nature research

Corresponding author(s): Pavel Flegontov, David Reich, Stephan Schiffels

Last updated by author(s): 13.01.2019

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	Genotype data were generated from sequencing reads using the following tools: the SeqPrep tool (https://github.com/jstjohn/SeqPrep) for read processing and bwa v.0.6.1 for aligning reads to the human reference genome.
Data analysis	A central piece of software substantially updated for this manuscript is Rarecoal and accessory tools, first published by Schiffels et al. (Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nature Communications 7, 10408, 2016). Up-to-date versions of Rarecoal and accessory tools are deposited at https://github.com/stschiff/rarecoal and https://github.com/stschiff/rarecoal-tools. Custom scripts for calculating rare allele sharing statistics were also used (https://github.com/TCLamnidis/RAStools). Here is a list of published tools used: ShapeIt v.2.20, PLINK v.1.90b3.36, ADMIXTURE v.1.23, ALDER v.1.03, EIGENSOFT v.6.1.4, AdmixTools v.4.1 (qpWave v.310, qpAdm v.401, qpGraph v.5052, qpDstat v.701), ChromoPainter v.1 and v.2, fineSTRUCTURE v.2.0.7, fineSTRUCTURE GUI v.0.1.0, GLOBETROTTER (May 27, 2016).

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

Raw sequence data (bam files) from the 48 newly reported ancient individuals are available from the European Nucleotide Archive under accession number PRJEB30575. The genotype data for the Iñupiat were obtained through informed consent that is not consistent with public posting of the data, analyses of phenotypic traits, or commercial use of the data. In order to protect the privacy of participants and ensure that their wishes with respect to data usage are followed, researchers wishing to use data from the Iñupiat samples should contact Geoffrey Hayes (ghayes@northwestern.edu) and Deborah Bolnick

(deborah.bolnick@uconn.edu), who can then arrange to share the data with researchers who can formally affirm that they will abide by these conditions. The newly reported SNP genotyping data for West Siberians (Enets, Kets, Nganasans, Selkups) is publicly available at https://reich.hms.harvard.edu/datasets.

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	Sample sizes for ancient populations depended solely on the availability of archaeological material and on ancient DNA preservation. Sample sizes for extant populations genotyped in this study were typical for studies in genetic history: from 3 to 35 individuals.
Data exclusions	During the dataset construction procedure present-day and ancient individuals with a high proportion of missing loci were removed. The missing rate thresholds varied from dataset to dataset and are shown in Supplementary Table 4. First- and second- degree relatives were excluded from all populations genotyped in this study as an assumption of most analytical methods is that individuals are not related closely. For some analyses (qpWave, qpAdm, etc), American individuals with noticeable European, or East Asian or African ancestry (>1%) were excluded, as described in Methods ("Dataset preparation for present-day genomes"). The exclusion criteria mentioned above were pre-established. In addition, for analyses relying on groups rather then on independent individuals (qpAdm, qpWave, qpGraph, f4-statistics) a complex genetic outlier removal procedure was implemented. This procedure is described in detail in Methods ("Supplementary Information section 4"), in Supplementary Information section 4, and its outcome is illustrated in Tables S4.1 and S4.2.
Replication	We studied unique entities (past and present populations) and did not perform experiments or study various treatments, so replication is not applicable.
Randomization	We studied unique entities (past and present populations) and did not perform experiments or study various treatments, so randomization is not applicable.
Blinding	We studied unique entities (past and present populations) and did not perform experiments or study various treatments, so blinding is not applicable.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Human research participants belonged to various ethnic groups: Inupiat from Alaska, USA (35 individuals); Enets (3 ind.), Kets (19 ind.), Nganasans (22 ind.), and Selkups (14 ind.) from the Krasnoyarsk Region, Russia. Age and gender did not influence participant selection. Information on participant age is not presented in this study for privacy reasons, and information on gender is presented in Supplementary Table 2.
Recruitment	Participants were selected using the following criteria: healthy adults whose ancestors within three generations all belonged to the same ethnic group. First- and second-degree relatives were excluded. Since ethnicity of participants and their ancestors was self-reported, potential bias and mistakes cannot be excluded. To take care of the self-reporting bias, genetic outliers were removed from each population prior to the analysis stage.
Ethics oversight	New SNP array genotyping data are reported here for individuals sampled and analyzed in two previous studies: Raff et al.

Ethics oversight	(Mitochondrial diversity of Iñupiat people from the Alaskan North Slope provides evidence for the origins of the Paleo- and Neo-Eskimo peoples. Am. J. Phys. Anthropol. 157, 603–614, 2015) and Flegontov et al. (Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. Scientific Reports 6, 20768, 2016). In the case of the Iñupiat, the current study was approved by Northwestern University's Institutional Review Board, after consultation with the Ukpeagvik Iñupiat Corporation, the Native Village of Barrow, and Senior Advisory Council of Barrow (Elders). In the case of the West Siberians, genetic sampling and research on these participants was approved by the ethical committee of the Lomonosov Moscow State University (Russia). All volunteers have signed informed consent forms. The sampling was also approved by local administrations of the Taymyr and Turukhansk districts (Krasnoyarks region, Russia) and discussed with local committees of small Siberian nations for observance of their rights and traditions.

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