

# 9,000 years of genetic continuity in southernmost Africa demonstrated at Oakhurst rockshelter

Received: 31 January 2024

Accepted: 2 August 2024

Published online: 19 September 2024

 Check for updates

Joscha Gretzinger<sup>1</sup>, Victoria E. Gibbon<sup>2</sup>✉, Sandra E. Penske<sup>1</sup>, Judith C. Sealy<sup>3</sup>, Adam B. Rohrlach<sup>1,4</sup>, Domingo C. Salazar-García<sup>5,6</sup>, Johannes Krause<sup>1,7</sup> & Stephan Schiffels<sup>1,7</sup>✉

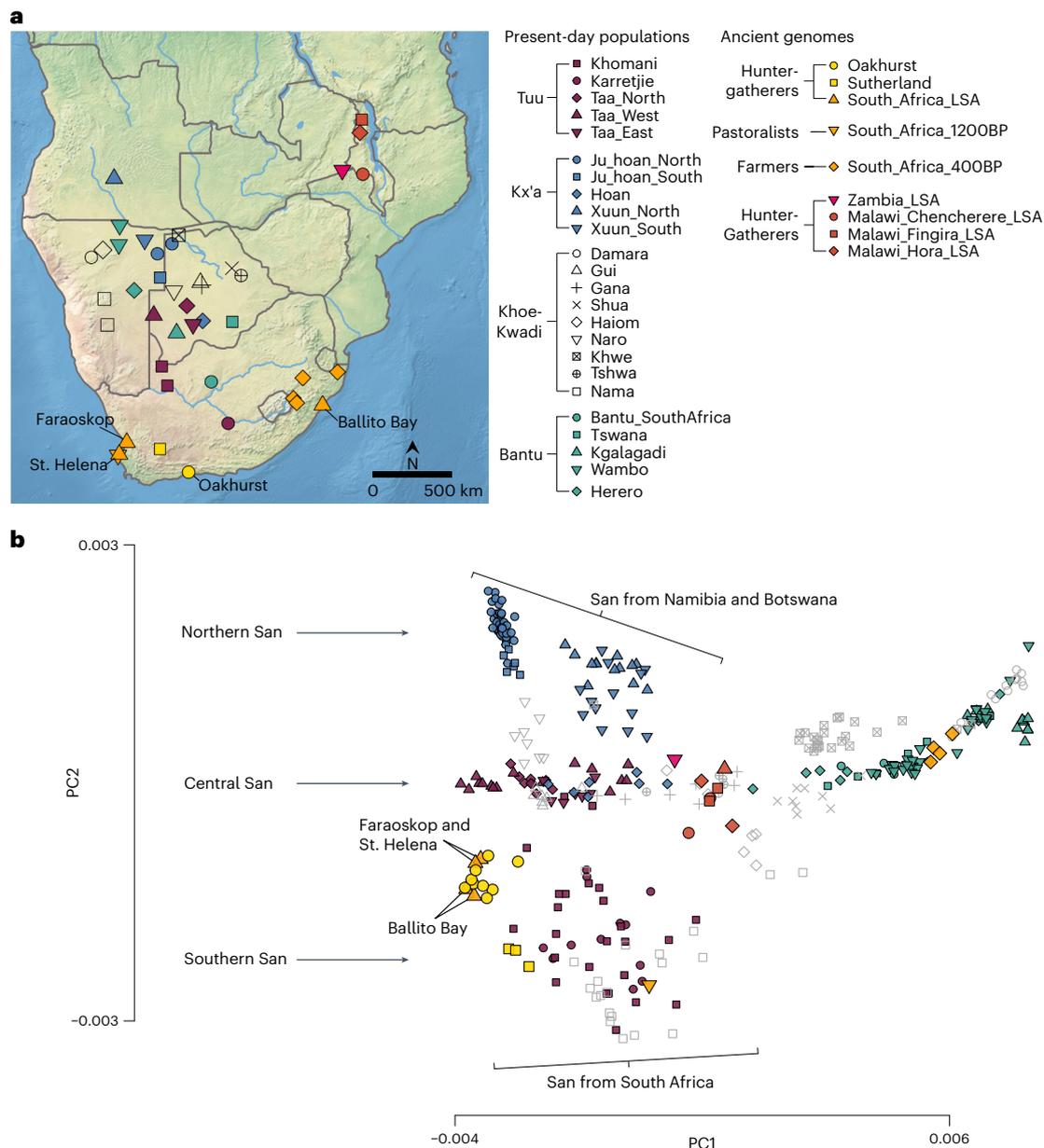
Southern Africa has one of the longest records of fossil hominins and harbours the largest human genetic diversity in the world. Yet, despite its relevance for human origins and spread around the globe, the formation and processes of its gene pool in the past are still largely unknown. Here, we present a time transect of genome-wide sequences from nine individuals recovered from a single site in South Africa, Oakhurst Rockshelter. Spanning the whole Holocene, the ancient DNA of these individuals allows us to reconstruct the demographic trajectories of the indigenous San population and their ancestors during the last 10,000 years. We show that, in contrast to most regions around the world, the population history of southernmost Africa was not characterized by several waves of migration, replacement and admixture but by long-lasting genetic continuity from the early Holocene to the end of the Later Stone Age. Although the advent of pastoralism and farming substantially transformed the gene pool in most parts of southern Africa after 1,300 BP, we demonstrate using allele-frequency and identity-by-descent segment-based methods that the †Khomani San and Karretjiemense from South Africa still show direct signs of relatedness to the Oakhurst hunter-gatherers, a pattern obscured by recent, extensive non-Southern African admixture. Yet, some southern San in South Africa still preserve this ancient, Pleistocene-derived genetic signature, extending the period of genetic continuity until today.

Southern African populations today harbour genetic variation that traces deep human population history<sup>1,2</sup>, reflected also in the archaeological record with fossils of archaic *Homo sapiens* dating back to 260 thousand years ago (ka) before present (BP) (uncalibrated) and evidence for the presence of anatomically modern humans in South Africa from at least -120 ka BP onwards<sup>3,4</sup>. While genetic investigations have extensively explored the significance of southern African population structure in human evolution, there is a noticeable gap in our understanding of the more recent demographic trajectories during

the Holocene (the last 11,700 years), which remain relatively understudied genetically.

During the Holocene, major transformations in lithic industries and subsistence practices probably also reflect demographic shifts<sup>5,6</sup>. In the last 2,000 years, the spread of pastoralism and farming have resulted in repeated admixture events visible in genetic complexity in both ancient and contemporary populations<sup>1,2,7,8</sup>. First, the spread of herders contributed northeast African, Levantine-enriched ancestry to the genetic make-up of southern African hunter-gatherers<sup>2</sup>.

A full list of affiliations appears at the end of the paper. ✉e-mail: [victoria.gibbon@uct.ac.za](mailto:victoria.gibbon@uct.ac.za); [stephan\\_schiffels@eva.mpg.de](mailto:stephan_schiffels@eva.mpg.de)



**Fig. 1 | Population structure in present-day southern Africa. a**, Approximate locations of present-day populations and ancient individuals mentioned in the article. Present-day populations are coloured according to linguistic affiliation, as indicated in the legend. **b**, PCA on 212,000 SNPs with ancient individuals

projected onto PC1 and PC2. Shown are the positions of each individual along the first and second axes of genetic variation, with symbols denoting the individual's population and linguistic affiliation using the same colour coding as in **a**.

Second, the influx of farmers closely related to present-day Bantu-language speakers introduced western African ancestry to San and Khoe populations<sup>2</sup>. Consequently, all contemporary San and Khoe groups exhibit at least 9% genetic admixture from non-San sources outside modern-day South Africa, Namibia and Botswana<sup>1,2,7,8</sup>, obscuring the population structures of the Later Stone Age (LSA) San population. To provide insights into early Holocene San population structure, we sampled and recovered genome-wide data from a series of individuals unearthed from the Oakhurst rockshelter in South Africa, offering a chronological spectrum spanning most of the Holocene.

Oakhurst rockshelter is located close to George in southernmost Africa, ~7 km from the coast (Fig. 1a). Excavated in the first half of the twentieth century<sup>9</sup>, it is especially noteworthy for its substantial accumulation of deposit that spans 12,000 years. Its Early Holocene macro-lithic stone artefact assemblage is characteristic of the period, similar

to those found at many sites across South Africa and is today regarded as a distinctive technocomplex termed the 'Oakhurst Complex', named after the site<sup>5,6,10–13</sup>. At ~8,000 BP, the lithics change to microlithic 'Wilton' assemblages which persist through the Middle and Late Holocene, albeit with some temporal shifts, notably the addition of ceramics in the last 2,000 years (refs. 6,9,14,15). The site also preserves the complete and partial burials of 46 juvenile and adult individuals, spanning the complete period of occupation of the site<sup>16,17</sup> and providing a valuable resource for the study of LSA San population structure. Here, we present genome-wide data from 13 individuals, including the oldest DNA from South Africa, dating back to ~10,000 years (calibrated) cal BP.

## Results

We sampled skeletal remains from 13 individuals, each radiocarbon dated on bone collagen with dates ranging between approximately

10,000 and 1,300 cal BP (Supplementary Table 1). Nine of these <sup>14</sup>C dates were previously reported; for four samples we generated new <sup>14</sup>C dates. We prepared powder from skeletal material, extracted ancient DNA (aDNA) and converted it into double- or single-stranded libraries (Methods; Supplementary Table 1). We selected 11 double-stranded and 15 single-stranded libraries for hybridization DNA capture to enrich for sequences that overlapped 1.24 million single-nucleotide polymorphisms (SNPs). For all 13 individuals, we determined the genetic sex and classified mitochondrial DNA and Y chromosome haplogroups for nine and five individuals, respectively, all of which are common in ancient and contemporary San and Khoe populations<sup>1,2,18–24</sup> (Supplementary Table 1). After quality filtering (Methods) and merging of duplicate libraries, we recovered genome-wide data sufficient for population genetic analysis for nine individuals, featuring on average 368,359 SNPs of the 1,240k panel, a mean mtDNA contamination of 1.7% and a mean X chromosome contamination of 1.5% (Supplementary Table 1).

### Genetic affinities between Oakhurst and contemporary San and Khoe

To place the nine individuals from Oakhurst into a pan-African evolutionary context, we first constructed a population tree based on allele frequencies in ancient and present-day populations, using TreeMix<sup>25</sup> (Methods). Similar to contemporary San individuals<sup>1,2,7</sup>, the nine Oakhurst individuals diverge basal to all other human lineages (Extended Data Fig. 1a). To investigate their genetic ancestry in detail, we performed principal component analysis (PCA) on a set of 24 contemporary San, Khoekhoe and Bantu-speaking populations from Namibia, Botswana and South Africa and projected new and published ancient genomes onto the first two principal components (Fig. 1 and Supplementary Tables 26 and 27). In line with previous analysis of San and Khoekhoe fine-scale population structure, we observed marked genetic differentiation between San and Khoe populations along PC2, reflecting the geographic separation between groups living north and south of the Kalahari Desert<sup>7,8,26–29</sup>. In total, we observe three principal clusters with the Kx<sup>ˆ</sup>a-speaking Jul'Hoan (genetic group label in figures and tables: Ju\_hoan) and !Xuun (Xuun) representing the northern San ancestry component, the Khoe-Kwadi-speaking Nama as well as Tuu-speaking ǀKhomani (Khomani) and Karretjiemense (self-identification of these San descendants from the Karoo region of South Africa, the Afrikaans term translates to 'the people of the cart') (Karretjie) forming the southern San ancestry component and the Tuu-speaking Taa, Kx<sup>ˆ</sup>a-speaking ǀHoan (Hoan) and Khoe-Kwadi-speaking G!ui (Gui) and G!lana (Gana) corresponding to the central San ancestry component. In this context, eight of the Oakhurst individuals cluster closely together with four previously published LSA hunter-gatherers from South Africa<sup>1,2</sup> within the diversity of the southern San and Khoekhoe cluster (Fig. 1b). We also considered a slightly different PCA with a larger SNP overlap among fewer analysed populations (Extended Data Fig. 1b), which proves useful for specific signals. Here, we see that the oldest individual in our dataset, OAK006, which dates between 9,900 and 10,500 cal BP (95% confidence interval (CI)), is slightly shifted in the direction of the northern San cluster (Extended Data Fig. 1b). Unsupervised ancestry decomposition using DISTRUCT<sup>30</sup> (Extended Data Fig. 2) shows overall a similar pattern as PCA (with  $K = 6$  clusters): present-day southern San and Khoekhoe are assigned the same major ancestral component (shown in orange), which is maximized in the LSA individuals from Oakhurst, St. Helena, Faraoskop and Ballito Bay. In contrast, northern San exhibit a different component (shown in blue) that is maximized in northern Jul'Hoan. Finally, a third component (shown in magenta) is maximized in Taa groups and also represents the largest San ancestry component in most remaining Khoe-Kwadi populations such as the ǀHoan, G!ui, G!lana and Tshwa (Extended Data Fig. 2).

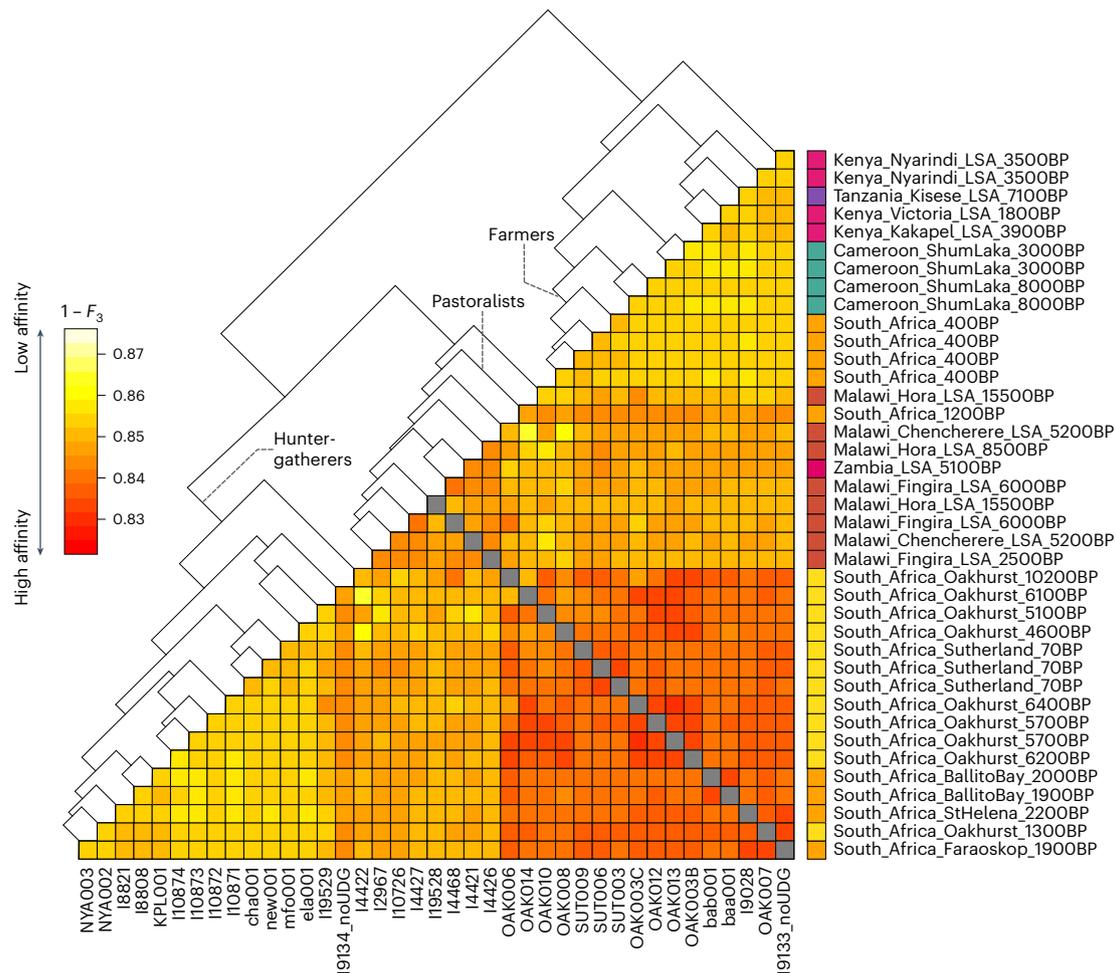
To quantitatively test whether the observations from the PCA and ancestry clustering are consistent with patterns of shared genetic drift, we compute outgroup  $F_3$ -statistics of the form  $F_3(\text{Archaic};$

Oakhurst, Test) between the Oakhurst individuals and present-day San and Khoe populations (Extended Data Fig. 3a and Supplementary Table 2). We also calculate the fixation index ( $F_{ST}$ ) by pairs and compare the two measures (Extended Data Fig. 3a and Supplementary Table 3). Although outgroup  $F_3$  and  $F_{ST}$  point estimates are significantly associated (Pearson's correlation;  $t = -13.472$ , d.f. = 27,  $P = 1.683 \times 10^{-13}$ ,  $r = -0.933$ ) (Extended Data Fig. 3a), the outgroup  $F_3$  signal is mostly correlated with the proportion of indigenous San ancestry (measured using qpAdm; see Methods and analysis around Figs. 3 and 4) in the present-day populations (Pearson's correlation;  $t = 16.67$ , d.f. = 9,  $P = 8.478 \times 10^{-13}$ ,  $r = 0.968$ ).  $F_{ST}$  appears less affected by varying percentages of non-San ancestry (evidenced by a reduced correlation between  $F_{ST}$  and San ancestry; Pearson's correlation;  $t = -6.252$ , d.f. = 19,  $P = 5.278 \times 10^{-6}$ ,  $r = -0.82$ ) and is consequently able to detect subtle population structure that was obscured by later admixture events. We find the highest genetic affinity between Oakhurst and groups of the southern San cluster, namely the Karretjiemense, ǀKhomani and Nama (Extended Data Fig. 3b). In general,  $F_{ST}$  between the Oakhurst individuals and present-day San and Khoe-Kwadi-speaking groups is strongly correlated with latitude (Pearson's correlation;  $t = 2.828$ , d.f. = 23,  $P = 0.009528$ ,  $r = 0.508$ ), demonstrating that San and Khoekhoe groups who live closer to Oakhurst rockshelter are still today more closely related to its LSA inhabitants than groups from further north. This is furthermore supported by the sharing of identity-by-descent (IBD) segments between the most recent individual, OAK007 (dating to -1,344 cal BP; 1,400–1,300 cal BP 95% CI) and present-day southern Africans. On average, OAK007 shares more and longer IBD segments with the Karretjiemense and ǀKhomani than with any other tested population, demonstrating direct genetic relatedness between the ancient hunter-gatherer and modern San and Khoe groups from South Africa (Extended Data Fig. 4 and Supplementary Table 24).

