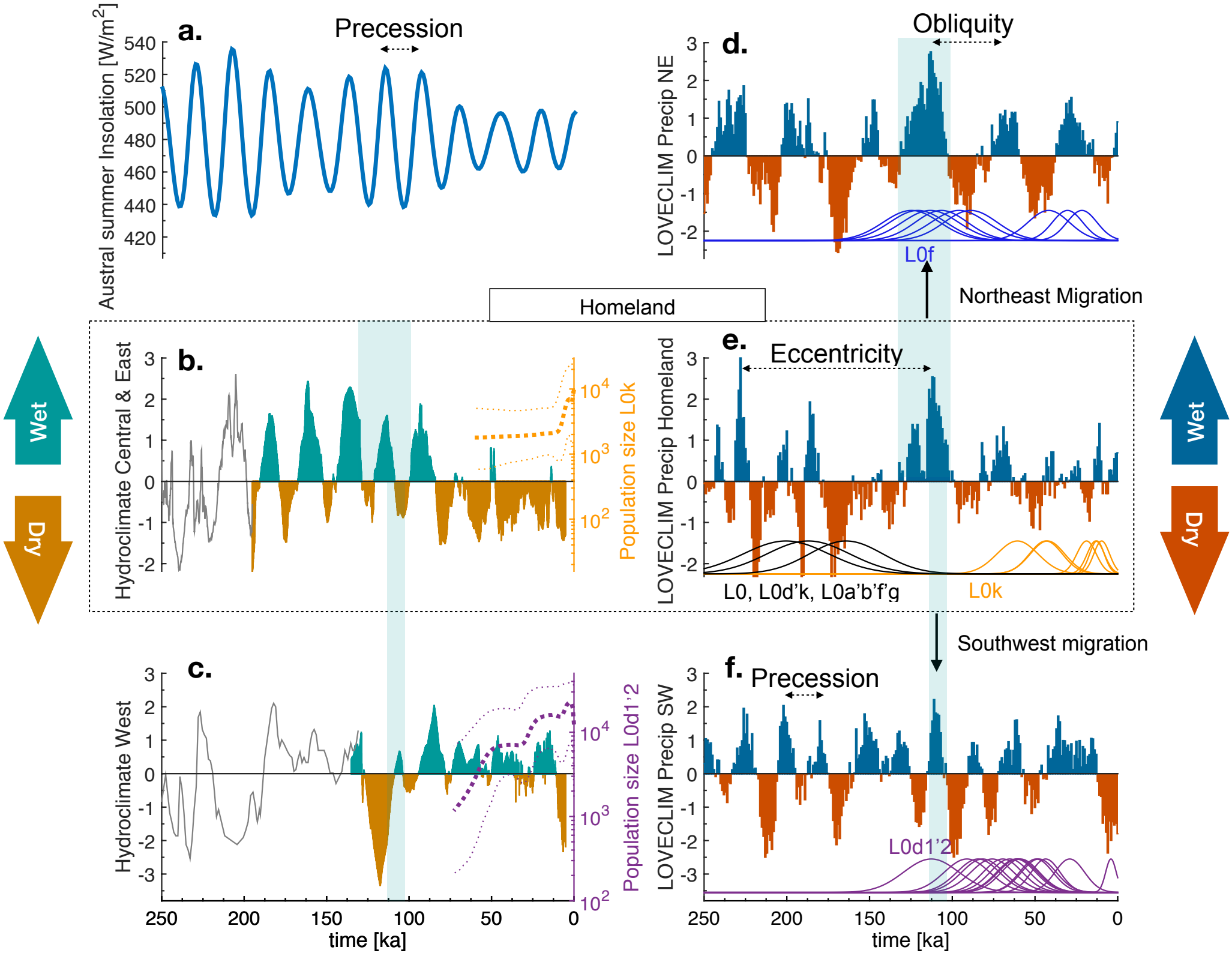


b. Geographic Distribution of L0d, L0k, and L0f from 1,217 L0 Mitogenomes



c. Migrations




a.
Precession
Austral summer Insolation [W/m^2]
540
520
500
480
460
440
420
d.
Obliquity
LOVECLIM Precip NE
L0f
Northeast Migration
Homeland
Wet
Dry
b.
Hydroclimate Central & East
Population size L0k
e.
Eccentricity
LOVECLIM Precip Homeland
L0, L0d'k, L0a'b'f'g
L0k
Southwest migration
c.
Hydroclimate West
Population size L0d1'2
f.
Precession
LOVECLIM Precip SW
L0d1'2
time [ka]

METHODS

Statement on population identifiers. The authors acknowledge that population identifiers (or ethnic labels) have different meanings for 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, with recognizing themselves as a Republic with a national flag[37].

In this study, the authors have used linguistics, supported by ethnicity, to provide population identification, with further historical, geographic 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 over the entire southern African region, KhoeSan languages are today restricted largely to populations residing in Namibia and Botswana (and southern Angola), while two Tanzanian isolates, Sandawe and Hadza, are believed to be linguistically-related click-languages (or East African KhoeSan)[38]. Meaning literally 'forager' (San) 'person' (Khoe), culturally the KhoiSan identifier refers to forager hunters (San) or herders (Khoi). At times linguistic and cultural identities clash. For example, Nama and Hai//om peoples both speak the Khoe-Kwadi based language Nama, while culturally and historically these two

389 populations are quite different, representing a herder and hunter-gatherer ancestry,
390 respectively. Additionally, autosomal genetic data has been used to provide further insights
391 into KhoeSan admixture and substructures, highlighting, for example at a genetic-level
392 historical differences between the Nama and Hai//om[39]. The authors have attempted to capture
393 both ethnic and linguistic identifiers that best reflect population ancestry. In contrast to
394 KhoeSan languages, Bantu languages are not traditionally click-based, yet again exceptions
395 exist within southern African Bantu languages, for example isiXhosa and isiZulu languages
396 have borrowed click consonants from their KhoeSan neighbours. Spoken across the entire sub-
397 Saharan Africa (up to 500 groups), the Guthrie classification of languages further identifies the
398 S-zone or Southern Bantu (South Africa, Zimbabwe, southern Mozambique and most of
399 Botswana) and the R-zone or Southwest Bantu languages (northern Namibia, southern Angola
400 and northwest Botswana)[40], which are of relevance to this study.
401 **Ethics statement and recruitment.** The study was performed in accordance with the ethical
402 standards of the overseeing human research ethics committees and local governance, as per the
403 1964 Helsinki Declaration. The study was reviewed and approved by the Ministry of Health
404 and Social Services (MoHSS) in Namibia (#17-3-3 2008, 2014 and 2019), with additional local
405 approvals from community leaders, the University of Pretoria Human Research Ethics
406 Committee (HREC #43/2010 and HREC #280/2017), including US Federal-wide assurance
407 (FWA00002567 and IRB00002235 IORG0001762), as well as the South African National
408 Blood Service (SANBS) HREC (HREC #2012/11). Participants were recruited within the
409 borders of Namibia and South Africa, with self-reported ethno-linguistic population identifiers
410 recorded. Blood samples were taken after receiving written and/or recorded informed consent.
411 Isolated DNA was shipped under the Republic of South Africa Department of Health Export
412 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 inhabiting the inland semi-desert Kalahari region of Botswana and Namibia include the Kw'a (Ju/'hoan or Hoan, and !Xun or !Xuun,), Tuu (or Taa) and Khoe-Kwadi (Naro, ǂAni, Khwe, Buga, Gǁana, Gǁui, ǂXokhoe, Tshwa and Shua) click-speakers. Indigenous KhoeSan inhabiting the west coastal region of Namibia speak a Khoe-Kwadi or Nama language and include the Nama, Damara, Topnaar (#Aonin) and Haiǁom click-speakers[41,42] 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 (#Aonin, n=6) from Namibia. Southwest Bantu (non-KhoeSan) non-click speakers 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[43], we speculate in this study a probable Kalahari KhoeSan heritage for the three Herero mitogenomes.

While indigenous KhoeSan are arguably absent from the coastal regions of South Africa, KhoeSan skeletal remains spread across the region[44]. 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 ~2.2 kya, migrating along the southwest coast to the southern Cape by ~2 kya[12,16,20,44]. 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 contribution. Along the east coast, southward

438 migrating non-click-speaking Bantu farmers (Southern Bantu) enter South Africa ~1,500 ya,
439 while a second wave of Bantu migrants (Southwest Bantu) cross central Africa into Namibia
440 ~800 ya[12]. Maternal contributions to the South African Southern Bantu speaking populations
441 (n=43, this study) may therefore either be of Bantu origin (in this case L0a lineages and
442 therefore non-KhoeSan) or of east-coastal KhoeSan-ancestry. The arrival of European colonists
443 and Dutch-east Indian slaves to the Cape in the mid-1600's, gave rise to multi-ethnic
444 (European, Asian, KhoeSan and Bantu) Cape population, the ancestors of the South African
445 Coloured (n=90, this study) and Namibian Basters (n=24. this study), speaking historically a
446 Dutch-derived language, known as Afrikaans[13,45]. Emerging from a common historical
447 background to the Coloured, the Baster population have since the late 1800's distinguished
448 themselves as independent from the Coloured, migrating to the Baster nation of Rehoboth in
449 Namibia[46]. While the vast majority of L0-mitogenomes represented in the Baster and Coloured
450 population are of Cape KhoeSan heritage (100% and 94.4%, respectively), we observe a
451 percentage of non-KhoeSan (Bantu) L0a lineages within the Coloured.

452 **L0-Haplogroup pre-screening.** Subjects were selected for whole mitogenome sequencing
453 based on pre-screening for specific L0 markers using direct amplicon-specific Sanger
454 sequencing. Specifically, a 2,673 bp region (rCRS position 3322–5995) was amplified and
455 initially screen for the L0 variant T5442C. L0 samples were further screened to delineate into
456 L0d (T4232C), L0d1 (G3438A), L0d1b (T3618C), L0d1c (C4197T), L0d1'2 (A3756G), L0d2
457 (A3981G, C205T, A4044G), L0d2a (A5153G), L0d2d (G5147A, G5231A), L0d2C (A4038g,
458 T4937C) and L0d3 (G5460A, G5773A). This identified a 188 samples carrying a rare L0-
459 haplogroup: L0d1b (n=21), L0d1c (n=13), L0d2a (n=30), L0d2b (n=7), L0d2c (n=15), L0d2d
460 (n=6), L0d3 (n=29), L0a1 (n=6), L0a2 (n=6), L0f (n=5) and L0k (n=5); as well as 55 samples
461 that could not be unambiguously assigned to major L0 sub-lineage: L0d1a'c (n=2) L0a'b'f'k
462 (n=5), L0a'b (n=2), L0d2 (n=1), and L0d1 (n=45, assumed L0d1a) (Supplementary Table 1).

463 **Whole mitogenome sequencing.** Mitogenomes were isolated using two overlapping
464 amplicons as previously described[6,47]. Specifically, two primer pairs to isolate and amplify
465 fragments 12,250–3,005 (7.2 kb) and 2,583–12,337 (9.7 kb) of the circular mitogenome. This
466 pair of primers has been demonstrated to effectively capture of the mitogenome with high
467 specificity whilst minimizing off-target capture of nuclear copies of mitochondrial-derived
468 DNA (NuMTs). Following touchdown long-range amplification with the Platinum™ Taq DNA
469 Polymerase High Fidelity (Invitrogen), the two amplicons were purified using the AMPure XP
470 beads (Agencourt) and combined in a 7:13 ratio of short to long fragments. Sequencing was
471 performed on the Ion Torrent PGM platform. In brief, 200 bp single-end sequencing libraries
472 were prepared using the Ion Xpress Plus Fragment Kit and Ion Xpress™ Barcode Adaptors
473 (ThermoFisher), and 4 to 16 samples (barcodes) were pooled and sequenced on 314v2 Ion
474 Chips. Using the Ion Torrent suite v5.0.2.1, sequencing reads were quality trimmed, and
475 aligned to the human mitochondrial revised Cambridge Reference Sequence (rCRS; accession
476 NC_012920.1). Consensus mitogenome sequences were derived by first identifying variants
477 relative to rCRS, using samtools (v1.3.1) mpileup (with parameters -d 10000 -L 1000 -Q 7 -h
478 50 -o 10 -e 17 -m 4)[48] and bcftools (v1.3.1) call (with parameters -c -M)
479 (http://www.htslib.org/doc/bcftools.html), then converting to fasta format using samtools'
480 vcfutils.pl vcf2fq program.

