nature research

Corresponding author(s): Vanessa M. Hayes

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

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

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

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

Climate model: LOVECLIM earth system model

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

Data

Policy information about availability of data

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

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

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

Field-specific reporting

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

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

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

Life sciences study design

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

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

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Palaeontology

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

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

Human research participants

Policy information about studies involving human research participants

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

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