### Genomic continuity since the early Holocene

We then proceeded to investigate individual changes in ancestry through time. First, we assessed the extent of genetic similarity between the Oakhurst individuals and previously published prehistoric genomes from South Africa, Cameroon<sup>31</sup>, Kenya<sup>32,33</sup>, Malawi<sup>2,34</sup>, Tanzania<sup>34</sup> and Zambia<sup>34</sup> by means of outgroup  $F_3$ -statistics (Supplementary Table 4). All LSA genomes from South Africa are more like one another than any other tested prehistoric ancient African (Fig. 2). Yet, some fine-scale population stratification is evident, with the most recent sample OAK007 clustering together with two 2,000-year-old hunter-gatherers from St. Helena and Faraoskop. Together these samples form a sister clade with two contemporaneous samples from Ballito Bay, located in KwaZulu-Natal on the eastern coast of South Africa, within the diversity of the older Oakhurst samples. Visualizing the transformed pairwise-distance  $F_3$ -matrix through multidimensional scaling, we find the Oakhurst individuals older than 1,300 cal BP shifted along coordinate 1 away from the younger LSA genomes and three historical San samples from Sutherland, Western Cape province (Extended Data Fig. 5a). On the other hand, consistent with the affinities detected in PCA and DISTRUCT analysis, the genomes of a 1,200-year-old pastoralist and four Iron Age farmers from South Africa cluster in the diversity of LSA genomes from Malawi and Cameroon, respectively, highlighting the impact of admixture events after 1,300 cal BP that contributed varying fractions of non-San ancestry to all populations in southern Africa<sup>1,2,7,8</sup> (Fig. 2).

To test whether the Oakhurst individuals already exhibit subtle excess affinity to non-San ancestries, we calculated  $F_4$ -statistics<sup>35</sup> of the form  $F_4(\text{Archaic}; \text{Test}; \text{Ju\_hoan\_North}, \text{Tanzania\_Luxmanda\_3000BP})$  (Fig. 3a and Supplementary Table 6) and  $F_4(\text{Archaic}; \text{Test}; \text{Ju\_hoan\_North}, \text{Cameroon\_SMA})$  (Extended Data Fig. 5b and Supplementary Table 5). We find that none of the Oakhurst individuals, including the most recent individual OAK007, shares significantly more genetic drift



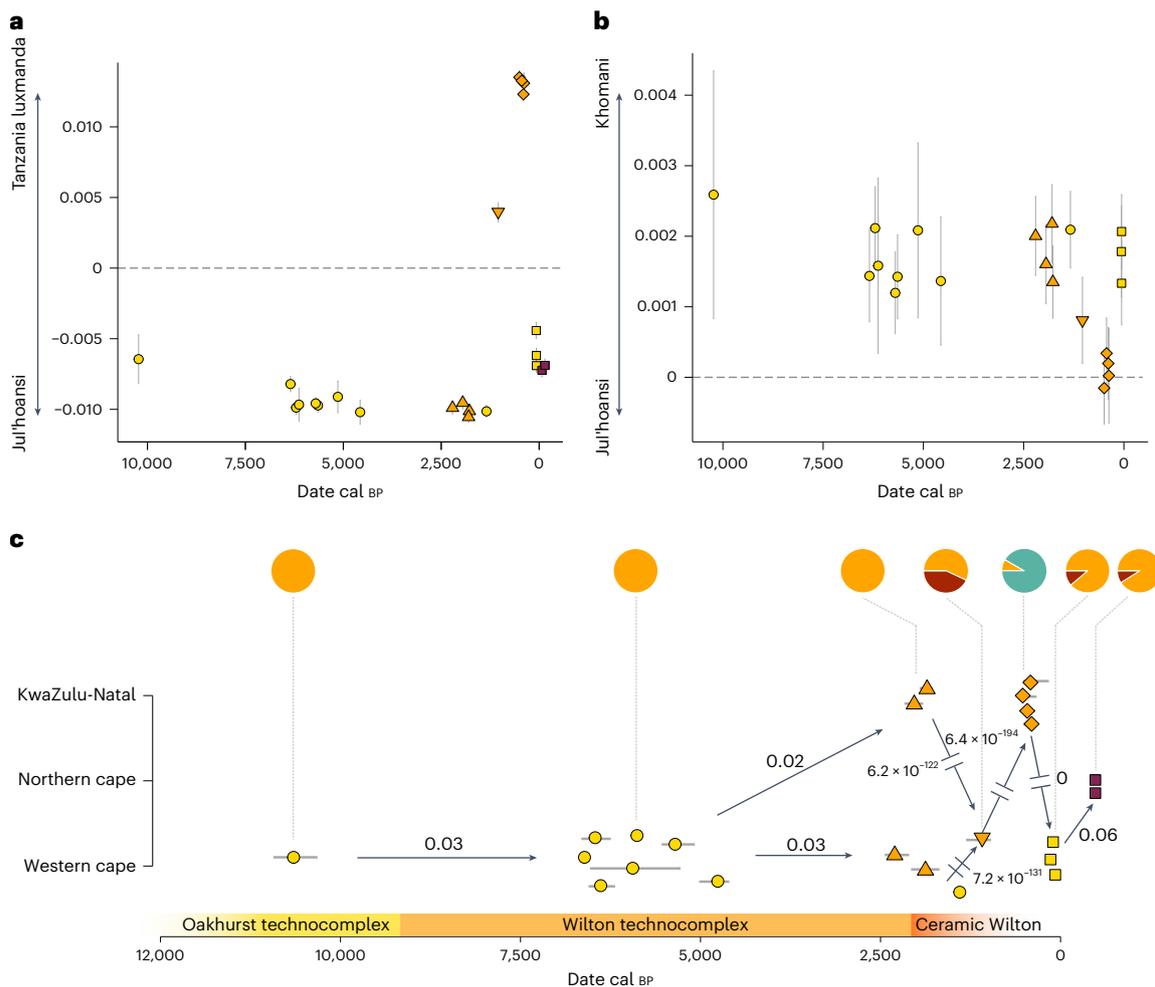
**Fig. 2 | Genetic affinities among LSA individuals from sub-Saharan Africa.** Shown is a heat-map matrix of pairwise outgroup  $F_3$ -statistics of the form  $F_3(X, Y; \text{Chimp})$ . Hierarchical cluster analysis applying Ward's minimum variance method to the rows is added as a dendrogram. Data can be found in Supplementary Table 4.

with Tanzania\_Luxmanda\_3000BP (associated with East African pastoralist ancestry) or Cameroon\_SMA (representing Central and Western African ancestry) than the published LSA genomes, providing an early bound for the date of the arrival of non-San ancestry at the southern coast only after 1,300 cal BP. For all LSA samples from South Africa, we observe significantly higher affinity ( $Z > 3$ ) to present-day  $\ddagger$ Khomani than to Jul'Hoan, confirming an old split age for the northern and southern San ancestries before 20,000 BP (refs. 7,8,27,36) and, thus, before the lake Makgadikgadi palaeo-wetland dried up<sup>29,37</sup> (Fig. 3b and Supplementary Table 7).

To evaluate whether chronological groups in South Africa were consistent in sharing the same genetic make-up as the preceding and succeeding populations, we used qpWave<sup>38–40</sup>, a generalization of  $F_4$ -statistics, testing for significant evidence of continuity (that is, we tested whether they were consistent with forming a clade at  $P > 0.01$ ) (Supplementary Table 8). We find that groups of individuals from Oakhurst dating to 10,000 cal BP and 6,000 to 4,000 cal BP as well as groups from the western and eastern coast (St. Helena, Faraoskop, OAK007 and Ballito Bay, respectively; dating between 2,200 and 1,300 cal BP) were all genetically indistinguishable (Fig. 3c). On the other hand, we observe significant discontinuity between 1,300 and 1,200 cal BP, as well as between 1,200 and 400 cal BP, consistent with the independent arrivals of non-San East African pastoralist and West African farmer ancestry in South Africa (Fig. 3c). To assess these demographic changes quantitatively, we used qpAdm<sup>39,40</sup> to successively model these groups as mixtures between local LSA, pastoralist

and farmer ancestry components (Methods; Supplementary Table 9). We find no evidence of West African ancestry within the three Sutherland individuals<sup>41</sup> (dating to the second half of the nineteenth century), yet, we detect small amounts of Tanzania\_Luxmanda\_3000BP-related ancestry ( $11\% \pm 0.9\%$ ) comparable to the proportions measured in present-day  $\ddagger$ Khomani from the Northern Cape Province ( $9\% \pm 1\%$ ) (Fig. 3c).

Overall, these observations indicate that between 10,000 and 1,300 cal BP, no ancestry from outside present-day South Africa arrived at Oakhurst rockshelter, demonstrating a remarkable degree of relative genetic continuity over a time range of nearly 9,000 years. Such a demographic pattern is exceptional in the global archaeogenetic record, yet, the Oakhurst samples do not exhibit signs of genetic isolation. While the conditional nucleotide diversity (CND; Methods) of the Oakhurst individuals is lower than in LSA populations from Malawi, Kenya and Cameroon, it is comparable to the diversity measured in the published hunter-gatherers from the Western Cape and KwaZulu-Natal and higher than the CND in ancient hunter-gatherers from Serbia<sup>42</sup>, Japan<sup>43</sup> or Brazil<sup>44</sup> (Extended Data Fig. 6a). Furthermore, we find the average heterozygosity levels within the three highest-coverage individuals (OAK007, OAK012 and OAK013) to be higher than in most present-day San and Khoe populations, disagreeing with a model of prolonged genetic isolation, yet supporting recent findings of a continuous, substantial reduction of effective population size in southern San and Khoe after 1,300 BP (ref. 45) (Extended Data Fig. 6b and Supplementary Table 25).



**Fig. 3 | Long-term genetic continuity and abrupt disruption in South Africa.** **a**, Individual  $F_4$ -statistics of the form  $F_4(\text{Archaic, Test}; \text{Ju\_hoan\_North, Tanzania\_Luxmanda\_3000BP})$  through time for 21 ancient and 2 current-day Khomani genomes from South Africa. Error bars represent 2 s.e. Data can be found in Supplementary Tables 6 and 7. **b**, Individual  $F_4$ -statistics of the form  $F_4(\text{Archaic, Test}; \text{Ju\_hoan\_North, Khomani\_San})$  through time for 21 ancient genomes from South Africa. **c**, Overview about population genetic changes in South Africa from

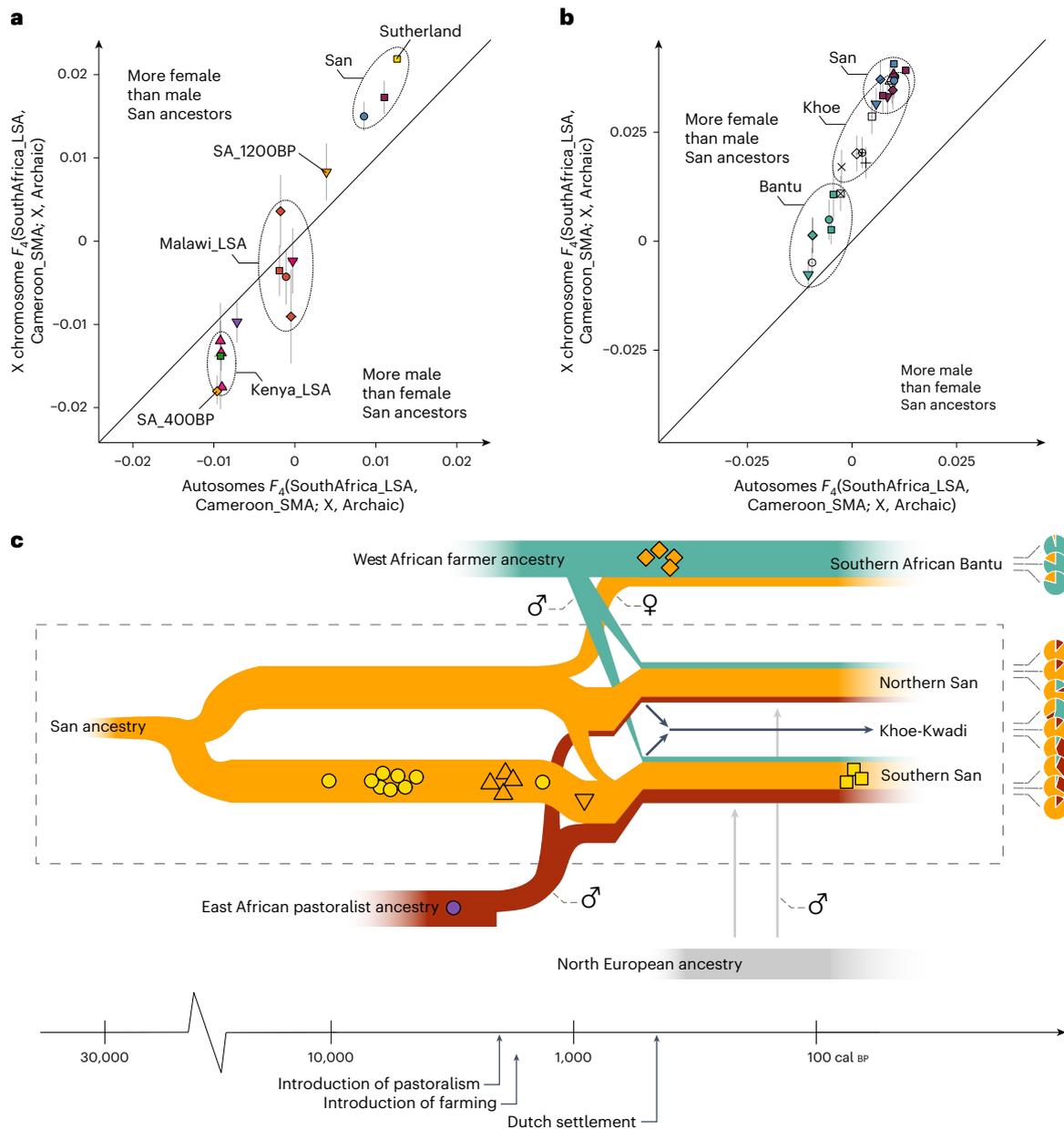
10,000 cal BP to the present-day. Arrows indicate  $P$  values from generalized-likelihood ratio tests implemented in qpWave testing for genetic continuity between temporally preceding and succeeding groups in the Northern Cape, Western Cape and KwaZulu-Natal, respectively. Discontinuities are explicitly marked as interrupted arrows. Pie charts depict the ancestry composition for each group derived from qpAdm modelling. Symbols and colours correspond to Fig. 1. Data can be found in Supplementary Tables 8 and 9.

