nature research

Corresponding author(s): Eske Willerslev Nature 2017-04-05008C

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used for data collection. All software used in this study is listed below.
Data analysis	All software used in this work is publicly available. Corresponding publications are cited in the main text and supplementary material. List of software and respective versions: CASAVA v1.8.2 AdapterRemoval v2.1.3 bwa v0.7.10 bwa mem 0.7.10 picard tools v1.127 bamUtil v1.0.14 samtools v1.3.1 GATK v3.3.0 and v 3.6* pysam 0.7.4 (python module) bedtools 2.27.1 mapDamage2.0 contamMix v1.0-5 SHRiMP 2.2.3 YFitter v0.2 Haplogrep 2.0

ANGSD v0.915
PRANK v.150803
BEASTv1.8.2
schmutzi
admixtools v4.1
NGSrelate
plink v1.07* and v1.9
ADMIXTURE v1.3
RAxML-8.1.15
SnpEff
SPAdes-3.9.0
R 3.2.3
python 2.7.12
perl v5.22.1
CALIB
NOTE: Versions marked with * were used for Y chromosome analyses.

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 upon request. 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

Sequence data were deposited in the European Nucleotide Archive (ENA) under accession PRJEB20658. SNP data for present-day populations are available after ethical validation in the European Genome-Phenome Archive (EGA) under accession: EGAS00001002926.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-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 in ancient population genetic studies are limited by the number of samples yielding endogenous DNA proportions amenable to whole genome sequencing.
Data exclusions	We selected 137 samples for whole-genome sequencing, out of all screened samples, based on their endogenous content and low contamination estimates. These criteria are described in detail in Supplementary Section 3.1. Furthermore, closely related individuals were excluded from analyses requiring population allele frequencies.
Replication	We did not attempt to specifically replicate experimental findings. But we note that samples from the same population carry similar genetic signatures. Moreover, genome-wide data allows for the analysis of multiple realisations of the sample history, by studying hundreds of thousands of SNP sites.
Randomization	No experimental groups or effect sizes were measured in this study, thus we did not implement any random group assignment.
Blinding	No blinding techniques were implemented, as exprimental group assignment is not relevant for popluation genetic history studies of this kind.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Human research participants

Policy information about studies involving human research participants

Population characteristics	No experimental procedures were carried out on human participants. We genotyped 502 individuals from 16 self-reported ethnicities from Altai, Central Asia, Siberia and the Caucasus. Sampling procedures are detailed in Supplementary Section 5.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging