

An ancient DNA perspective on the Russian Conquest of Yakutia

Corresponding Author: Professor Ludovic Orlando

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)

A. Summary of the key results:

The manuscript by Crubézy and colleagues describes ancient DNA analyses of Yakut individuals from northeastern Siberia spanning four archaeological periods from 1500 – early 1900s CE. Analyses of human DNA suggest that these ancient Yakut individuals originated from admixture between local populations and trans-Baikal groups migrating with the spread of the Mongol Empire. The authors find strong genetic continuity between ancient Yakut populations and modern Yakut individuals, with marginal contribution from Russian individuals. Analyses of oral microbiome profiles suggest no major changes in the oral taxonomic and functional composition, despite the introduction of carbohydrate-rich foods after Russian contact. The authors recovered two partial *Variola major* (smallpox) genomes from the Yakut individuals; phylogenetic analyses suggest that these ancient Yakut strains are distinct from those previously identified among ancient European populations as well as those isolated from 20th century individuals.

B. Originality and significance:

The study represents the first ancient DNA analysis of Yakut individuals. The sample size is large, and the data generated by this study will be of great interest to the field of ancient DNA.

C. Data & methodology:

The study methodologies are mostly sound (see below). I do not have concerns regarding data quality or presentation.

D. Appropriate use of statistics and treatment of uncertainties:

I am not an expert on the admixture and human population genetics analyses, so I will restrict my comments to the microbiome and pathogen screening analyses (see below).

E. Conclusions:

Most of the conclusions seem sound to me.

F. Suggested improvements

See below for major and minor comments.

G. References:

I have no concerns with the References cited. I have noted one additional study that could be included (see below).

H. Clarity and context:

The manuscript is overall well-written, well-contextualized, and it was a pleasure to read. The Introduction and Discussion sections are especially well-written.

Major Comments:

In general, the microbiome section lacks a description of the Yakut oral microbiome composition. I recommend at least a

couple of sentences describing the abundant taxa be included in the Main Text. The MetaPhlAn4 results should be included in the Supplementary Tables.

The authors state that they used default parameters for the MetaPhlAn4 and HUMAnN3 taxonomic and functional assessment. Using the default parameters, collapsed reads shorter than 70 bp would be excluded from the analysis. I recommend the authors double-check the parameters used and consider using a minimum read length more appropriate for ancient DNA (such as 30 bp).

Line 214: The authors state that the stability in abundances of the bacterial complexes over time is “indicating no apparent changes in oral health in relation to Russian conquest”. There have been multiple studies showing that the presence and/or abundance of bacterial complexes in dental calculus, such as the Red Complex, cannot be used to determine oral health status of ancient individuals (Velsko et al. 2018 and others). Is there any paleopathological information available for these individuals to suggest changes in oral disease prevalence in relation to the Russian conquest?

Along similar lines, do any of the individuals show paleopathological signs of diseases (eg. TB)?

The authors state “In cases where a tooth and a calculus sample from the same individual passed filters, the profile maximizing oral microbial sources was kept, resulting in final data set 74 individual oral microbiomes”. Does this mean that oral profiles from tooth are being compared to those from dental calculus? It may not be such a big deal in this case since none of the microbes or complexes are found to be significantly differentially abundant, but nevertheless, I do not think it is an appropriate comparison.

By only focusing on the abundances of microbial complexes, the authors have missed an opportunity to study strain-level variation in the Yakut oral microbiome over time. An attempt could be made to reconstruct the genomes of some of the abundant oral microbes to determine if the Russian conquest introduced new strains to the Yakut, similar to what has been found with the opportunistic oral pathogen, *Tannerella forsythia*, in the Americas (see Bravo-Lopez et al. 2020, Honap et al. 2023).

It would be good to specify what material(s) yielded the Variola DNA for the three positive samples. From S Table A, it seems the Rassoloda individual had tooth and dental calculus sampled – was Variola DNA recovered from both? I cannot remember it being recovered from dental calculus in previous studies, so this would be an interesting finding. I would also recommend adding a table showing the mapping and coverage statistics of the Yakut samples to the Variola genome.

Minor Comments:

I am curious if the USER-treated libraries still showed significant terminal damage, given that the treatment was conducted for the full 3 hours.

It may be unclear to a non-specialist reader what the different bacterial complexes (red, orange, yellow, green and purple) represent and why their abundances would be different across the archaeological stages.

Line 110: Specify that shotgun sequencing was conducted.

Line 279: Change “originated” to “originating”.

Line 313: Please cite Ottoni et al. 2021, in addition to Quagliariello et al. 2022, and contextualize the results of this study accordingly.

Line 655 in the Main Manuscript states that the cutoff used for determining if a sample is “positive” for a pathogen was “a minimal number of 100 high-quality alignments and average edit distance was inferior or equal to 0.01” whereas the Supplementary Methods state that “samples with more than 100 reads and a mean distance below 0.1 were considered positive”. Please clarify.

Figure 2A: This figure would look neater if the labels for non-Yakut samples were removed.

In Figure M1_ 11, the caption states the mean distance is calculated as “mean distance of per-individual reads aligned against the pathogen’s genome to the reference genome”. Can this be rephrased to add clarity?

Supplementary Table A is difficult to digest in its present form. It would be better if it were broken down into multiple tables – for eg. one table for the archaeological metadata, one for the lab-work and sequencing, and another for the results of the human DNA mapping. It would be good to include the sequencing statistics on a per sample type basis, so that readers can see what information could be recovered from bone vs tooth vs calculus. It would also be good to have a column showing the average length of reads, % damage at terminal bases, etc.

Throughout the manuscript and supplementary methods, please ensure that “MetaPhlAn4” is written correctly (not MetaPhlan4).

Referee #2

(Remarks to the Author)

The study by Crubezy et al. is focused on the effects of the Russian colonization of Yakutia on the genetic makeup, microbiome, and health of the indigenous population. While West European colonialism receives a lot of attention from many disciplines, including archaeogenetics, the Russian colonialism remains understudied and even ignored, which makes the study timely and important.

Major issues:

My main concern is that the manuscript just presents the findings, but they are not discussed in a broader context of colonial processes from various parts of the world and historical periods. Even previous archaeogenetic studies in Yakutia are not discussed much in the paper. Such a discussion is needed since the findings are rather surprising, showing no genetic admixture with Russian colonists and even no effects of the new diet on oral microbiomes and of the new lifestyle on DNA methylation patterns. I am not an expert on microbiomes and methylation, and I cannot judge the robustness of these negative findings. Concerning human genetics, I do not question the main analytical approaches (based on both allele frequencies and autosomal haplotypes) and the treatment of uncertainties, but the selection of samples for the study raises questions since Christian burials were mostly left out (see below). I am afraid that random archaeogenetic sampling across cemeteries could have uncovered more extensive genetic admixture between Yakuts and Russians. I suggest that the authors describe their sampling approach in much greater detail than few sentences in the supplement, provide a map of both sampled and unsampled cemeteries in the region (and maps of individual large cemeteries), and acknowledge these issues in the manuscript.

Lines 90-91: "Human tombs displaying obvious external Christian attributes were deliberately avoided while traditional funerary constructions were targeted"

This is potentially a problem since by design the study is not based on an unbiased sample of the local population, and thus the level of Russian-Indigenous cultural and genetic admixture may be underestimated. And Russian cultural influence has direct consequences for oral microbiomes and for pathogen prevalence. As I understood from Suppl. section 2.2, this avoidance of obviously Christian burials (although not complete according to Fig. 1c) affected the sampling of chronological stage 4 (19th century and later) and maybe of stage 3 (2nd half of the 18th century) since hallmarks of Christian burials are simply absent in earlier periods. Anyway, that is a problem that should be discussed more in the main text.

Another source of archaeological bias in sampling of the ancient genetic landscape is beyond researchers' control: that is the high frequency of tree burials (apparently leaving no preserved remains) during the chronological stage 2, as mentioned in Suppl. section 2.2. This bias may affect genetic results on the ancient Yakut community since apparently only members of the elite were inhumated, as also illustrated by frequent finds associated with wealth, leadership, and shamanism at this stage (Fig. 1c).

Lines 281-285: "The persistence of a substantial fraction of local genetic ancestry (62.0-84.3%; Table S1B) contrasts, however, with both cultural and linguistic evidence pointing to a primarily central Asian turko-mongol influence. Therefore, the Yakut history provides an example of a local population merging both genetically and culturally with a group entering the region through migration."

In the context of stepping-stone and isolation by distance landscapes, which is arguably a more accurate model of reality than admixture graphs, qpAdm generates rather imprecise estimates of admixture fractions (<https://www.biorxiv.org/content/10.1101/2023.04.25.538339v5>). That is why I would not trust this high estimate of local ancestry if it relies solely on qpAdm results. Moreover, there are no clear guidelines on choice of reference populations for qpAdm, and this choice influences estimated admixture fractions and p-values strongly. I suggest that the authors either discuss these caveats or estimate local pre-Yakut ancestry using alternative methods.

Minor comments:

I am glad to see a catalogue of archaeological descriptions for burials (Suppl. section 2.3) and photos of some artefacts in Figs. 1d and 4d since this crucial information is usually omitted from archaeogenetic studies. I recommend adding also photos/plans of each burial and associated artefacts, if possible, as this information is valuable for archaeologists. I understand that complete documentations of burials are often published in specialized Russian-language journals, but since they are poorly accessible to the international community, I see value in republishing. Moreover, reproduction of images without explicit permission is allowed under the Russian law if it is done for scientific purposes.

Line 87: "the western Vilyuy, the northern Verkhoyansk, and the eastern Indigirka"

Please mention that the former and the latter regions are river basins and indicate the type of the Verkhoyansk region.

Lines 119, 120: "using pseudo-haploid data for the 1.24 million SNPs comprised in the Allen Ancient DNA Resource"

I recommend mentioning a version of the AADR used in the paper as data and archaeological/geographical annotations are constantly updated in this database.

Lines 125-127: PC1 scores separated European from American and Asian populations, with ancient and modern Yakuts clustering together and alongside Asian populations.

Please clarify if modern Yakuts were projected on the other modern groups or not since the shrinkage artefacts are not corrected completely in the existing versions of the smartPCA software. That is why comparing the positions of mapped and non-mapped individuals at fine scale is not recommended.

Lines 140-141: "Combined, these analyses suggest that Yakuts from the ~1500 CE onwards are not the direct descendants of people inhabiting Yakutia prior to ~280 BCE (i.e. the radiocarbon age of Mokp)."

The term “direct descendant” is vague, in my opinion, since children, grandchildren, etc., are direct descendants, but obviously they may look very different from their ancestor on PCA etc. I prefer to use terms like “genetically continuous” instead. And, strictly speaking, the analyses cited here do not exclude genetic continuity of Yakuts with the Iron Age population and some incoming gene flow, although we do know from other genetic analyses and from historical evidence that Yakuts migrated to their current territory relatively recently and replaced the local population at least partially, as discussed further in the paper. Another caveat to be considered in the paper is that two Iron Age and two Late Neolithic individuals (strangely, I see only one of those individuals mapped on the CPA in Fig. 2a) are not representative of the respective genetic landscapes over vast territories. That is a mistake common in the archaeogenetic literature.

Lines 143-144: “and an ADMIXTURE profile resembling that of the Trans-Baikal Evenk (Fig. 2b)”
I don't see any modern Evenk individuals analyzed in Fig. 2b.

Lines 160-163: “Except for rare genetic outliers (Table S1B), the genetic makeup of all other Stage 1 to Stage 4 Yakuts could be modelled as a two-way admixture between populations inhabiting Yakutia in the Iron Age (Yakutia_IA, 62-84%), and the Baikal region between the 13th and 15th centuries CE (Baikal_his, 16-38%; Table S1B)”
I cannot easily find the composition of these reference groups (Yakutia_IA and Baikal_his). If they include just 1-2 individuals, I find this analysis problematic.

Lines 169-173: “Neither of the Yakut groups from Stage 2 to modern times showed a statistically significant excess of genetic sharedness with Europeans, including from Western Russia, relative to Stage 1 (Extended Data Fig.2d). Additionally, no Yakut group from a given archaeological stage was enriched in European-like genetic variation relative to those from previous archaeological stages (Extended Data Fig.2d).”
Looking at EDF 2d, I see that the authors compared stages to each other, but not groups within stages. Please clarify that since the text implies otherwise.

Lines 232-233: “with four cases involving first and second-degree relatives buried at the same site (AC, At Daban, Oktom and Oulakh; Fig. 4a)”
Please spell out the abbreviation “AC”.

Lines 247-249: “Combined, these results social practices conducive to preferential unions, contacts and microbial exchanges intra-regionally, between close neighbors.”
A word is missing in this sentence.

Lines 310-311: “Russian contact, however, also introduced a wealth of new infectious agents, such as but not limited to the smallpox identified in this study, with devastating demographic consequences for immunologically naive Indigenous populations”
Does not the stable Ne estimated throughout stages 1-4 (Table S1C) contradict this? Or did pathogens affect Ne just in the very beginning of Russian contact in stage 1? Please clarify. The bottleneck analysis in the paper is focused on the formation of the Yakut gene pool in the Middle Ages, but not on the Russian conquest and contact period.

Lines 596-597: “Each ancient and modern individual from Yakutia was tested for every combination of 2 or 3 populations putting the non-used population in the right group (Table S1B).”
Please show the standard list of “right” groups used.

Suppl. section 2.1
“North Eastern Federal University” – please correct to “North-Eastern Federal University”.

References 2, 21
A highly unusual transliteration of Russian is used, please use a more standard one, although there are many slightly different systems (https://en.wikipedia.org/wiki/Romanization_of_Russian).
Reference 10
The title is duplicated.

Reference 19: Crubézy, E., Melnichuk, O., and Alexeev, A. (2020). Archaeology , Genetics and History: 15 years of research in Yakutia (2002 – 2017). *estnik arkheologii, antropologii i etnografii* 4, 110–119.
<https://doi.org/10.1128/microbiolspec.PoH-0004-2014.Th>.
Here there is a typo in the word “Vestnik” (the V is dropped), and the DOI link points to a completely different paper on smallpox.

Reference 35: the doi link is duplicated.

I prefer to sign my reviews.

Pavel Flegontov

Referee #3

(Remarks to the Author)

Originality and significance: if not novel, please include reference / Suggested improvements: experiments, data for possible revision / References: appropriate credit to previous work?:

Although bringing DNA analysis to the history of colonialism in Siberia is novel, the article could benefit from engaging with well-established scholarship in environmental history. In Alfred Crosby's work, for instance, the author may find many relevant parallels: Crosby, Alfred W. 2004. *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*. 2nd edition. Cambridge and New York: Cambridge University Press.

I also found the use of the ethnonym Yakut problematic, especially as it spans several centuries in the article. As historians and anthropologists, we understand that processes of identification are fluid, often shaped by relativistic understandings of social and cultural practices, including interactions with animals and the environment. Furthermore, shifts in identity are always influenced by political and economic regimes. I would recommend elaborating on why the authors chose to use the colonial ethnonym Yakut rather than Sakha, and why they portray the migrations of groups as if their identities remain static. For further reading on identity politics in North Eurasia, you may consult the following works:

- Anderson, David G., Dmitry V. Arzyutov, and Sergei S. Alymov, eds. 2019. *Life Histories of Etnos Theory in Russia and Beyond*. Cambridge, UK: Open Book Publishers.

- Anderson, David G. 2000. *Identity and Ecology in Arctic Siberia: The Number One Reindeer Brigade*. Oxford: Oxford University Press.

Another issue I found problematic is the assumption that alcohol was introduced through Russian colonialism and Chinese/Mongolian trade. Ethnographic studies of southern Siberian Indigenous groups (Altai, Khakas, Buryat, Tyva, and others) suggest that the tradition of making alcohol from milk is likely Indigenous. This practice appears to be rooted in animal husbandry rather than agriculture, as is the case with vodka, which is made from wheat, potatoes, or other agricultural ingredients.

Ethnographically, the use of the term exogamy, as discussed in the article, would benefit from additional references that consider its relationship to territorial identity or neighborhood, as this understanding may differ from the classic kinship-based approaches in anthropology.

Referee #4

(Remarks to the Author)

Summary

The manuscript of Crubézy and coauthors presents the analysis of genetic data of 107 individuals, 21 methylation datasets, samples of calculus analyzed for oral microbiome and 2 soft tissue samples. Population-genetic, pathogen, microbial and methylation patterns have been investigated and compared with markers from material culture. These individuals were uncovered during a French-Russian archaeological expedition in the Yakuts region and the individuals are associated with the local Yakut population. They were dated to historical periods between 16th and 20th century and they were chronologically aimed to predate and postdate historically-documented Russian conquest of the region. The results show no major differences between the 4 periods investigated and in particular no changes in population-genetic, microbial or methylation patterns that can be attributed to Russian conquest (even though it is linked to changes of lifestyle and diet). Based on the interpretations presented, Yakut population was formed in one discrete event in the 13th century by immigration of Baikal-related populations to the region and by admixture with the locals that form the majority of their ancestry. The marital habits of the population are described as non-strict patrilocality with strong ties of the population to their land. That is because methylation, genetic and oral microbiome patterns do not seem to be linked to chronology, social status, assumed religion or wealth but they do seem to be linked to differences between regions.

The genetic analysis of indigenous populations from the Yakutsk region is an important pursuit as such groups are generally underrepresented in both modern and ancient datasets. Additionally, these groups are also often understudied in terms of archaeology, rarely appear in historical sources and genetic research thus provides a crucial perspective on their, even quite recent, past. The research strategy focusing on recent ancient individuals associated with these groups is sound - it helps to circumvent issues related to large-scale demographic events in the 20th and 21st centuries in the region and by sampling individuals that can be divided to a period before and after Russian conquest of the region, the effects of inter-population contacts can be studied. Additionally, the results can be linked to a wealth of archaeological information obtained during the excavations. The whole-genome data from this region will have high value for future studies, that is especially the case for higher coverage individuals (e.g. a sample beyond 65x coverage if I read right on the line 112). It is also valuable that microbial, methylation, pathogen patterns are studied jointly with population-genetic results and material culture. The study highlights limits of assessing known historical events with biomolecular methods as Russian conquest is practically invisible in the data and showcases a very detailed view into a life history of certain individuals e.g. female shaman or an individual interpreted as a founding female.

As such, the manuscript will be of interest to researchers in various disciplines studying this region, in particular scholars from genetics, archaeology, anthropology, biology and history, but also those interested in comparing various sources of data and those interested in the impact of colonial powers on an Indigenous society. The disciplines are well connected in the text, often describing archaeological and genetic features of the burials jointly. The various analyses are also relatively well connected and the manuscript is easily readable. In my understanding, the manuscript also speaks respectfully about Yakuts, their lifestyle and their history. The sampling and joint French-Russian research seemed to have been done ethically and respectfully and several co-authors are Yakuts though it is not clear if the wider Yakut community has been consulted regarding the results. Community engagement is important here not only because the study could be misused as to invalidate experiences of Yakuts and other indigenous populations during Russian (or other colonial) conquests, but also because direct descendants of the studied people could be traceable via IBD (see both of these in more detail in the major

points mentioned below). Methodologically, I believe that the number of samples was relatively modest for a large period of history and region covered and I think some facets of assessment of the social structure were thus overinterpreted (see major points). Altogether, I however like the study and I think the study could be suitable for publication in Nature pending revisions.

The major points:

1) I have a few ethical issues that I hope can be remedied. I would suggest including a section in the SI discussing these. In general, it is often not straightforward how to proceed with the analysis of data attributed to individuals from Indigenous communities whose agency regarding their past and the archaeological and anthropological remains that are associated with them might have been taken away. That is even more the case if these are recent individuals whose direct relatives could be still living and genetically traced through open access to such data. In this case, there was a longterm collaboration with researchers from the region of interest and several co-authors are Yakuts. By my (limited) assessment, care has been taken to represent this population in the manuscript carefully and archaeological remains have been described in detail. However, it is unclear if communities have been contacted prior, during and/or after the study. This should be remedied by the co-authors that are in the region (though this might be difficult to do for other co-authors due to the current political situation). Community outreach is needed for 2 reasons:

1a) The community members should understand that these individuals might have living relatives today and that it might be their grandparents or great grandparents that are being investigated. Related to that, in my understanding, no effort seemed to have been made to trace living relatives of the individuals that were analysed and dated to the 19th and 20th century. The authors should at least try to trace them and receive their consent when possible. If it is not possible to outreach to the communities in Russia in person, a contact to expat communities or online should be at least attempted. It should be thoroughly considered if the genetic data of individuals from the 19th and 20th century should be made publicly available or if other sort of access should be prepared instead (EGA). Upon publication, such data will be immediately uploaded to GEDmatch and other databases by third parties and the living relatives might be discovered (it is to the discretion of the authors if that could be used to trace some of the relatives). I understand that it is a large undertaking to try to trace relatives for such individuals but given that these are personal data of very recently living individuals, it would mean disregarding their rights if ignored.

1b) The community outreach is important also because the communities might be faced with a lot of questions regarding the absence of impact of Russian conquest. That can be even generalised for other communities that were under Russian and other colonial power. It would be good to discuss this issue in this section also and possibly use the literature to describe the impact of Russian conquest historically more clearly in general. This is also important to fully understand the impact of this study and the relationship of biomolecular methods and historical evidence. At the moment, the manuscript suggests that because of the lack of genetic impact, Yakutia was a middle ground, a region that was not conquered demographically or militarily - it should be heavily noted that there might be strong effects on livelihoods and the populations as a whole even without the genetic impact.

2) I am not persuaded by the statements about social practices.

2a) There are 4 cases of close relatives and 2 of these contain a case that is an exception to patrilocality. The patrilocality hence has to be more established, e.g. by exploring differences between IBD sharing between and within sites between females and males explicitly and exploring if it is significantly different (uniparental markers are not sufficient to claim patrilocality). Patrilocality is also potentially questioned by the findings of important female individuals in the society. If anthropological/ethnological studies are used as a basis for the patrilocality statement, they and this assumption need to be clearly described and discussed.

2b) The exogamy is considered to be relatively local as not many connections are found on close relatedness between the regions - however given the time and space, how likely is it to find such relatives? Also, IBD sharing of all individuals within regions is considered higher within a region than between regions and a test that the individuals vary from panmixia is used to support this to showcase that there are preferential unions and contacts between close neighbours. This could be stated for almost any population from any species as there are clear patterns of genetic differentiation and relatedness being mainly driven by geography (isolation by distance). If the claim is that this population is special in this as having an increased rate of within region connectivity when compared to other populations (as this is then connected in the text to the statement that there is a high connection to the land of their ancestors), it needs to be substantiated: at least a departure from a panmictic population simulated in space should be shown.

3) A clear presentation of historical debate and various theories regarding Yakut origins should be also included in the SI. While it is well written in the main text, upon looking through the literature, I got the impression that the question is even more complex and that the evidence (oral, linguistic, cultural, historical) does not align so clearly with the presented results as it can appear from the main text. Also while reading the main text, I was left wondering what evidence does and what does not support the results (e.g. slightly enigmatic mentions of oral histories). Showcasing how debated this question is (even in the main text) would strengthen the impact of this study.