### Demographic changes in the last 2,000 years

We use the increased availability of LSA data to quantify and characterize the demography of transitions in southern Africa during the last 2,000 years. Yet, these admixture events are challenging to reconstruct because of additional gene flow from at least two immigrant populations during prehistoric times<sup>1,2,7,8,27,46–49</sup> and additional inter- and intra-continental admixture following European settlement from the 1650s onwards<sup>46,50–52</sup>. This complex history hinders inferences about the timing and mode (for example, involving sex bias) of admixture events. To circumvent these issues, we focused on groups with only two of the various ancestries present in the region today and compared the resulting patterns to identify putative trajectories of non-LSA ancestries through Southern Africa. Specifically, we used qpAdm to test 1-source, 2-source or 3-source models (excluding individuals with European admixture for now) for present-day San, Khoe and Bantu-speaking populations. As sources, we used (1) Stone Age hunter-gatherers from South Africa (SA\_LSA), (2) Tanzania\_Luxmanda\_3000BP and (3) present-day Mende, reflecting the local LSA, pastoralist and farming ancestries<sup>2</sup>, respectively (Methods; Extended Data Fig. 7 and Supplementary Tables 10–12).

On the basis of the estimated admixture proportions in the best-fitting model using the lowest number of sources, we grouped populations into

primarily West African- or East African-admixed categories, excluding ambiguous cases with both non-San ancestries being present substantially (Supplementary Tables 13–16). Within these categories, we computed admixture dates for the West African component (Extended Data Fig. 8a and Supplementary Table 14) and the East African component (Extended Data Fig. 8b and Supplementary Table 13) in the relevant target groups. We find East African, Tanzania\_Luxmanda-related ancestry gene flow into San and Khoe populations to be consistently older than the gene flow from West African-related groups. For the Tanzania\_Luxmanda-related gene flow, we identify a mean admixture date of 1,068 BP among San and Khoe populations, agreeing with the observed East African ancestry in the 1,200-year-old pastoralist from Kasteelberg and the admixture date estimated for the nineteenth century Sutherland samples ( $1,228 \pm 278$  BP). In contrast, West African admixture in San, Khoe and Bantu groups is dated consistently younger than the Tanzania\_Luxmanda-related ancestry gene flow and also exhibits a difference between Bantu and San/Khoe. Specifically, the mean date for admixture of West African ancestry among Bantu-speaking populations in southern Africa (for example, Herero, Tswana and Kgalagadi) is estimated at 808 BP on average, remarkably agreeing with the dating of that admixture in the 400 cal BP Iron Age farmers from KwaZulu-Natal ( $832 \pm 139$  BP). In contrast, the



**Fig. 4 | Demographic changes in the San and Khoe populations of southern Africa.** **a**, Scatter plot of group-based jackknife point estimates for  $F_4$ -statistics of the form  $F_4$ (Archaic, X; Cameroon\_LSA, SA\_LSA) on loci of the X chromosome (y axis) and the autosomes (x axis), where X represents the ancient genomes ( $n = 17$ ) and two present-day San populations (Khomani\_San and Ju\_Hoan\_North). Error bars, 1 s.e. Data can be found in Supplementary Table 18. **b**, As **a** but for

present-day San, Khoe and Bantu-speaking populations ( $n = 25$ ). Error bars, 1 s.e. Data can be found in Supplementary Table 17. **c**, Summary of the inferred population history of the San and Khoe in southern Africa. Sex symbols indicate male- and female-biased reproduction. Note that pastoralism and farming both appeared in present-day South Africa at about the same time, 2,000 years ago. Symbols and colours correspond to Fig. 1.

estimated date of West African ancestry in San and Khoe groups is more recent (578 BP). This discrepancy might suggest successive waves of Bantu immigration<sup>53</sup> or continuing gene flow from Bantu-related groups into San and Khoe populations after the initial admixture event that contributed San and Khoe ancestry to Bantu-speaking groups (Extended Data Fig. 8).

For the mode of interaction between locals and newcomers, we find evidence for sex bias in most present-day San, Khoe and Bantu populations, with stronger signals in the San and Khoe populations compared to the Bantu-speaking groups (Fig. 4b and Supplementary Table 17). In general, both San and Khoe as well as Bantu groups exhibit significantly more SA\_LSA ancestry on the X chromosome than the autosomes and (congruently) share more drift with SA\_LSA on the X chromosome than on the autosomes (Fig. 4b). This suggests that

substantially more female than male San ancestors were involved in the admixture events following the spread of East African pastoralist and West African farmer ancestry, which is consistent with previous studies of uniparental and genome-wide markers<sup>19,20,22,28,29,47,48,54–57</sup>. Assuming a single admixture event (using the dates obtained from DATES analysis), we explicitly compared autosome to X chromosome ancestry to determine female ( $sf$ ) and male contributions ( $sm$ ) to the gene pools of selected Khoe and Bantu populations, using a previously described method<sup>58,59</sup> (Supplementary Table 20). For the Damara, we estimate that for each San man  $\sim 1.4$  San women contributed to the gene pool, for the #Hoan  $\sim 2.28$  San women per San man, for the Shua  $\sim 4$ , for the Haiiom (Haiom)  $\sim 5.2$  and for South African Bantu  $\sim 2.1$  San women per San man.

Although this signal of female-biased admixture is also evident in the historical Sutherland genomes and the 1,200-year-old pastoralist from Kasteelberg, we observe that the four Iron Age KwaZulu-Natal samples (SA\_400BP) share more drift with SA\_LSA on the autosomes than on the X chromosome, indicative of more male than female San ancestors (Fig. 4a and Supplementary Tables 18 and 19). This contradicts the pattern observed in most present-day Bantu or San and Khoe groups in South Africa and Botswana and might be related to changes in interaction between Bantu and San/Khoe groups after 400 BP. Overall, our results show that, despite an overarching trend of female-biased gene flow from San and Khoe populations into Bantu-speaking groups, the modes of interaction and reproduction were strongly influenced by locally and temporally defined factors after the initial arrival of the first farmers<sup>47</sup>.

Finally, we detect a comparatively recent admixture date corresponding to male-biased<sup>28,46,50–52</sup> gene flow from Europe into San/Khoe and mixed<sup>60</sup> groups from Colesberg and Wellington<sup>7</sup> (Supplementary Tables 21–23). This ancestry is best approximated by north-western Europeans, as shown by admixture  $F_3$ -statistics of the form  $F_3$  (Test, Sutherland, Target) (Supplementary Table 23). For instance, for South Africans of mixed ancestry from Colesberg, we observed that  $F_3$  values among 40 West Eurasian populations are minimized for Irish, Icelandic and Norwegian people, followed by English (all  $Z < -69$ ). We date the arrival of northern European ancestry among these populations to 199 yr BP (Extended Data Fig. 8 and Supplementary Table 15), postdating the settlement of South Africa by Dutch and British immigrants from the mid-1600s onwards, a development that led ultimately to the demise of most San and Khoe genetic, linguistic and cultural diversity in the region<sup>61</sup> and lastingly affected the demographic trajectories in southern Africa<sup>62</sup> (Fig. 4c). While population structure in South Africa partly collapsed, new extracontinent ancestries were introduced to the region, increasing the heterogeneity of the admixture landscape. To quantitatively estimate this influx from outside Africa and its impact on genetic diversity, we decomposed admixture sources using a supervised clustering approach implemented in the software ADMIXTURE<sup>63</sup> (Extended Data Fig. 9 and Supplementary Tables 21 and 22). For example, in South Africans of mixed ancestry from Colesberg, we observe on average  $24.4\% \pm 3.2\%$  South Asian ancestry and  $2.8\% \pm 0.5\%$  East Asian ancestry besides  $8.2\% \pm 1.4\%$  North European ancestry, yet only  $35.5\% \pm 2.9\%$  San-related ancestry. Besides South Africans of mixed ancestry, we furthermore detect European ancestry in the Karretjiemense (5.61%),  $\ddot{z}$ Khomani (9.45%) and Nama (6.83%), indicating that the southern San and Khoe were especially affected by admixture with European sources. Using the output from ADMIXTURE, we proceeded to measure the variability of ancestry components in the southern African groups via FSTruct<sup>64</sup>. We find that the relative levels of variation among southern San, namely the Karretjiemense,  $\ddot{z}$ Khomani and Nama are significantly higher than in any other present-day San or Khoe population because of the frequent presence of variable European ancestry components, comparable to the variability measured in South Africans of mixed ancestry from Colesberg and Wellington (Supplementary Tables 21 and 22 and Extended Data Fig. 9). This heterogeneity in non-Southern African admixture across individuals obscures the high genetic affinity to the ancient Oakhurst samples, as measured using  $F_{ST}$  and IBD metrics and highlights the necessity of further sampling of local communities to adequately assess the effect of non-southern African admixture on the current genetic landscape of San populations in southernmost Africa. On the other hand, northern and central San feature significantly lower variability, which is similar to the diversity observed in neighbouring Bantu-speaking people, who also do not exhibit substantial proportions of non-African ancestry (Extended Data Fig. 9).

## Discussion

The question of population continuity or discontinuity during the LSA of southern Africa has been the focus of anthropological research for well over a century. Archaeogenetic research of the last two decades

has revealed that the Holocene demographic histories of Stone Age Europe<sup>39,65–71</sup>, Asia<sup>43,72–77</sup> and North Africa<sup>78–80</sup> were characterized by several episodes of large-scale migrations, either in the form of admixture with newcomers or by total replacement of the established inhabitants. While these biological transformations modified the genetic make-up of the local populations, they were also vectors for technological innovation, such as the introduction of new technologies, raw material uses or subsistence strategies. In contrast, for South Africa, we demonstrate that the local gene pool was characterized by a prolonged period of genetic continuity with no (detectable) gene flow from outside southern Africa. The earliest individual in our study that yielded aDNA showed a genetic make-up indistinguishable from the later inhabitants of Oakhurst rockshelter, suggesting that this local 'southern' San gene pool was formed more than 10,000 years ago and remained isolated from admixture with neighbouring 'central' and 'northern' San populations<sup>78,26–29</sup> or with more distant sources to the northeast, which admixed with San groups in Malawi and Tanzania<sup>2</sup>.

Consequently, the sequence of cultural change at Oakhurst, for example from the Oakhurst to Wilton technocomplexes<sup>6,9,14,15</sup>, appears to result from local development initiated by the indigenous inhabitants<sup>14</sup>, highlighting the role of in situ innovations followed by acculturation. Our data also demonstrate that subtle fluctuations in the craniofacial size of South African LSA coastal inhabitants<sup>81,82</sup> (for example, between 4,000 and 3,000 BP) were not the product of genetic discontinuity but probably related to changes in environmental factors or population size<sup>81–83</sup>. Yet, we highlight that the inhabitants of Oakhurst were not a small, bottlenecked population. Genomic measurements of diversity indicate a degree of genetic variation comparable to other African hunter-gatherer populations and higher than Stone Age foragers from Europe or America. Furthermore, the current resolution of our methods and limited reference dataset in sub-Saharan Africa restricts our ability to detect subtle changes in group size or small-scale immigration of people from within southern Africa. However, our data are congruent with a population in reproductive isolation from other San (and non-San) populations over the whole period of occupation of the site.

This period of -9,000 years of genetic continuity ends rather abruptly in the migration events which introduced East and West African-related ancestry to South Africa, accompanied by the spread of herding and farming<sup>1,2,7,8</sup>. On the basis of present available data, it appears that non-southern African ancestry reached the southernmost parts of South Africa only after 1,300 cal BP. There is, however, abundant archaeological evidence of marked changes in subsistence and settlement patterns among coastal and near-coastal communities in this region from -2,000 cal BP. These changes have previously been interpreted as resulting from the disruption of hunter-gatherer communities by the emergence of herding<sup>84–87</sup>. Notably, a similar temporal discrepancy was observed during the Mesolithic–Neolithic transition in Europe, where the admixture between hunter-gatherers and incoming farmers postdates the emergence of agriculture by almost 2,000 years (ref. 88). This indicates that hunter-gatherers and farmers resided in close geographic proximity for a considerable time before mixing<sup>88</sup> and demonstrates that migration can precede any subsequent population admixture substantially. Alternatively, the practice of pastoralism may have spread to Southern Africa through a process of cultural diffusion in advance of substantial population expansion<sup>89,90</sup>, explaining the absence of any East African-related ancestry in South Africa before 1,300 cal BP.

Yet, the events of the last 1,300 years had a substantial impact on the local gene pool of South Africa. Today, all San and Khoe populations are admixed with one or both of East African Pastoralist and West African Farmer ancestry<sup>1,2,7,8</sup>. The collapse of the LSA population structure was accelerated by the arrival of European settlers in the mid-seventeenth century<sup>62</sup>. Together with the continuous loss of oral traditions, these issues contribute to our poor understanding

of the prehistoric southern African population structure. Using allele-frequency and IBD segment-based analyses, we were able to show that the present-day San and Khoe inhabitants of South Africa are, despite recent periods of disruption under Dutch and British rule, still directly related to the ancient Oakhurst individuals of the last 10,000 years. Especially among the ǀKhomani, Karretjiemense and Nama, who belong to the most admixed San/Khoe groups in southern Africa, some individuals still trace most of their ancestry back to these LSA hunter-gatherers. This also applies to the three historic San individuals from Sutherland dating to the late nineteenth century, who show only minor ancestry contribution from outside southern Africa and otherwise close autosomal and mitochondrial similarity to the LSA Oakhurst population<sup>41</sup>, demonstrating that the early Holocene gene pool of the Western Cape persisted in some regions throughout the last 2,000 years without major changes and that in some parts of southern Africa the long-lasting population continuity was not completely disrupted.

## Methods

### Study design and ethics

The human remains from the Oakhurst rockshelter site are housed in the University of Cape Town (UCT) human skeletal repository. The approach for permission to use these samples for aDNA was followed according to ref. 91, which included consultation with representatives from the San community in accordance with the South African Heritage Resources Agency and permission from the repository research committee. The Oakhurst samples were approved by the UCT human research ethics committee under ethics no. 715/2017 and Heritage Western Cape permit no. 17071302AS0718E.

The sampling strategy used was twofold: to be minimally invasive and serial sample through time in the occupation of the site. We selected only individuals with loose, broken or previously glued petrous bones so that upon return the samples could be re-glued back to their original state (only featuring a small, unnoticeable sampling hole). Additionally, a single tooth per individual was sampled. For the DNA libraries analysed in this study, 13 individuals were sampled (including 11 petrous bones and 12 teeth). OAK003.B and OAK003.C were initially thought to belong to the same individuals but in fact represent two distinct individuals. These small bone and tooth elements were shipped to Germany for sampling and (after processing) returned to South Africa. Collection of bone powder for aDNA extraction was performed as described in the section on 'Ancient DNA work' below.

New radiocarbon dates for this study were measured on the bone and tooth fragments sampled for DNA. These dates were obtained at the Curt-Engelhorn-Center Archaeometry gGmbH, Mannheim, using MICADAS-AMS. Collagen was extracted from the previously sampled bones, purified by ultrafiltration (fraction >30 kDa) and freeze-dried. The <sup>14</sup>C ages were normalized to  $\delta^{13}\text{C} = -25\%$ . The calibration was done using the SHCal20 calibration curve for the Southern Hemisphere<sup>92</sup>.

### Site background

Oakhurst rockshelter (33° 59' 00" S–34° 00' 00" S and 22° 35' 00" E–22° 43' 00" E) is important in the history of LSA studies in southern Africa. Excavations by John Goodwin from 1932 to 1935 produced many artefacts and some human skeletons. The site has substantial deposits (>2 m deep) extending over the last 10,000 years. It was one of the first in South Africa to be excavated in accordance with professional standards, with Goodwin and his team carrying out meticulous excavation and detailed recording. In the 1930s, a main goal was a better understanding of the stratigraphic (and thus temporal) relationships between different stone artefact assemblages, seen at that time as different 'cultures'.