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

483 **Public data.** An exhaustive search for publicly available L0-mitogenomes was performed
484 between 2015 and 2017, identifying 26 studies containing a total of 6,334 mitogenomes. L0
485 status for all mitogenomes was deduced, either directly from the original publication or by
486 downloading the nucleotide sequences from NCBI and evaluating their haplogroup using
487 HaploGrep2 (v2.1.13)[49] based on PhyloTree Build 17[50]. From this, a subset of 1,019 L0-

mitogenomes were identified and included in this study (Extended Data Table 1, Supplemental Table 2). Public 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[49] was used to type all 1,217 sequences against PhyloTree Build 17[50]. 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 public mitogenomes (Supplementary Table 2).

Phylogenetic inference. Multiple sequence alignment was performed across all 1,217 mitogenomes along with seven Neanderthal genomes (Supplemental Table 10), using MUSCLE v3.8.31[51] with parameters -maxiters 3 -diags1. Phylogenetic inference was performed using FastTree v2.1.7 (SSE3)[52] using the generalised 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 re-rooted to the seven Neanderthal genomes.

Bayesian phylogenetic inferences and divergence times were calculated using BEAST2 v2.4.2 with BEAGLE 2.0[53]. Due 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[54,55] were excluded); (ii) All 198 novel mitogenomes from this study; (iii) All 121 L0-mitogenomes from our previous studies, Chan *et al*. 2015[6] (n=77), Morris *et al*. 2014[15] (StHe, defining the new haplogroup L0d2c1c), McCrow *et al*. 2016[47] (n=37), Schuster *et al*. 2010[56] (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

513 mitogenomes that could be unambiguously typed by HaploGrep2[49] (n=14; none from this
514 study); and (vi) A random subset of mitogenomes for all remaining sub-lineages not already
515 represented.

516 Multiple sequence alignment of the subset of 461 AMH and seven Neanderthal
517 mitogenomes was converted to NEXUS format using the convert function of seqmagick v0.6.1
518 (https://fhcrc.github.io/seqmagick) with parameter --alphabet dna-ambiguous. This provided
519 the input to BEAST2. Specifically, BEAUTi v2.4.2 was used to set up the phylogenetic model,
520 assuming: (i) the Gamma Site Model with 6 gamma categories and no invariant sites; (ii) the
521 generalized time reversible substitution model; (iii) a strict constant clock model with a normal
522 prior of with $\mu = 1.665 \times 10^{-8}$ and $\sigma = 1.479 \times 10^{-9}$ based on Soares *et al*. 2009[57]; and (iv) a
523 Coalescent Constant Population. Times were calibrated on the seven *Homo neanderthalensis*
524 mitogenomes with tip dates set to their reported archeological dating estimates (in years before
525 present): Feldhofer1=40000, Vindija=38000, ElSidron=39000, Feldhofer2=40000,
526 Mezmaiskaya=65000[58], Croatia=38310[58], Altai=50000[59] (see Supplementary Table 10). No
527 prior was set on the most recent common ancestor of this taxon set, and calibration was applied
528 to the leaves instead of the most recent common ancestor. Further, a normal prior, $N(\mu =$
529 $200000, \sigma = 50000)$, was set on the coalescent time of the AMH genomes, and a tip date of
530 2,330 ybp was set for StHe[15].

531 Five BEAST replicates were performed, each with 100 million MCMC iterations,
532 sampling every 10,000. Tracer v1.6 was used to evaluate BEAST trace files (Supplementary
533 Table 11), ensuring all runs had converged. The five replicates were combined using
534 LogCombiner v2.4.2, discarding 10% of the samples as burn-in for each replicate and without
535 re-sampling states at a lower frequency.