4) From the archeological description, it is clear that the Christian graves have been excavated not systematically but only accidentally when (usually child) solitary graves have been unearthed under the assumption that these are from predating periods. It should be discussed if there are any issues related to this bias in the dataset: these for-their-time unusually buried individuals might come from particular subgroups and might not be representative of the whole population and their oral

microbiome might be also affected by their 'special' status and by the fact that they were children (though their oral microbiomes are likely linked to those of their close relatives and their methylation patterns are likely not affected). It would be good to discuss in the discussion if there are any other potential technical/sampling biases that could be linked to the absence of chronological patterns in the dataset (the most surprising finding in my assessment).

5) Use of reference data.

5a) In the main text, only AADR paper is cited and the authors of papers where the data are actually from are not acknowledged for the data they produced. Also, in the main text the authors use abbreviations (like the enigmatic `Baikal_sib` and `Baikal_his`) that are not described there and the same goes for `Yakut_IA`. Since these are crucial datasets for modelling the population studied, they should be clearly acknowledged and it should be described to the reader where and when these samples come from (in the main text at least shortly, in the SI the logic for their grouping should be described).

5b) Also: for example, 'Mongolia_Dornod_LateMedieval' (as in AADR) was published in Jeong et al. 2020 though this paper is not cited for this data in the main text and in the SI another paper (Kilinc et al. 2019) has been cited for this data instead (probably something happening to the citations).

5c) I am also not happy with the interchangeable use of `Baikal_his` and `Baikal_sib` in the main text. I understood after reading the methods and SI but when reading the body of the text, I thought it was a typo. I do not understand why not to use both datasets even if the interval for DATES estimate for one group was wide (which did not look so wide to me when looking at the SI table).

5d) Additionally, the `Yakut_AI` group (described also only in the main text and with the same issues with the citations in SI as above) consists only of two individuals. Except for Admixture analysis, it is not justified that they should be grouped. As such it is not clear if these samples were used individually, the same results would be obtained and it is not discussed how the low number of samples might affect the results.

5e) Related to this: Given the lack of possible sources, more distal modelling of the studied individuals would help to confirm that there are no differences between them - from Table S1B it seems only simple combinations of Russian, `Yakutia_IA`, `Baikal_his` and `Baikal_sib` has been attempted. Finestructure shows on an intercontinental scale some slight differences between individuals (e.g. `Oktom1` `Sujet`, `Lepsei1`, `AtDaban4`) so I am wondering if there is some unexplored diversity (though there is a lack of modern and ancient reference data so it might be hard to explore).

6) Some statements are hard to evaluate given how the data has been presented:

6a) A claim is made that relatives within the sites are buried close together but the cemetery plan is not shown (or referenced if I did not miss it). This can be also statistically evaluated by actually comparing GIS coordinates of the related individuals to unrelated individuals.

6b) Relationships like father-son or grand-mother, daughter and grand-children are discussed but the genetic relatedness is showcased as a network and not as pedigrees and these relationships are not readable from that figure (it would from pedigrees) - this should be changed so the major relationships that were resolved are presented.

7) It is a bit of detective work to figure out how many samples have been used for what analysis. It is also unclear how many and what individuals with what types of datasets have been generated. Please show that clearly in a figure (e.g. as an extension of Fig 1c) and describe the numbers clearly in the text. The absence of this leads to questions about how representative the assessment of methylation of 21 individuals between the periods was, if the same individuals have been compared or how much of the comparison relies on different sets of individuals.

Minor issues:

1) In some cases (e.g. line 102), 'Yakut' is not capitalised.

2) It is not always clear in the methods if imputed or not imputed data was used (eg. PCA, hapROH), please clarify.

3) Unclear how the individuals used for IBD were filtered, usually coverage or n. of SNPs is used and the results are then filtered by the length and number of IBD segments.

4) Line 672: Manually corrected alignments. Is that an established practice? It does not sound good to me.

5) Soft tissue samples: What were they used for? Any results from them worth mentioning? Did I miss that? It is methodologically interesting and a lot of space is dedicated to their description so it is a bit anticlimactic.

6) Fig 1a - unclear what are the colours of the regions, no legend and very similar to other colours on the plot

7) Fig 1c: mark Russian conquest and potentially other important historical events mentioned in the SI

8) Fig 1c: caption - refer to where in SI described the variables (Wealth, Leadership..)

9) Fig2a: use the same scale for both axis

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

Overall, Crubezy et al. have done a good job of addressing my comments on the previous version of the manuscript. I am restricting my comments to the oral microbiome, pathogen screening, and phylogenetic analyses, as the other analyses are beyond my expertise. To me, the revised manuscript seems to be much improved. I have a couple of minor concerns which should be addressed.

Are there any cases where StrainPhlAn identified closely related strains within the same metagenome for the six bacterial species? How are such instances of multiple strains present within the same individual being accounted for when building the bacterial phylogenetic trees? We often see such examples of multiple strains in oral metagenome data and their presence will impact the placement of those samples in the tree.

For the six bacterial phylogenies, please specify how many clade-specific genes each tree is based on, how many data-points i.e. number of sites in each alignment, and percentage of missing data allowed.

The circular trees shown in figures S2 15-17 are a bit difficult to interpret. I recommend showing these as regular bracket-style trees. For some of the trees, not many of the branches seem to have more than 95% bootstrap support, which is the recommended threshold when using ultrafast bootstrapping. This should be mentioned while interpreting the results.

Referee #2

(Remarks to the Author)

I am happy with the extensive changes made in response to my comments and the thorough explanations provided. It is especially nice to see the maps of historical Indigenous and Russian settlements as they help to clarify the demographic situation (Figure S1_1), and to see the very extensive collection of primary archaeological data posted at Zenodo. This is a great resource for future researchers!

I agree that all genetic analyses for present-day Yakuts (published and unpublished ones, cited in answer 2.2) demonstrate very low levels of Russian ancestry, supporting the authors' claim that almost no admixture happened between ancient Yakuts and Russian colonizers (at least outside of the Russian settlements in the south of Yakutia, which were not excavated). This, and the much more detailed supplemental section on the sampling strategy, addresses my main concern (point 2.2).

However, I'm still unconvinced that the preferred qpAdm models for the admixture between the migrating Yakuts and the local Iron Age population of Yakutia are robust. Namely, I am puzzled by the rather low fraction of trans-Baikal ancestry estimated, 16-38% (point 2.12), and the high fraction of indigenous IA ancestry (62-84%). The relevant preprint which I referred to is now published (Flegontova et al. 2025, Performance of qpAdm-based screens for genetic admixture on graph-shaped histories and stepping stone landscapes. doi: 10.1093/genetics/iyaf047), and I'd like to highlight few points below.

1. Crucial proxy source groups used in the analysis (Yakutia_IA and Baikal_his) are very small, comprising just 2 or 4 individuals (answer 2.12), which limits the accuracy of admixture fraction estimates. I also see that other alternative source populations were very small (answer 4.13): Russia_AngaraRiver_Medieval.SG (Kılınç et al., 2021) (N=1), Mongolia_Sukhbaatar_Xiongnu (N=1) and Mongolia_Khuvsgul_MLBALateMedieval (Jeong et al., 2020) individuals (N=2) and modern Buryat (Buryat.SG (Raghavan et al., 2015; de Barros Damgaard et al., 2018), N=4). I suggest increasing the samples size for the proxy sources, if possible.

2. The "right" population set (answer 2.17) includes no group even remotely close to the region/period of interest and to the most plausible ancestry sources, which may also contribute to biased admixture fraction estimates (even if their SEs reported by qpAdm are low). The "right" (or reference) set includes: an African group (N=4), a Southeast Asian group (N=15), a Native American group (N=6), an Upper Paleolithic Siberian (N=1), a Neolithic Levantine group (N=4), a Neolithic Anatolian farmer group (N=16), a Paleolithic individual from West Europe (N=1), and a Paleolithic individual from the East European Plain (N=1). In our study of qpAdm performance on admixture-graph-shaped histories and stepping stone landscapes (Flegontova et al. 2025), we conclude that qpAdm results are much more robust when sampling is dense: when all the groups participating in the analysis in any role are sampled from a relatively small region such as West Eurasia or East Eurasia, rather than just few highly non-contemporaneous groups around the world. I suggest that the authors replicate their qpAdm analysis on several different sets of "right" populations (larger and more regionally and chronologically-focused sets); that would make the results much more robust!

3. I agree that the ADMIXTURE results supporting the high IA ancestry estimates for Yakuts are suggestive (answer 2.4), but we showed that both ADMIXTURE and qpAdm often detect ancestry from the same false sources (see Suppl. text 2 in Flegontova et al. 2025). All analyses based on allele frequencies are somehow correlated (PCA, f-statistics, qpAdm, ADMIXTURE) and should be interpreted with caution if they are not supported by other data types such as autosomal

haplotypes.

4. Answer 2.4 "All admixture proportion estimates obtained with qpAdm were within the [0 – 1] interval, whereas estimates outside these bounds are expected under the Isolation-By-Distance model from Harney et al. (Harney et al., 2021)."
This is not true since Harney et al. (2021) explored unrealistic one-dimensional stepping stone landscapes and only briefly, while Flegontova et al. focused on two-dimensional landscapes sampled in random and systematic ways. On 2D landscapes, estimated admixture fractions between 0 and 1 are typical for symmetric source arrangement around the target, but are also relatively common for non-optimal sources (those far away from the target in the geographic and genetic space) and for highly asymmetric source arrangements, which drives high false discovery rates typical for qpAdm-based admixture screens (Flegontova et al. 2025). In general, the qpAdm method is very noisy (especially p-values, but admixture fraction estimates too) and should be applied cautiously, especially if the vast parameter space of potential "right" or rotating deme sets is explored just a tiny bit (see my recommendation above).

5. On the other hand, I do appreciate that no extensive exploration of hundreds of alternative qpAdm models was attempted by the authors who instead chose to focus on the most historically plausible models (answer 4.13). High-throughput model exploration is dangerous since prestudy odds are very low, and false discovery rates of such protocols often reach 90% (Flegontova et al. 2025).

Referee #4

(Remarks to the Author)

Responses: Referee #4 (Remarks to the Author):

I am happy with the improved manuscript. Only: With more information, I am a bit less sure about the ethical oversight (see my comment below to answer 4.3). While improved, I would also still suggest toning down the language regarding social structure and mention the caveats: given the vast spaces and regions, the sampling is very sparse and there are no sizable sites/cemeteries that would allow study complete communities. It is thus hard to maintain that patrilocality is strongly present in the data and the differences within/between regions metrics could still contain mobility between regions. The findings are interesting and impactful even if it is admitted that more research is needed to elucidate these issues (answers 4.5 and 4.6).

REACTION OF R4 TO ANSWER 4.1: I am glad these efforts have been made and I am glad that they are now featured in the SI. I understand the need not to over-self-promote (and it can be also stated in the SI that this was a result of the review process if the authors so wish) but I do think that this new addition also serves as a good example for studies where such efforts were not made (and should be made) potentially raising standards in archeological/archeogenetic studies in general. It also highlights the amount of efforts of local researchers that, while having limited resources, often spearheaded them.

REACTION OF R4 TO POINT 4.2: Great, see my reaction above. Though the actual discussions of genetic results that worried me are a bit hidden within a large scale archaeological outreach and I would suggest to separate them and highlight them specifically.

REACTION OF R4 TO ANSWER 4.3: I disagree that 'no genetic relatedness could be expected' (even when it might be unlikely given how many modern and ancient DNA samples were analysed) and I am surprised the authors think so. I personally know people living (and researching) today that have their grandparent born in 1850s and with IBD fragments, the detection of relatedness is well possible beyond 4th degree. So it does not seem to be at all impossible that individuals dated to stage 4 (1800-1922) and potentially even earlier stages could be found to have direct links to living people. Given the newly-stated facts in the SI, I agree that legally everything seems correct. But one could discuss public availability of the recent data related to a historically-exploited minority in a situation where one cannot be sure about governmental or private company misuse of private data when this is only loosely supported by an ethical permission dating to 2012, quite earlier before any large-scale autosomal aDNA data existed and many issues regarding DNA and its privacy have not yet been established in genetics (as is also evidenced in the wording of the 2012 permission that mentions bacteria and viruses and 'influence of socio-ecological factors on movement, adaptation, and dynamics of ancient populations' but does not touch on the issues of human DNA directly). I have seen decisions from ethical commissions at European/African institutions in recent years for issues related to recently-dated aDNA samples that prefer to protect this data (EGA – data is available but upon request and does not allow automatic use for example forensic or genealogical companies and searches) and I would like to ask that the authors consider getting an updated opinion from the ethical commission from a French institution (as getting it from both countries might be difficult and time-intensive and importantly, it might be good to avoid an impression of double-standards and ethics dumping).

REACTION OF R4 TO ANSWER 4.4: Thank you for updating the text, as far as I am concerned, this is fine (though some other reviewers from history will be more knowledgeable in this regard and raised similar issues).

REACTION OF R4 to ANSWER 4.5: I appreciate the increased precision in the text and the additional tests. I would still mention the caveats of low sample size, sparse sampling, archaeological biases (more men buried at a site, more men likely to be determined as related at a site → driving the signal). So I would be careful about the sweeping statements of how much in line the findings are with the expectations as the significance of the tests is low and as mentioned previously because within the 3 pedigrees resolved (extended fig. 5), there can be actually a case made for 2 possible matrilineal situations (not only the mother who is also a daughter in AC but also in Oktiom – a woman that has a brother at the site) and not even one patrilineal case (father with adult son might be a son prior to marriage). I do not argue that the society was not patrilineal. I argue that this could be argued from genetics: this is mainly a historical statement to which genetics finds exceptions rather than confirms it really. If the historical records spoke about matrilineality, I am sure that then matrilineality would be also claimed as in line with the evidence due to the matrilineal cases. Genetics has history of over-interpreting social practices, for example from mtDNA/Y-chrom patterns (despite the issues with effective population size). I am not suggesting to delete these statements about social practices but I would suggest being clear about the limitations and minimize statements like 'Social organization remained mainly structured around patrilineal clans showing strong ties to the land of their ancestors' as that then sounds as inferred from genetics.

REACTION OF R4 to ANSWER 4.6: It still feels a bit overinterpreted, I would be very careful about statements about social structure giving the sampling. For example, as I have mentioned, how likely it would be to find such relatives between the regions given the space and time covered? How much migration and mobility there could be that would not be visible? I would say quite a lot. This observation should be stated but I would be clear about the increased precision in the text and the additional tests. More work in future with further samples and other sampling strategies might answer these questions

RESPONSE TO ANSWER 4.9: OK.

RESPONSE TO ANSWER 4.10: OK.

RESPONSE TO ANSWER 4.11: OK..

RESPONSE TO ANSWER 4.12: OK..

RESPONSE TO ANSWER 4.13.1: OK..

RESPONSE TO ANSWER 4.13.2: Thank you for the addition.

RESPONSE to ANSWER 4.14: OK, if the intra-site analysis was not the goal, I would suggest rewording the statement in the discussion: 'Additionally, kinship analysis revealed that first- and second-degree relatives were primarily buried in geographically close locations.' as this to me rather suggests that comparison of related and unrelated individuals within a site was made. Thank you for sharing the detailed archaeological information with the community, this is a great addition and allows further usability of the results of this study.

RESPONSE to ANSWER 4.15: I think both presentations are helpful and great to have, thanks.

RESPONSE to ANSWER 4.16: OK.

RESPONSE to ANSWER 4.17: OK.

RESPONSE TO ANSWER 4.18: OK.

RESPONSE TO ANSWER 4.19: Actually there is an additional filter recommended by Ringbauer et al. 2023 that can be expressed in SNPs or in coverage. Namely it is 0.25x coverage for WGS data (as stated in the paper but it is likely better for most samples and also for IBD fragments longer than 12cm that were used here). It is unclear how low quality samples have been used in the analysis (e.g. I would have checked if such samples are the ones that show extreme numbers of IBD fragments shared, e.g. Extended fig. 5 for pairs of males from the same regions).

RESPONSE TO ANSWER 4.20: OK, I find the analysis reproduced without the manual alignment convincing.

RESPONSE TO ANSWER 4.21: OK. I would actually mention these observations in the SI but I leave that to the authors' discretion.

RESPONSE TO ANSWER 4.22: OK

RESPONSE TO ANSWER 4.23: OK

RESPONSE TO ANSWER 4.24: Yes, only reference was needed. Thanks.

RESPONSE TO ANSWER 4.25: OK

New small point:

Axis label in extended Fig. 5 is missing D in IBD.

I was asked to add reactions to the comments of reviewer 3:

RESPONSE of R4 to ANSWER 3.1

I found this new addition rather too positive for Russian settlers. Indeed the other cases of colonialism might have been harsher in some ways though one could question the need to relativize one injustice with another. Also, I am not keen on the wording from the perspective of Russia ('hindered the development of demographically-*thriving* Russian colonies' - this could be reworded from the perspective of Yakuts, for example population replacement of Yakuts or similar). Otherwise ok as far as I can judge.

RESPONSE of R4 to ANSWER 3.2

I understand but I suggest that this section is referred to in the main text - not the section only but the fact that there is a terminological issue (e.g. "for discussion of the terminology of ethnonyms, see Section..") so it is clear to the readers that it is not so simple.

RESPONSE of R4 to ANSWER 3.3

OK I think, though the idea of the reviewer 3 seems to be also that alcohol could get to Yakut communities via trade with other Steppe people not only through European/Chinese/Mongolian route.

RESPONSE of R4 to ANSWER 3.4

As stated in my comments, I found this sentence overinterpreted but this comment (exogamy) seemed to have been addressed.

Version 2:

Reviewer comments:

Referee #1

(Remarks to the Author)

I am satisfied with the edits made by Crubézy and co-authors in response to my comments on the previous version of the manuscript and I am happy to recommend this manuscript for publication.

Referee #4

(Remarks to the Author)

I am happy with the manuscript after the 2nd review, congratulations to the authors! I have only 2 additions to the points previously raised (below): except for a small typo, all my points were sufficiently addressed and I recommend the article to publication.

---RESPONSE of R4 to Response 4.4

I am OK with this even though I would like to state that I have not recommended EGA, I have given an example of what ethical commissions in other similar cases preferred and my recommendation was to get an ethical assessment which I would have preferred. But if I was on such a commission, this would seem sufficient and it is in line with similar decisions in similar cases in the past.

---RESPONSE of R4 to Response 4.10

I think the old extended Fig 5 was copy-pasted here and it is also still incorrect in the submitted file for extended fig. 5. In the subfigure A, the Y axis reads IB>12cM [cm] and I think it should read IBD (also it might need to read [cM]).

Open Access This Peer Review File 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 license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <https://creativecommons.org/licenses/by/4.0/>

Referees' comments:

Referee #1 (Remarks to the Author):

A. Summary of the key results:

The manuscript by Crubézy and colleagues describes ancient DNA analyses of Yakut individuals from northeastern Siberia spanning four archaeological periods from 1500 – early 1900s CE. Analyses of human DNA suggest that these ancient Yakut individuals originated from admixture between local populations and trans-Baikal groups migrating with the spread of the Mongol Empire. The authors find strong genetic continuity between ancient Yakut populations and modern Yakut individuals, with marginal contribution from Russian individuals. Analyses of oral microbiome profiles suggest no major changes in the oral taxonomic and functional composition, despite the introduction of carbohydrate-rich foods after Russian contact. The authors recovered two partial *Variola major* (smallpox) genomes from the Yakut individuals; phylogenetic analyses suggest that these ancient Yakut strains are distinct from those previously identified among ancient European populations as well as those isolated from 20th century individuals.

B. Originality and significance:

The study represents the first ancient DNA analysis of Yakut individuals. The sample size is large, and the data generated by this study will be of great interest to the field of ancient DNA.

C. Data & methodology:

The study methodologies are mostly sound (see below). I do not have concerns regarding data quality or presentation.

D. Appropriate use of statistics and treatment of uncertainties:

I am not an expert on the admixture and human population genetics analyses, so I will restrict my comments to the microbiome and pathogen screening analyses (see below).

E. Conclusions:

Most of the conclusions seem sound to me.

F. Suggested improvements

See below for major and minor comments.

G. References:

POINT 1.1

I have no concerns with the References cited. I have noted one additional study that could be included (see below).

ANSWER 1.1:

The study suggested, Ottoni et al. (2021), is now cited directly in the main text. See ANSWER 1.14.

H. Clarity and context:

The manuscript is overall well-written, well-contextualized, and it was a pleasure to read. The Introduction and Discussion sections are especially well-written.

Major Comments:

POINT 1.2:

In general, the microbiome section lacks a description of the Yakut oral microbiome composition. I recommend at least a couple of sentences describing the abundant taxa be included in the Main Text.

ANSWER 1.2:

The most abundant bacterial species present in the historical Yakut samples is *Actinomyces dentalis*, a species that was originally isolated from human dental abscess (Hall et al., 2005), but that in fact represents a standard member of the oral bacterial flora. Other oral bacteria that are commonly found belong to the Actinobacteria and the Proteobacteria class, such as *Olsenella sp. oral taxon 80*, *Desulfobulbus oralis* and *Desulfomicrobium orale*. Members of these classes are known to be part of the multi-species bacterial communities that colonize the oral cavity, and play key roles in the development of dental biofilms and plaque (Modi et al., 2023). We have integrated the contextual information discussed above as a figure in the Supplementary Information Section 2.14, and in the main text (page 4, lines 207-209), as follows:

“The most abundant bacterial species identified was *Actinomyces dentalis*, followed by other species such as *Desulfomicrobium orale*, *Desulfobulbus oralis*, and *Olsenella sp. oral taxon 80*, all of which known to be prevalent in past oral environments and to participate to dental biofilms and plaque³⁸ (Fig. S2_11).”

We have also prepared a new Supplementary Figure (Fig. S2_11), showing the microbial composition of the different samples analysed in the four archaeological stages investigated. This figure is duplicated below.