Now that more sites have been excavated, we recognize that the large, relatively unstandardized stone artefacts from the lower part of the Oakhurst sequence form part of a widespread artefact-making

tradition in the terminal Pleistocene/early Holocene, extending across South Africa and into Zimbabwe and southern Namibia. Although there are regional variations, these assemblages are sufficiently similar that they are generally grouped as the 'Oakhurst technocomplex', acknowledging their early recognition at this site. At Oakhurst, this technocomplex extends through approximately the lower half of the deposits and dates to 9,000–8,000 BP. At about 8,000 BP there was a shift to microlithic (Wilton) assemblages, also very widely distributed across the subcontinent (and into East Africa). Goodwin distinguished 'Smithfield C', 'Wilton', 'Developed Wilton' and 'Wilton with pottery' but today we see these as an evolving microlithic tradition. Selection of fine-grained stone raw materials facilitated the making of tiny artefacts, with materials probably sourced over considerable distances. We note that the inhabitants of Oakhurst continued to make microlithic artefacts into the last 2,000 years (when they also made pottery), as seen at the sites of Boomplaas, further inland and Die Kelders, on the coast further to the west. Along most of the southern Cape coast, however, preferences in the last ~3,500 years shifted back to macrolithic artefacts with very little retouch, often made on locally available quartzite. A greater degree of spatial heterogeneity in the last few millennia is consistent with higher population densities and more territorial settlement patterns, as seen amongst hunter-gatherers in coastal and riverine areas elsewhere in the world.

Excavations at Oakhurst were made considerably more challenging by the many burials, with some grave shafts and even graves intersecting others. Some individuals could be recovered in their entirety, sometimes with rich grave goods, for example grave VIa (UCT 204). Others had been dispersed by disturbances in antiquity or by the roots of plants or burrowing animals, making it difficult to assess how many individuals are represented in the remains recovered.

### Ancient DNA work

**Collection of bone powder.** Sampling of 23 bone and teeth samples took place in clean-room facilities dedicated to aDNA work at the Max Planck Institute for Science of Human History in Jena (MPI-SHH). The sampling workflow included documenting and photographing the provided samples. For teeth, we either cut along the cementum/enamel junction and collected powder by drilling into the pulp chamber or accessed the pulp chamber by drilling the tooth transversally. For the petrous bones, we cut the petrous pyramid longitudinally to drill the dense part directly from either side<sup>93</sup>. We collected 28–178 mg of bone or tooth powder per sample for DNA extractions.

For four bone samples, more bone powder was obtained for <sup>14</sup>C dating at the Curt-Engelhorn-Center Archaeometry gGmbH.

**DNA extraction.** The aDNA was extracted following a modified protocol of ref. 94, as described in ref. 95, where we replaced the extended-MinElute-column assembly for manual extractions with columns from the Roche High Pure Viral Nucleic Acid Large Volume Kit<sup>96</sup> and for automated extraction with a protocol that replaced spin columns with silica beads in the purification step<sup>97</sup>.

**Library construction.** We generated 23 double-indexed<sup>98</sup>, double-stranded libraries using 25 µl of DNA extract and following established protocols<sup>99</sup>. We applied the partial UDG (UDG half)<sup>100</sup> protocol to remove most of the aDNA damage while preserving the characteristic damage pattern in the terminal nucleotides. Additionally, we generated 15 double-indexed single-stranded libraries<sup>101</sup> using 20 µl of DNA extract and applied no UDG treatment.

**Shotgun screening, capture and sequencing.** Libraries were sequenced inhouse on an Illumina HiSeq 4000 platform to an average depth of 5 million reads and after demultiplexing processed through EAGER<sup>102</sup>. After an initial quality check based on the presence of aDNA damage and endogenous DNA >0.1%, we subsequently selected and enriched

11 double-stranded and 15 single-stranded libraries using in-solution capture probes synthesized by Agilent Technologies for ~1,240k SNPs along the nuclear genome<sup>103</sup>. The captured libraries were sequenced for ~34 million reads on average (minimum, 17 million; maximum, 52 million) using a single end (1 × 75 base pair (bp) reads) configuration. Taking all double- and single-stranded libraries together, we generated 40–139 million reads for the 13 individuals (on average 62 million reads).

### aDNA data processing

**Read processing and aDNA damage.** After demultiplexing based on a unique pair of indexes, raw sequence data were processed using EAGER<sup>102</sup>. This included clipping sequencing adaptors from reads with AdapterRemoval (v.2.3.1)<sup>104</sup> and mapping of reads with BWA (Burrows–Wheeler Aligner) v.0.7.12 (ref. 105) against the human reference genome hg19, with seed length (-l) disabled, maximum number of differences (-n) of 0.01 and a quality filter (-q) of 30. We removed duplicate reads with the same orientation and start and end positions using DeDup v.0.12.2 (ref. 102). Terminal base deamination damage calculation was done using mapDamage v.2.0.6 (ref. 106), specifying a length (-l) of 100 bp. For the ten libraries that underwent UDG half treatment, we used BamUtil v.1.0.14 (<https://genome.sph.umich.edu/wiki/BamUtil:trimBam>) to clip two bases at the start and end of all reads for each sample to remove residual deaminations, thus removing genotyping errors that could arise as a result of aDNA damage.

**Sex determination.** To determine the genetic sex of the ancient individuals, we calculated the coverage on the autosomes as well as on each sex chromosome and subsequently normalized the X and Y reads by the autosomal coverage<sup>107</sup>. For that, we used a custom script (<https://github.com/TCLamnidis/Sex.DetERRmine>) for the calculation of each relative coverage as well as their associated error bars<sup>108</sup>. Females are expected to have an X rate of 1 and a Y rate of 0, whereas males are expected to have a rate of 0.5 for both X and Y chromosomes.

**Contamination estimation.** We used the ANGSD (analysis of next generation sequencing data) package<sup>109</sup> (v.0.923) to test for heterozygosity of polymorphic sites on the X chromosome in male individuals, applying a contamination threshold of 5% at the results of method 1. For male and female samples, we estimated contamination levels on the mtDNA using Schmutzi<sup>110</sup> (v.1.5.4) by comparing the consensus mitogenome of the ancient sample to a panel of 197 worldwide mitogenomes as a potential contamination source, applying a contamination threshold of 5%.

**Genotyping.** We used the program pileupCaller (v.1.4.0.2) (<https://github.com/stschiff/sequenceTools.git>) to genotype the trimmed BAM files of ten UDG half libraries. A pileup file was generated using samtools mpileup with parameters -q 30 -Q 30 -B containing only sites overlapping with our capture panel. From this file, for each individual and each SNP on the 1,240k panel<sup>39,40,111</sup>, one read covering the SNP was drawn at random and a pseudohaploid call was made; that is, the ancient individual was assumed homozygous for the allele on the randomly drawn read for the SNP in question. For the 15 single-stranded libraries that underwent no UDG treatment, we used the parameter -SingleStrandMode, which causes pileupCaller to ignore reads aligning to the forward strand at C/T polymorphisms and at G/A polymorphisms to ignore reads aligning to the reverse strand, which should remove postmortem damage in aDNA libraries prepared with the non-UDG single-stranded protocol. To maximize our resolution, we filled missing data in the single-stranded libraries with additional genotypes present in the trimmed, double-stranded libraries but not in the single-stranded libraries.

**Mitochondrial and Y chromosome haplogroup assignment.** To process the mtDNA data, we extracted reads from 1,240k data using samtools (v.1.3.1)<sup>112</sup> and mapped these to the revised Cambridge reference sequence. We subsequently called consensus sequences using

Geneious R9.8.1 (ref. 113) and used HaploGrep 2 (v.2.4.0)<sup>114</sup> (<https://haplogrep.uibk.ac.at/>); with PhyloTree v.17-FU1) to determine mitochondrial haplotypes. For the male individuals, we used pileup from the Rsamtools package to call the Y chromosome SNPs of the 1,240k SNP panel (mapping quality  $\geq 30$  and base quality  $\geq 30$ ). We then manually assigned Y chromosome haplogroups using pileups of Y-SNPs included in the 1,240k panel that overlap with SNPs included on the ISOGG SNP index v.15.73 (Y-DNA Haplogroup Tree 2019–2020; 2020.07.11).

**Identity-by-descent.** We imputed and phased individuals with >500,000 SNPs (OAK003B, OAK007, OAK012 and OAK013) using GLIMPSE<sup>115</sup> (v.2.0.0) (<https://github.com/odelaneau/GLIMPSE>), applying the default parameters and using the 1,000 genomes reference panel. Samples with >600,000 SNPs exhibiting a genotype posterior of  $\geq 0.99$  after imputation were included in downstream IBD analysis. Subsequently, we used BEAGLE<sup>116,117</sup> (v.5.2) to phase the newly imputed genotypes. Following ref. 118, the window and overlap lengths were set as wider than any chromosome (window length 380 cM and overlap length 190 cM) to maximize the information used for phasing the genomes. The 1,000 genomes phase 3 dataset ([https://bochet.gcc.biostat.washington.edu/beagle/1000\\_Genomes\\_phase3\\_v5a](https://bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a)) and GRCh37 genomic maps ([https://bochet.gcc.biostat.washington.edu/beagle/genetic\\_maps/](https://bochet.gcc.biostat.washington.edu/beagle/genetic_maps/)) provided by BEAGLE were used for phasing. The identification of IBD segments was done using RefinedIBD<sup>119</sup>. The window size was set to 3 cM. The minimal size for a segment to be considered shared by IBD is 1 cM, the same threshold used in refs. 118,120. Finally, we removed gaps between IBD segments that have at most one discordant homozygote and that are <0.6 cM in length and aggregated the sum and number of IBD segments between each pair of ancient and present-day individuals.

**Kinship estimation.** We calculated the pairwise mismatch rate<sup>121</sup> in all pairs of individuals from our pseudohaploid dataset to double-check for potential duplicate individuals and to determine first-, second- and third-degree relatives. However, no relatives were identified.

**Diversity estimation.** CNV was estimated by counting the differences between the ascertained pseudohaploid genotype calls present in one pair of individuals from the same population as described in the section on ‘Kinship estimation’ above. For these comparisons we grouped the individuals per geographic origin and time period. Results are reported as boxplots, where each dot corresponds to the CNV value for a unique pair of individuals. We also estimate average heterozygosity levels for the imputed genomes of OAK007, OAK012 and OAK013 as well as for two published Iron Age individuals from Botswana (XAR001 and XAR002) by taking the fraction of the number of heterozygous sites over the total number of sites across 22 autosomes<sup>122</sup>. Subsequently, we compared these estimates with the heterozygosity levels observed in 562 individuals belonging to 49 present-day sub-Saharan African populations.

### Population genetic analysis

**Dataset.** We merged our aDNA data with previously published datasets of ancient individuals reported by the Reich Lab in the Allen Ancient DNA Resource v.54.1 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>) (1,240k SNP panel)<sup>123</sup> (Supplementary Table 26). Present-day data from primarily sub-Saharan Africans were assembled from refs. 7,8,36,111,124 (human origins SNP panel and human origins-Schlebusch SNP panel) (Supplementary Table 27). We excluded recently admixed individuals for PCA, DYSTRUCT and qpAdm analysis from ref. 7 (see Data Availability).

**Naming.** Within tables and figures, we refer to populations by the names given in the Allen Ancient DNA resource v.54.1 (ref. 123).

Additionally, we refer to individuals from Oakhurst, St. Helena<sup>2</sup>, Faraoskop<sup>2</sup> and Ballito Bay<sup>1</sup> grouped together as South Africa LSA or SA\_LSA. We follow the San Council recommendations in using population-specific terms whenever possible and alternatively use the terms San for Tuu and K'xaa language-speaking hunter-gatherer groups and Khoe for Khoe-Kwadi speakers. Within the text, we spell names of San and Khoe groups using click consonant symbols. Within figures and tables we refer to populations using the labels from the original publications of the genotype data. When necessary we collectively refer to groups with indigenous southern Africa-specific ancestry as having San-related ancestry. We used the label 'coloured' for some groups in the figures and supplementary tables following the labelling in the original publications of these genomes<sup>1,7</sup>. This bureaucratic denotation refers to South Africans of mixed ancestry, who represent a biologically heterogeneous group with variable and complex admixture from indigenous San and Khoe, European, Bantu-speaking African, Asian and Madagascar Cape slaves or migrants<sup>60</sup>. In the text, we use the label South Africans of mixed ancestry for these individuals, acknowledging that many are genetically homogenous to one origin yet were classified during apartheid under a single racial label, which is still used today<sup>60</sup>.

**Principal components analysis.** We carried out PCA using the smartpca software v.16000 from the EIGENSOFT package (v.6.0.1)<sup>125</sup>. We computed principal components on two different sets of southern African populations and projected ancient individuals using Isqproject: YES and shrinkmode: YES. Dataset (1) includes 24 San, Khoe and Bantu-speaking populations from three sources (refs. 7,8,111) as well as 212,000 SNPs (Fig. 1b); dataset (2) includes only 22 San, Khoe and Bantu-speaking populations from two sources (refs. 25,111) but 597,000 SNPs (Extended Data Fig. 1b). We highlight that the PCA computed on dataset (1) better reflects the genetic diversity within southern San (which is under-represented in dataset (2) because of the lack of samples from South Africa).

**F-statistics.**  $F_3$ - and  $F_4$ -statistics were computed with ADMIXTOOLS v.3.0 (ref. 35) (<https://github.com/DReichLab>).  $F_3$ -statistics were calculated using qp3Pop (v.435). For  $F_4$ -statistics, we used the qpDstat (v.755) and with the activated  $F_4$ -mode. Significant deviation from zero can be interpreted as rejection of the tree population typology ((Outgroup, X);(Pop1, Pop2)). Under the assumption that no gene flow occurred between Pop1 and Pop2 and the Outgroup, a positive  $F$ -statistic suggests affinity between X and Pop2, whilst a negative value indicates affinity between X and Pop1. Standard errors were calculated with the default blockjackknife 5 cM in size. As outgroup for  $F_3$ - and  $F_4$ -statistics, we used either diploid genotypes from two archaic human genomes (a Neanderthal<sup>126</sup> and a Denisovan<sup>127</sup>) or haploid genotypes from one chimpanzee genome (the chimpanzee genome is required for technical reasons as an outgroup to all humans).

**Fixation index.** We calculated  $F_{ST}$  using smartpca software v.16000 from the EIGENSOFT package (v.6.0.1)<sup>125</sup> with the fstonly, inbreed and fsthprecision options set to YES.

**Inference of mixture proportions.** We estimated ancestry proportions using qpWave<sup>39,128</sup> (v.410) and qpAdm<sup>39</sup> (v.810) from ADMIXTOOLS v.3.0 (ref. 35) with the allsnps: YES and inbreeding: YES options and one basic set of 11 outgroups modified from ref. 33: Mbuti, Dinka, Ju\_hoan\_North, Turkey\_N<sup>129</sup>, Iran\_GanjDareh\_N<sup>40</sup>, French, Sardinian, Punjabi, Ami, Papuan and Karitiana.

For group-based qpAdm analysis, we tested for each ancient and present-day population 1-, 2- and 3-way admixtures scenarios between SA\_LSA (consisting of Oakhurst, without OAK006, St. Helena, Faraoskop and Ballito Bay), Tanzania\_Luxmanda\_3000BP and Mende.DG<sup>2</sup>. We selected for each population the admixture model with  $P > 0.01$  featuring the lowest number of sources.

To analyse potential sex bias in the admixture process, we used qpAdm to estimate SA\_LSA admixture proportions on the autosomes (default option) and on the X chromosome (option "chrom:23") using the abovementioned outgroups. Following the approach established by ref. 42, Z-scores were calculated for the difference between the autosomes and the X chromosome using the formula  $Z = \frac{pA - pX}{\sqrt{\sigma A^2 + \sigma X^2}}$  where

$pA$  and  $pX$  are the SA\_LSA admixture proportions on the autosomes and the X chromosome and  $\sigma A$  and  $\sigma X$  are the corresponding jackknife standard deviations<sup>42</sup>. Thus, a negative Z-score means that there is more SA\_LSA admixture on the X chromosome than on the autosomes, indicating that the SA\_LSA admixture was female-biased.

