

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

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ 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. 14C ages were normalized to $\delta^{13}C = -25\text{‰}$. 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.