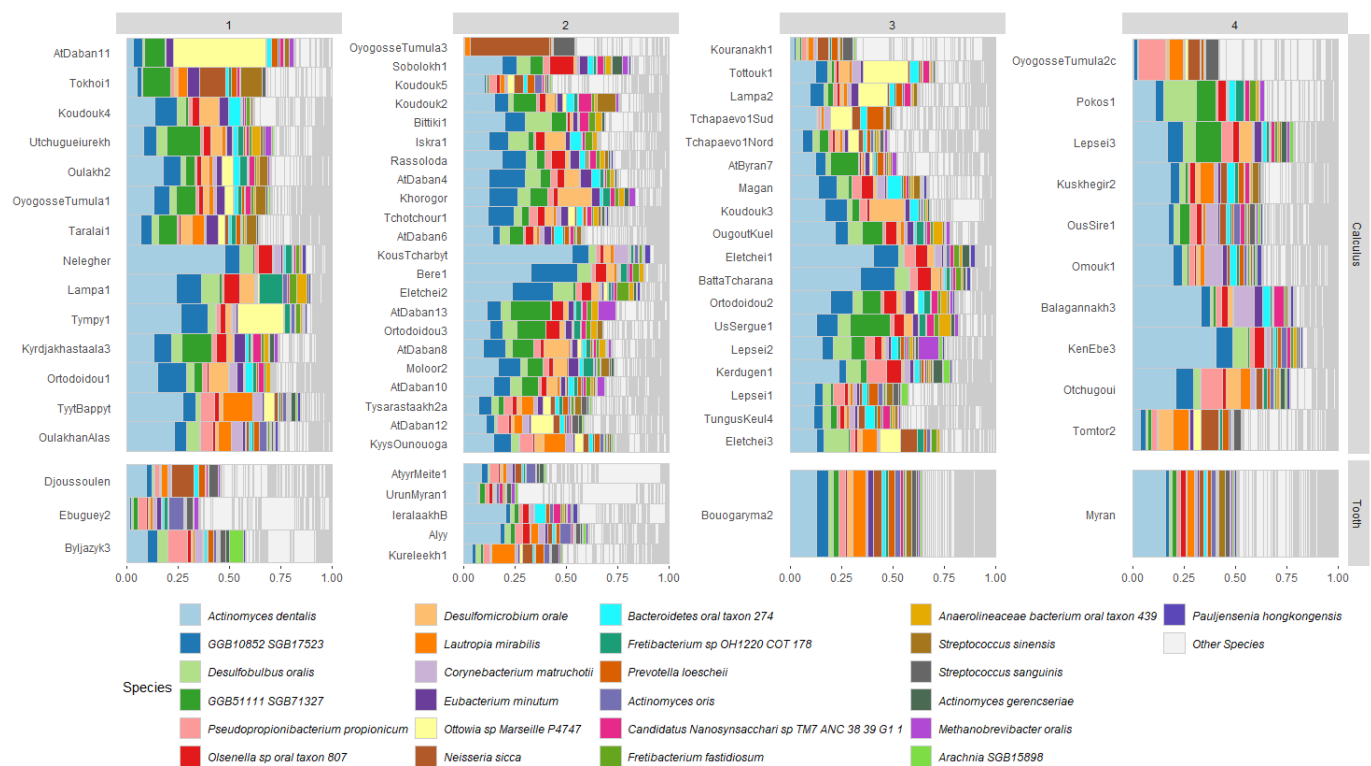


Figure S2_11. Microbial composition at the species level of the samples selected for analyses. The 25 taxa that are most abundant across all plotted samples are displayed in colors, while the other less abundant species identified are shown in light grey. Results are organized in four columns, representing each archaeological stage (1 to 4), and separated by sample type (ie dental calculus *vs* tooth).

POINT 1.3:

The MetaPhlAn4 results should be included in the Supplementary Tables.

ANSWER 1.3:

We now provide as Supplementary Table 1l the exact output of the MetaPhlAn4 analysis for each sample. For consistency, we have also prepared the new Supplementary Table 1o, which presents the HUMAnN3 output underlying the functional analysis of oral microbiomes. The captions of the two tables are copied below.

Supplementary Table 1Q. MetaPhlAn4 taxonomic profiles per sample.

The table combines all Metaphlan4 individual outputs, with identifications down to species levels. Samples are displayed per column, with their corresponding read counts of each microbial taxonomic group provided per row. The various microbial taxonomic groups identified are displayed per row, with their respective classification. Taxonomic names are prefixed to help indicate their rank, with k__ indicating Kingdom, p__ Phylum, c__ Class, o__ Order, f__ Family, g__ Genus, s__ Species, and t__ species.

Supplementary Table 1R. HUMAnN3 functional profiles per sample.

The table combines all HUMAnN3 individual outputs, with the taxon associated with the pathway reported identified at the Genus g__ level. Individual fields indicate the number of read counts normalized by count per millions identified in a given sample (column) for the pathways indicated (row).

POINT 1.4:

The authors state that they used default parameters for the MetaPhlAn4 and HUMAnN3 taxonomic and functional assessment. Using the default parameters, collapsed reads shorter than 70 bp would be excluded from the analysis. I recommend the authors double-check the parameters used and consider using a minimum read length more appropriate for ancient DNA (such as 30 bp).

ANSWER 1.4:

We apologize that the method section did not describe the custom parameters that were used. Microbial read counts were obtained using MetaPhlAn4, discarding unclassified reads and reads that were too short, using a dynamic read length filter to minimize the inclusion of low quality reads. The threshold was equal to the most frequent read length value of the sample minus ten and never going lower than 30bp and higher than 70bp.

The online Materials and Methods section has been updated (page 13, line 694-697):

“Microbial read counts were obtained using MetaPhlAn4 (Blanco-Míguez et al., 2023) (Supplementary Table 1l), discarding unclassified reads and reads too short. We used as a minimal read length filter equal to the most frequent read length value (visually checked) minus ten, never going lower than 30 bp and higher than 70 bp (Supplementary Information Section 2.14)”

And the Supplementary Information Section 2.14 has been updated (p. 44, line 441-449):

“Microbial read counts were obtained using MetaPhlAn4 (Blanco-Míguez et al., 2023), discarding unclassified reads and reads too short. To ensure robust taxonomic profiling of microbial communities, we implemented a dynamic read length filtering strategy while running MetaPhlAn4. This approach accounted for variability in read length distributions across samples, enabling the selection of an appropriate minimum read length threshold for each sample, to avoid using low quality reads and thus false positives. For each sample, we first determined the most frequent read length (peak) and visually checked the distribution to account for anomalous read length distributions, where the primary peak did not correspond to the actual dominant read length. We then used as a minimal read length filter equal to the peak value minus ten, never going lower than 30 bp and higher than 70 bp.”

POINT 1.5:

Line 214: The authors state that the stability in abundances of the bacterial complexes over time is “indicating no apparent changes in oral health in relation to Russian conquest”. There have been multiple studies showing that the presence and/or abundance of bacterial complexes in dental calculus, such as the Red Complex, cannot be used to determine oral health status of ancient individuals (Velsko et al. 2018 and

others). Is there any paleopathological information available for these individuals to suggest changes in oral disease prevalence in relation to the Russian conquest?

ANSWER 1.5:

MAFSO excavations (2004-2019) have revealed an extensive corpus of 209 skeletons. These included 96 adults with unique dental features characterized by a low prevalence of cavities, of which only a fraction proved lethal (Crubézy et al., 2021). Previous work indicated subtle changes in dental health following contacts, but these became more frequent during Archeological Stage 4, as the local population became Christianized. More specifically, subjects from Archeological Stage 4 showed reduced tooth wear but increased tooth loss. Historical textual evidence rejects increasing sucrose consumption at the time. The deprecation of dendrophagy (wood consumption), which is associated with increasing smoking practices and cereal consumption, but decreasing meat consumption, represents the most-commonly assumed driving factor. We have integrated the evidence discussed above in the main text (page 5, lines 226-228):

“Previous work reported a shift in carbohydrate consumption shifting during Russian conquest (Seroshevsky, 1993), from pine sapwood flour (dendrophagy) to barley and rye, alongside increased risks of dental infection linked to smoking, particularly in Stage 4 (Macé et al., 2021).”

And page 5 and line 235-237 : “These findings suggest that oral health remained largely unchanged during this period, consistent with the low prevalence of dental cavities observed archaeologically (Crubézy et al., 2021).”

And we also now mentioned the limited performance of bacterial complexes in ancient calculus to infer oral health of ancient individuals page 5 and line 232-235:

“Additionally, strain-level analysis of six oral pathogens or abundant oral species, which offer improved resolution into the oral health of ancient individuals relative to bacterial complexes (Velsko et al., 2019), indicated different strains circulating in Yakutia and the rest of the world, and no shifts during the Russian conquest (Fig. S2_15-S2_17).”

POINT 1.6:

Along similar lines, do any of the individuals show paleopathological signs of diseases (eg. TB)?

ANSWER 1.6:

Previous work involving some of the coauthors of our study has described the presence of bone tuberculosis in a subset of individuals, on the basis of various lytic lesions affecting the vertebrae, long bones and joints. So-called ‘cold abscesses’, potentially also indicative of the same disease, were also reported on other skeletal bones (Dabernat et al., 2014). The tuberculosis diagnostic was confirmed with molecular evidence on two subjects also analyzed in the present study (Okhtoubout2 and Bakhtakh3), through the amplification and sequencing of IS6110 and SNP genotyping, which indicated European origins for the pathogenic strain. Additionally, previous molecular work (also involving some of the coauthors of our study) identified subjects infected by smallpox at the Arbre Chamanique (AC) site. However, no obvious skeletal and dermatological evidence were present (Biagini et al., 2012). Smallpox diagnostic was revealed following (1) DNA amplification and sequencing of smallpox-specific barcodes and (2) histopathological analyses of the lung tissue. The sequence data generated in the present study confirmed the presence of smallpox in some of the AC subjects analyzed (including from lung tissue, eg AC1num2). However, it did not confirm the presence of *M. tuberculosis* in Okhtoubout2 and Bakhtakh3, most likely due to the focus on petrosal bones, also prone to increased false negative rates when attempting to detect bacterial pathogens (Margaryan et al., 2018). We have integrated the evidence discussed above in the main text (page 7, lines 360-363):

“Despite the absence of significant genetic impact, Russian conquest profoundly affected Yakut livelihoods and demographics. Russian contact introduced numerous new infectious agents, including the smallpox identified here and *Mycobacterium tuberculosis* (Dabernat et al., 2014), with devastating demographic effects on immunologically naive Indigenous populations (Dabernat et al., 2014).”

POINT 1.7:

The authors state “In cases where a tooth and a calculus sample from the same individual passed filters, the profile maximizing oral microbial sources was kept, resulting in final data set 74 individual oral microbiomes”. Does this mean that oral profiles from tooth are being compared to those from dental calculus? It may not be such a big deal in this case since none of the microbes or complexes are found to be significantly differentially abundant, but nevertheless, I do not think it is an appropriate comparison.

ANSWER 1.7:

We apologize for the confusion. We indeed co-analyzed tooth and calculus metagenomic data together. We have now carried out a complementary analysis, restricting comparisons to data derived from dental calculus samples only. The results of this analysis were added to the Supplementary Information (Supplementary Information Section 2.14 and 2.16), and were almost identical to those previously reported, apart from the statistical difference in taxonomical composition between regions, which is now only marginally significant (PERMANOVA p -value = 0.053; likely due to the more limited statistical power resulting from data exclusion). We mentioned these analyses in the main text (page 5, line 222-223):

“Restricting all the analyses above on the subset of 66 dental calculi did not alter the conclusions (Supplementary Information Section 2.14 and 2.16, Fig. S2_13 and S2_19, Supplementary Table 1m).”

We have added two new paragraphs in the Supplementary Information Section 2 that read as follows:
For Taxonomical analysis, page 44-45, line 470-477:

“These analyses were repeated on a dataset restricted to calculus samples only. Here, we removed the individuals for whom only tooth samples were analyzed, and kept only the calculus data for those individuals associated with both tooth and calculus data. Alpha and beta diversity indexes, as well as abundance levels of bacterial complexes and pathogenic species showed the same temporal trends as analyses included both tissue types (Figure S2_13). Statistical assessment of differences in composition between region, period, sex and cultural group also returned similar results, except for the significance between region, which was only marginally significant (PERMANOVA, p -value = 0.0532), probably due to reduced power (Supplementary Table 1m).”

For functional analysis, page 53, line 575-580:

“These analyses were repeated on a dataset restricted to calculus samples only. Here, we removed the individuals for whom only tooth samples were analyzed, and kept only the calculus data for those individuals associated with both tooth and calculus data. Beta-diversity and abundance levels of carbohydrate pathways showed the same temporal trends as analyses included both tissue types (Figure S2_19). Statistical assessment of differences in composition between region, period, sex and cultural group also returned similar results, as none were significant (Supplementary Table 1m).”

We have also added two new figures in the Supplementary Information: Figure S2_13 and S2_19, copied below:

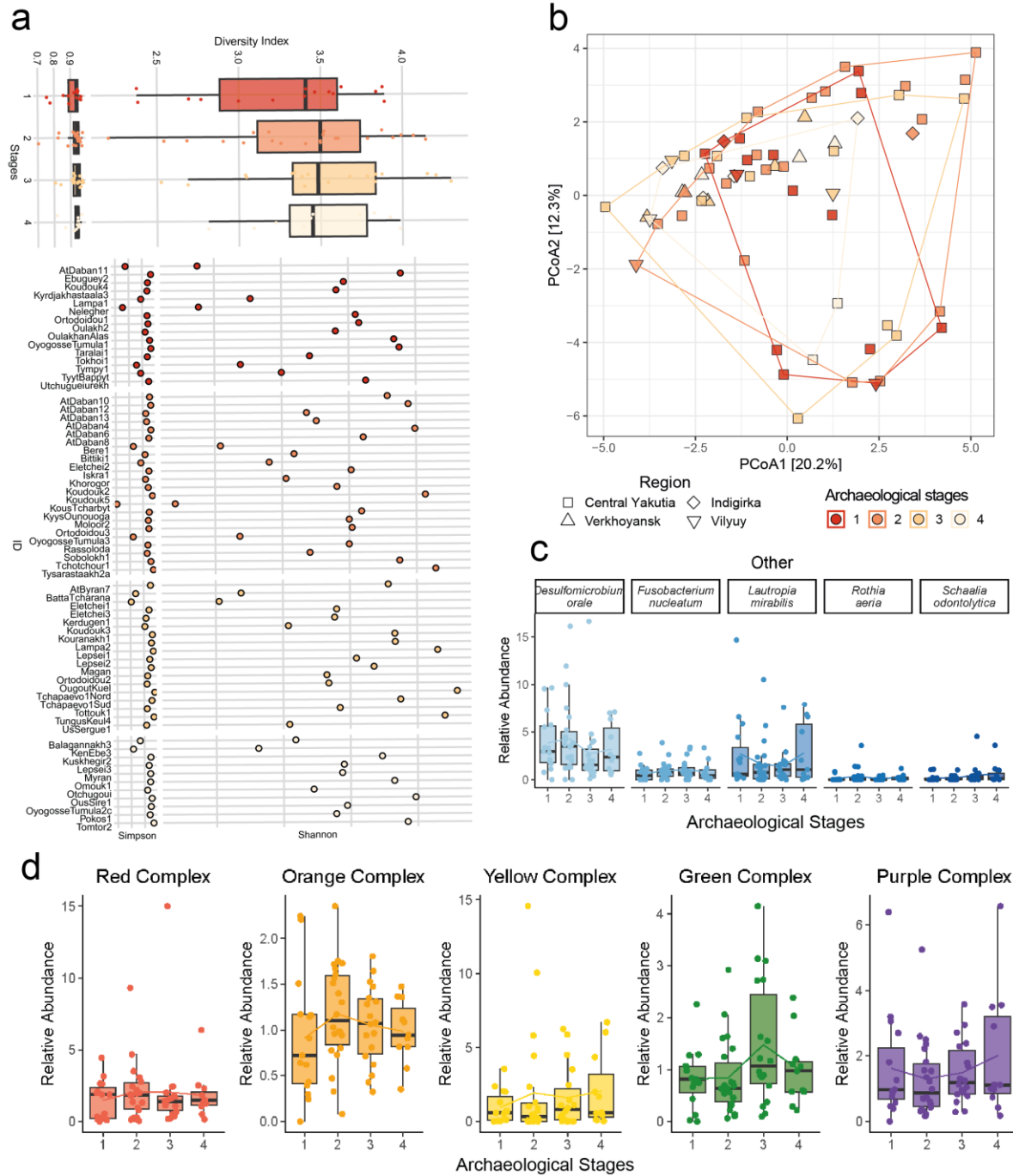


Figure S2_13. Taxonomical analyses for the calculus only dataset.

a) Diversity indexes for taxonomic diversity across the four archaeological stages. Top: Boxplot of both Shannon and Simpson diversity indexes for each Stage. Bottom: Per-individual diversity indexes. b) PCoA of 66 calculus samples carried out on Aitchison distances calculated on high-quality counts for 448 unique taxonomic species (abundance>1%) from the Metaphlan4 database. c) Relative species abundance of five oral pathogens. d) Relative species abundance of five bacterial complexes. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over 75th or under 25th percentiles, respectively.

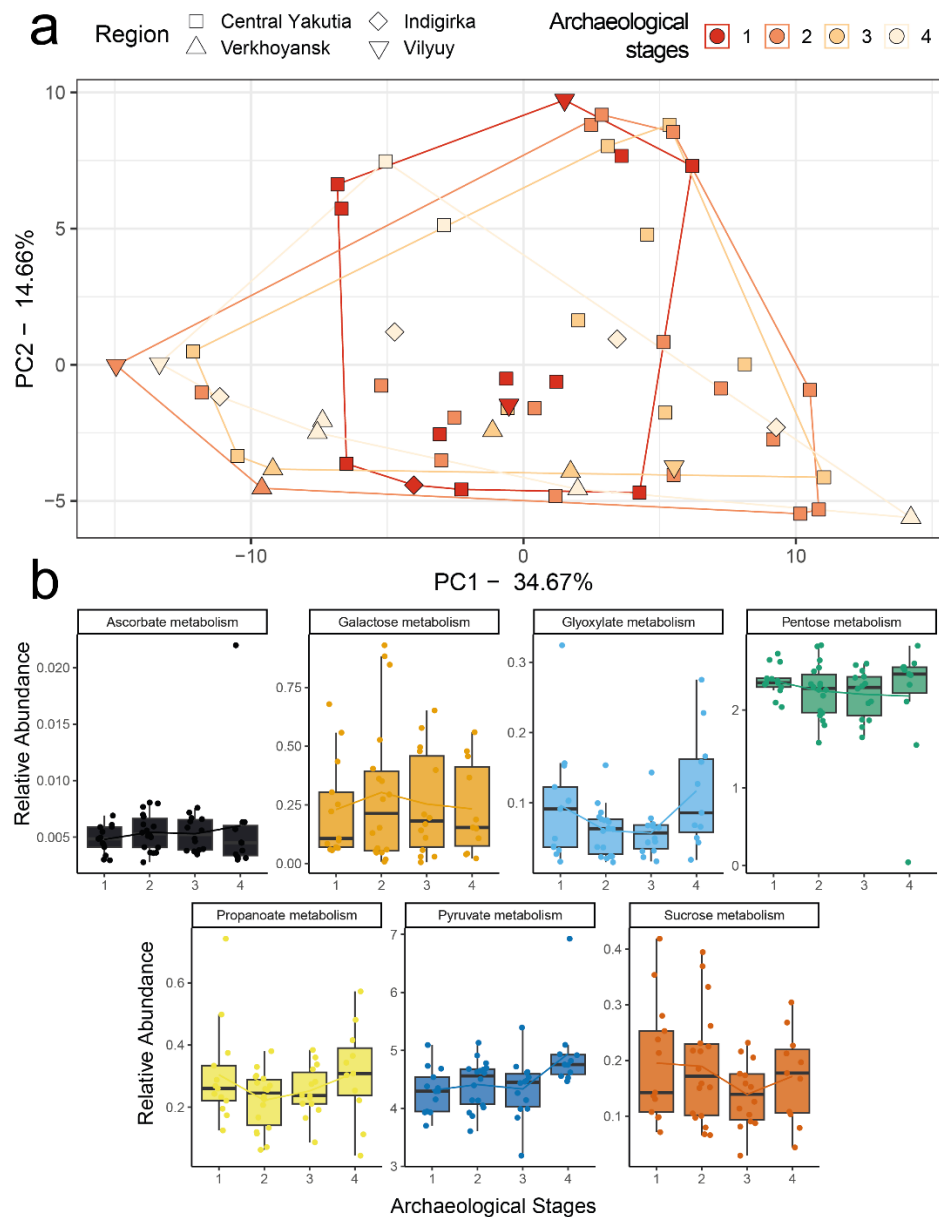


Figure S2_19. Functional analyses for the calculus only dataset. a) PCA of 58 calculus samples carried out on Aitchison distances calculated on high-quality counts for 303 pathways, after a centered log-ratio transformation. b) Relative abundance of the pathways underlying carbohydrate metabolism. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over 75th or under 25th percentiles, respectively.

POINT 1.8:

By only focusing on the abundances of microbial complexes, the authors have missed an opportunity to study strain-level variation in the Yakut oral microbiome over time. An attempt could be made to reconstruct the genomes of some of the abundant oral microbes to determine if the Russian conquest introduced new strains to the Yakut, similar to what has been found with the opportunistic oral pathogen, *Tannerella forsythia*, in

the Americas (see Bravo-Lopez et al. 2020, Honap et al. 2023).

ANSWER 1.8:

We appreciate the reviewer's suggestion regarding the possibility to investigate strain-level variation in the Yakut oral microbiome. We have followed this suggestion and carried out a strain-level analysis for several bacterial pathogens identified as the most prevalent in Yakut oral microbiomes. This analysis leveraged the StrainPhlAn, a strain-level population genomics tool (Truong et al., 2017; Blanco-Míguez et al., 2023), and included an additional reference dataset from medieval Europe and Africa as well as Neanderthal to use as outlier, which are part of the AncientMetagenomeDir (Fellows Yates et al., 2021) repository and provided interpretative context.

For six pathogenic or abundant oral species, we successfully reconstructed phylogenies spanning the four archaeological stages. However, no significant shifts in strain composition could be observed before and after colonization. These analyses could not be carried out on *Tannerella forsythia*, as it identified in a too restrictive number of individuals and archaeological stages. These findings suggest that the Russian conquest did not introduce detectable new strains into the Yakut oral microbiome, for the six pathogenic species analyzed. Future work is necessary to assess whether or not this extends to *Tannerella forsythia*, for which substantial strain turnover was identified in the course of the colonization of the Americas.

We have added these results in the main text (page 5, lines 232-235):

“Additionally, strain-level analysis of six oral pathogens or abundant oral species, which offer improved resolution into the oral health of ancient individuals relative to bacterial complexes⁴⁵, indicated different strains circulating in Yakutia and the rest of the world, and no shifts during the Russian conquest (Supplementary Information Section 2.15 and Fig. S2_15-S2_17).”