536 Sampled trees from BEAST were summarized into a single Maximum Clade
537 Credibility target tree using TreeAnnotator v2.4.2 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 v2.4.2, again discarding the first 10% as burn-in from each replicate, but re-sampling at a lower frequency of 50,000 (five replicates of 10,000 samples). The combined, re-sampled trees were then summarized with TreeAnnotator v2.4.2 as for the individual replicate BEAST results.

FigTree v1.4.2+ (Rambaut 2012; http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize all resulting trees.

Bayesian Skyline Plot analysis. BSP analyses were performed to estimate the demographic history of each maternal haplogroup. While maternal haplogroups do not necessarily equate to population data, it has been suggested that the signal associated with a haplogroup can still provide insight into the demographic processes in the populations carrying it[60,61].

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 1.665×10^{-8} based on Sores *et al.*[57]; 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. Results of this analysis are summarized in Supplementary Table 12.

Geographic history of the palaeo-wetland Makgadigadi. Initiated around 2 million years ago, palaeo-lake Makgadikgadi originally covered an area of ~170,000 km^2 at its highest lake

562 stand, bounded by a shoreline of ~995 m[7]. A degraded sand ridge (the Deception Ridge), was
563 associated with the 995 m shore in the southwest of the lake. This lake would have covered
564 more than twice the area of modern Lake Victoria, and like the latter, would have caused a
565 significant climatic feedback, with locally enhanced rainfall. We previously proposed that this
566 was, in turn, responsible for initiation of the surrounding (now-fossil) drainages, creating a
567 well-watered environment and very favorable habitat for mammals, including hominids[7].
568 Smaller lakes, now represented by residual wetlands, also formed on the upper Zambezi and
569 the modern Kafue Flats on the Kafue River, resulting in an archipelago of palaeo-lakes in
570 south-central Africa during the early and Middle Pleistocene.

571 Palaeo-Makgadikgadi bounded the 995 m shoreline was originally sustained by a major
572 drainage line, which included the Chambeshi as headwaters, connected to the upper Zambezi
573 via the Upper Kafue River. Severance of original links between the Chambeshi River and
574 upper Kafue, and the latter and the Upper Zambezi resulted in a sequential contraction of the
575 Makgadikgadi to a much smaller water body. This is reflected in a series of fossil shorelines,
576 associated with breaks in slope, at progressively lower levels (945 m, 936 m and 922 m). The
577 Gidikwe Ridge was associated with the 945 m shoreline. However, contraction of the lake was
578 accompanied by the development of the modern Okavango Delta. Timing of the contraction of
579 the lake and initiation of the Okavango Delta is not tightly constrained, but by the time we
580 propose that modern humans emerge within the region, some 200 ka, we speculate that the
581 former extensive Makgadikgadi palaeo-lake had contracted to a much less extensive deltaic-
582 lacustrine system, though nevertheless a favourable habitat for hominid occupation. It is this
583 palaeo-wetland region that we propose as the homeland for AMH's founder population.

584 **Climate Model Simulations and palaeo-climate data.** To place the coalescence time
585 estimates of the L0 branch into a climatic context and to test the robustness of simulated
586 hydroclimate responses in South Africa to orbital-scale conditions, we use the LOVECLIM

587 earth system model of intermediate complexity[28]. It is based on a 3-layer atmosphere, a 20-
588 level ocean general circulation model, a dynamic-thermodynamic sea-ice model and a
589 terrestrial vegetation model. A transient simulation covering several glacial/interglacial cycles
590 was conducted by using time-dependent boundary conditions. The experiment[28] (covering past
591 784 ka) uses time-varying boundary conditions for orbital parameters, CO_2 and other
592 greenhouse gas concentrations obtained from Antarctic ice cores, and an estimate of northern
593 hemispheric ice-sheet orography and albedo changes (data are used in **Fig. 3** and Extended
594 Data Fig. 6). The forcings are applied with an acceleration factor of 5: one coupled model year
595 corresponds to five orbital calendar years. Our analysis focuses on the past 250 ka in both
596 simulations. The climate sensitivity of this model to CO_2 variations was modified to capture
597 the range of reconstructed global mean surface temperature changes in response to radiative
598 forcing[62]. The transient LOVECLIM model simulations has previously been validated against
599 other palaeo-climate records from around the world[28,62,63]. Our analysis here focuses on the
600 simulated precipitation as well as tree and grass fraction changes in central eastern Africa and
601 western southern Africa (data used in **Fig. 3 d-f** and Extended Data Fig. 6 b,d).