**Ancestry decomposition.** We performed model-based clustering analysis using two different approaches: (1) We applied DYSTRUCT<sup>30</sup>, with a cluster number ( $K$ ) ranging between 2 and 10. Mean radiocarbon ages for the ancient individuals included were converted to generation ages (assuming a generation time of 30 yr; refs. 130,131) and provided for the analysis. (2) We applied ADMIXTURE<sup>63</sup> in supervised mode using modern reference populations at  $K = 7$ . This analysis was run on haploid data with the parameter -haploid set to all ("=\*"). To obtain point estimates for populations, we averaged individual point estimates and calculated the s.e.m. As modern references we used the following groupings: San (Ju\_hoan\_North.DG, Khomani\_San.DG), West Africa (YRI.SG, ESN.SG), East Africa (Somali, Masai, Sandawe), South Europe (TSL.SG, IBS.SG), North Europe (CEU.SG, GBR.SG), South Asia (PJI.SG, GIH.SG), East Asia (CHB.SG, JPT.SG)<sup>111,132,133</sup>. The  $Q$  matrix of this ADMIXTURE analysis was also used as input for FSTruct as described by the authors<sup>64</sup>.

**Potential ascertainment bias.** Recent analyses have shown that co-modelling more than one sub-Saharan African and/or archaic human groups (Neanderthals and Denisovans) using  $F$ -statistics in a non-outgroup-ascertained SNP panel can lead to false rejection of true demographic histories and failure to reject incorrect models in  $F_4$ -derived methods like qpAdm<sup>134</sup>. However, most  $F_4$ -statistics themselves remain minimally affected by ascertainment<sup>134</sup>. Regarding qpAdm analysis, we compared our results calculated on the complete 1,240k or human origins SNP panels with the published results by ref. 2. That study<sup>2</sup> obtained an outgroup-ascertained set of 814,242 transversion SNPs polymorphic between the Denisovan and Neanderthal genomes to minimize the bias on  $F$ -statistics<sup>2</sup>. We find our qpAdm estimates to be highly correlated with the ones reported by Skoglund et al.<sup>2</sup> (Extended Data Fig. 7a) (Pearson's product-moment correlation;  $t = 22.244$ , d.f. = 17,  $P = 5.23 \times 10^{-14}$ ,  $\text{cor} = 0.983$ ), suggesting a negligible effect of ascertainment bias on our results. The qpWave analysis was restricted to A/T and G/C sites in the 1,240k SNP panels as recommended by ref. 134.

**Admixture dating.** Admixture dates between SA\_LSA and Tanzania\_Luxmanda, Mende or English as sources were calculated using DATES (distribution of ancestry tracts of evolutionary signals) (v.4010)<sup>88</sup> using standard settings. A default bin size of 0.001 M is applied in our estimates (flag "binsize:0.001" added). We used a standard of 29 years per generation to convert the generation times in years since admixture.

**Maximum likelihood tree.** We constructed maximum likelihood trees using TreeMix (v.1.12)<sup>25</sup>. For each tree, we performed a round of global rearrangements after adding all populations (-global) and calculated 100 bootstrap replicates to assess the uncertainty of the fitted model (-bootstrap). Sample size correction was disabled.

## Reporting summary

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

## Data availability

Raw sequence data (fastq files) and mapped data (bam files) from the 13 newly reported ancient individuals will be available before publication from the European Nucleotide Archive under accession no. [PRJEB77188](https://www.ebi.ac.uk/ena/record/PRJEB77188). A Poseidon package of the genotype data analysed in this paper is available on the Poseidon Community Archive ([https://www.poseidon-adna.org/#/archive\\_explorer](https://www.poseidon-adna.org/#/archive_explorer)). Owing to ethical prescriptions of this research under UCT human research ethics no. 715/2017, DNA sequencing libraries, both before and after SNP capture, are available for replication upon request to the corresponding authors and the UCT Human Skeletal Repository Committee at [uctskeletalrepository-group@uct.ac.za](mailto:uctskeletalrepository-group@uct.ac.za) as aliquots, pending consultation, approval and permission by the UCT Skeletal Repository Committee and consulted San communities who granted the original sample access. Previously published genotype data for ancient and present-day individuals were reported by the Reich Lab in the Allen Ancient DNA Resource v.54.1 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). Additional previously published genotype data for the present-day San and Khoe samples from ref. 7 are available at the Arrayexpress database (<https://www.ebi.ac.uk/arrayexpress/>) under accession no. E-MTAB-1259. The Genome Reference Consortium Human Build 37 (GRCh37/hg19) is available via the National Center for Biotechnology Information under accession no. [PRJNA31257](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.1). The revised Cambridge reference sequence is available via the National Center for Biotechnology Information under reference sequence [NC\\_012920.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_012920.1).

## Code availability

All software used in this work is publicly available. List of software and respective versions: AdapterRemoval (v.2.3.1), Burrows–Wheeler Aligner (v.0.7.12), DeDup (v.0.12.2), mapDamage (v.2.0.6), BamUtil (v.1.0.14), EAGER (v.1), Sex.DetERRmine (v.1.1.2) (<https://github.com/TCLamnidis/Sex.DetERRmine>), ANGSD (v.0.915), Schmutzi (v.1.5.4), PMDtools (v.0.50), pileupCaller (v.1.4.0.2), samtools (v.1.3.1), Geneious (R9.8.1), HaploGrep 2 (v.2.4.0), READ (<https://bitbucket.org/tgunter/read>) (v.f541d55), PLINK (v.1.90b3.29), Picard tools (v.2.27.3), smartpca (v.16000; EIGENSOFT v.6.0.1), qp3Pop (v.435; ADMIXTOOLS v.3.0), qpDstat (v.755; ADMIXTOOLS v.3.0), qpWave (v.410), qpAdm (v.810), DATES (v.4010), ADMIXTURE (v.1.3), GLIMPSE (<https://github.com/odelaneau/GLIMPSE>) (v.2.0.0), DyStruct (<https://github.com/tyjo/dystruct>) (v.2.0.0), BEAGLE (v.5.4), RefinedIBD (v.17Jan20.102), FSTruct (<https://github.com/MaikeMorrison/FSTruct>) (d39827e) and TreeMix (v.1.12). Data visualization and descriptive statistical tests were performed in R (v.4.1.1). The following R packages were used: Rsamtools (v.2.12.0), vegan (v.2.6-2), factoextra (v.1.0.7), ggplot2 (v.3.3.6), ggExtra (v.0.10.0), ggforce (v.0.3.3), rnatlearn (v.0.1.0), sf (v.1.0-8), raster (v.3.5-21), rgdal (v.1.5-32), spatstat (v.2.3-4), maptools (v.1.1-4), gstat (v.2.0-9), sp (v.1.5-0), labdsv (v.2.0-1), rcarbon (v.1.5.1), magrittr (v.2.0.3), dplyr (v.1.0.9), reshape 2 (v.1.4.4) and tidyverse (v.1.3.2). Y chromosome and mtDNA haplogroups were determined using the ISOGG SNP index (v.15.73) and PhyloTree (v.17-FU1) reference databases, respectively.

## References

- Schlebusch, C. M. et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
- Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71 (2017).
- Dusseldorp, G., Lombard, M. & Wurz, S. Pleistocene *Homo* and the updated Stone Age sequence of South Africa. *S. Afr. J. Sci.* **109**, 7 (2013).
- Mounier, A. & Mirazón Lahr, M. Deciphering African late middle Pleistocene hominin diversity and the origin of our species. *Nat. Commun.* **10**, 3406 (2019).
- Wadley, L. Legacies from the Later Stone Age. *Goodwin Ser.* **6**, 42–53 (1989).
- Sealy, J. in *Africa from MIS 6-2. Population Dynamics and Paleoenvironments* (eds Jones, S. C. & Stewart, B. A.) 65–75 (Springer Nature, 2016).
- Schlebusch, C. M. et al. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012).
- Pickrell, J. K. et al. The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
- Goodwin, A. J. H. Archaeology of the Oakhurst Shelter George, Part I: Course of the Excavation. *Trans. R. Soc. S. Afr.* **25**, 229–245 (1937).
- Wadley, L. & Laue, G. Adullam Cave, eastern Free State, South Africa: test excavations at a multiple-occupation Oakhurst Industry site. *South Afr. Humanit.* **12**, 1–13 (2000).
- Low, M. Continuity, variability and the nature of technological change during the Late Pleistocene at Klipfonteinrand Rockshelter in the Western Cape, South Africa. *Afr. Archaeol. Rev.* **36**, 67–88 (2019).
- Forssman, T. A review of hunter-gatherers in Later Stone Age research in Southern Africa. *Goodwin Ser.* **12**, 56–68 (2019).
- Lombard, M. et al. The South African stone age sequence updated (II). *South Afr. Archaeol. Bull.* **77**, 172–212 (2022).
- Fagan, B. M. The Glentyre Shelter and Oakhurst re-examined. *South Afr. Archaeol. Bull.* **15**, 80–94 (1960).
- Schrire, C. Oakhurst: a re-examination and vindication. *South Afr. Archaeol. Bull.* **17**, 181–195 (1962).
- Drennan, M. R. Archaeology of the Oakhurst Shelter, George. Part III: The Cave-Dwellers. *Trans. R. Soc. S. Afr.* **25**, 259–280 (1937).
- Patrick, M. K. *An Archaeological, Anthropological Study of the Human Skeletal Remains from the Oakhurst Rockshelter, George, Cape Province, Southern Africa* (Univ. of Cape Town, 1989).
- Underhill, P. A. et al. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
- Wood, E. T. et al. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* **13**, 867–876 (2005).
- Schlebusch, C. M., de Jongh, M. & Soodyall, H. Different contributions of ancient mitochondrial and Y-chromosomal lineages in ‘Karretjie people’ of the Great Karoo in South Africa. *J. Hum. Genet.* **56**, 623–630 (2011).
- Barbieri, C. et al. Ancient substructure in early mtDNA lineages of southern Africa. *Am. J. Hum. Genet.* **92**, 285–292 (2013).
- Schlebusch, C. M., Lombard, M. & Soodyall, H. MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evol. Biol.* **13**, 56 (2013).
- Rito, T. et al. The first modern human dispersals across Africa. *PLoS ONE* **8**, e80031 (2013).
- Naidoo, T. et al. Y-chromosome variation in Southern African Khoe-San populations based on whole-genome sequences. *Genome Biol. Evol.* **12**, 1031–1039 (2020).
- Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- Uren, C. et al. Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. *Genetics* **204**, 303–314 (2016).
- Montinaro, F. et al. Complex ancient genetic structure and cultural transitions in Southern African populations. *Genetics* **205**, 303–316 (2017).
- Vicente, M. et al. Male-biased migration from East Africa introduced pastoralism into southern Africa. *BMC Biol.* **19**, 259 (2021).

29. Pfennig, A., Petersen, L. N., Kachambwa, P. & Lachance, J. Evolutionary genetics and admixture in African populations. *Genome Biol. Evol.* **15**, evad054 (2023).
30. Joseph, T. A. & Pe'er, I. Inference of population structure from time-series genotype data. *Am. J. Hum. Genet.* **105**, 317–333 (2019).
31. Lipson, M. et al. Ancient West African foragers in the context of African population history. *Nature* **577**, 665–670 (2020).
32. Prendergast, M. E. et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* **365**, eaaw6275 (2019).
33. Wang, K. et al. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183 (2020).
34. Lipson, M. et al. Ancient DNA and deep population structure in sub-Saharan African foragers. *Nature* **603**, 290–296 (2022).
35. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
36. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
37. Barbieri, C. et al. Unraveling the complex maternal history of Southern African Khoisan populations. *Am. J. Phys. Anthropol.* **153**, 435–448 (2014).
38. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
39. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
40. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
41. Gibbon, V. E. et al. Confronting historical legacies of biological anthropology in South Africa—restitution, redress and community-centered science: the Sutherland Nine. *PLoS ONE* **18**, e0284785 (2023).
42. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
43. Wang, C.-C. et al. Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).
44. Posth, C. et al. Reconstructing the deep population history of Central and South America. *Cell* **175**, 1185–1197 (2018).
45. van Eeden, G. et al. The recombination landscape of the Khoe-San likely represents the upper limits of recombination divergence in humans. *Genome Biol.* **23**, 172 (2022).
46. Choudhury, A. et al. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nat. Commun.* **8**, 2062 (2017).
47. Choudhury, A., Sengupta, D., Ramsay, M. & Schlebusch, C. Bantu-speaker migration and admixture in southern Africa. *Hum. Mol. Genet.* **30**, R56–R63 (2021).
48. Sengupta, D. et al. Genetic substructure and complex demographic history of South African Bantu speakers. *Nat. Commun.* **12**, 2080 (2021).
49. Fortes-Lima, C. A. et al. The genetic legacy of the expansion of Bantu-speaking peoples in Africa. *Nature* **625**, 540–547 (2024).
50. Patterson, N. et al. Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* **19**, 411–419 (2010).
51. Petersen, D. C. et al. Complex patterns of genomic admixture within southern Africa. *PLoS Genet.* **9**, e1003309 (2013).
52. Hollfelder, N. et al. Patterns of African and Asian admixture in the Afrikaner population of South Africa. *BMC Biol.* **18**, 16 (2020).
53. Huffman, T. N. *Handbook to the Iron Age: The Archaeology of Pre-Colonial Farming Societies in Southern Africa* (Univ. KwaZulu-Natal Press, 2007).
54. Coelho, M., Sequeira, F., Luiselli, D., Beleza, S. & Rocha, J. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.* **9**, 80 (2009).
55. Quintana-Murci, L. et al. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am. J. Hum. Genet.* **86**, 611–620 (2010).
56. 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 (2012).
57. Bajić, V. et al. Genetic structure and sex-biased gene flow in the history of southern African populations. *Am. J. Phys. Anthropol.* **167**, 656–671 (2018).
58. Goldberg, A., Verdu, P. & Rosenberg, N. A. Autosomal admixture levels are informative about sex bias in admixed populations. *Genetics* **198**, 1209–1229 (2014).
59. Goldberg, A. & Rosenberg, N. A. Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. *Genetics* **201**, 263–279 (2015).
60. Tawha, T., Dinkele, E., Mole, C. & Gibbon, V. E. Assessing zygomatic shape and size for estimating sex and ancestry in a South African sample. *Sci. Justice* **60**, 284–292 (2020).
61. Sinclair-Thomson, B. & Challis, S. Runaway slaves, rock art and resistance in the Cape Colony, South Africa. *Azania Archaeol. Res. Afr.* **55**, 475–491 (2020).
62. Schlebusch, C. M., Prins, F., Lombard, M., Jakobsson, M. & Soodyall, H. The disappearing San of southeastern Africa and their genetic affinities. *Hum. Genet.* **135**, 1365–1373 (2016).
63. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
64. Morrison, M. L., Alcalá, N. & Rosenberg, N. A. FSTruct: an FST-based tool for measuring ancestry variation in inference of population structure. *Mol. Ecol. Resour.* **22**, 2614–2626 (2022).
65. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
66. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
67. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
68. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
69. Villalba-Mouco, V. et al. Survival of Late Pleistocene hunter-gatherer ancestry in the Iberian Peninsula. *Curr. Biol.* **29**, 1169–1177 (2019).
70. Posth, C. et al. Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature* **615**, 117–126 (2023).
71. Villalba-Mouco, V. et al. A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. *Nat. Ecol. Evol.* **7**, 597–609 (2023).
72. Siska, V. et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3**, e1601877 (2017).
73. Lipson, M. et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
74. Yang, M. A. et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020).
75. Ning, C. et al. Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* **11**, 2700 (2020).
76. Carlhoff, S. et al. Genome of a middle Holocene hunter-gatherer from Wallacea. *Nature* **596**, 543–547 (2021).