We have added a new section of the Supplementary Information entitled “Section 2.15. Strains analyses”, which reads as follows:

“To investigate strain-level variation in the Yakut oral microbiome, we used StrainPhlAn4 (Truong et al., 2017; Blanco-Míguez et al., 2023), a population genomics tool undertaking metagenomic strain-level analysis. This analysis was carried out on the species belonging to the red complex and the eight most prevalent bacterial species in our dataset. In addition to the Yakut individuals presented in this study, we also included set of reference samples from the AncientMetagenomeDir (Fellows Yates et al., 2021) repository v24.09 to provide interpretative context. More specifically, we included Neanderthal oral metagenomic data as outgroup (Yates et al., 2021), as well as oral metagenome data for 50 additional individuals dating back the last 500 years at most, and originating from Eurasia and Africa (Weyrich et al., 2017; Velsko et al., 2019, 2022; Eisenhofer et al., 2020; Yates et al., 2021; Kazarina et al., 2021). The description of the comparative dataset can be found in Supplementary Table S1n. The sequence data underlying our comparative panel were pre-processed following the same pipeline as that described above (Supplementary Methods section 2.14) for each Yakut extracts. StrainPhlAn4 was run with default parameters to extract species-specific MetaPhlAn4 markers from each extracts. Those reads aligning against these markers were combined to form a multifasta alignment for phylogenetic reconstruction, adding the respective reference genome sequence. Phylogenetic reconstructions were carried out in the maximum-likelihood framework of in IQtree 1.6.12 (Nguyen et al., 2015), with the best substitution model estimated using the Akaike Information Criterion (-m MFP), and node support assessed from 1,000 ultrafast bootstrap pseudo-replicates (UFBoot) (each bootstrap tree optimized using a hill-climbing nearest neighbor interchange (NNI) search; -bb 1000 -bnni). Trees were then rooted using Neanderthal samples as outgroups when Neanderthal alignments could be characterized (*Actinomyces dentalis*, *Desulfobulbus oralis*, *Desulfomicrobium orale*, and *Lautropia mirabilis*), or using the tree midpoint otherwise (*Treponema denticola*, *Porphyromonas gingivalis*). In total, we successfully reconstructed phylogenies for six bacterial species providing strain-level information for at least one Yakut individual per archaeological stage, which helped test whether significant shifts in strains accompanied the Russian Conquest.

The individual strains identified amongst ancient Yakut individuals did not cluster by archaeological stages, nor formed monophyletic clades organized per individual region (Figs. S2_15-S2_17). For the four species in which sufficient sequence alignments were detected amongst our comparative non-Yakut historical human samples,

these clustered together with strong node support. This indicates different strains circulating in those regions and historical Yakutia. The absence of stratification and significant clustering revealed no significant phylogenetic shifts during Russian Conquest, suggesting that, for the six species investigated, no new strain was introduced and ended up replacing those present locally.”

We have added three new figures, displaying the phylogenetic trees of the strains for each species: Figure S2_15, S2_16 and S2_17, that we reproduce below:

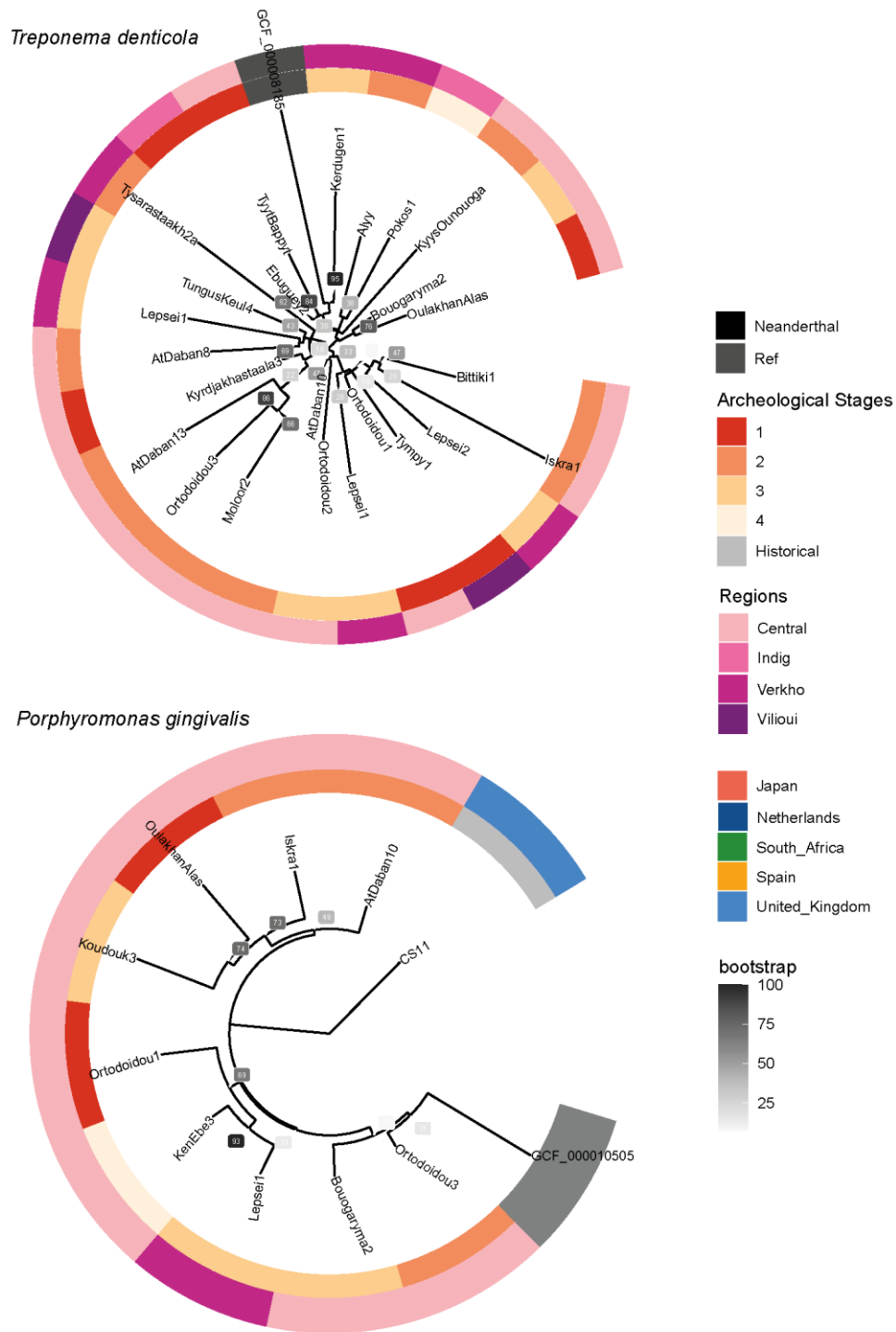


Figure S2_15 Strain-level analysis of species *Treponema denticola* and *Porphyromonas gingivalis*. The Maximum Likelihood phylogenetic trees presented was reconstructed using IQtree 1.6.12 (Nguyen et al., 2015), considering positive hits identified in 23 and 9 ancient Yakut individuals respectively, supplemented with the data from one historical human for *P. gingivalis*, plus the sequence of the reference genome of each species. Trees are rooted at mid-point.

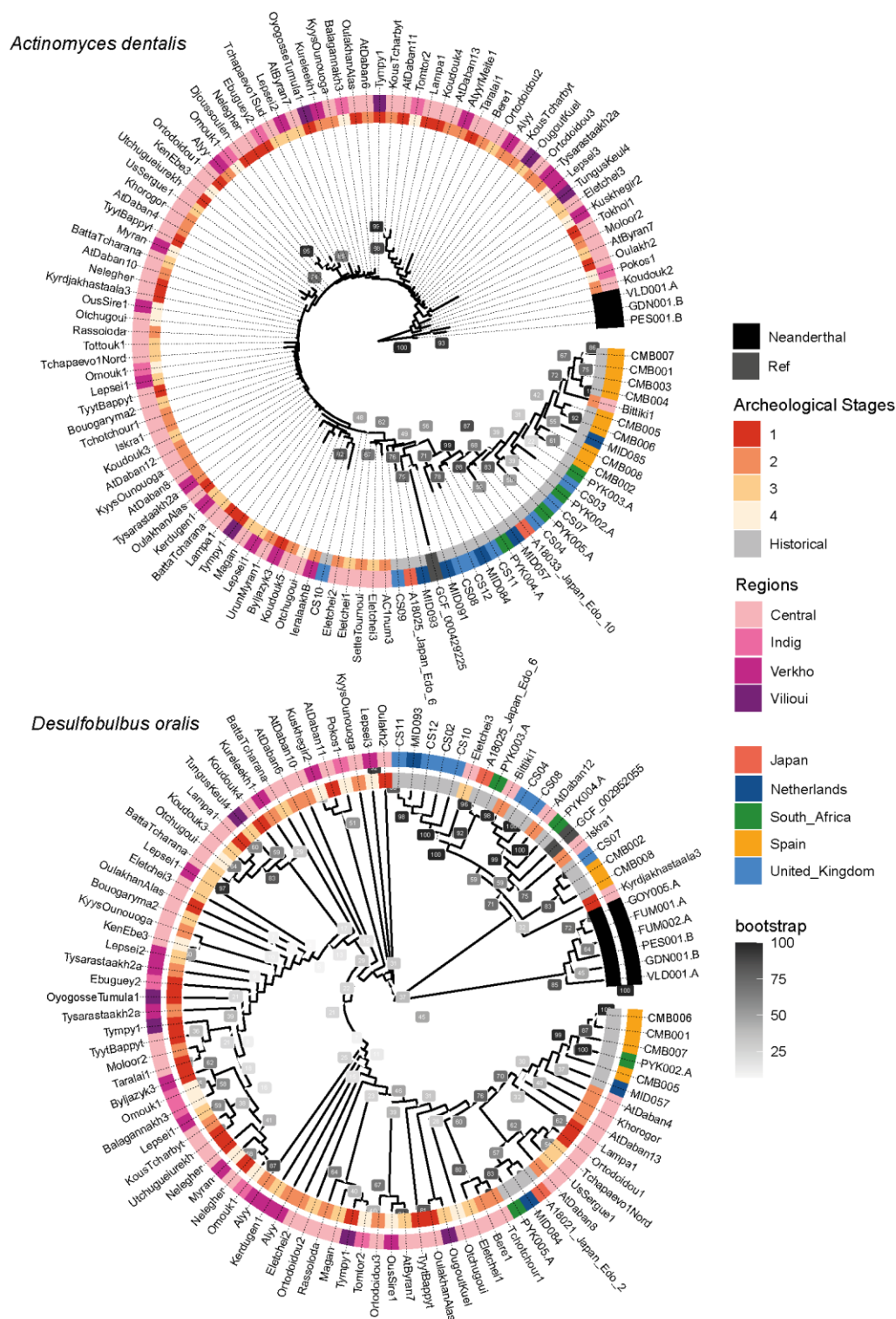


Figure S2_16. Strain-level analysis of *Actinomyces dentalis* and *Desulfobulbus oralis*. The Maximum Likelihood phylogenetic trees presented was reconstructed using IQtree 1.6.12 (Nguyen et al., 2015), considering positive hits identified in 84 and 73 ancient Yakut extracts respectively, supplemented with the data from three and six Neanderthal outgroups respectively and 19 and 22 historical human respectively, plus the sequence of the reference genome of each species. Trees are rooted using the Neanderthal samples as outgroups.

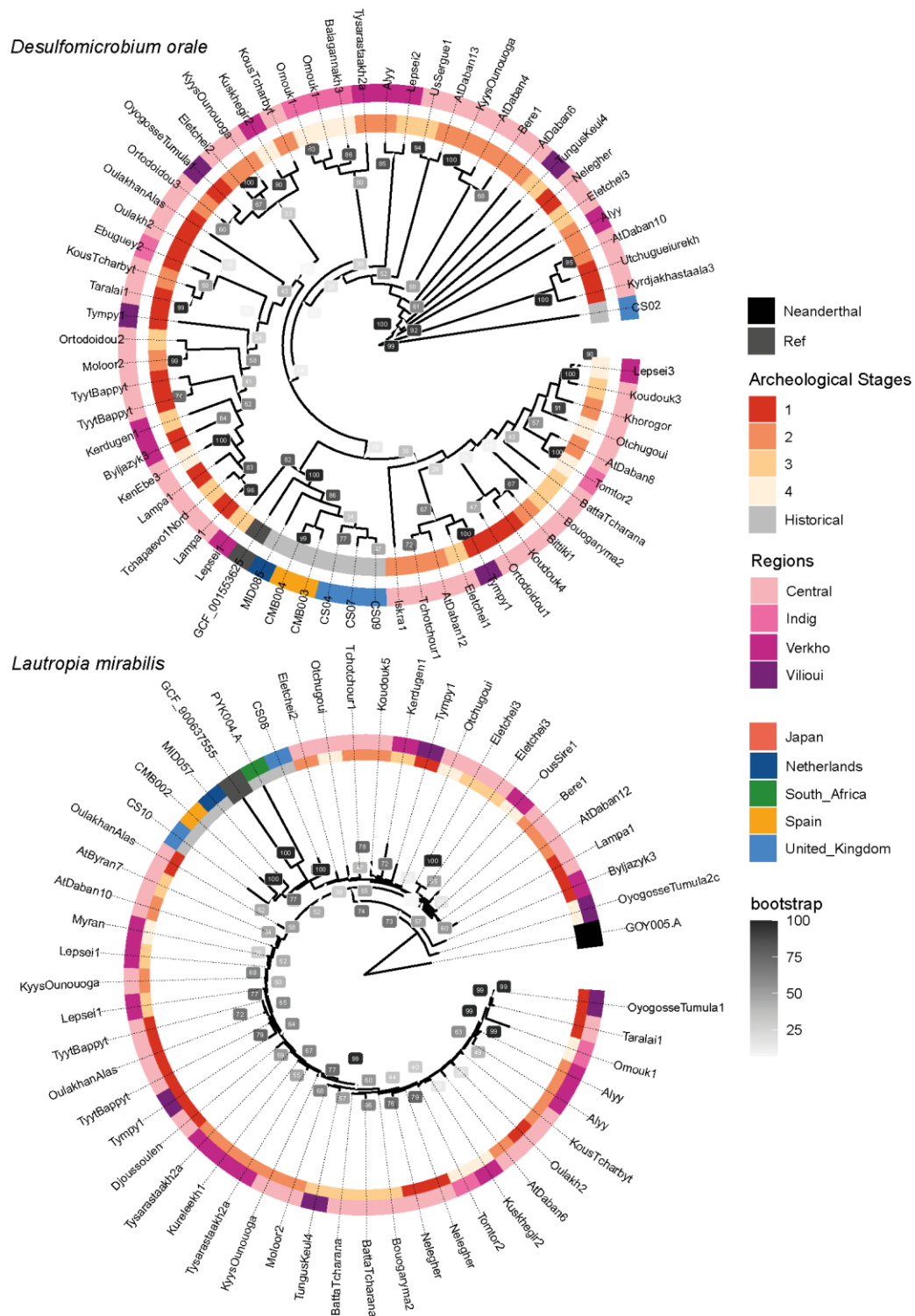


Figure S2_17. Strain-level analysis of *Desulfomicrobium orale* and *Lautropia mirabilis*. The Maximum Likelihood phylogenetic trees presented was reconstructed using IQtree 1.6.12 (Nguyen et al., 2015), considering positive hits identified in 58 and 48 ancient Yakut extracts respectively, supplemented with the data from one Neanderthal outgroup for *L. mirabilis* and seven and six historical human respectively, plus the sequence of the reference genome of each species. *D. orale* tree is rooted at mid-point and *L. mirabilis* tree is rooted using the Neanderthal sample as outgroup.

POINT 1.9:

It would be good to specify what material(s) yielded the Variola DNA for the three positive samples. From S Table A, it seems the Rassoloda individual had tooth and dental calculus sampled – was Variola DNA recovered from both? I cannot remember it being recovered from dental calculus in previous studies, so this would be an interesting finding. I would also recommend adding a table showing the mapping and coverage statistics of the Yakut samples to the Variola genome.

ANSWER 1.9:

We have prepared a new Supplemental Table (Supplementary Table 1p), breaking down sequencing efforts per library and tissue, as well as other statistics pertaining to post-mortem DNA damage, and human and smallpox genome coverage (the latter for the three positive individuals identified, to avoid highlighting false positives). Smallpox identification for the Rassoloda individual derives from the sequencing data of both calculus and tooth extracts.

POINT 1.10:

Minor Comments:

I am curious if the USER-treated libraries still showed significant terminal damage, given that the treatment was conducted for the full 3 hours.

ANSWER 1.10:

The newly added Supplementary Table 1b provides estimates of post-mortem DNA damage, as assessed from mapDamage2 (Jónsson et al., 2013) and PMDtools (Skoglund, Northoff, et al., 2014). Nucleotide mis-incorporation profiles, conditioning on CpG dinucleotides or not, are also shown in Figure S2_1, while Figure S2_2 provides post-mortem fragmentation profiles. DNA libraries constructed on USER-treated DNA extracts show the expected increase of C→T (and complementary G→A) mis-incorporation rates towards template starts (ends) when conditioning the analyses on CpG dinucleotides. This is due to USER enzymes cleaving DNA at Uracil residues (ie deaminated Cytosines), but not at those methylated Cytosines that were affected by post-mortem DNA deamination (as these are in fact Thymine residues; (Orlando et al., 2021)). For those DNA libraries constructed on DNA extracts that were not USER-treated, the expected increase of C→T (and complementary G→A) mis-incorporation rates towards template starts (ends) is found regardless of conditioning on CpG di-nucleotides, as expected ((Pedersen et al., 2014).

Similarly, the base composition at the position in the reference genome located right before sequencing starts was enriched in Cytosine residues for those DNA libraries constructed on DNA extracts that were USER-treated, confirming the efficacy of the USER treatment, which specifically targets unmethylated Cytosine residues that were deaminated post-mortem. The other DNA libraries show instead inflated percentages of purines, as expected due to depurination representing the most prevalent hydrolytic reaction cleaving DNA templates post-mortem (Briggs et al., 2007). Combined, the sequencing data generated in our study show all expected evidence of post-mortem DNA damage.

We have added the material above, including underlying methodology, in a new section of the Supplementary Information entitled '2.4. Post-mortem DNA damage' (page 29, lines 114-146), which reads as follows:

“The sequencing data underlying individual libraries constructed on DNA extracts from bones, teeth and soft tissues, were analyzed for the presence of post-mortem DNA damage (Supplementary Table 1b). Base compositional profiles within (1) the first and last ten nucleotide positions of read alignments against the human reference genome, and (2) the five reference positions preceding and following read alignments, were estimated using mapDamage2 v2.2.1 (Jónsson et al., 2013), with the following parameters: -l 50 -b 5 -a 5 -Q 20 (Fig. S2_1). Considering the sequencing data generated from DNA extracts that were not treated by the USER enzymatic cocktail, the genomic position preceding (following) read alignments was generally enriched in purine

(pyrimidine) residues. This aligns with depurination representing the major process driving DNA fragmentation post-mortem (Briggs et al., 2007). In contrast, the sequencing data generated from DNA extracts that were treated by the USER enzymatic cocktail showed an enrichment of Cytosine (Guanine) residues at the genome reference position preceding (following) alignment starts. This is consistent with the cleavage of unmethylated Cytosine residues that were deaminated into Uracil residues post-mortem, which represents the specific target of the Uracil DNA Glycosylase, which is part of the USER enzymatic cocktail (Briggs et al., 2009).

Nucleotide mis-incorporation profiles were estimated using PMDtools v0.60 (Skoglund, Northoff, et al., 2014), with the following parameters (-m 25 -q 20 -d --CpG --range 10 --platypus). The sequencing data generated from DNA extracts that were not treated by the USER enzymatic cocktail showed an increase of C→T mis-incorporation rates towards alignment starts, mirrored by an increase of G→A mis-incorporation rates towards alignment ends. This aligns with Cytosine deamination representing one of the most important chemical drivers of DNA decay post-mortem (Briggs et al., 2007). This pattern was considerably reduced when considering the sequencing data generated from DNA extracts that were treated by the USER enzymatic cocktail, in line with this treatment largely reducing the sequencing errors resulting from post-mortem DNA damage (Briggs et al., 2009; Rohland et al., 2015) (Fig. S2_2). Conditioning the analyses on CpG dinucleotide sites revealed the presence of an increase of C→T mis-incorporation rates towards alignment starts (and of G→A mis-incorporation rates towards alignment ends), including when considering the sequencing data generated following USER enzymatic treatment (Fig. S2_2). This is in line with the deamination of methylated Cytosine residues, which are predominantly found in CpG contexts, forming Thymine residues (instead Uracil residues when affecting unmethylated Cytosine residues), which do not represent USER targets. Combined, base compositional and nucleotide mis-incorporation profiles show the expected signatures of post-mortem DNA damage (Orlando et al., 2021), and, thus, indicate that the sequence data generated are authentic and ancient.”