602 As a result of its coarse horizontal atmospheric resolution (5.6^{o}) and the use of only
603 parameterized ageostrophic wind components, LOVECLIM has several deficiencies. Among
604 the most noteworthy are the lack of realistic El Nino-Southern Oscillation variability and the
605 fact that annual mean freshwater flux corrections have been applied to mimic the atmospheric
606 moisture transport from the Atlantic to the Pacific and to stabilize the Atlantic Meridional
607 Overturning Circulation.

608 There exist only a few long-term hydroclimate datasets from Southern Africa that cover the
609 last >120-ka. Here we compare the simulated LOVECLIM precipitation (normalized) in
610 central southern Africa with a southern central African hydroclimate composite, obtained by
611 averaging the normalized orbitally-tuned rainfall reconstruction from the Pretoria salt pan[29] and

612 the normalized Fe/K river runoff proxy obtained from marine sediment core CD154-1006P[31]
613 (**Fig. 4b**). The composite index emphasizes the joint variability in both records. We find that
614 some of the overall features in the observations, in particular the fact that rainfall is modulated
615 by the precessional cycle of austral summer insolation[64] (**Fig. 3a**), are well captured by the
616 LOVECLIM model experiment. However, we also find some discrepancies in the central part
617 of southern Africa, such as in the phase of the precessional signal and the difference in overall
618 wet and dry conditions during the Homeland period from 200-120 ka. The overall glacial
619 drying in the central part of southern Africa from 100 ka to 20 ka is, however, captured in both,
620 model simulation and palaeo-proxy reconstructions (**Fig. 3b,e**). Orbital-scale hydroclimate
621 variations in southern Africa are clearly not spatially homogenous (**Fig. 3b-f**). To get a better
622 understanding of the spatial patterns of hydroclimate variability we compared the model
623 simulation with a composite index from southwestern Africa, obtained by averaging a
624 normalized aridity index reconstructed from sediment core MD96-2094[33] and the normalized
625 $\delta^{13}C$ isotope ratio data of leaf waxes extracted from the South Atlantic sediment core MD08-
626 3167[32] (**Fig. 3c** and Extended Data Fig. 6 c,d). The results show a good correspondence
627 between model and reconstructions on the western side of southern Africa, and in particular
628 reproduce a major drought period peaking ~120 ka and a subsequent increase of rainfall
629 towards the last glacial period. This gradual increase in humidity/rainfall corresponds to an
630 overall increase in lineage-splitting of the L0d1'2 haplogroup (**Fig. 3f**) and growth of its
631 population (**Fig. 3c**). This result further highlights the possibility that climate shifts may have
632 played an important role in the southwestward migration of L0d1'2 descendants (**Fig. 2**).
633 To further test the fidelity of LOVECLIM in reproducing interhemispheric orbital rainfall
634 shifts across Africa, we also compared the simulated vegetation changes with a leaf-wax index
635 from stable hydrogen isotope data extracted from a sediment core in the Gulf of Aden[30], which
636 is indicative of hydroclimate and vegetation changes in the north-eastern Horn of Africa

637 (Extended Data Fig. 6b). The comparison shows a good qualitative correspondence for the
638 precessional-scale timing of rainfall/vegetation maxima and minima as well as of the
639 eccentricity modulated amplitude of these changes, lending further support to the credibility of
640 the simulated rainfall patterns across Africa. It should be noted that regional patterns of paleo
641 rainfall changes are in general difficult to simulate. In response to Last Glacial Maximum
642 boundary conditions different Coupled General Circulation models simulate widely varying
643 responses in rainfall over Africa[28].