77. Wang, K. et al. Middle Holocene Siberian genomes reveal highly connected gene pools throughout North Asia. *Curr. Biol.* **33**, 423–433 (2023).
78. Fregel, R. et al. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc. Natl Acad. Sci. USA* **115**, 6774–6779 (2018).
79. van de Loosdrecht, M. et al. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
80. Simões, L. G. et al. Northwest African Neolithic initiated by migrants from Iberia and Levant. *Nature* **618**, 550–556 (2023).
81. Stnyder, D. D., Ackermann, R. R. & Sealy, J. C. Craniofacial variation and population continuity during the South African Holocene. *Am. J. Phys. Anthropol.* **134**, 489–500 (2007).
82. Irish, J. D., Black, W., Sealy, J. & Ackermann, R. R. Questions of Khoesan continuity: dental affinities among the indigenous Holocene peoples of South Africa. *Am. J. Phys. Anthropol.* **155**, 33–44 (2014).
83. Pfeiffer, S. & Harrington, L. Growth, mortality, and small stature. *Curr. Anthropol.* **52**, 449–461 (2011).
84. Sealy, J. C. & Van der Merwe, N. J. Social, spatial and chronological patterning in marine food use as determined by  $\delta^{13}\text{C}$  measurements of Holocene human skeletons from the south-western Cape, South Africa. *World Archaeol.* **20**, 87–102 (1988).
85. Parkington, J. et al. In *The Archaeology of Prehistoric Coastlines* (eds Bailey, G. & Parkington, J. E.) 22–41 (Cambridge Univ. Press, 1988).
86. Sealy, J. Diet, mobility, and settlement pattern among Holocene hunter-gatherers in Southernmost Africa. *Curr. Anthropol.* **47**, 569–595 (2006).
87. Jerardino, A. Large shell middens and hunter-gatherer resource intensification along the west coast of South Africa: the Elands Bay case study. *J. Isl. Coast. Archaeol.* **7**, 76–101 (2012).
88. Chintalapati, M., Patterson, N. & Moorjani, P. The spatiotemporal patterns of major human admixture events during the European Holocene. *eLife* **11**, e77625 (2022).
89. Jerardino, A., Fort, J., Isern, N. & Rondelli, B. Cultural diffusion was the main driving mechanism of the Neolithic transition in southern Africa. *PLoS ONE* **9**, e113672 (2014).
90. Sadr, K. Livestock first reached southern Africa in two separate events. *PLoS ONE* **10**, e0134215 (2015).
91. Gibbon, V. E. African ancient DNA research requires robust ethics and permission protocols. *Nat. Rev. Genet.* **21**, 645–647 (2020).
92. Hogg, A. G. et al. SHCal20 Southern Hemisphere calibration, 0–55,000 years calBP. *Radiocarbon* **62**, 759–778 (2020).
93. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
94. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
95. Velsko, I., Skourtanioti, E. & Brandt, G. Ancient DNA Extraction from Skeletal Material. *protocols.io* [www.protocols.io/view/ancient-dna-extraction-from-skeletal-material-baksicwv](http://www.protocols.io/view/ancient-dna-extraction-from-skeletal-material-baksicwv) (2020).
96. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
97. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
98. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
99. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
100. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
101. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
102. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
103. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
104. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
105. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
106. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
107. Mittnik, A., Wang, C.-C., Svoboda, J. & Krause, J. A molecular approach to the xexing of the triple burial at the Upper Paleolithic site of Dolní Věstonice. *PLoS ONE* **11**, e0163019 (2016).
108. Lamnidis, T. C. et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* **9**, 5018 (2018).
109. Korneliusson, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356 (2014).
110. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
111. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
112. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
113. Kears, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
114. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
115. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
116. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
117. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
118. Morez, A. et al. Imputed genomes and haplotype-based analyses of the Picts of early medieval Scotland reveal fine-scale relatedness between Iron Age, early medieval and the modern people of the UK. *PLoS Genet.* **19**, e1010360 (2023).
119. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
120. Margaryan, A. et al. Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).
121. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).

122. Wang, K. et al. High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry. *Cell Genomics* **3**, 100377 (2023).
123. Mallick, S. et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data* **11**, 182 (2024).
124. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
125. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
126. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
127. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
128. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
129. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
130. Moorjani, P. et al. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc. Natl Acad. Sci. USA* **113**, 5652–5657 (2016).
131. Wang, R. J., Al-Saffar, S. I., Rogers, J. & Hahn, M. W. Human generation times across the past 250,000 years. *Sci. Adv.* **9**, eabm7047 (2023).
132. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
133. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
134. Flegontov, P. et al. Modeling of African population history using *f*-statistics is biased when applying all previously proposed SNP ascertainment schemes. *PLoS Genet.* **19**, e1010931 (2023).

## Acknowledgements

We acknowledge and thank the descendants of Oakhurst and all those who have contributed to the global aDNA data. A special note of thanks to the individuals whose remains rest in the repositories assessed, without whom this research would not be possible. We thank the consulted communities, UCT human research ethics committee and Heritage Western Cape for permitting this research. We thank the curators and staff of the UCT human skeletal repository. V.E.G. acknowledges financial support for this research by the South African National Research Foundation (grant nos. 115357 and 120816) and J.C.S. by the South African Research Chairs Initiative (grant no. 84407). Opinions expressed and conclusions arrived at are those of the authors and not necessarily to be attributed to the South African National Research Foundation. We thank T. C. Lamnidis for his help in uploading the aDNA data to the European Nucleotide Archive.

## Author contributions

V.E.G., J.C.S., D.C.S.-G., S.S. and J.K. conceived the study. V.E.G., J.C.S. and D.C.S.-G. provided archaeological samples and context. S.E.P. performed laboratory analyses. J.G., S.E.P. and A.B.R. analysed data. V.E.G., J.C.S., S.S. and J.K. supervised the project. J.G., V.E.G., J.C.S. and S.S. wrote the paper with input from all coauthors.

## Funding

Open access funding provided by Max Planck Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-024-02532-3>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02532-3>.

**Correspondence and requests for materials** should be addressed to Victoria E. Gibbon or Stephan Schiffels.

**Peer review information** *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.

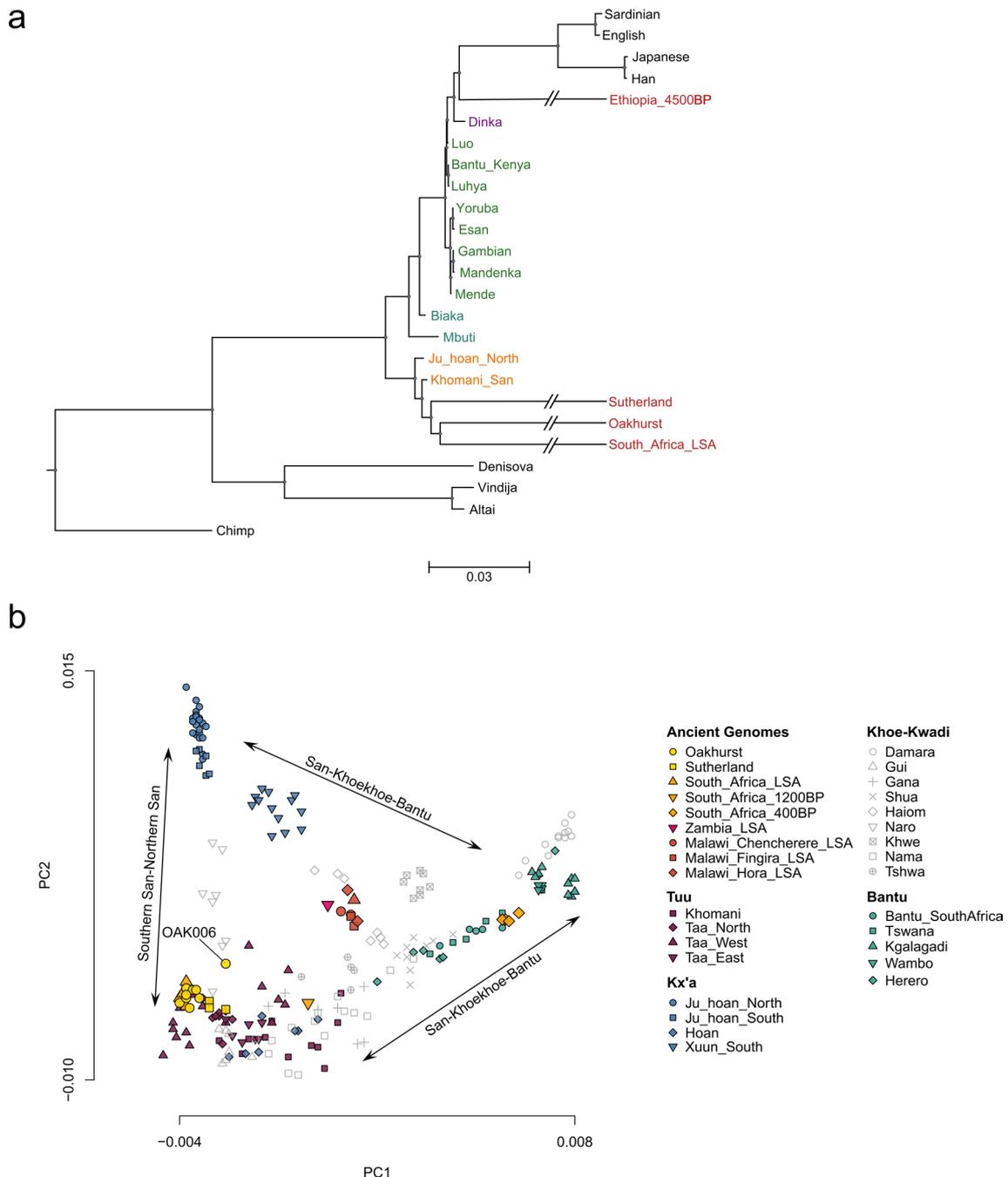
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

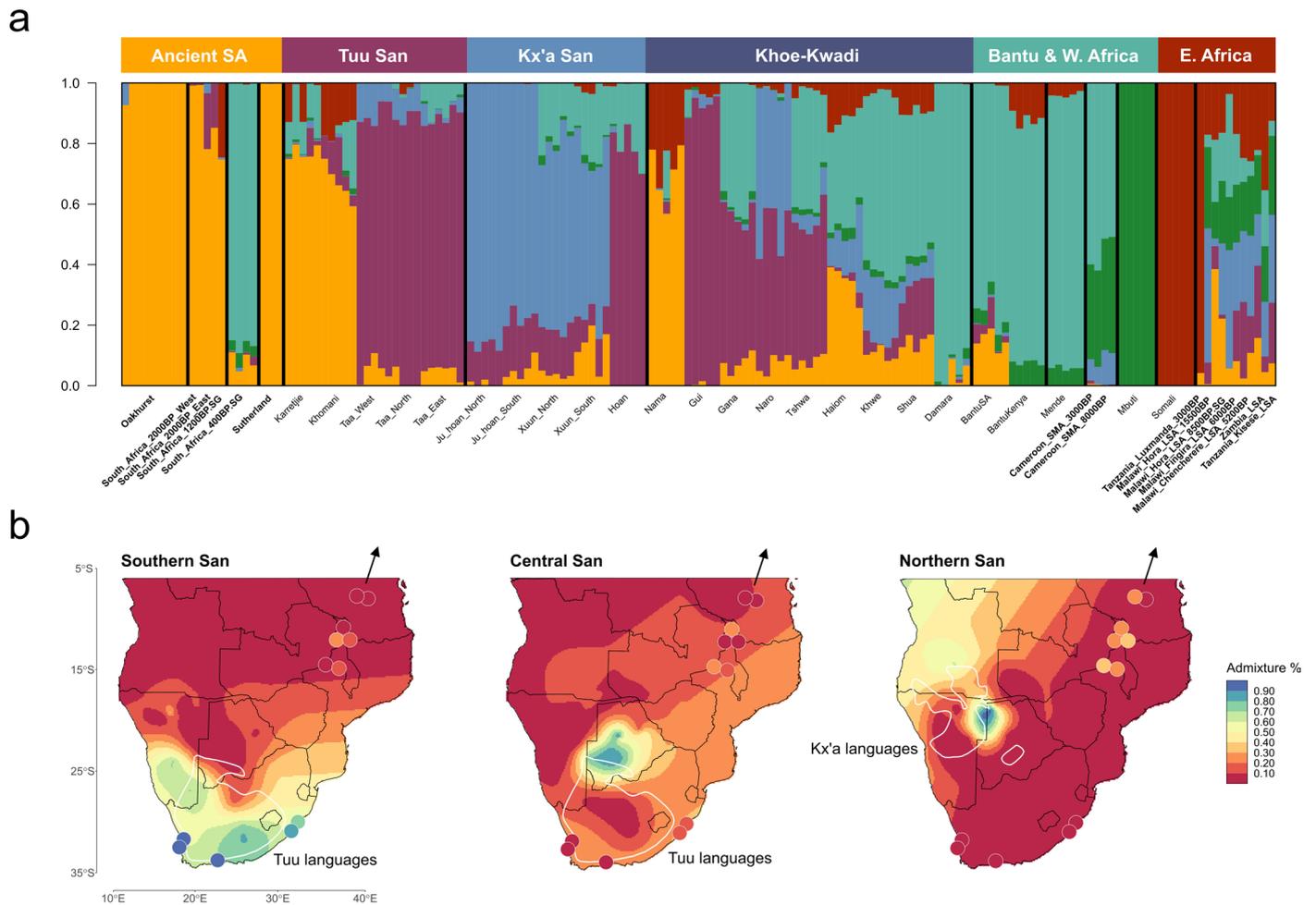
© The Author(s) 2024

<sup>1</sup>Max Planck Institute for Evolutionary Anthropology, Department of Archaeogenetics, Leipzig, Germany. <sup>2</sup>Division of Clinical Anatomy and Biological Anthropology, Department of Human Biology, University of Cape Town, Cape Town, South Africa. <sup>3</sup>Department of Archaeology, University of Cape Town, Cape Town, South Africa. <sup>4</sup>School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, South Australia, Australia. <sup>5</sup>Department of Geological Sciences, University of Cape Town, Cape Town, South Africa. <sup>6</sup>Departament de Prehistòria, Arqueologia i Història Antiga, Universitat de València, València, Spain. <sup>7</sup>These authors contributed equally: Johannes Krause, Stephan Schiffels. ✉ e-mail: [victoria.gibbon@uct.ac.za](mailto:victoria.gibbon@uct.ac.za); [stephan\\_schiffels@eva.mpg.de](mailto:stephan_schiffels@eva.mpg.de)



**Extended Data Fig. 1 | Genetic affinities between ancient and present-day southern Africans. a)** Maximum likelihood tree, generated using TreeMix (see Methods) of genome sequences from present-day and ancient populations, excluding populations with evidence of asymmetrical allele sharing with non-Africans indicative of recent gene flow. Branches of ancient individuals/groups are truncated for better readability. **b)** Alternative PCA computed using

22 present-day populations from southern Africa and 597,000 SNPs. This PCA drops two populations from the 24 shown in Fig. 1b, for the benefit of using a larger number of SNPs. The arrows indicate the three distinct clines of genetic variation between Bantu groups and Northern or Southern San, respectively. Khoekhoe populations cluster in between the three poles.

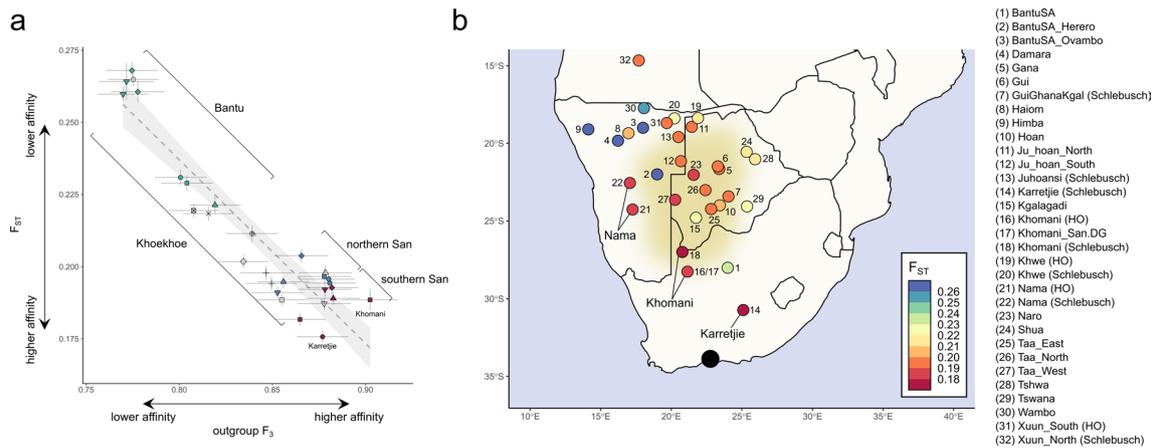


**Extended Data Fig. 2 | Sample age-informed ancestry decomposition.**

**a)** Clustering of 162 ancient and present-day sub-Saharan African individuals (~212,000 SNPs) at  $K = 6$  using DYSTRUCT. Major groups are indicated on top of the bar plots based on chronological, geographical and linguistic classifications.