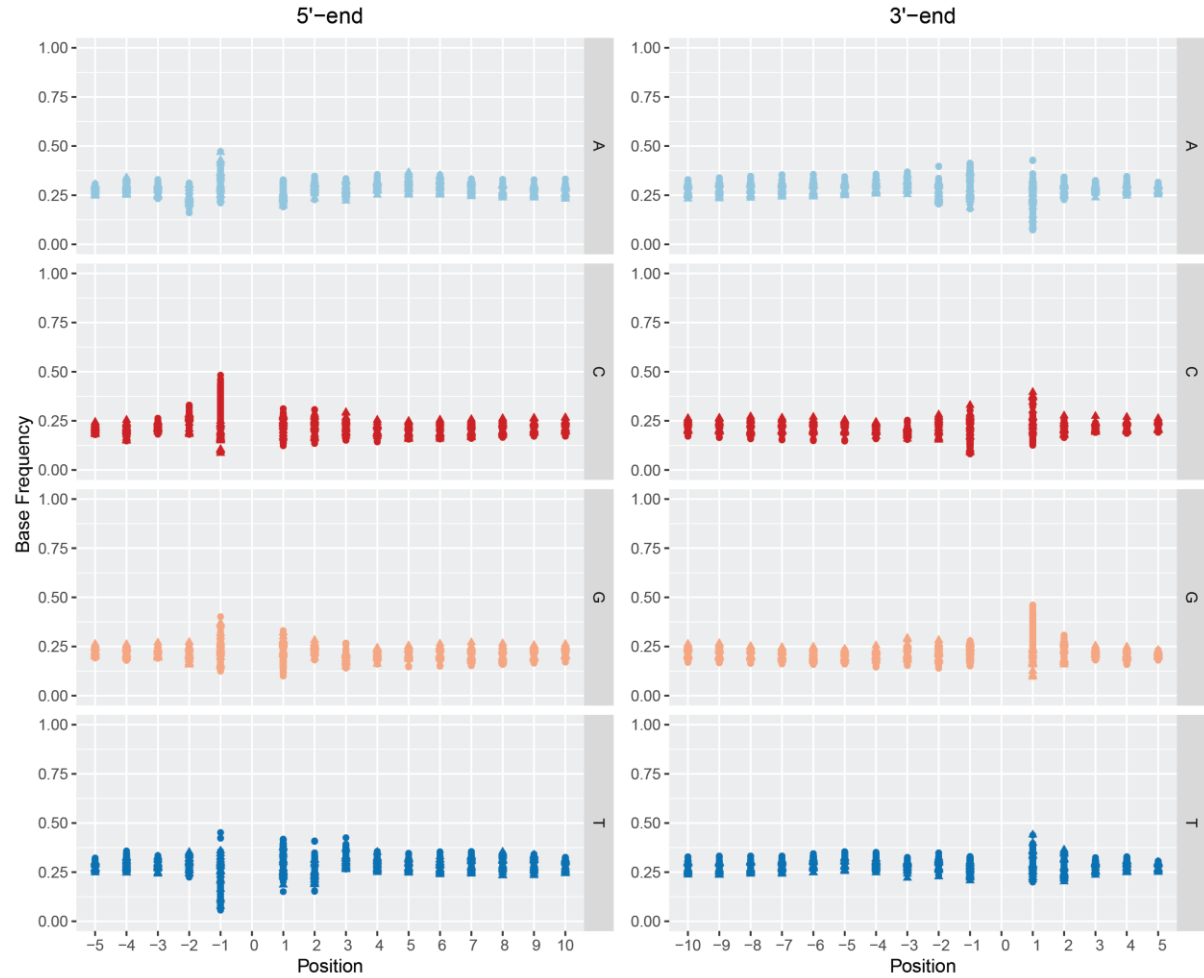
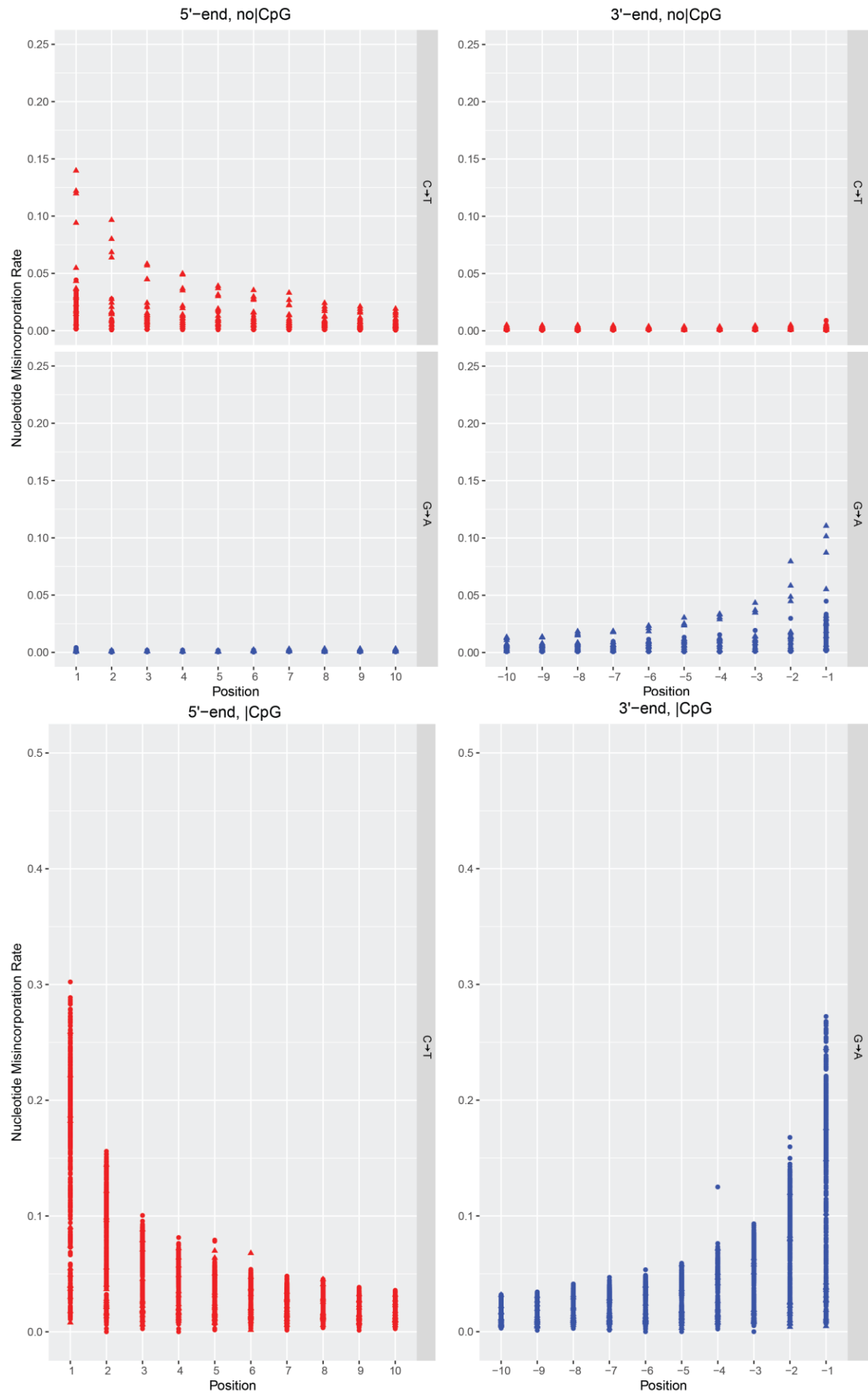


Figure S2_1. Base composition profiles.

Left: The frequency of each of the four nucleotides is calculated for the first ten (1 to 10) positions of read-to-reference alignments, as well as the five (-5 to -1) reference positions preceding alignment starts. Right: Same as Left, excepting that the last ten (-10 to -1) positions of read-to-reference alignments are shown, together with the five (1 to 5) reference positions following alignment starts. Circle: USER treatment. Triangle: no USER treatment.

Figure S2_2. (next page) Nucleotide mis-incorporation profiles.

Left: $C \rightarrow T$ (top, red) and $G \rightarrow A$ (middle, blue) mis-incorporation rates within the first 10 nucleotide positions of read alignments. The $C \rightarrow T$ mis-incorporation rates shown at the bottom are estimated within CpG dinucleotide contexts. Right: $C \rightarrow T$ (top, red) and $G \rightarrow A$ (middle, blue) mis-incorporation rates within the last 10 positions of read alignments. The $G \rightarrow A$ mis-incorporation rates shown at the bottom are estimated within CpG dinucleotide contexts. Circle: USER treatment. Triangle: no USER treatment.



POINT 1.11:

It may be unclear to a non-specialist reader what the different bacterial complexes (red, orange, yellow, green and purple) represent and why their abundances would be different across the archaeological stages.

ANSWER 1.11:

Sigmund Socransky was the first to categorize oral bacterial according to the severity of periodontal diseases they cause (Socransky et al., 1998). Bacteria forming the green and yellow complexes represent early colonizers, adhering to the pellicle and forming a biofilm favorable to other bacteria associated with periodontal diseases that are part of the other complexes (especially those from the red complex, including the most pathogenic). Recent work has suggested more nuanced models for the development of periodontal diseases, in which bacteria forming the red, orange, yellow, green and purple complexes are commensal, and that diseased states develop when some bacteria transition from commensalism to pathogenic behavior. (Van Dyke et al., 2020). These more nuanced models extended the list of bacteria that are considered potentially pathogenic, eg *Filifactor alocis* (Antezack et al., 2023). In this study, we have considered the most inclusive list of bacteria forming the five complexes, in an attempt to detect compositional (and functional) shifts accompanying the Russian conquest. We had hypothesized that such shifts may have resulted from the dietary and sanitary changes introduced by the Russian conquest. For example, Archaeological Stage 4 was associated with increasing tooth loss (and decreasing tooth wear) (Crubézy et al., 2021), which could have resulted from a prominence of members of the orange complex known to be associated with increased tooth pocket depth and progressive attachment loss (Socransky et al., 1998). We have integrated the evidence discussed above in the main text (page 5, lines 226-230):

“Previous work reported a shift in carbohydrate consumption shifting during Russian conquest (Seroshevsky, 1993), from pine sapwood flour (dendrophagy) to barley and rye, alongside increased risks of dental infection linked to smoking, particularly in Stage 4 (Macé et al., 2021). We investigated whether these changes affected oral health by analyzing the abundance of five bacterial complexes involved in biofilm formation and periodontal disease (Van Dyke et al., 2020), plus eight oral pathogens (Supplementary Information Section 2.14).”

POINT 1.12:

Line 110: Specify that shotgun sequencing was conducted.

ANSWER 1.12:

Done (on first occurrence, page 3, lines 117-120):

“Shotgun sequencing of 44.77 billion DNA templates from 627 libraries yielded genome-scale data suitable for downstream analyses in 61 males and 46 females, with a median depth-of-coverage of 2.15-fold (range = 0.017-69.85), and nuclear contamination estimates $\leq 0.5\%$ for males.”

POINT 1.13:

Line 279: Change “originated” to “originating”.

ANSWER 1.13:

The sentence has been modified as followed page 6 line 313-314 : “First, we found that the Yakut gene pool emerged through admixture between a local population with roots in the Iron Age and another group likely from the trans-Baikal region”

POINT 1.14:

Line 313: Please cite Ottoni et al. 2021, in addition to Quagliariello et al. 2022, and contextualize the results of this study accordingly.

ANSWER 1.14:

We agree with the reviewer that the results presented by Ottoni and colleagues (2021) align with those presented in our study. Briefly, these authors report that the Neolithic transition from hunting/foraging to farming was not associated with drastic shifts in the overall composition of the human oral microbiome. Detectable changes were only found for one species, *Anaerolineaceae* bacterium oral taxon 439, at the strain level. While the composition of Paleolithic, Mesolithic and Neolithic oral microbiomes were largely similar, that of modern individuals stood as an outlier, suggesting that modern lifestyles (including the use of antibiotics), more than dietary preferences, drive the composition of oral microbiomes. The suggested reference was added, and the discussion slightly expanded as follows (page 7, lines 373-376):

“This is also in agreement with other studies reporting stable oral microbiomes over deep phylogenetic scales, including between Neanderthals and anatomically modern humans (Yates et al., 2021), and during the Neolithic transition from hunting and foraging to farming (Ottoni et al., 2021).”

POINT 1.15:

Line 655 in the Main Manuscript states that the cutoff used for determining if a sample is “positive” for a pathogen was “a minimal number of 100 high-quality alignments and average edit distance was inferior or equal to 0.01” whereas the Supplementary Methods state that “samples with more than 100 reads and a mean distance below 0.1 were considered positive”. Please clarify.

ANSWER 1.15:

We apologize for the confusion. We confirm that the parameter used was 0.01, and not 0.1. The parameter value has been edited in the Supplementary Information.

POINT 1.16:

Figure 2A: This figure would look neater if the labels for non-Yakut samples were removed.

ANSWER 1.16:

We agree with the reviewer and have removed the sample labels for all non-Yakut samples (we have also changed the scale of the y-axis, following one of the requests from another reviewer). See ANSWER 2.9, explaining why the two Neolithic and the single previously published Iron Age individual are also annotated in this figure. The resulting figure looks as follows:

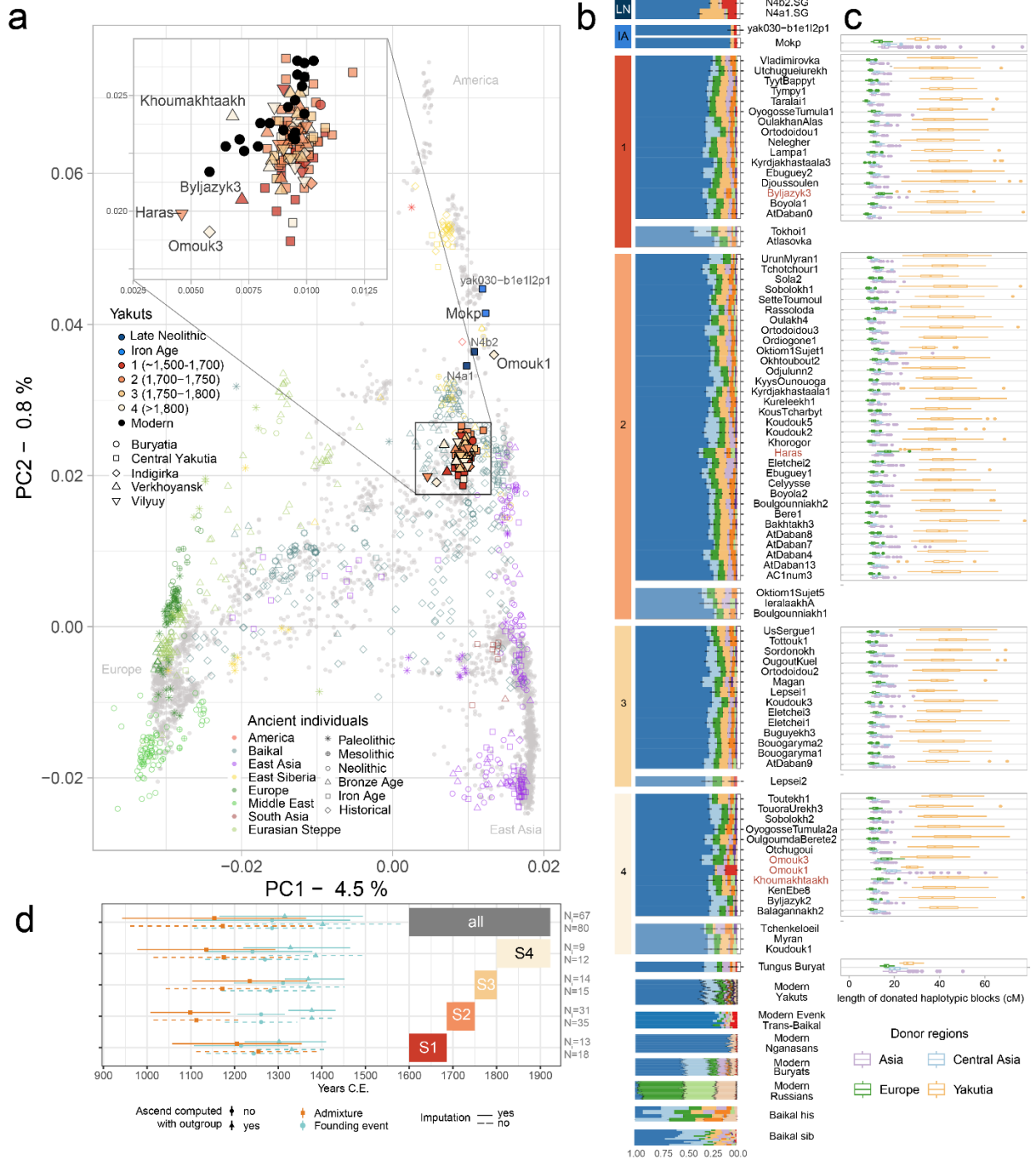


Figure 2: Population structure and history.

a) PCA conducted on 597,573 SNPs for 2,761 modern Eurasian and American individuals, with 892 ancient Eurasian and American individuals and 106 ancient Yakut individuals projected. For clarity, all the modern individuals have been removed on the zoom panel, except the modern Yakuts (black dots). The percentage indicated on each PC reflects the explained fraction of genetic variation. **b)** Unsupervised ADMIXTURE²⁶ genetic ancestry profiles ($K=10$) calculated for 327,582 unlinked SNPs and, 3,639 Eurasian modern and ancient individuals. Low-coverage individuals (<0.35 -fold) are shown with transparency. In addition to the newly published Yakuts, two Late Neolithic (LN) and one Iron Age (IA) Yakuts from²⁸ are shown. **c)** Length of haplotypic blocks donated by modern individuals (Asia refers to East, North-East and South-East Asian populations) to imputed ancient Yakut individuals (coverage ≥ 0.35 -fold; $MAF \geq 0.01$). Genetic outliers

discussed in the text are in red. **d)** Time estimates for admixture and Yakut founder event (generation time=29 years). N_i and N refer to the number of imputed and non-imputed individuals, respectively, used in the analyses.

POINT 1.17:

In Figure S1_11, the caption states the mean distance is calculated as “mean distance of per-individual reads aligned against the pathogen’s genome to the reference genome”. Can this be rephrased to add clarity?

ANSWER 1.17:

We apologize for the confusion. For each alignment, we have calculated the ratio between the edit distance (ie number of mismatches to the orthologous reference sequence), and read length. Our calculation, thus, measures the relative distance at the read level. The figure shows the average of these calculations across all the high-quality alignments identified in each of the Yakut individuals with reads aligning against the smallpox genome. We have now edited the legend of Figure S2_20 (formally M1_11) to: ‘Meandist is the read-to-reference edit distance normalized by read length. Nbreads is the number of per-individual reads aligned against the pathogen’s genome.’.

POINT 1.18:

Supplementary Table A is difficult to digest in its present form. It would be better if it were broken down into multiple tables – for eg. one table for the archaeological metadata, one for the lab-work and sequencing, and another for the results of the human DNA mapping.

ANSWER 1.18:

We agree with the reviewer and have now broken down the main table previously submitted into 3 individual tables, as suggested. We have followed this principle and have extended the number of Supplementary Tables to a total of 22 (Tables S1a-S1v).

Supplementary Table 1a. Metadata describing the newly sequenced samples and summarizing information on sequencing efforts, error rates, coverage and haplogroups of uniparental markers.

The table indicates the broad type of analyses conducted for each sample (metagenomic, genomic, epigenetic), their geographical and chronological labels, and the type of tissues sequenced. It also provides information human (mito)genome mapping, contamination estimates and mitochondrial and Y-chromosomal haplogroups.

Supplementary Table 1b. Sequencing efforts, genome coverage and post-mortem DNA damage estimates.

The table breaks down the total number of sequencing read pairs (Pairs) and/or single reads and/or collapsed read pairs per DNA library. The type of material extracted is indicated, including petrosal bones (petrous), bones, teeth and calculus. Estimates for the post-mortem DNA damage parameters defined in the mapDamage2 (Jónsson et al., 2013) statistical model are reported. These include dd (the rate of Cytosine deamination within double stranded DNA) and ds (within single stranded overhangs), calculated using a random selection of up to 50,000 reads across the genome. The first number indicates the median of the posterior distribution. It is followed by the 95% confidence range, shown between parentheses. Estimates based on less than 50,000 reads are highlighted in grey. C>T and G>A nucleotide mis-incorporation rates at the first (5) and last (3) positions within each DNA template are estimated with PMDtools (Skoglund, Northoff, et al., 2014), conditioning or not at CpG dinucleotide sites. Estimates are calculated for a random selection of up to 100,000 read alignments against chromosome 1. PE: Paired end sequencing. SE: Single end sequencing. !HQ: High-quality and PCR duplicate filtered.

Supplementary Table 1c. Archaeological information on the Yakut individuals presented in this study.

The table summarizes the items of material culture and social status of the various individuals analyzed in this study. The position of items inside (interior) or outside (exterior) the coffin is indicated, together with the presence and type of offerings and/or animals also buried, such as horses.

POINT 1.19:

It would be good to include the sequencing statistics on a per sample type basis, so that readers can see what information could be recovered from bone vs tooth vs calculus. It would also be good to have a column showing the average length of reads, % damage at terminal bases, etc.

ANSWER 1.19:

We agree with the reviewer. The requested statistics are now provided in Supplemental Supplementary Table 1b.

POINT 1.20:

Throughout the manuscript and supplementary methods, please ensure that “MetaPhlAn4” is written correctly (not MetaPhlan4).

ANSWER 1.20:

Done, for each and every ‘MetaPhlAn4’ instance.

Referee #2 (Remarks to the Author):

The study by Crubezy et al. is focused on the effects of the Russian colonization of Yakutia on the genetic makeup, microbiome, and health of the indigenous population. While West European colonialism receives a lot of attention from many disciplines, including archaeogenetics, the Russian colonialism remains understudied and even ignored, which makes the study timely and important.

POINT 2.1:

Major issues:

My main concern is that the manuscript just presents the findings, but they are not discussed in a broader context of colonial processes from various parts of the world and historical periods. Even previous archaeogenetic studies in Yakutia are not discussed much in the paper. Such a discussion is needed since the findings are rather surprising, showing no genetic admixture with Russian colonists and even no effects of the new diet on oral microbiomes and of the new lifestyle on DNA methylation patterns.

ANSWER 2.1:

The reviewer is right that the previously submitted version of our work (1) did not elaborate on previous archaeogenetic results in the region, and (2) did not compare the specificities of the colonization process of Yakutia.

Regarding (1), we now provide in Supplemental Information Section 1.1.10 a summary of previous archaeogenetic work in Yakutia. Previous findings include the identification of founder event(s) from uniparental marker data (Zerjal et al., 1997; Pakendorf, 2007; Zlojutro et al., 2009; Fedorova & Barashkova, 2022), the timing of which remained unprecise for maternal lineages, or loosely associated with the Great Mongolian expansion for paternal lineages. The contribution of local populations to the genetic makeup of Yakuts also remained debated. The present work not only confirms that the historical Yakut population formed following a single, strong founder event that occurred between ~1,200-1,400 CE, but also establishes the homogeneity of the Yakut genetic makeup following foundation, and until the present day. The extent of admixture with other populations, including local Evenks, or Russian colonizers, was therefore limited.

Regarding (2), our findings support colonization models in which the autochthonous population was not replaced by the colonizers, but was instead mobilized as a working force that could hunt the extremely rich local zibeline resources and, thus, could feed an economically-wealthy fur trade. Additionally, historical sources (Jochelson, 1933) have mapped the expansion of Russian colonizers no further north than the alluvial plains of the Olyominsk region, in southern Yakutia. Mortality rates indeed remained rather high until the 20th century in the region, including as a consequence of epidemiological outbreaks (Kabanovitch, 1952). The hard climatic conditions further north of the Olyominsk region were also not favorable to the development of an agriculture mainly based on cereal cultivation. Combined, the Russian power could not easily establish thriving colonies, and privileged instead a system of local alliances with Yakut leaders (Supplementary section 1.1.9). This colonization framework stands in striking contrast to other systems, such as those in the Americas (ie. settler and plantation colonialism), or India (ie. indirect rulers) (Crosby, 2009). The conquest of Yakutia was largely deprived of colonial settlements, and those Yakut leaders in charge of the gathering of furs, as all Yakuts, became direct subjects of the Russian Tsar, and thus, could bypass the local power and request support from authorities in Saint Petersburg, or Moscow, depending on the exact time period.