644

645 **Additional References**

646 37. Orizio, R. *Lost White Tribes: The End of Privilege and the Last Colonials in Sri Lanka,*
647 *Jamaica, Brazil, Haiti, Namibia, and Guadeloupe*. (Free Press, 2001).
648 38. *African Languages: An Introduction*. (Cambridge University Press, 2000).
649 39. Montinaro, F. *et al.* Complex Ancient Genetic Structure and Cultural Transitions in
650 Southern African Populations. *Genetics* **205**, 303–316 (2017).
651 40. Guthrie, malcolm. *The classification of the Bantu languages*. (Oxford University Press,
652 1948).
653 41. Honken, H. & Heine, B. The Kx'a Family: A New Khoisan Genealogy. *J. Asian Afr. Stud.*
654 **79**, 5–36 (2010).
655 42. Güldemann, T. & Elderkin, E. D. On external genealogical relationships of the Khoe
656 family. in *Khoisan languages and linguistics: proceedings of the 1st International*
657 *Symposium January 4-8, 2003, Riezlern/Kleinwalsertal* (eds. Brenzinger, M. & König, C.)
658 15–52 (Rüdiger Köppe, 2010).
659 43. Stockton, R. The Herero Genocide: Germany's First Mass Murder. *All That's Interesting*
660 (2017).

44. Smith, A. B. *Excavations at Kasteelberg and the Origins of the Khoekhoen in the Western Cape, South Africa*. (Archaeopress, 2006).
45. Patterson, N. *et al.* Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* **19**, 411–419 (2010).
46. van der Ross, R. E. *Up from slavery: slaves at the Cape : their origins, treatment and contribution*. (Ampersand Press in association with the University of the Western Cape, 2005).
47. McCrow, J. P. *et al.* Spectrum of mitochondrial genomic variation and associated clinical presentation of prostate cancer in South African men. *The Prostate* **76**, 349–358 (2016).
48. 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).
49. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
50. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386-394 (2009).
51. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
52. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
53. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
54. Kivisild, T. The Role of Selection in the Evolution of Human Mitochondrial Genomes. *Genetics* **172**, 373–387 (2006).

685 55. Herrnstadt, C. *et al.* Reduced-Median-Network Analysis of Complete Mitochondrial DNA
686 Coding-Region Sequences for the Major African, Asian, and European Haplogroups. *Am.*
687 *J. Hum. Genet.* **70**, 1152–1171 (2002).
688 56. Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature*
689 **463**, 943–947 (2010).
690 57. Soares, P. *et al.* Correcting for Purifying Selection: An Improved Human Mitochondrial
691 Molecular Clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
692 58. Green, R. E. *et al.* The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**,
693 2494–2502 (2009).
694 59. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai
695 Mountains. *Nature* **505**, 43–49 (2014).
696 60. Gandini, F. *et al.* Mapping human dispersals into the Horn of Africa from Arabian Ice Age
697 refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).
698 61. Soares, P. *et al.* The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol.*
699 *Biol. Evol.* **29**, 915–927 (2012).
700 62. Friedrich, T., Timmermann, A., Tigchelaar, M., Elison Timm, O. & Ganopolski, A.
701 Nonlinear climate sensitivity and its implications for future greenhouse warming. *Sci. Adv.*
702 **2**, e1501923–e1501923 (2016).
703 63. Stockhecke, M. *et al.* Millennial to orbital-scale variations of drought intensity in the
704 Eastern Mediterranean. *Quat. Sci. Rev.* **133**, 77–95 (2016).
705 64. Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth.
706 *Astron. Astrophys.* **428**, 261–285 (2004).
707 65. Barbieri, C. *et al*. Unraveling the complex maternal history of Southern African Khoisan.
708 *Am J Phys Anthropol*. **153**, 435–448 (2014).

66. 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).

67. Barbieri, C. *et al.* Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol*. **29**, 1213-1223 (2012).

68. Barbieri, C. *et al*. Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS One*. **9**, e99117 (2014).

69. Batini, C. *et al*. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol Biol Evol*. **28**, 1099-1110 (2011).

70. Eaaswarkhanth, M. *et al*. Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur J Hum Genet*. **18**, 354-363 (2010).

71. 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).

72. 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 U S A*. **92**, 532-536 (1995).

73. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature*. **408**, 708-713 (2000).

74. 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**, e45-48 (2008).

75. 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).

76. 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).