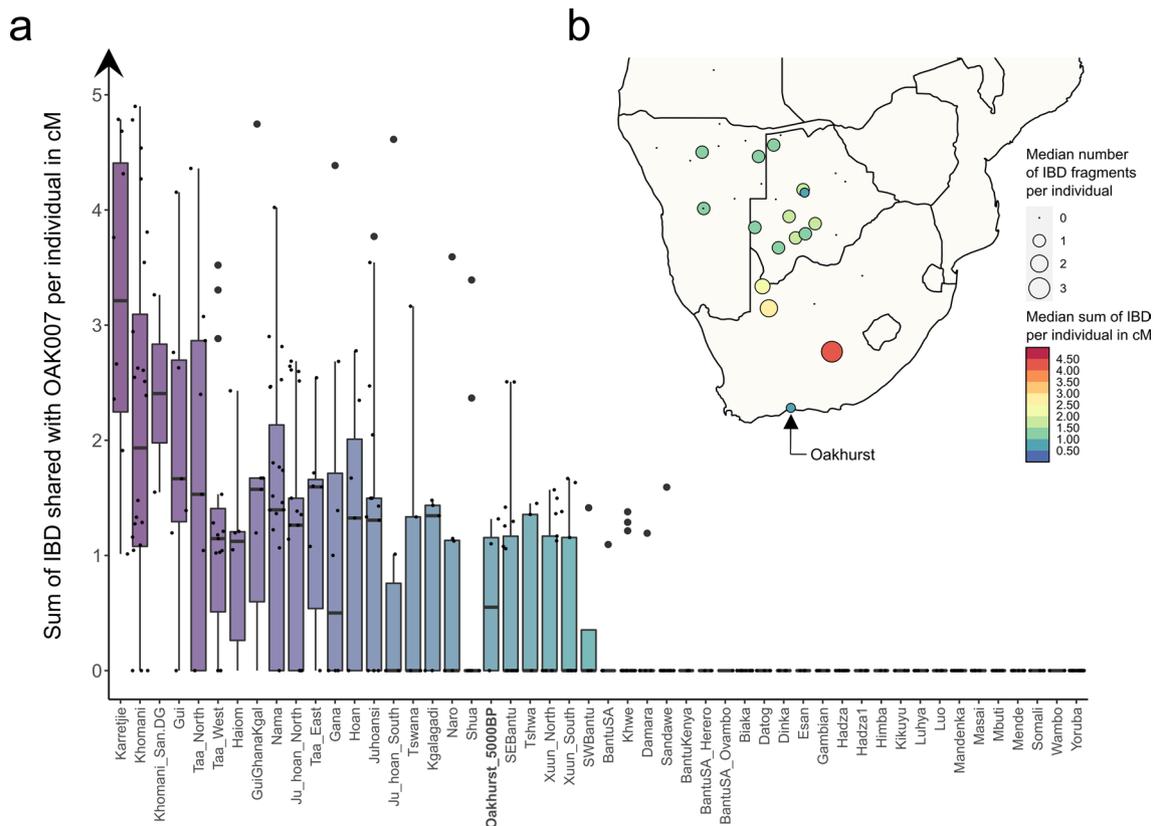
**b)** Kriging interpolated distributions of southern (orange component), central

(magenta component) and northern (blue component) San ancestry in present-day populations. Proportions for ancient LSA individuals are indicated as dots. White outlines indicate the present-day distributions of the Tuu (left and centre) and Kx'a (right) language groups.



**Extended Data Fig. 3 | Genetic affinities of Oakhurst individuals to present-day southern Africans in F-statistics. a** Scatter plot of group-based jackknife point estimates from  $F_{ST}$  (Y axis) and outgroup  $F_3$  (Archaic; Oakhurst, X), where X represents the present-day populations ( $n = 32$ ). Error bars represent 2 standard errors. The error band indicates the 95% confidence interval of the linear regression. Symbols and colours correspond to Fig. 1.

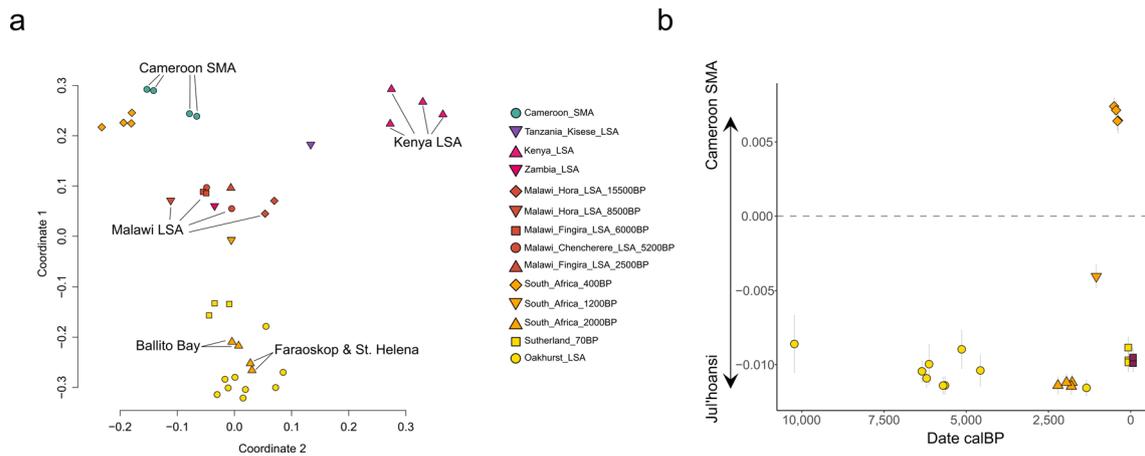
Data can be found in Supplementary Table 2 & 3. **b**  $F_{ST}$  distances visualized on a map of southern Africa. The populations exhibiting the highest genetic affinity to Oakhurst are highlighted. The location of Oakhurst is plotted in black. The beige-coloured region is the Kalahari semi-desert. Data can be found in Supplementary Table 3.



**Extended Data Fig. 4 | Identity-by-descent segment sharing between imputed Oakhurst genomes and present-day sub-Saharan Africans.**

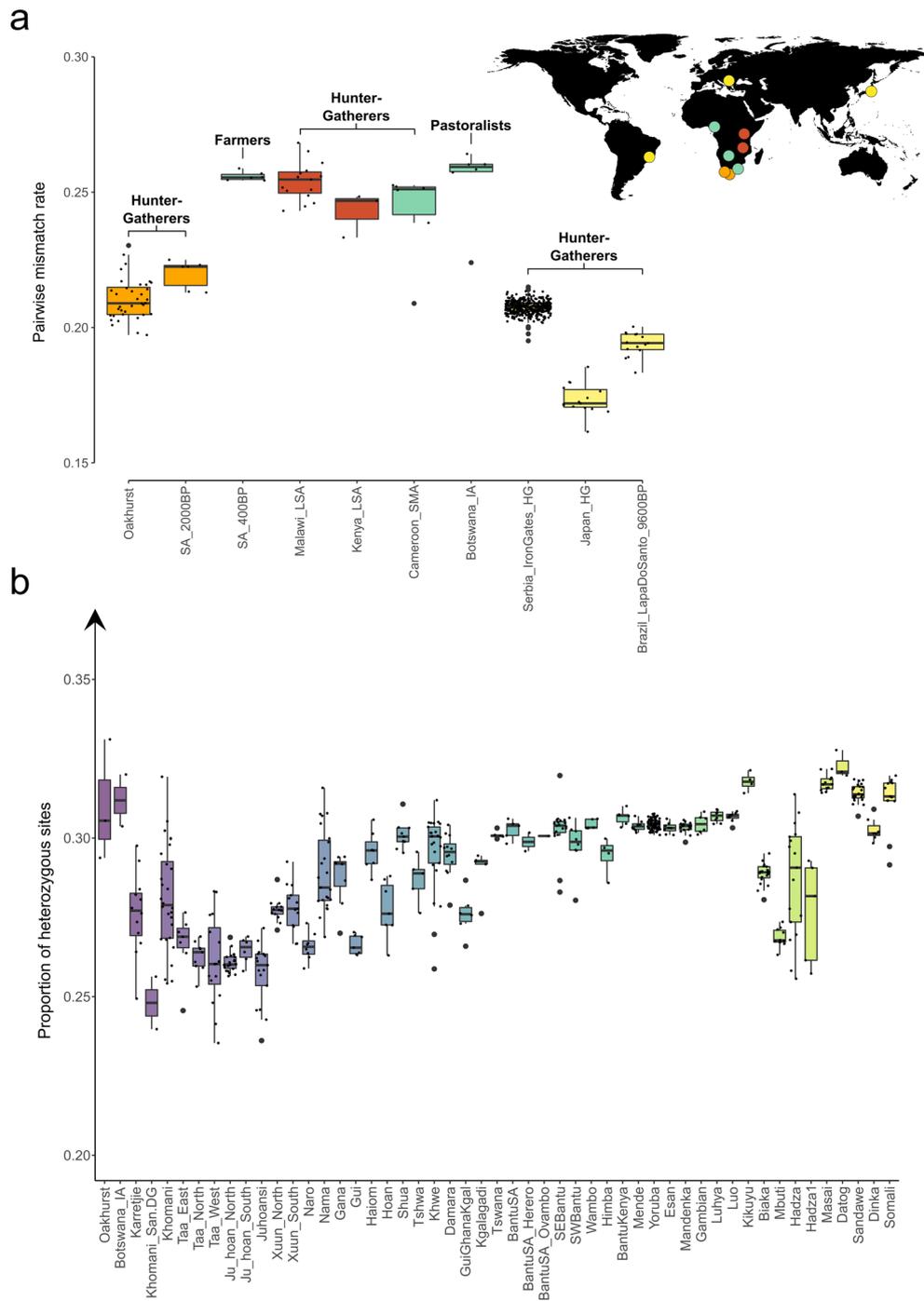
**a** Length of IBD segments shared between OAK007 and present-day Africans. The bounds of the box represent the 25<sup>th</sup> and 75<sup>th</sup> percentile, the centre represents the median, and Whiskers represent the smallest value greater than the 25<sup>th</sup> Percentile minus 1.5 times the interquartile range and largest value less than the 75<sup>th</sup> Percentile plus 1.5 times the interquartile range of the data, respectively. Outliers present the minimum and maximum values in the data and are depicted as large dots. Each dot corresponds to the summed IBD for a unique pair of OAK007 and an individual from the respective population ( $n = 31$ ). The

boxplots are ordered according to the averaged length of IBD segments shared between all individuals of a population and OAK007 (including pairs with no IBD sharing). Populations on the left feature the longest mean IBD segment sharing with OAK007, populations on the right the shortest. Lengths of segments are given in centiMorgan (cM). Data can be found in Supplementary Table 24. **b** Median IBD sharing between OAK007 and present-day populations visualized on a map of southern Africa. The colour of each dot corresponds to the median length of summed IBD segments shared with OAK007 per individual, the size corresponds to the median number of IBD segments shared per individual.



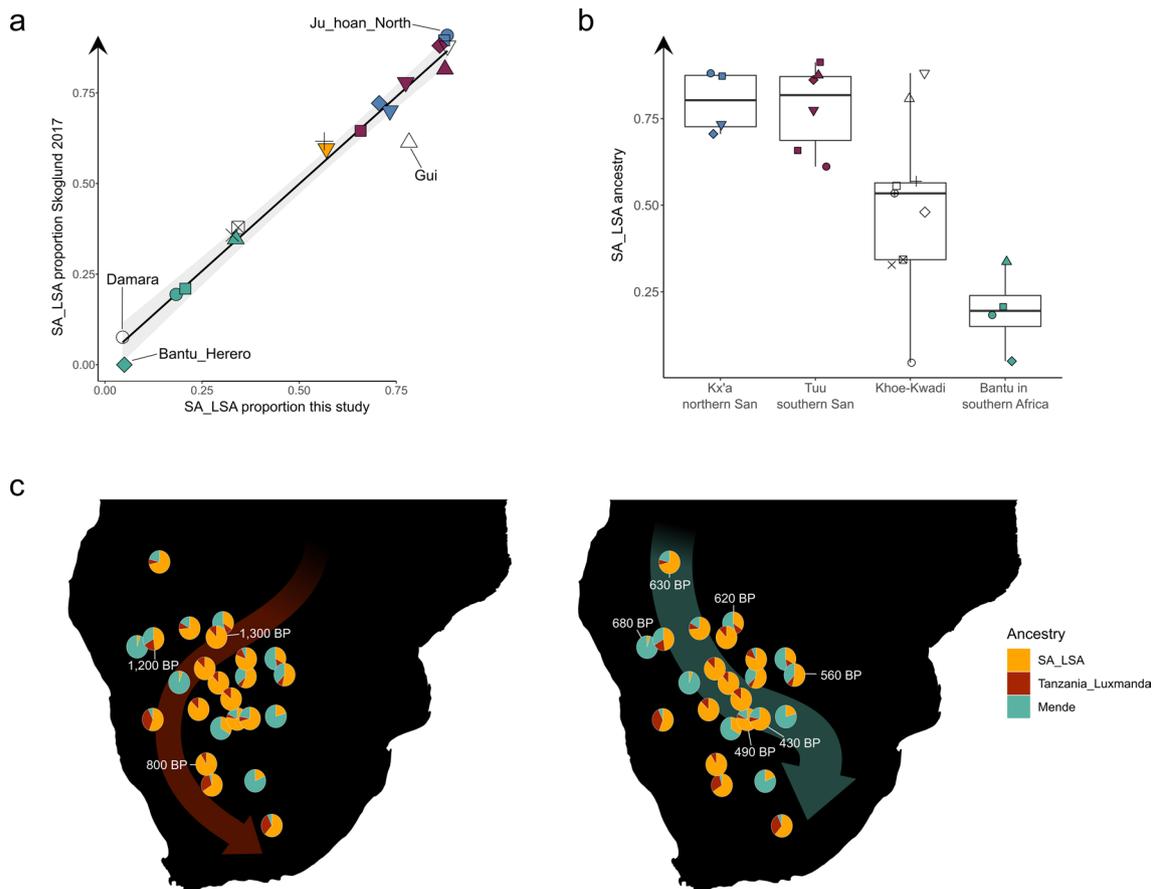
**Extended Data Fig. 5 | Population structure in ancient Africa. a)** MDS plot of the pairwise  $F_3$ -matrix of the form  $F_3(\text{Chimp}; X, Y)$  (shown in Fig. 2) transformed into distances using the formula  $1 - F_3$ . Data can be found in Supplementary Table 4. **b)** Individual  $F_4$ -statistics of the form  $F_4(\text{Archaic, Test}; \text{Ju\_hoan\_North,$

Cameroon\_SMA) through time for 21 ancient and 2 current-day Khomani genomes from South Africa. Error bars represent 2 standard errors. Data can be found in Supplementary Table 5.