We notably added in the Supplementary section 1.1.9, maps showing the evolution of the number of Yakut tribes and Russian settlements, reproduced below with its legend:

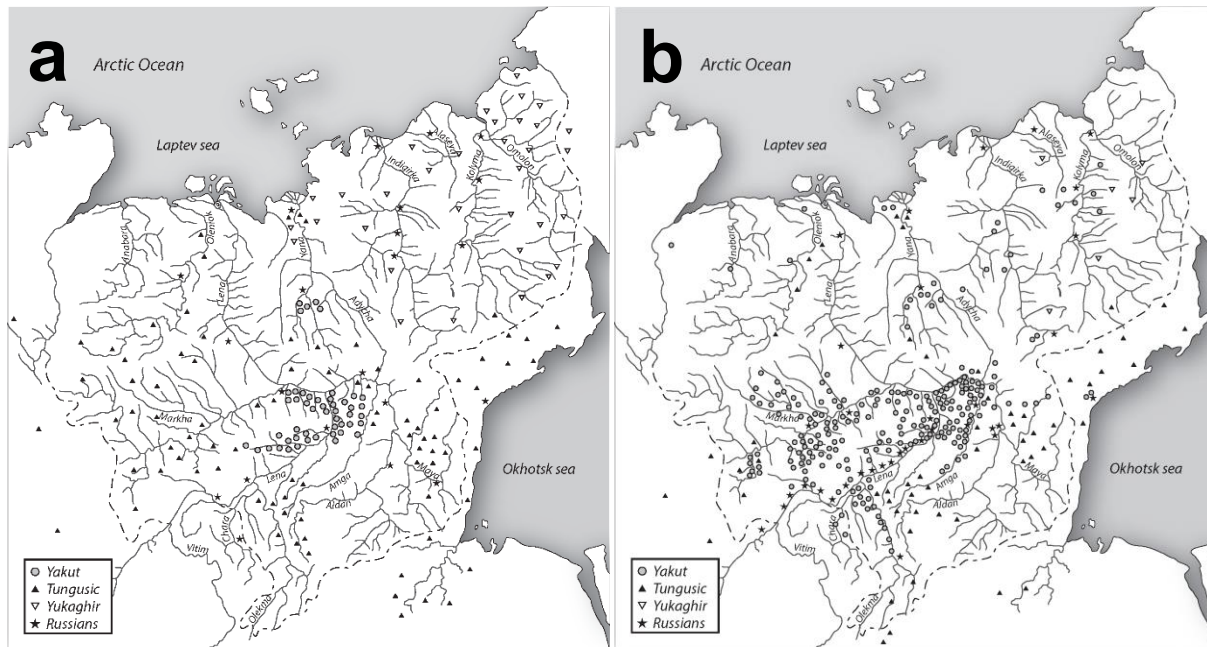


Figure S1_1. Maps showing the expansion of tribes and Russian settlements, according to the censuses of 1676 (a) and 1897 (b). We can observe that, except for a few forts during the period in question, Russian settlements always remained to the south of Yakut settlements. According to Jochelson, W. (1933) *The Yakuts*. The American Museum of Natural History. Volume XXXIII, 225p (Jochelson, 1933).

In addition to developing Supplementary sections describing the above, the main text of the revised manuscript provides more context to our findings, including with specific reference to previous work and colonialism. This reads as follows:

(page 7, lines 348-359)

“Therefore, our work portrays Yakutia as a middle-ground (White, 2010) where Yakuts and Russians engaged in an economic partnership rather than a site of military or demographic conquest (Crubézy & Nikolaeva, 2017b). Historical sources highlight Yakut peoples as pivotal to the success of the then-expanding fur trade (Crubézy & Nikolaeva, 2017b), which not only motivated Russian conquest (Fischer, 1943; Forsyth, 1992), but also provided prominent Yakut clans with opportunities for considerable wealth and power (Zvéni gorosky et al., 2020). Additionally, the harsh environment, often lethal to European settlers and unsuitable for large-scale cereal agriculture, hindered the development of demographically-thriving Russian colonies (Kabanovitch, 1952) (Figure S1_1). Consequently, Russian expansion in Yakutia relied heavily on local populations, representing a form of colonialism distinct from the settler colonialism of the Americas, which proceeded through Indigenous displacement and eradication (Crosby, 2009). It also diverged from the indirect rule characteristic of colonial India, where colonial authority was maintained through local leaders without administrating the territory. In contrast, Russian rule actively controlled taxation, and granted citizenship to local Yakut elites, integrating them as full subjects of the empire.”

POINT 2.2:

I am not an expert on microbiomes and methylation, and I cannot judge the robustness of these negative findings. Concerning human genetics, I do not question the main analytical approaches (based on both allele frequencies and autosomal haplotypes) and the treatment of uncertainties, but the selection of samples for the study raises questions since Christian burials were mostly left out (see below). I am afraid that random archaeogenetic sampling across cemeteries could have uncovered more extensive genetic admixture between Yakuts and Russians. I suggest that the authors describe their sampling approach in much greater detail than

few sentences in the supplement, provide a map of both sampled and unsampled cemeteries in the region (and maps of individual large cemeteries), and acknowledge these issues in the manuscript.

Lines 90-91: “Human tombs displaying obvious external Christian attributes were deliberately avoided while traditional funerary constructions were targeted”

This is potentially a problem since by design the study is not based on an unbiased sample of the local population, and thus the level of Russian-Indigenous cultural and genetic admixture may be underestimated.

And Russian cultural influence has direct consequences for oral microbiomes and for pathogen prevalence.

As I understood from Suppl. section 2.2, this avoidance of obviously Christian burials (although not complete according to Fig. 1c) affected the sampling of chronological stage 4 (19th century and later) and maybe of stage 3 (2nd half of the 18th century) since hallmarks of Christian burials are simply absent in earlier periods.

Anyway, that is a problem that should be discussed more in the main text.

ANSWER 2.2:

Agreed. At the archaeological level, we have now prepared a new Supplemental section (“**Section 1.2 Archaeological context: temporal stages and sampling**”) presenting the archaeological context, and especially detailing the diversity of funeral complexes identified across the four Archaeological Stages, as well as a companion manuscript (<https://doi.org/10.5281/zenodo.15017495>) providing several maps of the investigated cemeteries. The Supplementary Information Section 1.2 section also provides the rationale underlying expedition and excavation design, as the project started and further developed. It also displays a map of the surveyed areas by the Archaeological team, reproduce below with its legend:

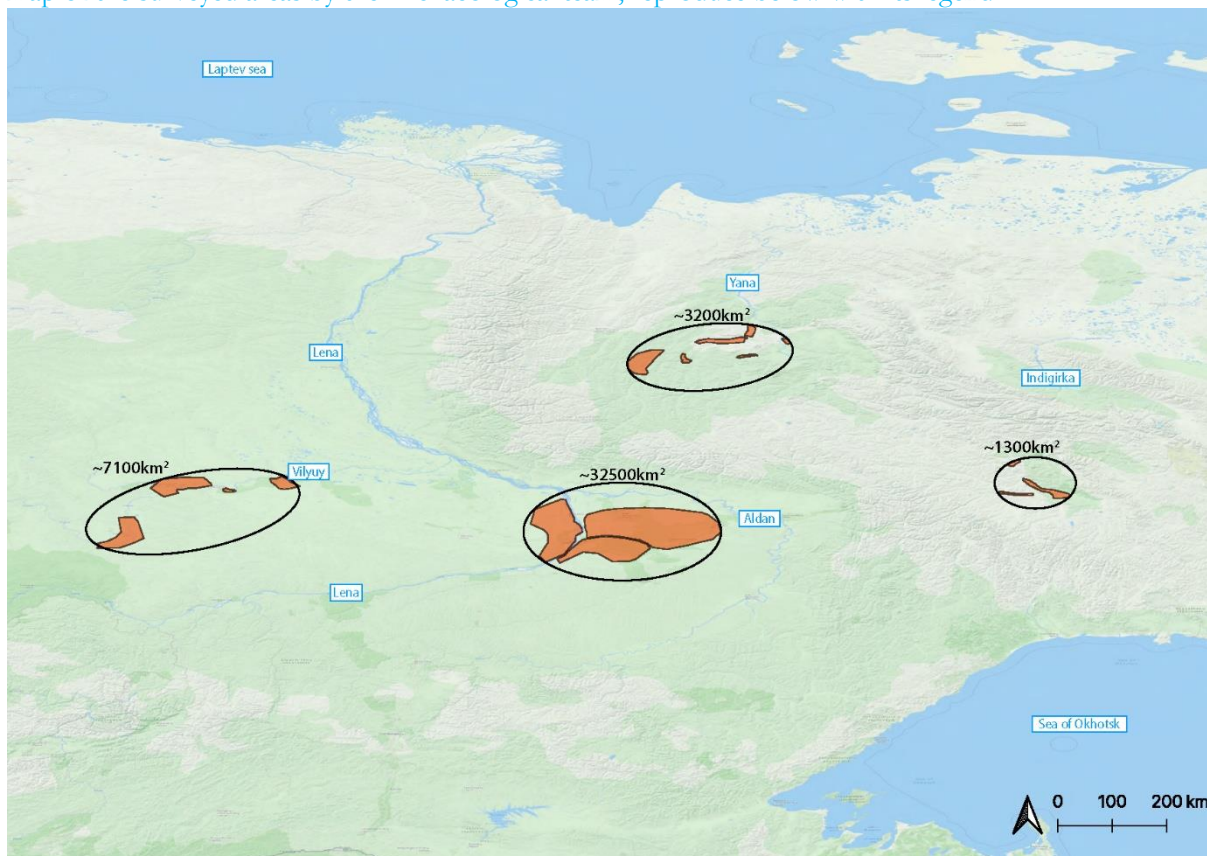


Figure S1_2. Map of Surveyed Areas in the Republic of Sakha. Prospected zones are in orange, with cumulated area in ellipses. Historical texts indicated Yakut settlements in central Yakutia, particularly along the Lena River, Vilyuy, Indigirka, and Verkhoyansk, within the Yana River basin. Analyzing old maps and historical documents has identified key early settlement areas. Experienced researchers have employed a systematic approach in exploring all these regions on foot to locate potential burial sites.

While our experimental design largely dismissed Christian cemeteries in the field, it is important to underline that the material analyzed for Archaeological Stage 4 encompasses all the Christian funerary assemblages known to date, including those associated with visible attributes reminiscent of European beliefs (eg. stelae found at Urekh2 and Urekh3), or those that yielded Christian crosses after excavation (Omouk1 and Omouk3, OulgoumdaBerete2, and OyogosseTumula2). The genetic fraction of European ancestry is comparable between the subjects from Archaeological Stage 4, and those from earlier Archaeological Stages, supporting a minimal impact of the sampling scheme in the genetic results obtained. We, thus, conclude that the strong male-bias associated with the burials from Archaeological Stage 1 (only a single subject, Atlaskovka, turned out to be a woman) did not impact our genomic analyses, likely due to the fact that they spun a whole array of economical and/or social statuses (as demonstrated through the grave goods associated, and the diversity of funerary practices). Additionally, and until Archaeological Stage 4, it is true that most Yakuts were not buried, but deposited on open-air platforms, leaving only a subfraction of the population in the funerary assemblages analyzed in this study. Those individuals buried, however, also comprise an entire diversity of economic and social statuses, ranging from small group leaders, shamans, and suicides, to major Clan leaders. Previous isotopic work also demonstrated that the diet of the archaeological subjects analyzed here is on par with that described in historical sources, such as ethnographic reports across Yakutia (Romanova et al., 2019). Overall, the archaeological sampling strategy successfully represented most of the main layers of the Yakut societies, limiting bias in the analyses presented.

Moreover, the Reviewer suggests that the preferential sampling of tombs without Christian attributes may have favored the sampling of ancient Yakuts with low levels of West Eurasian-related genetic ancestry, biasing our results. We would like to point out that if West Eurasian-related ancestry was higher in Christianized ancient Yakuts, and if this social stratum contributed substantially to the gene pool of the next generations, then present-day Yakuts should show detectable levels of Russian-related ancestry. However, we found that the 20 present-day Yakuts studied here show very low proportions of West Eurasian-related ancestry (mean = 5%, range = [0.2% - 12%]). Notably, some of the authors of the present study have recently generated genome-wide data for an additional sample of 250 Northern Siberians and have found, again, that West Eurasian genetic ancestry is ~4.4% in modern Yakuts (SOURCEFIND estimates; unpublished data), confirming the results presented here. Nevertheless, we agree with the Reviewer that we cannot totally exclude that substantial admixture between Yakuts and Russian settlers did occur during the studied period, but was not captured by our sample. We have now revised several sentences in the main text, accordingly (page 7, lines 341-347):

“Since the Yakuts traditionally deposited their deads on open platforms rather than burying them, the extent of Russian admixture within the broader population not represented archaeologically remains unknown. However, we note that the individuals buried span a wide range of economic and social statuses, from small group leaders, shamans, and suicides to major clan leaders, suggesting minimal sampling bias. While the extent of Russian admixture in Christian cemeteries is also uncharacterized, the Yakut genetic pool appears stable throughout all archaeological stages, and into the present-day, indicating limited admixture during Russian conquest.”

POINT 2.3:

Another source of archaeological bias in sampling of the ancient genetic landscape is beyond researchers' control: that is the high frequency of tree burials (apparently leaving no preserved remains) during the chronological stage 2, as mentioned in Suppl. section 2.2. This bias may affect genetic results on the ancient Yakut community since apparently only members of the elite were inhumated, as also illustrated by frequent finds associated with wealth, leadership, and shamanism at this stage (Fig. 1c).

ANSWER 2.3:

See ANSWER 2.2. We apologize that the previously submitted manuscript was unclear about the social status of the subjects analyzed. They do not belong to societal elites only, but rather encompass a whole range of economical wealth, as assessed from grave goods and funerary practices. This was, for example, summarized in the 46 archaeological each coded systematically for each subject, along with their precise location and detailed

typology - e.g., type of coffin or chest, the significance of silk in clothing, etc.- (tabulated in Supplemental Supplementary Table 1c). The diversity of subjects analyzed is now fully apparent in the 150 pages submitted to Zenodo (*Beyond the Permafrost: Excavated Tombs of Eastern Siberia. Archaeological data from the French-Sakha missions* <https://doi.org/10.5281/zenodo.15017495>), providing a synthetic description of 113 graves (including individual and multiple graves), and documenting all of the associated material with photographs. As an illustrative example, individual Eletchei3 (Archaeological Stage 4) was associated with no single grave goods and was found wearing only one Iron belt buckle. He was buried in a half-trunk coffin, as were AtDaban4, who displayed food offerings, weaponry (knife, sword, and bow), and a saddle. The following page replicates the documentation provided for individual Eletchei3 in the Zenodo companion manuscript.

Name: Eletchei 3 - (ELET3)

Location: Central Yakutia (Uluus: Menigho-Khangalass)

GPS: 61.700193°N 131.316031°E

Period: 1,750 - 1,800 CE

Chest or coffin: squared half-trunk coffin.

Associated goods (in the chest or on the chest): nothing.

Individual: Male, 30-50 years old.

The only clothing evidence is an iron belt buckle and remnants of leather around the hips.



Iron belt buckle



POINT 2.4:

Lines 281-285: “The persistence of a substantial fraction of local genetic ancestry (62.0-84.3%; Supplementary Table 1B) contrasts, however, with both cultural and linguistic evidence pointing to a primarily central Asian turko-mongol influence. Therefore, the Yakut history provides an example of a local population merging both genetically and culturally with a group entering the region through migration.”

In the context of stepping-stone and isolation by distance landscapes, which is arguably a more accurate model of reality than admixture graphs, qpAdm generates rather imprecise estimates of admixture fractions (<https://www.biorxiv.org/content/10.1101/2023.04.25.538339v5>). That is why I would not trust this high estimate of local ancestry if it relies solely on qpAdm results. Moreover, there are no clear guidelines on choice of reference populations for qpAdm, and this choice influences estimated admixture fractions and p-values strongly. I suggest that the authors either discuss these caveats or estimate local pre-Yakut ancestry using alternative methods.

ANSWER 2.4:

The reviewer suggests that the local genetic ancestry proportions in the Yakuts are overestimated because, under a stepping-stone and isolation-by-distance models, *qpAdm* generates biased estimates of admixture proportions. Nevertheless, several pieces of evidence support the view that our admixture proportion estimates are unbiased: (1) Archaeological records and oral traditions strongly suggest that Yakuts are descended from a population that migrated from the Lake Baikal region (Okladnikov, 1970; Crubézy & Nikolaeva, 2017b), in agreement with the pulse admixture model between previously isolated populations assumed by *qpAdm*; (2) ADMIXTURE estimates of local ancestry proportions were similar to those estimated by *qpAdm* (57-78% vs 62.0-84.3%), even though the former analyses are unsupervised (and, thus, do not depend on the definition of source populations), and; (3) All admixture proportion estimates obtained with *qpAdm* were within the [0 – 1] interval, whereas estimates outside these bounds are expected under the Isolation-By-Distance model from Harney et al. (Harney et al., 2021).

POINT 2.5:

Minor comments:

I am glad to see a catalogue of archaeological descriptions for burials (Suppl. section 2.3) and photos of some artefacts in Figs. 1d and 4d since this crucial information is usually omitted from archaeogenetic studies. I recommend adding also photos/plans of each burial and associated artefacts, if possible, as this information is valuable for archaeologists. I understand that complete documentations of burials are often published in specialized Russian-language journals, but since they are poorly accessible to the international community, I see value in republishing. Moreover, reproduction of images without explicit permission is allowed under the Russian law if it is done for scientific purposes.

ANSWER 2.5:

We thank the reviewer for their suggestion. MAFSO excavations in Yakutia took place between 2004 and 2019 and have revealed an extensive archaeological material that could not be fully presented in the two monographs published in 2007 (Crubézy and Alexeev, 2007), and 2012 (Crubézy and Alexeev, 2012). We have, thus, prepared a companion catalog that compiles all the archaeological information associated with the 124 individuals analyzed in the present study. This catalog was submitted together with these revisions and is available online through Zenodo, and is led by those archaeologists in charge of field excavations, including local and non-local experts (*Beyond the Permafrost: Excavated Tombs of Eastern Siberia. Archaeological data from the French-Sakha missions*, <https://doi.org/10.5281/zenodo.15017495>).

POINT 2.6:

Line 87: “the western Vilyuy, the northern Verkhoyansk, and the eastern Indigirka”

Please mention that the former and the latter regions are river basins and indicate the type of the Verkhoyansk region.

ANSWER 2.6:

All three regions represent river basins: the Vilyuy river, the Yana river, and the Indigirka river, respectively. This is now indicated in the main text (page 2, lines 93-96):

“Excavations by MAFSO between 2002 and 2018 (Crubézy et al., 2020), teaming together Russian and French archaeologists, uncovered an outstanding archeological record across four key regions of Yakutia: Central Yakutia, and the river basins of the western Vilyuy, northern Verkhoyansk (Yana), and eastern Indigirka (Fig. 1a; MAFSO stands for the French Archeological Mission in Eastern Siberia).”

POINT 2.7:

Lines 119, 120: “using pseudo-haploid data for the 1.24 million SNPs comprised in the Allen Ancient DNA Resource”

I recommend mentioning a version of the AADR used in the paper as data and archaeological/geographical annotations are constantly updated in this database.

ANSWER 2.7:

The AADR version 5 was used. This methodological information, and the relevant citation (Mallick & Reich, 2023; Mallick et al., 2024) were added in the Main text (page 3, lines 126-128) : “Population genetic analyses were conducted on both genotypes imputed from the 1000 Genomes Project catalog (The 1000 Genomes Project Consortium et al., 2015), and pseudo-haploid data for the 1.24 million SNPs comprised in the Allen Ancient DNA Resource v5 (Mallick & Reich, 2023; Mallick et al., 2024).” and the Material and Methods section (page 11, lines 588-589): “Genotypes were downloaded from the Allen Ancient DNA Resource v5 (AADR) website (Mallick & Reich, 2023; Mallick et al., 2024).”

POINT 2.8:

Lines 125-127: PC1 scores separated European from American and Asian populations, with ancient and modern Yakuts clustering together and alongside Asian populations.

Please clarify if modern Yakuts were projected on the other modern groups or not since the shrinkage artefacts are not corrected completely in the existing versions of the smartPCA software. That is why comparing the positions of mapped and non-mapped individuals at fine scale is not recommended.

ANSWER 2.8:

The genome data from modern Yakuts were not projected but used instead for the calculation of the main Principal Components. Additionally, we have now added to the Supplementary Information a new PCA plot, which was generated in the total absence of projection (Figure S2_5). The latter showed similar individual affinities to those previously identified (Fig. 2a), ruling out any possible impact of shrinkage compensation in our analyses. This figure is duplicated below.

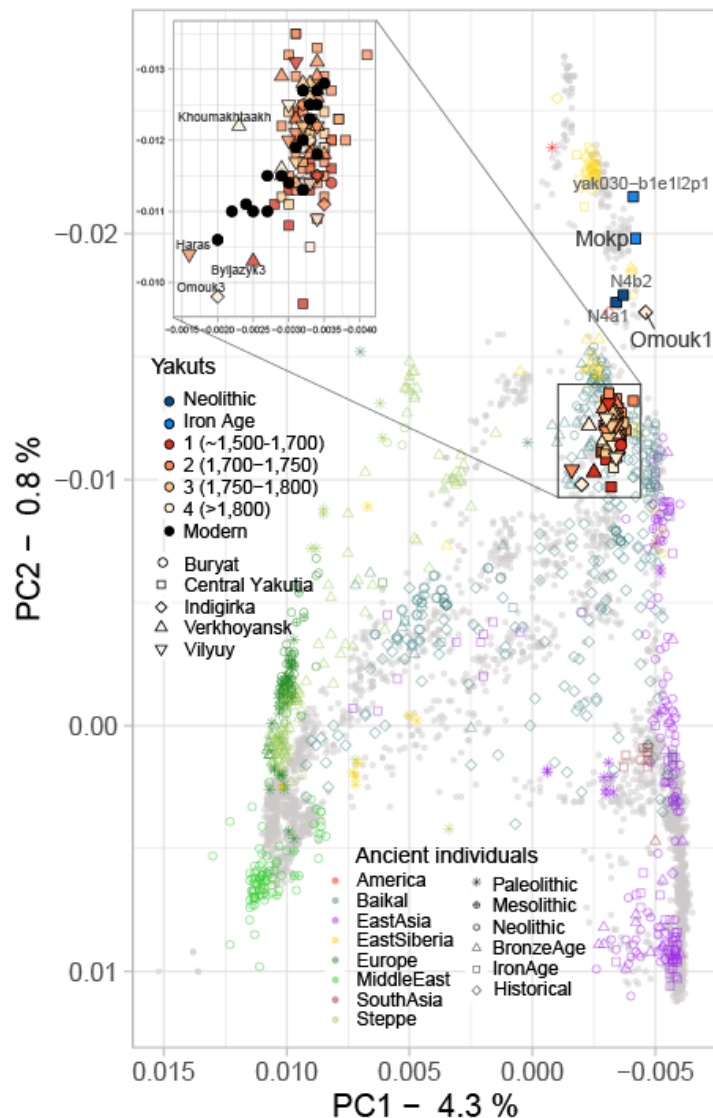


Figure S2_5: PCA performed on 597,573 SNPs for 2,761 modern Eurasian and American individuals, 892 ancient Eurasian and American individuals and 106 ancient Yakut individuals without projection. All the modern individuals have been removed in the zoom panel, except the modern Yakuts (black dots). The percentage indicated on each PC reflects the explained fraction of genetic variation.

POINT 2.9:

Lines 140-141: “Combined, these analyses suggest that Yakuts from the ~1500 CE onwards are not the direct descendants of people inhabiting Yakutia prior to ~280 BCE (i.e. the radiocarbon age of Mokp).”

The term “direct descendant” is vague, in my opinion, since children, grandchildren, etc., are direct descendants, but obviously they may look very different from their ancestor on PCA etc. I prefer to use terms like “genetically continuous” instead. And, strictly speaking, the analyses cited here do not exclude genetic continuity of Yakuts with the Iron Age population and some incoming gene flow, although we do know from other genetic analyses and from historical evidence that Yakuts migrated to their current territory relatively recently and replaced the local population at least partially, as discussed further in the paper.