77. Macaulay, V. *et al*. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*. **308**, 1034-1036 (2005).

78. Margaryan, A. *et al*. Eight Millennia of Matrilineal Genetic Continuity in the South Caucasus. *Curr Biol*. **27**, 2023-2028.e7 (2017).

79. Olivieri, A. *et al*. Mitogenome Diversity in Sardinians: A Genetic Window onto an Island's Past. *Mol Biol Evol*. **34**, 1230-1239 (2017).

80. 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).

81. 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).

Extended Data

Extended Data Table 1. L0-mitogenomes included in this study. Footnotes: * Our previously published data; † Mitochondrial 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.

Extended Data Fig. 1. Phylogenetic tree of all 1,217 L0-mitogenomes. Phylogeny was inferred using FastTree v2.1.746, 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 indicated.

Extended Data Fig. 2. Detailed phylogenetic branching of L0k, L0d3, L0f, and L0g. Shown are the expanded sections of the phylogenetic tree depicted in **Fig. 2a**, including: **a**, 34 (of 113 total) L0k, **b**, all 40 L0d3, **c**, all 27 L0f, and **d**, all nine L0g mitogenomes. Each mitogenome is represented as a tip and coloured based on their broad ethno-linguistic classification, if known: KhoeSan in orange, non-KhoeSan in grey, and Cape multi-ethnic (KhoeSan-ancestral) in green. Public 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 further detailed in Supplementary Tables 7, 8, and 9.

773

774 **Extended Data Fig. 3. Detailed phylogenetic branching of L0d2.** Shown are the expanded
775 branches of the phylogenetic tree depicted in Fig. 2a, including: **a**, 51 (of 118 total) L0d2a, **c**,
776 25 (of 53) L0d2c, and **d**, all 11 L0d2d mitogenomes. For L0d2b, an additional BEAST analysis
777 was performed using an alternate subset of 441 mitogenomes that included all 43 L0d2b
778 samples, as opposed to the n=461 subset (Fig. 2a) that included only 13 L0d2b. The same
779 model parameters were used for both data subsets. In all panels, each mitogenome is
780 represented as a tip and coloured based on their broad ethno-linguistic classification, as in
781 Extended Data Fig. 2. The previously defined L0d2c1c haplogroup, containing the Coastal
782 KhoeSan StHe skeleton[6] and other newly proposed sub-lineages are indicated by red node
783 labels (Supplementary Tables 4, 5, and 6).

784

785 **Extended Data Fig. 4. Detailed phylogenetic branching of L0d1.** Shown are the expanded
786 branches of the phylogenetic tree depicted in Fig. 2a, including: **a**, 54 (of 91 total) L0d1a, **b**, 45
787 (of 174) L0d1b, and **c**, 33 (of 184) L0d1c mitogenomes. Each mitogenome is represented as
788 tips and coloured here based on their broad ethno-linguistic classification as in Extended Data
789 Fig. 2.

790

791 **Extended Data Fig. 5. Detailed phylogenetic branching of L0a.** Shown is the L0a branch of
792 the phylogenetic tree displayed in Fig. 2a, which includes a subset of 114 (of 294 total) L0a
793 mitogenomes. Each mitogenome is represented as tips and coloured here based on their broad
794 ethno-linguistic classification as in Extended Data Fig. 2.

Extended Data Fig. 6. Paleo-data/model comparison. a, Relative rainfall changes (%) between Eemian (125 ka) and pre-industrial conditions simulated by the Community Earth System model using 20 km atmosphere and 10 km ocean resolution (see Methods). The simulated rainfall differences are mostly due to enhanced (reduced) northern (southern) hemisphere summer insolation. Locations of key sites that are used for paleo model/data comparison in this study are highlighted in red. **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 waxes, corrected for ice volume contributions from the Gulf of Aden marine sediment core RC09-166[31], indicating changes in hydroclimate. **c**, Relative precipitation changes (%) simulated by LOVECLIM transient model (all forcings) for 11^{o}E, 19^{o}S and grain-size aridity index reconstructed from sediment core MD96-2094[34]. **d**, Grass fraction changes simulated by LOVECLIM transient model (all forcings) at 11^{o}E, 14^{o}S-17^{o}S (dark blue, grey shading) 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 Namibia desert and further inland[33].