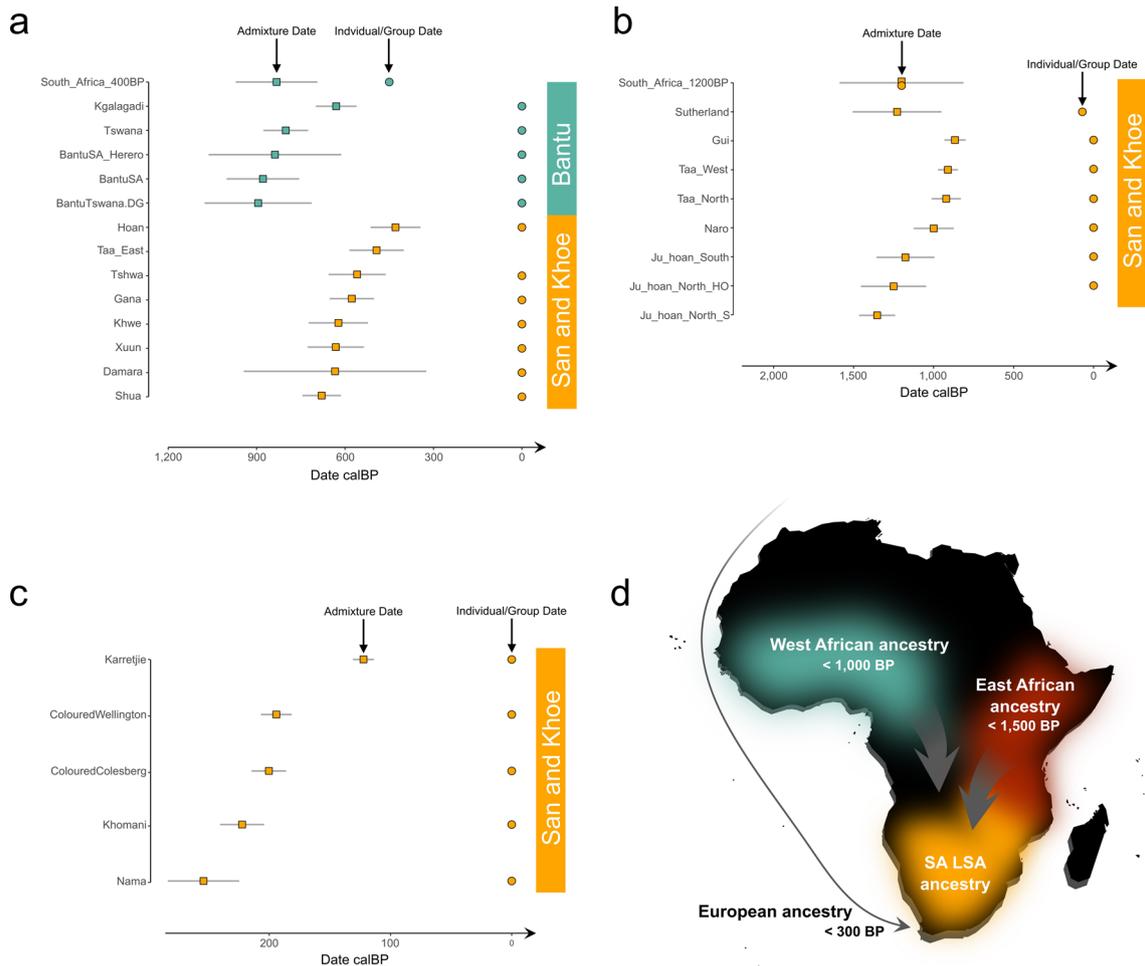
**Extended Data Fig. 6 | Genetic diversity in ancient and present-day sub-Saharan Africans. a)** Conditional nucleotide diversity (measured as the intra-population pairwise mismatch rate) in ancient pseudo-haploid populations ( $n = 10$ ). Each dot depicts the pairwise mismatch rate between a pair of individuals from the same group. The bounds of the box represent the 25th and 75th percentile, the centre represents the median, and Whiskers represent the smallest value greater than the 25th Percentile minus 1.5 times the interquartile range and largest value less than the 75th Percentile plus 1.5 times the interquartile range, respectively. Outliers present the minimum and maximum values in the data and are depicted as large dots. The geographical origin of each

ancient group is indicated on the world map. **b)** Average heterozygosity in 5 imputed ancient genomes from South Africa and Botswana and 561 individuals from 49 present-day African populations. Each dot corresponds to the fraction of heterozygous sites over the total number of genotyped sites across 22 autosomes. The bounds of the box represent the 25th and 75th percentile, the centre represents the median, and Whiskers represent the smallest value greater than the 25th Percentile minus 1.5 times the interquartile range and largest value less than the 75th Percentile plus 1.5 times the interquartile range, respectively. Outliers present the minimum and maximum values in the data and are depicted as large dots. Data can be found in Supplementary Table 25.



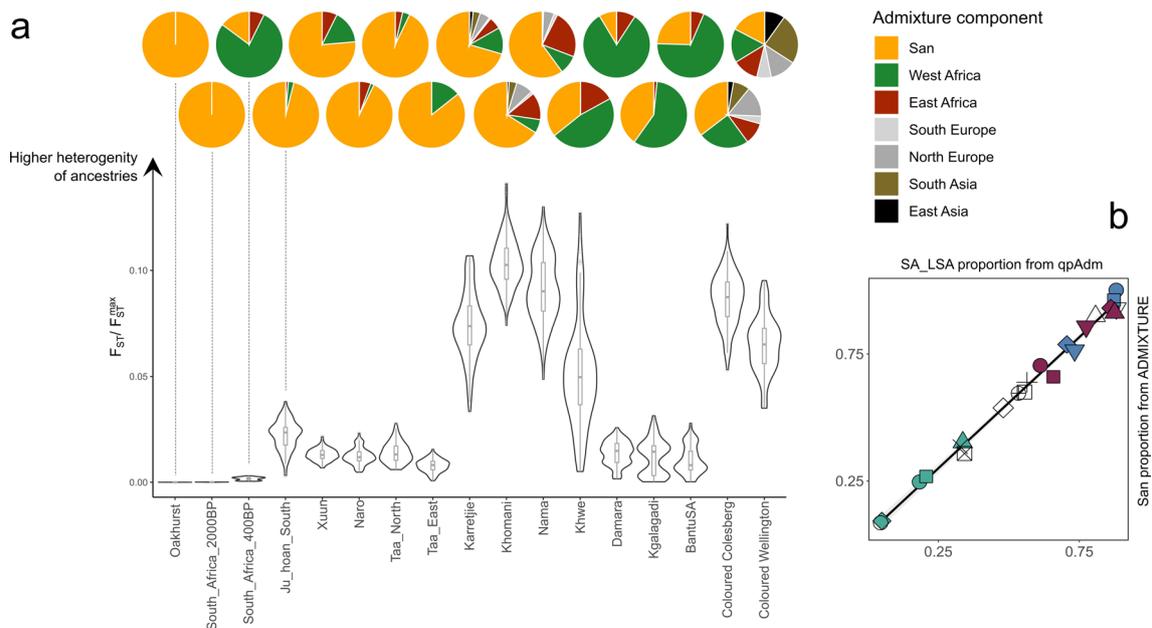
**Extended Data Fig. 7 | Ancestry composition of present-day San, Khoe and Bantu populations.** **a**) Scatter plot of the proportions of SA\_LSA ancestry measured by Skoglund et al. 2017 (represented by the St. Helena and Faraoskop genomes) and this study (represented by the Oakhurst, St. Helena, Faraoskop and Ballito Bay genomes) in present-day San, Khoe and Bantu-speaking groups ( $n = 19$ ) using qpAdm. The error band indicates the 95% confidence interval of the linear regression. Symbols and colours correspond to Fig. 1. **b**) Fractions of SA\_LSA ancestry in San, Khoe and Bantu groups stratified by language affiliation. The bounds of the box represent the 25th and 75th percentile, the centre represents the median, and Whiskers represent the smallest value greater than the

25th Percentile minus 1.5 times the interquartile range and largest value less than the 75th Percentile plus 1.5 times the interquartile range, respectively. **c**) Ancestry compositions retrieved from 2- and 3-way qpAdm admixture modeling of present-day San, Khoe and Bantu groups ( $n = 23$ ) visualized on a map of southern Africa. Mean admixture dates from DATES analysis are indicated for selected populations. On the left for populations admixed between SA\_LSA and Tanzania\_Luxmanda\_3000BP and on the right for populations admixed between SA\_LSA and Mende. Arrows on maps indicate a general direction of influences rather than discrete routes of migration. Data can be found in Supplementary Table 10-12.



**Extended Data Fig. 8 | Admixture dates in ancient and present-day San, Khoe and Bantu populations.** **a**) Admixture date point estimates for the admixture (squares) between SA\_LSA and Mende in groups with primarily West African-related admixture (based on qpAdm) ( $n = 9$ ). For ancient individuals, the mean radiocarbon date is indicated (circles). Error bars represent 1 standard error. **b**) Admixture date point estimates for the admixture (squares) between SA\_LSA and Tanzania\_Luxmanda\_3000BP in groups with primarily East African-related

admixture (based on qpAdm) ( $n = 14$ ). Error bars represent 1 standard error. **c**) Admixture date point estimates for the admixture (squares) between SA\_LSA and English in groups with European-related admixture ( $n = 5$ ). Error bars represent 1 standard error. **d**) Schematic representation of the geographic origin and potential admixture date of non-Southern African ancestry found in present-day San and Khoe populations. Data can be found in Supplementary Table 13-16.



**Extended Data Fig. 9 | Variability in admixture proportions among ancient and present-day San, Khoe, Bantu and Coloured populations.**

**a)** Pie charts depict the averaged ancestry composition for each group derived from supervised ADMIXTURE modeling at  $K = 7$ . Box- and violin plots depict the Bootstrap ( $n = 1000$ ) distributions of the ancestry variability measure,  $F_{ST}/F_{ST}^{max}$ , obtained from FStruct (methods) calculated on the Q matrix from supervised ADMIXTURE for each population. The bounds of the box represent the 25th and 75th percentile, the centre represents the median, and Whiskers represent the smallest value greater than the 25th Percentile minus 1.5 times the interquartile

range and largest value less than the 75th Percentile plus 1.5 times the interquartile range, respectively. **b)** Scatter plot of the proportions of SA\_LSA ancestry measured using qpAdm and the San proportion measured using supervised ADMIXTURE in present-day San, Khoe and Bantu-speaking groups. Both estimates are significantly correlated (Pearson's correlation;  $t = 49.826$ ,  $df = 20$ ,  $p = 2.2e-16$ ,  $cor = 0.996$ ). The error band indicates the 95% confidence interval of the linear regression. Symbols and colours correspond to Fig. 1. Data can be found in Supplementary Table 21 & 22.

## Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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** No specific software was used for sample collection. Genotype data were generated from sequencing reads. All software used in this study is listed below.

**Data analysis** All software used in this work is publicly available. List of software and respective versions: AdapterRemoval (v2.3.1), Burrows-Wheeler Aligner (v0.7.12), DeDup (v0.12.2), mapDamage (v2.0.6), BamUtil (v1.0.14), EAGER (v1), Sex.DetERRmine (v1.1.2) (<https://github.com/TCLamnidis/Sex.DetERRmine>), ANGSD (v0.915), Schmutzi (v1.5.4), PMDtools (v0.50), pileupCaller (v1.4.0.2), samtools (v1.3.1), Geneious (R9.8.1), HaploGrep 2 (v2.4.0), READ (<https://bitbucket.org/tguenther/read>) (vf541d55), PLINK (v1.90b3.29), Picard tools (v2.27.3), smartpca (v16000; EIGENSOFT v6.0.1), qp3Pop (v.435; ADMIXTOOLS v3.0), qpDstat (v.755; ADMIXTOOLS v3.0), qpWave (v410), qpAdm (v810), DATES (v4010), ADMIXTURE (v1.3), GLIMPSE (<https://github.com/odelaneau/GLIMPSE>) (v2.0.0), DyStruct (<https://github.com/tyjo/dystruct>) (v1.1.0), BEAGLE (v5.4), RefinedIBD (v17Jan20.102), FSTruct (<https://github.com/MaikaMorrison/FSTruct>) (d39827e), TreeMix (v1.12). Data visualisation and descriptive statistical tests were performed in R (v4.1.1). The following R packages were used: Rsamtools (v2.12.0), vegan (v2.6-2), factoextra (v1.0.7), ggplot2 (v3.3.6), ggExtra (v0.10.0), ggforce (v0.3.3), rnaturalearth (v0.1.0), sf (v1.0.-8), raster (v3.5-21), rgdal (v1.5-32), spatstat (v2.3-4), maptools (v1.1-4), gstat (v2.0-9), sp (v1.5-0), labdsv (v2.0-1), rcarbon (v1.5.1), magrittr (v2.0.3), dplyr (v1.0.9), reshape 2 (v1.4.4), and tidyverse (v1.3.2). Y-chromosome and mtDNA haplogroups were determined using the ISOGG SNP index (v15.73) and PhyloTree (v17-FU1) reference databases, respectively.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw sequence data (fastq files), mapped data (bam files) from the 13 newly reported ancient individuals will be available prior publication from the European Nucleotide Archive under accession number PRJEB77188. A poseidon package of the genotype data analysed in this paper is available on the Poseidon Community Archive ([https://www.poseidon-adna.org/#/archive\\_explorer](https://www.poseidon-adna.org/#/archive_explorer)). Due to ethical prescriptions of this research under UCT Human Research Ethics #715/2017, DNA sequencing libraries, both before and after SNP capture, are available for replication upon request to the corresponding authors and the UCT Human Skeletal Repository Committee at [uctskeletalrepository-group@uct.ac.za](mailto:uctskeletalrepository-group@uct.ac.za) as aliquots, pending consultation, approval and permission by the UCT Skeletal Repository Committee and consulted San communities who granted the original sample access. Previously published genotype data for ancient and present-day individuals was reported by the Reich Lab in the Allen Ancient DNA Resource v.54.1 (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). Additional previously published genotype data for the present-day San and Khoe samples from Schlebusch et al. 2012 are available at [www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/data/](http://www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/data/). The Genome Reference Consortium Human Build 37 (GRCh37/hg19) is available via the National Center for Biotechnology Information under accession number PRJNA31257. The revised Cambridge reference sequence is available via the National Center for Biotechnology Information under NCBI Reference Sequence NC\_012920.1.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Does not apply. This study does not include novel data from present-day humans or any present-day human participants.
Reporting on race, ethnicity, or other socially relevant groupings	Does not apply. This study does not include novel data from present-day humans or any present-day human participants.
Population characteristics	Does not apply. This study does not include novel data from present-day humans or any present-day human participants.
Recruitment	Does not apply. This study does not include novel data from present-day humans or any present-day human participants.
Ethics oversight	Does not apply. This study does not include novel data from present-day humans or any present-day human participants.

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

## 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](https://www.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	We did not rely on statistical methods to predetermine sample sizes. Sample sizes for ancient populations depended solely on the availability of archaeological material and on ancient DNA preservation. For present-day populations, sample sizes are predefined by the availability of published data. For both ancient and present-day populations, we aim to maximize sample sizes by including all genomes that fulfil our quality criteria mentioned in the Methods section and the quality criteria described in their respective source publications. The reported standard errors used to describe uncertainty ranges of our statistical analyses often reflect both sample size and data quality per sample.
Data exclusions	Four samples (OAK004, OAK009, OAK011, OAK015) were excluded from genome-wide analyses since the authenticity of the autosomal ancient DNA data could not be ensured. The quality criteria forming the basis of this decision are mentioned in the Methods section and pre-established by various previous publications.
Replication	We studied unique entities (past and present populations) and did not perform experiments or study various treatments, so replication is not applicable. But we note that samples from the same population carry similar genetic signatures. For 11 bone samples, we produced complementary to the partial UDG-treated, double-stranded DNA libraries also non UDG-treated single-stranded libraries to increase the genome-wide coverage. While the proportion of authentic ancient DNA obtained from the single-stranded libraries is generally comparable to the proportion measured in the double-stranded libraries, the combination of genotypes from both sources substantially increases the coverage of the respective genomes. Testing of pairwise mismatches in 1,24 Million SNP sites between two libraries of the same bone

specimen confirmed that DNA sequences from both double- and single-stranded libraries are indeed identical in all 11 cases. Also, genome-wide data allows for the analysis of multiple realisations of the sample history, by studying hundreds of thousands of SNP sites.

**Randomization** We studied unique entities (past and present populations) and did not perform experiments or study various treatments, so randomization is not applicable. However, many of our analyses based on f-statistics involve a block jackknife to obtain uncertainty ranges through analysis of uncorrelated segments of the genome.

**Blinding** Blinding is not applicable for ancient specimens as the sampling locations and dates are known as prior. In downstream data analysis blinding is also not relevant since the newly generated ancient genomes are co-analysed with previously published present-day and ancient human genomes.

## 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                                 | Included in the study   |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants                                   |

### Methods

- | n/a                                 | Included in the study                           |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Palaeontology and Archaeology

**Specimen provenance** All sampled bone material belongs to the University of Cape Town, South Africa. This donating partner institution is represented in the author list. Permits for destructive analyses were given explicitly by the UCT Skeletal Repository Committee.

**Specimen deposition** Specimens were returned to the owning institutions after laboratory analyses.

**Dating methods** New radiocarbon dates for 4 samples reported in this study were measured on the bone and tooth fragments sampled for DNA. These dates were obtained at the Curt-Engelhorn-Center Archaeometry gGmbH, Mannheim, using MICADAS-AMS. Collagen was extracted from the previously sampled bones, purified by ultrafiltration and freeze-dried.  $^{14}\text{C}$  ages were normalized to  $\delta^{13}\text{C} = -25\text{‰}$ . The Radiocarbon Lab. IDs are: MAMS 49254, MAMS 37907, MAMS 37908, MAMS 37909. Ages were calibrated using the rcarbon package (v1.5.1) in R using the SHCal20 revised calibration curve.

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

**Ethics oversight** The human remains from the Oakhurst rockshelter site are housed in the University of Cape Town (UCT) human Skeletal repository. The approach for permission to use these samples for ancient DNA was followed according to Gibbon 2020, which included consultation with representatives from the San community in accordance with the South African Heritage Resources Agency and permission from the repository research committee. The Oakhurst samples were approved by the UCT Human Research ethics committee under ethics #715/2017 and Heritage Western Cape permit #17071302AS0718E.

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

## Plants

**Seed stocks** Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

**Novel plant genotypes** Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

**Authentication** Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.