ANSWER 2.9:

Previous studies characterized three Individuals pre-dating the Yakut historical expansion: yak030 from the Iron Age, and N4a1 and N4b2 from the Neolithic period (Kilinc et al. 2021). These three individuals were included in our analyses. The positions of these individuals in the PCA space are now indicated on Fig. 2A (see ANSWER 1.16), and largely overlap with that of the individual Mokp (Iron Age), analyzed for the first time in this study. This confirms a rather homogeneous gene pool in the region prior to the Yakut expansion. Historical and modern Yakut, however, do not overlap with these individuals, confirming our previous conclusions of an absence of genetic continuity in the region between the Iron Age and Archaeological Stage 1. We agree with the reviewer that ‘genetically continuous’ better refers to our findings than ‘direct descendants’. The sentence in the main text was, thus, edited accordingly. It reads as follows (page 3, lines 149-151):

“Combined, these analyses suggest that Yakuts from ~1500 CE onwards are not genetically continuous with the populations inhabiting Yakutia until ~280 BCE (the radiocarbon age of Mokp).”

POINT 2.10:

Another caveat to be considered in the paper is that two Iron Age and two Late Neolithic individuals (strangely, I see only one of those individuals mapped on the CPA in Fig. 2a) are not representative of the respective genetic landscapes over vast territories. That is a mistake common in the archaeogenetic literature.

ANSWER 2.10:

Our study takes advantage of all the genetic data available (and even extends it) for the region in these two time periods. We agree with the reviewer that the number of individuals (N=4) scales poorly with the size of Yakutia. Nonetheless, we do find that individuals from the Iron Age provide acceptable sources in qpAdm modelling, suggesting that some of the local genetic background was admixed with another source (likely from Mongolia) while forming the Yakut population. DATES analyses further indicate that such an admixture took place well after the Iron Age (therefore, the founding population did not exist as already admixed during the Iron Age). We do not conclude anything regarding the diversity of the local genetic pool during the Iron Age, which should be the focus of future paleogenetic studies in the region. See ANSWER 2.9. We have nuanced the main text to reflect the scarcity of the possible sources available in the region (page 6, lines 318-320):

“The precise contribution of the local population requires further investigation, as current models rely on a limited number of Iron Age individuals, likely failing to fully capture the genetic diversity across the vast territory of Yakutia at the time of admixture.”

POINT 2.11:

Lines 143-144: “and an ADMIXTURE profile resembling that of the Trans-Baikal Evenk (Fig. 2b)”

I don’t see any modern Evenk individuals analyzed in Fig. 2b.

ANSWER 2.11:

Modern Evenk individuals were in fact considered in our ADMIXTURE analyses. Due to space limitations, the version of Fig. 2B that was previously submitted, however, only included a sub-selection of the individuals analyzed, which did not include Evenks. The whole analysis is shown as Supplementary Figure S2_8. Modern Evenks are now included in the revised version of Fig. 2B, as requested. See ANSWER 1.16 for the new figure.

POINT 2.12:

Lines 160-163: “Except for rare genetic outliers (Supplementary Table 1B), the genetic makeup of all other Stage 1 to Stage 4 Yakuts could be modelled as a two-way admixture between populations inhabiting Yakutia in the Iron Age (Yakutia_IA, 62-84%), and the Baikal region between the 13th and 15th centuries CE (Baikal_his, 16-38%; Supplementary Table 1B)”

I cannot easily find the composition of these reference groups (Yakutia_IA and Baikal_his). If they include just 1-2 individuals, I find this analysis problematic.

ANSWER 2.12:

The submitted version of our work did provide a detailed description of the Yakutia_IA and Baikal_his groups. This was, however, provided in the Supplementary Information, and not directly in the Materials and Methods. We apologize for this oversight. We now provide the exact composition of both groups in the Materials and Methods section, which includes all individuals previously reported for the relevant time periods and geographic regions. We also display the group ADMIXTURE profile in Fig. 2b (see ANSWER 1.16). This information is available in the Supplementary Information Section 2.12) and Materials and Methods (page 12, lines 645-655):

“Admixture modeling and dating

Admixture models for ancient Yakut individuals (coverage $\geq$ 0.1) were assessed using the pseudo-haploid 1240K data set and qpAdm from Admixtools (Patterson et al., 2012) (version 5.056). qpAdm models were aimed at testing whether the Yakut genomic makeup was compatible with a 2-ways admixture from a local Siberian background (Yakutia_IA, N=2, comprising Mokp and yak03041), and another source, potentially from the Baikal region (Baikal_his (N=4) or Baikal_sib (N=11)), or Russia (Russian (Bergström et al., 2020)) (Supplementary Table 1f). Baikal sources were defined as Baikal_sib (N=11) and Baikal_his (N=4), with the former consisting of Mongolia_Khuvsgul_LateMedieval (Jeong et al., 2020) (N=2), Mongolia_Dornod_LateMedieval (Jeong et al., 2020) (N=7), and Mongolia_Khentii_LateMedieval (Jeong et al., 2020) individuals (N=2), and the later comprising Russia_AngaraRiver_Medieval.SG (Kılınç et al., 2021) (N=1), Mongolia_Sukhbaatar_Xiongnu (N=1) (Jeong et al., 2020) and Mongolia_Khuvsgul_MLBALateMedieval (Jeong et al., 2020) individuals (N=2).”

See also ANSWER 4.13.

POINT 2.13:

Lines 169-173: “Neither of the Yakut groups from Stage 2 to modern times showed a statistically significant excess of genetic sharedness with Europeans, including from Western Russia, relative to Stage 1 (Extended Data Fig.2d). Additionally, no Yakut group from a given archaeological stage was enriched in European-like genetic variation relative to those from previous archaeological stages (Extended Data Fig.2d).”

Looking at EDF 2d, I see that the authors compared stages to each other, but not groups within stages. Please clarify that since the text implies otherwise.

ANSWER 2.13:

We apologize for this confusion and have clarified our statement. It now reads in the main text (page 4, lines 182-185):

“Compared to Stage 1, neither the groups forming each subsequent stage, nor modern Yakuts, showed a statistically significant excess of genetic sharedness with Europeans, including Western Russians (Extended Data Fig. 3d). Furthermore, no enrichment was detected in any given archaeological stage relative to earlier stages (Z-scores <3 ; Extended Data Fig. 3d).”

POINT 2.14:

Lines 232-233: “with four cases involving first and second-degree relatives buried at the same site (AC, At Daban, Oktiom and Oulakh; Fig. 4a)”

Please spell out the abbreviation “AC”.

ANSWER 2.14:

AC stands for ‘Arbre Chamanique’, a French expression translated into the Shamanic tree. This information is now provided in the main text (page 5, lines 259-262):

“Kinship analyses identified 61 individuals genetically related up to the third degree, including four cases of first- and second-degree relatives buried at the same site (AtDaban, Oktiom, Oulakh, and Arbre Chamanique,

which translates to “Shamanic Tree” in English; Fig. 4a, Supplementary Table 1r-1t; Supplementary Information Sections 2.8).”

POINT 2.15:

Lines 247-249: “Combined, these results social practices conducive to preferential unions, contacts and microbial exchanges intra-regionally, between close neighbors.”

A word is missing in this sentence.

ANSWER 2.15:

Fixed (page 5, lines 278-279):

“Combined, these results suggest social practices conducive to preferential unions, contacts and microbial exchanges intra-regionally, between close neighbors.”

POINT 2.16:

Lines 310-311: “Russian contact, however, also introduced a wealth of new infectious agents, such as but not limited to the smallpox identified in this study, with devastating demographic consequences for immunologically naive Indigenous populations”

Does not the stable N_e estimated throughout stages 1-4 (Supplementary Table 1C) contradict this? Or did pathogens affect N_e just in the very beginning of Russian contact in stage 1? Please clarify. The bottleneck analysis in the paper is focused on the formation of the Yakut gene pool in the Middle Ages, but not on the Russian conquest and contact period.

ANSWER 2.16:

We agree with the Reviewer that there is an apparent contradiction between the historical records, indicating that the Russian conquest was accompanied by a mortality crisis in the Yakuts, likely caused by newly introduced pathogens, and the relatively stable N_e estimated by HapROH at the different stages. Considering that the N_e estimation method implemented in HapROH is accurate (see Supplement of (Fernandes et al., 2020)), it is possible that the epidemics triggered by Russian colonization resulted in bottlenecks with a sufficiently short duration or weak intensity that they were not detected by HapROH. We have now rephrased the sentence accordingly. Furthermore, we now comment the HapROH estimates in the main text, suggesting that the effects of this mortality crisis on the genetic diversity of Yakuts are limited (page 7, lines 360-365):

“Despite the absence of significant genetic impact, Russian conquest profoundly affected Yakut livelihoods and demographics. Russian contact introduced numerous new infectious agents, including the smallpox identified here and *Mycobacterium tuberculosis* (Dabernat et al., 2014), with devastating demographic effects on immunologically naive Indigenous populations (Dabernat et al., 2014). The severe outbreaks documented in historical sources were not detected in HapROH, indicating limited, but stable, effective population sizes. This likely reflects the extremely rapid progression of these outbreaks, or other factors limiting detection power.”

POINT 2.17:

Lines 596-597: “Each ancient and modern individual from Yakutia was tested for every combination of 2 or 3 populations putting the non-used population in the right group (Supplementary Table 1B).”

Please show the standard list of “right” groups used.

ANSWER 2.17:

The description of the ‘right’ populations were already provided in the Supplementary Information Section 2.12 (formally 1.8), as follows (page 40, lines 364-367):

“In all models tested, we used Mbuti (N= 4), Ami (N=15), Mixe (N=6), modern individuals as outgroup (‘right’) populations, together with Malta Ma-1 (N=1), Natufian (N=4), Turkey Neo (N=16), Villabruna (N=1) and Kostenki14 (N=1).”

POINT 2.18:

Suppl. section 2.1

“North Eastern Federal University” – please correct to “North-Eastern Federal University”.

ANSWER 2.18:

Done (and also referred to as NEFU, to avoid repetitions in the Supplementary Information).

POINT 2.19:

References 2, 21

A highly unusual transliteration of Russian is used, please use a more standard one, although there are many slightly different systems (https://en.wikipedia.org/wiki/Romanization_of_Russian).

ANSWER 2.20:

Done as follows: We therefore used the ALA-LC standard for the Romanization of Russian Cyrillic: American Library Association and Library of Congress Romanization Tables for Russian, or the Library of Congress system.

POINT 2.21:

Reference 10

The title is duplicated.

ANSWER 2.21:

The corresponding title is now fixed.

POINT 2.22:

Reference 19: Crubézy, E., Melnichuk, O., and Alexeev, A. (2020). Archaeology , Genetics and History: 15 years of research in Yakutia (2002 – 2017). *estnik arkheologii, antropologii i etnografii* 4, 110–119.

<https://doi.org/10.1128/microbiolspec.PoH-0004-2014.Th>.

Here there is a typo in the word “Vestnik” (the V is dropped), and the DOI link points to a completely different paper on smallpox.

ANSWER 2.22

Thanks for spotting these typos, which are now fixed as follows

Vestnik arkheologii, antropologii i etnografii

<https://doi.org/10.20874/2071-0437-2020-51-4-11>

POINT 2.23:

Reference 35: the doi link is duplicated.

ANSWER 2.23:

This is now fixed as follows: Crubézy, E., Duchesne, S., Razafindrazaka, H., Romanova, L., Gérard, P., Alcouffe, A., Esclassan, R., Melnichuk, O., Ushnitsky, I., Ludes, B., et al. (2021). Sucrose Is Not the Whole Story: Risk Factors and Oral Health at the Contact (Yakutia, Siberia-16th/19th). *Biology (Basel)* 10, 974. <https://doi.org/10.3390/biology10100974>.

I prefer to sign my reviews.

Pavel Flegontov

Referee #3 (Remarks to the Author):

Originality and significance: if not novel, please include reference / Suggested improvements: experiments, data for possible revision / References: appropriate credit to previous work?:

POINT 3.1:

Although bringing DNA analysis to the history of colonialism in Siberia is novel, the article could benefit from engaging with well-established scholarship in environmental history. In Alfred Crosby's work, for instance, the author may find many relevant parallels: Crosby, Alfred W. 2004. *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*. 2nd edition. Cambridge and New York: Cambridge University Press.

ANSWER 3.1:

See ANSWER 2.1. We now explicitly refer to the work from (Crosby, 2009), and discuss the specificities of the Yakut conquest relative to other well-described colonization processes. In a nutshell, both the high mortality rates measured amongst Russian colonizers (Kabanovitch, 1952), and the poor suitability of the region located north to the alluvial plains of the Olyominsk region for cereal agriculture, limited the capacity of the Russian power to develop demographically-thriving colonies in Yakutia. As a result, the underlying process could not conform to the expectations of the so-called "settler colonialism", typical of the Americas. Colonialism could also not develop through "indirect ruling", characteristic of Colonial India, as Yakuts were conceded full Russian subjects, with equal rights across the whole Empire.

The European expansion across the world was facilitated by the organisms that accompanied the European settlers: livestock, domestic crops, and pathogens. The Europeans, thus, created new ecosystems that allowed them to extend their way of life into regions inhabited by other populations. These populations were brutally affected by the pathogens: 90% of Native American populations perished due to epidemics introduced by Western colonizers (Diamond, 1997). While many studies have been devoted to these processes in America, the biological dimensions of the conquest of Northern Asia by the Russian state have remained a largely unexplored area until now. Our work contributes to fill this gap by describing new smallpox strains in historical Yakutia of likely European origins.

The main text has been edited as follows (page 7, line 352-359:

"Additionally, the harsh environment, often lethal to European settlers and unsuitable for large-scale cereal agriculture, hindered the development of demographically-thriving Russian colonies⁵⁵ (Supplementary Information Section 1.1 and Fig. S1_1). Consequently, the Russian expansion in Yakutia relied heavily on local populations, representing a form of colonialism distinct from the settler colonialism of the Americas, which proceeded through Indigenous displacement and eradication⁵⁶. It also diverged from the indirect rule characteristic of colonial India, where colonial authority was maintained through local leaders without administrating the territory. In contrast, Russian rule actively controlled taxation, and granted citizenship to local Yakut elites, integrating them as full subjects of the empire."

POINT 32:

I also found the use of the ethnonym Yakut problematic, especially as it spans several centuries in the article. As historians and anthropologists, we understand that processes of identification are fluid, often shaped by relativistic understandings of social and cultural practices, including interactions with animals and the environment. Furthermore, shifts in identity are always influenced by political and economic regimes. I would recommend elaborating on why the authors chose to use the colonial ethnonym Yakut rather than Sakha, and why they portray the migrations of groups as if their identities remain static. For further reading on identity politics in North Eurasia, you may consult the following works:

- Anderson, David G., Dmitry V. Arzyutov, and Sergei S. Alymov, eds. 2019. *Life Histories of Etnos Theory in Russia and Beyond*. Cambridge, UK: Open Book Publishers.
- Anderson, David G. 2000. *Identity and Ecology in Arctic Siberia: The Number One Reindeer Brigade*. Oxford: Oxford University Press.

ANSWER 3.2:

We agree with the reviewer that the use of a single ethnonym to characterize a culturally dynamic and diverse population throughout five centuries has obvious limitations. For example, the Yakut traditionally refer to themselves as Saha, or Uraanhai. We have developed a full Supplementary Information section describing the multifaceted Yakut society, and its temporal evolution (Section 1.1.3. On the Use of the Term “Yakut”). We, however, decided to keep using the Yakut noun for consistency and clarity. It is noteworthy that the conventional term Yakut is used in historical sources and has become one of their official names in the Republic of Saha-Yakutia (Russian Federation). Despite their distance from other pastoral populations of Central Asia, the Yakut have maintained a unique cultural identity in the Arctic region over the centuries. They are distinguished from neighboring Arctic and Subarctic groups by their Turkic language, their mythology, their hierarchical social organization, their material culture, their cavalry tactics and their pastoral techniques, which combined connect them to the peoples of the southern steppes. They introduced the horse into the northernmost regions of the world (Okladnikov, 1970; Ferret, 2010). The new section reads as follows (page 3):

“1.1.3. On the use of the term “*Yakut*”

The concept of being a “*Yakut*”, along with other ethnic groups, did not exist amongst local groups at the time of first European contact. Terms such as “Dolgans,” “Yukaghirs,” or “Yakuts” were categories enforced by the Russian imperial power as tools of control. The earliest historical references to Yakutia as a territory with internal administrative subdivisions serve as a clear illustration of this process.

The Russian detachment led by Panteleimon Demidovich Pyanda, the first to reach the Lena River in 1620, vaguely referred to “*Indigenous populations*”, merely noticing the coexistence of nomadic and sedentary lifestyles (Okladnikov, 1955). In 1630, M. Vasiliev, in his report to the Tsar, elaborated more distinctions, with Tunguses inhabiting the Chona River, Nanagir the Vilyuy River (and forming a nomadic population of approximately 700 individuals), and Dolgans and Yakuts settled along the Lena River up to the mouth of the Vilyuy. The Yakut leader was named Dygintcha, possibly a mispronunciation or an abbreviation for Tygyn. By 1633, A.F. Palizyn, the voivode of Mangazeya (a merchant colony recently founded in northwestern Siberia), sent a report to Moscow further describing the Lena territory, indicating that the inhabitants “*belonged to no one*”, and recommending the establishment of an administration for a “*new Mangazeya*” to control the area and explore its surroundings (Ivanov, 1974). In 1634, the state requested additional information from the voivodes of Yenisei, Tobolsk, and Mangazeya, who confirmed that the Yakuts were under “*no jurisdiction*” (Crubézy & Nikolaeva, 2017a).

Between 1635 and 1639, the expedition of the “boyar’s son”, Voin Shakhov, collected yasak (tribute) at the mouth of the Vilyuy River from various groups. These included the Tagusy Yakuts, led by Chief Idykhtek, who had 50 men, as well as the pedestrian Yakuts of the Osugei, Kirik, Kokkui, and Kirikin clans—representing 125 Yakuts in total. They also took tribute from 12 members of the Tungus clans and 20 individuals from a Dolgan clan led by Chief Dygintcha (Crubézy & Nikolaeva, 2017a). This diversity of groups with distinct lifestyles challenges the idea of a unified nation in early 17th century Yakutia.

Moreover, it is highly likely that many fur trappers were attracted into the region during the fifteen years separating Pyanda’s (1620) and Shakhov’s (1635) expeditions. Local groups in this period started to appear as distinct yet interconnected entities, receiving a diversity of names, such as the “*Tagusy Yakuts*” and the “*pedestrian Yakuts*”. In fact, the term “*Yakut*” is thought to originate from the Tungus reindeer herders, who referred to them as “*eko*”, pronounced “*iako*”. The Cossacks transformed this to “*Yakut*” by altering the initial “*e*” sound, which is rarely found at the beginning of Russian words, to the more common “*Ya*”. The suffix “*out*” (“*ut*”) was subsequently added, as seen in other Siberian ethnonyms, such as the Telecuts and Lamouts.

Like for other populations, a single term—“*Yakut*”—cannot encapsulate the complex, heterogeneous, and dynamic populations inhabiting northeastern Siberia before and after the Russian conquest (Anderson, 2000; Anderson et al., 2019). It has, however, the practical advantage of improving readability and consistency throughout this study, as opposed to changing names for each period considered, as long as it is understood that the underlying group only started to become increasingly unified from 1689/1700 CE, and encompassed a diversity of cultural traits that changed through time. The individuals from the earliest phase, predating

1689/1700 CE, shared distinctive cultural characteristics such as horsehide clothing (as opposed to local hunter-gatherers, who predominantly wore reindeer hides) (Jochelson, 1926).

On April 27, 1922, the Yakut Autonomous Soviet Socialist Republic (YASSR) was established. Following the collapse of the Soviet bloc, YASSR adopted the status of a Sovereign Republic in 1990 that the Yakuts decided to call the “Sakha Republic (Yakutia)”, to emphasize reference to their own language, sakha tɬɬa. By extension, the term “Sakha” also refers to the people speaking this language.

Today, the region is officially known as the Republic of Sakha (Yakutia), commonly abbreviated as RS(Y). While Yakuts represent the most abundant fraction of the population, Sakha citizens include a more diverse cultural and ethnic mosaic, encompassing Yakuts, Slavs, and members of other Indigenous groups.”

POINT 3.3:

Another issue I found problematic is the assumption that alcohol was introduced through Russian colonialism and Chinese/Mongolian trade. Ethnographic studies of southern Siberian Indigenous groups (Altai, Khakas, Buryat, Tyva, and others) suggest that the tradition of making alcohol from milk is likely Indigenous. This practice appears to be rooted in animal husbandry rather than agriculture, as is the case with vodka, which is made from wheat, potatoes, or other agricultural ingredients.

ANSWER 3.3:

Many Southern Siberian groups did produce alcohol before colonization by distillation of fermented milk, however this technique had been long ago abandoned by the Yakut when the Russian settlers arrived in 17th century (Okladnikov 1970, p.242). We have now clarified our statement: “introduced tobacco, vodka and...”. (page 2, lines 74)

POINT 3.4:

Ethnographically, the use of the term exogamy, as discussed in the article, would benefit from additional references that consider its relationship to territorial identity or neighborhood, as this understanding may differ from the classic kinship-based approaches in anthropology.

ANSWER 3.4:

The reviewer is right that clan exogamy does not imply geographic distance. For example, one single village can be composed of two or more exogamous clans, in which case people respect exogamy when marrying someone from their own village but from the other clan. This pattern is well-documented in numerous kinship-based societies worldwide (Fox, 1967; Levi-Strauss, 1969). Regarding the Yakut, Vitashevskii (1910) states, “the poor take a wife from the same district as long as she is from a different *aga unsa* (patrilineal clan).” (Vitashevskii, 1910). This suggests the existence of exogamic practices.

The fact that we have “preferential unions, contacts and microbial exchanges intra-regionally, between close neighbors” is therefore not an argument against strict exogamy. Rather, it underscores that exogamy operates at the level of kinship structures rather than geographic distances. On the other hand, the study does demonstrate non-strict patrilocality.

However, the study does provide compelling evidence for non-strict patrilocality, a phenomenon also noted in other Eurasian pastoral and semi-pastoral societies, where economic and environmental factors influence post-marital residence patterns (Khazanov, 1994; Humphrey & Sneath, 1999). Further comparative research on Yakut kinship and spatial organization, drawing from sources such as Middendorf (1878) and Popov (1964), could provide additional insights into how exogamy and patrilocality intersect in this cultural context (Middendorf, 1878; Popov, 1964).

We now have modified the abstract, as followed:

“ Social organization remained mainly structured around patrilocal clans showing strong ties to the land of their ancestors, and maintaining low consanguinity – except for one female with the latest markers of traditional shamanism, who was the daughter of second-degree relatives.”

Referee #4 (Remarks to the Author):

Summary

The manuscript of Crubézy and coauthors presents the analysis of genetic data of 107 individuals, 21 methylation datasets, samples of calculus analyzed for oral microbiome and 2 soft tissue samples. Population-genetic, pathogen, microbial and methylation patterns have been investigated and compared with markers from material culture. These individuals were uncovered during a French-Russian archaeological expedition in the Yakuts region and the individuals are associated with the local Yakut population. They were dated to historical periods between 16th and 20th century and they were chronologically aimed to predate and postdate historically-documented Russian conquest of the region. The results show no major differences between the 4 periods investigated and in particular no changes in population-genetic, microbial or methylation patterns that can be attributed to Russian conquest (even though it is linked to changes of lifestyle and diet). Based on the interpretations presented, Yakut population was formed in one discrete event in the 13th century by immigration of Baikal-related populations to the region and by admixture with the locals that form the majority of their ancestry. The marital habits of the population are described as non-strict patrilocality with strong ties of the population to their land. That is because methylation, genetic and oral microbiome patterns do not seem to be linked to chronology, social status, assumed religion or wealth but they do seem to be linked to differences between regions.

The genetic analysis of indigenous populations from the Yakutsk region is an important pursuit as such groups are generally underrepresented in both modern and ancient datasets. Additionally, these groups are also often understudied in terms of archaeology, rarely appear in historical sources and genetic research thus provides a crucial perspective on their, even quite recent, past. The research strategy focusing on recent ancient individuals associated with these groups is sound - it helps to circumvent issues related to large-scale demographic events in the 20th and 21st centuries in the region and by sampling individuals that can be divided to a period before and after Russian conquest of the region, the effects of inter-population contacts can be studied. Additionally, the results can be linked to a wealth of archaeological information obtained during the excavations. The whole-genome data from this region will have high value for future studies, that is especially the case for higher coverage individuals (e.g. a sample beyond 65x coverage if I read right on the line 112). It is also valuable that microbial, methylation, pathogen patterns are studied jointly with population-genetic results and material culture. The study highlights limits of assessing known historical events with biomolecular methods as Russian conquest is practically invisible in the data and showcases a very detailed view into a life history of certain individuals e.g. female shaman or an individual interpreted as a founding female.

As such, the manuscript will be of interest to researchers in various disciplines studying this region, in particular scholars from genetics, archaeology, anthropology, biology and history, but also those interested in comparing various sources of data and those interested in the impact of colonial powers on an Indigenous society. The disciplines are well connected in the text, often describing archaeological and genetic features of the burials jointly. The various analyses are also relatively well connected and the manuscript is easily readable. In my understanding, the manuscript also speaks respectfully about Yakuts, their lifestyle and their history.

POINT 4.1:

The sampling and joint French-Russian research seemed to have been done ethically and respectfully and several co-authors are Yakuts though it is not clear if the wider Yakut community has been consulted regarding the results. Community engagement is important here not only because the study could be misused as to invalidate experiences of Yakuts and other indigenous populations during Russian (or other colonial) conquests, but also because direct descendants of the studied people could be traceable via IBD (see both of these in more detail in the major points mentioned below).

ANSWER 4.1:

The research teams developed a wide array of initiatives to engage the Yakut community over close to two decades. These include: local TV interviews; Whatsapp campaigning; lectures; meetings with local associations; publication of popular science books in both Russian and Yakut languages; TV documentaries; blogs; media interviews, and; three exhibitions, the first two of which consisted of photographs taken during field expeditions, and were inaugurated at the capital city of Yakutsk before touring in the two most populated regions of Yakutia (in 2015, and 2020); and another one as part of the CNRS 80th anniversary (2019), presenting in the new museum of Yakutsk a total of 100 artifacts excavated by MAFSO. We have summarized the broad range of such activities in a new Supplementary Information section entitled “1.3 Ethics and engagement with Indigenous Communities” (page 15-18).

Please note that the Supplemental section aimed for clarity, and not self-promotion. It, thus, does not provide an exhaustive list of the several dozens of initiatives developed between 2004 and 2020. We believe that the illustrated diversity of our actions, however, demonstrate full engagement with local communities.

Methodologically, I believe that the number of samples was relatively modest for a large period of history and region covered and I think some facets of assessment of the social structure were thus overinterpreted (see major points). Altogether, I however like the study and I think the study could be suitable for publication in Nature pending revisions.

The major points:

POINT 4.2:

1) I have a few ethical issues that I hope can be remedied. I would suggest including a section in the SI discussing these. In general, it is often not straightforward how to proceed with the analysis of data attributed to individuals from Indigenous communities whose agency regarding their past and the archaeological and anthropological remains that are associated with them might have been taken away. That is even more the case if these are recent individuals whose direct relatives could be still living and genetically traced through open access to such data. In this case, there was a longterm collaboration with researchers from the region of interest and several co-authors are Yakuts. By my (limited) assessment, care has been taken to represent this population in the manuscript carefully and archaeological remains have been described in detail. However, it is unclear if communities have been contacted prior, during and/or after the study. This should be remedied by the co-authors that are in the region (though this might be difficult to do for other co-authors due to the current political situation).

ANSWER 4.2:

See ANSWER 4.1. We have fully engaged with local communities, with a broad array of actions developed in Yakutia until 2020 involving European partners, and since (until 2024) through local colleagues. The political situation prevailing since 2020 has prevented the European collaborators contributing to our study to travel in Yakutia and personally engage with local communities. However, the local collaborators of the present study have kept informing the local communities about the project development, including our latest results. For example, Prof Liliya Alexeeva (NEFU) organized an exhibition at the Yakutsk University Museum in March 2023, presenting the rich repertoire of clothes that were excavated as part of the MAFSO field work (the poster of this exhibition is provided as an Appendix for the reviewers, at the end of this document; Appendix 1). Additionally Vassily Popov, curator of the Em. Yaroslavsky Museum in Yakutsk, regularly disseminates MAFSO-related findings in the Yakut newspaper, *Kyyym*, which is one of the mainstream newspapers in the region. These articles are written in Yakut, which broadens our capacity to reach out to the autochthonous populations. The following provides a non-exhaustive list of some of the articles that were published between 2022 and 2024:

2022

- **O iakutskom konskom sedle (po materialam SFAĖĖ)** [On the Yakut Horse Saddle (Based on SFAĖĖ Materials)]. June 16, 2022.
 - **Kenotaf (po materialam SFAĖĖ)** [Cenotaph (Based on SFAĖĖ Materials)]. January 27, 2022.
- 2023
- **Statia Niny Spiridonovoi "O rabote SFAĖĖ na territorii Iakutii"** [Nina Spiridonova's Article "On the Work of SFAĖĖ in Yakutia"]. November 2, 2023.
- 2024
- **Choron so vstanami nozhkami** [A Choron with Standing Legs]. January 11, 2024. Based on burial materials from Timir Bappyt, Churapchinsky ulus, 2005.

Additionally, other local media keep releasing reports on the MAFSO discoveries, for example, in 2023, in the online press called Ulus Media (<https://ulus.media/2023/08/15/chto-skryvaet-merzlota-unikalnye-eksponaty-iz-fonda-yakutskogo-muzeya-im-e-yaroslavskogo/>).

We have prepared a full section of the Supplementary Information that describes the broad and diverse activities that were developed to engage with local communities (see Supplementary Information Section 1.3, page 14-17), and including: (1) field work, ethics committee; (2) student training (including local students), and; (3) public outreach and community engagement. This section aims to provide an overview, and, thus, does not attempt to provide an exhaustive list of activities. The following provides more references, which together with the pictures taken on some of such events (provided as Appendix 3 and 4), demonstrate the materiality of our activities between 2016 and 2024.

- A French television documentary, a testament to the unity and shared commitment of French and Yakut researchers, was broadcasted twice on the Yakut television, including once on Christmas, a peak in yearly audience. It was focused on our excavation work, and was subtitled in Russian. This documentary made our expeditions familiar across the entire region of Yakutia. Jampolsky, M. *Les Mystères de Kyys La Chamane*. (Gédéon Programmes, France, 2007).
- Bravina, R., Popov V. 2008. In Russian, *Pokhoronnye obryady yakutov, pamyatniki i traditsii. Pamyatniki i traditsii XV–XIX vv.* [Funerary Rites of the Yakuts, Monuments and Traditions. Monuments and Traditions of the 15th–19th Centuries]. Izdatel': Akademiya nauk Respubliki Sakha i Muzei istorii i kul'tury narodov Severa imeni E.M. Yaroslavskogo, gorod Yakutsk [Publisher: Academy of Sciences of the Republic of Sakha and Museum of History and Culture of the Peoples of the North Named After E.M. Yaroslavski, City of Yakutsk]. 296 p
- *Drevnie yakutskie pogrebeniia (po materialam SFAĖĖ)* [Ancient Yakut Burials (Based on SFAĖĖ Materials)]. 10 aprilia 2008 g.
- *Obriad ochishcheniia ognem (po materialam SFAĖĖ)* [Fire Purification Ritual (Based on SFAĖĖ Materials)]. 11 fevralia 2010 g.
- *Arkheologicheskie raskopki v Verkhoyanskom uluse (po materialam SFAĖĖ)* [Archaeological Excavations in the Verkhoyansk Ulus (Based on SFAĖĖ Materials)]. 20 oktiabria 2011 g.
- *Nataznoe ukrasbenie (po materialam SFAĖĖ)* [Suspension Ornament (Based on SFAĖĖ Materials)]. 1 dekabria 2011 g.
- Crubézy, E., Alexeev, A. 2012. In Russian (dir.), *Mir Drevnikh Iakutov, Opyt Mezhdistsiplinarnykh Issledovanii* [The World of the Ancient Yakuts: A Transdisciplinary Approach from the French/Yakut Expedition], University of the North East (Yakutsk, 2012) https://vk.com/wall-341621_1426
- *Problemy arkheologii srednevekov'ia v Iakutii* [Issues of Medieval Archaeology in Yakutia]. 26 ianvaria 2012 g.
- *Ispolzovanie beresty v pogrebal'nom obriade sakha (po materialam SFAĖĖ)* [The Use of Birch Bark in Sakha Burial Rites (Based on SFAĖĖ Materials)]. 22 marta 2012 g.
- *Iakutskie ser'gi (po materialam SFAĖĖ)* [Yakut Earrings (Based on SFAĖĖ Materials)]. 22 noiabria 2012 g.
- *Itogi arkheologicheskikh raskopok v Oimiakonskom uluse* [Results of Archaeological Excavations in the Oymyakon Ulus]. 17 oktiabria 2013 g.

- Kiriānov N.S. *Drevnie zakhoroneniia Oïmiakonii XVII-XVIII vekov* // «Pod nehom Oïmiakonii» [*Ancient Burials of Oymyakonya, 17th–18th Centuries* // “Under the Sky of Oymyakonya”]. [pod red. S.S. Sleptsova]. – Iākutsk-Vladivostok: OAO «IPK Dal’press», 2016. – S. 39-43
- «Doch’ piānitsy-shamana ili arkheologicheskii detektiv» (Arina Iākovleva) // gazeta «Iākutsk vechernii», 08.04.2016 g. / o pogrebenii Lepsei II, Verkhoiānskii raion [“The Daughter of a Drunken Shaman or an Archaeological Detective” (Arina Yakovleva) // Newspaper *Yakutsk Evening*, 08/04/2016 / About the Burial of Lepsey II, Verkhoyansk District]
- *Pogrebenie cheloveka s berestianym pokryvalom v Khangalasskom uluse* [Burial of a Person with a Birch-Bark Cover in the Khangalass Ulus]. 3 noiabria 2016 g.
- *Pogrebenie konia na territorii Khangalasskogo ulusa v pozdnee srednevekov'e (po materialam SFAĖĖ)* [Horse Burial in the Khangalass Ulus During the Late Middle Ages (Based on SFAĖĖ Materials)]. 6 oktiabria 2016 g.
- *Arkheologicheskie raskopki v doline Erkeēni v 2016 g. (po materialam SFAĖĖ)* [Archaeological Excavations in the Erkeēni Valley in 2016 (Based on SFAĖĖ Materials)]. 22 sentiabria 2016 g.
- September 21st, 2016: Interview on local TV in Bakhsy (“Tolon TV”). Interview of Vincent Zvenigorosky by a local presenter, translated and assisted by Georgii Romanov and Aisen Solovov. Broadcast live on all televisions in the village, with questions from the audience, in preparation for our anthropological survey and collection of biological samples
- From September 21st to October 5th, 2016: Collection campaign in Tolon (Bakhsy), numerous discussions with volunteers and non-volunteers, our presence and purpose advertised through WhatsApp and posters.
- October 10th, 2016: Lecture at the North-Eastern Federal University (NEFU) in Yakutsk, “Genetic identification in archaeological contexts”. By Vincent Zvenigorosky, translated by Georgii Romanov.
- From October 31st to November 14th, 2016: Collection campaign in Khonuu (Momsky), numerous discussions with volunteers and non-volunteers, our presence and purpose advertised through WhatsApp and posters.
- November 7th, 2016: Presentation at the Cultural Center in Khonuu of the state of Yakut genetic studies and a more general lecture concerning ancient DNA. By Vincent Zvenigorosky, translated by Georgii Romanov.
- Franko-iakutskaiā ekspeditsiia: 15 let arkheologicheskogo sotrudnichestva (videoversiia) // www.youtube.com/watch?v=6cascHSrsVQ&t=223s (23 oktiabria 2017 g.) / Proekt SVFU «Nash Universitet. Lektīi» (2 sezon). [Franco-Yakut Expedition: 15 Years of Archaeological Collaboration (Video Version) // www.youtube.com/watch?v=6cascHSrsVQ&t=223s (October 23, 2017) / Project NEFU “Our University. Lectures” (Season 2).]
- *Drevnie pogrebal'nye obriady sakha (po materialam SFAĖĖ)* [Ancient Sakha Funeral Rituals (Based on SFAĖĖ Materials)]. 23 noiabria 2017 g.
- September 15th, 2017: Organized meeting at the cultural center in Tolon (Bakhsy), Churapchinsky, where volunteer participants from previous campaigns came to attend the presentation of the results by Vincent Zvenigorosky and Georgii Romanov.
- Popov- V.V. 2018. In Yakut *Sakha-Frantsuz arkheologicheskai non etnograficheskai ekspeditsiiei iŕdiiōnnērit, (2004-2017)* [Results of the Sakha-French Archaeological and Ethnographic Expedition (2004–2017)] Yakutsk, 2018- Publissert Kamuel Dokuusskay. 120 p
- «Po sledam predkov Omogoia» (Aleksei Evstaf'ev) // gazeta «Iākutsk Vechernii», 16 noiabria 2018 g., s. 19 (rubrika «Nauka») / o raskopkakh v Arylakhe, Ust'-Aldānskii raion [“In the Footsteps of Omogoy’s Ancestors” (Alexey Evstafyev) // Newspaper *Yakutsk Vecherny*, November 16, 2018, p. 19 (Science Section) / About Excavations in Arylakh, Ust-Aldan District]
- February 25th, 2019: Meeting and ethics discussion with the president and members of the Public Organization Association of Indigenous small-numbered peoples of the North of the Sakha Republic (Yakutia), and the World Association of Reindeer Herders. Discussion with Dr. Mikhail Pogodaev

(former president of the Even in Yakutia) regarding ethical and practical issues, with Anastasia Shestakova, Nicolas Bureau and Vincent Zvenigorosky.

- March 2nd, 2019: Meeting and discussion with the mayor-assistant of Sasyr (Momsky) with Anastasia Shestakova, Nicolas Bureau and Vincent Zvenigorosky.
- From March 2nd to March 13th, 2019: Collection campaign in Sasyr and the Magyr reindeer camp (Momsky), numerous discussions with volunteers and non-volunteers, our presence and purpose advertised through WhatsApp and posters.
- March 22nd, 2019: Participation in an interview at the Yakut Science Center of Complex Medical Problems for local (state) channel NVK Sakha. Vincent Zvenigorosky, Georgii Romanov and Nurgun Gotovtsev.
- From March 23rd to March 29th, 2019: Collection campaign in Tolon (Bakhsy), numerous discussions with volunteers and non-volunteers, our presence and purpose advertised through WhatsApp and posters. Nicolas Bureau, Nurgun Gotovtsev and Vincent Zvenigorosky made their presence at the cultural center known and received and answered questions from locals.
- *O iakutskom konskom sedle (po materialam SFAËË)* [On the Yakut Horse Saddle (Based on SFAËË Materials)]. 16 iunia 2022 g.
- *Kenotaf (po materialam SFAËË)* [Cenotaph (Based on SFAËË Materials)]. 27 ianvaria 2022 g.
- *Statia Niny Spiridonovoi "O rabote SFAËË na territorii Iakutii"* [Nina Spiridonova's Article "On the Work of SFAËË in Yakutia"]. 2 noiabria 2023 g.
- Popov- V.V. 2023. In Yakut *Süürbë ekspeditsiia (2004-2020)*. - [Twenty-Year Expedition (2004-2020)] Yakutsk, 2023. - 164 p
- Article in Russian: <https://ulus.media/2023/08/15/chto-skryvaet-merzlota-unikalnye-eksponaty-iz-fonda-yakutskogo-muzeya-im-e-yaroslavskogo/>
- *Choron so ustanami nozhkami* [A Choron with Standing Legs]. 11 ianvaria 2024 g. Po materialam pogrebenii v m. Timir Baptyt Churapchinskii ulus, 2005.
- Blog by Nicolai Baramygin (a blog on the history of Yakutia): <https://nikbara.ru/2021/08/23/o-proishozhdenii-i-passionarnosti-jakutskogo-naroda/>
- News about the mission: <https://ysia.ru/v-hangalasskom-i-namskom-ulusah-prohodit-franko-yakutskaya-arheologicheskaya-ekspeditsiya/>
- On the Yakutsk University website: <https://www.s-vfu.ru/universitet/rukovodstvo-i-struktura/strukturnye-podrazdeleniya/ums/resource-center-of-the-embassy-of-france-in-nefu/>
- Interview with E. Crubezy on the university website: https://www.s-vfu.ru/news/detail.php?SECTION_ID=2268&ELEMENT_ID=78569
- Article about the Kyys Ounyoga exhibition: <https://ysia.ru/v-yakutskom-muzee-vosstanovili-pogrebenie-xviii-veka-kyys-u-u-o-a/>
- Press article about the mission: <https://sakhapress.ru/index.php/archives/195561>

Community outreach is needed for 2 reasons:

POINT 4.3:

1a) The community members should understand that these individuals might have living relatives today and that it might be their grandparents or great grandparents that are being investigated. Related to that, in my understanding, no effort seemed to have been made to trace living relatives of the individuals that were analysed and dated to the 19th and 20th century. The authors should at least try to trace them and receive their consent when possible. If it is not possible to outreach to the communities in Russia in person, a contact to expat communities or online should be at least attempted. It should be thoroughly considered if the genetic data of individuals from the 19th and 20th century should be made publicly available or if other sort

of access should be prepared instead (EGA). Upon publication, such data will be immediately uploaded to GEDmatch and other databases by third parties and the living relatives might be discovered (it is to the discretion of the authors if that could be used to trace some of the relatives). I understand that it is a large undertaking to try to trace relatives for such individuals but given that these are personal data of very recently living individuals, it would mean disregarding their rights if ignored.

ANSWER 4.3:

Recent sociological surveys in Yakutia have demonstrated that young people in the city of Yakutsk are largely supportive of DNA testing (Kononova et al., 2021, entitled: “The attitude of young people in the city of Yakutsk to DNA-testing”, doi: 10.1080/22423982.2021.1973697). This survey confirms the general experience developed throughout the project, in which local villagers proved a strong interest in knowing whether some of their ancestors could be identified following DNA testing of the excavated graves (this, in fact, incentivized a majority of them to contribute to archaeological survey and work). Given the time difference to each of the archaeological stages analyses, no genetic relatedness could be expected within the first to fourth degree. This limitation was made clear to all the people we met, who understood that some of the individuals excavated might be at best one of their rather distant relatives. It is worth noting that, in preparation of these revisions, we consulted a native Yakut jurist (Natalia Egorova), who indicated that the Russian law regarding graves defines the rights of “close relatives”, which as surprising as it may seem in western constructs, excludes “cousins”. According to this legislation, only heirs can claim rights on graves and burials, and in the absence of an identified deceased person (as is the case for all the burials excavated by MAFSO due to the absence of names on the graves), no heirs persist to this day.

We also consulted local ethics committees while developing this project, in particular while launching the ANR Life Change project (2017-2022, ANR-17-CE 12-0018), which supported the genetic analyses presented in our study. The Yakut Ethics Committee granted authorization for the collection of human biological samples (saliva and blood) from 250 living individuals, all of whom provided informed consent. The Yakut Ethics Committee, formally named the “Interdisciplinary Scientific Institution 'Yakut Scientific Center'” approved our research objectives as stated in their official documentation (Protocol No. 49, dated March 25, 2018). The research objectives (translated from Russian) included: “The purpose of this international project is to examine the history of the population of Yakutia and track the social, genomic, and microbiological changes since the 16th century, the period marking Yakutia’s integration into Russia. The study is based on the analysis of biological samples collected from 250 individuals from various ethnic groups. The results will be compared with reference genetic data obtained from archaeological remains of ancient Yakutian burials.” Incidentally, the comparison of the sequence data presented here with that of the 250 modern Yakut individuals who were sequenced failed to identify matches at the first- to fifth-degree of kinship, as expected due to the underlying time difference (the sequence data from the 250 modern Yakut individuals represents the main material for another scientific article; no similar matches were evidenced between the ancient individuals and the 20 modern individuals included in HGDP and). Finally, it is important to recall that the identification of genetic matches between present-day and historical individuals constitutes a key objective for several Yakut-led research initiatives, including the study by Tikhonov et al. (2019), entitled: “On the origins of the Sakhas’ paternal lineages: Reconciliation of population genetic / ancient DNA data, archaeological findings and historical narratives.” (Tikhonov et al., 2019).

Given this context, which is widely favorable to DNA testing, we believe that the sequence data generated as part of this project should be made available in the public domain without restrictions, as is common to most genomic work. Therefore, the whole sequence data generated here, including fastq and bam files, were uploaded into the European Nucleotide Archive under Accession Number PRJEB66199. This follows previous work, also reporting their sequence material in relevant depositories (eg. The human Genome Diversity Project, HGDP, with individuals HGDP00951 and HGDP00956, see <https://www.internationalgenome.org/data-portal/population/YakutSGDP>; the Allen Ancient DNA Resources, AADR, which includes the 20 genome-wide data for 20 modern Yakut individuals published by Bergström et al. Nature 2020). The line of arguments developed above is provided in new sections of the Supplementary Information, including 1.3.2 *Legal and ethical aspects* and 1.3.3 *Fieldwork, Ethics committee and custom clearance* (pages 15-16, lines 597-623).

“1.3.2 Legal and ethical aspects

The CNRS (France) and the Université Paul Sabatier (Toulouse, France) supported the development of the COSIE International Associated Laboratory from 2011 to 2019 and in 2020 an International Research laboratory. COSIE aimed at studying the coevolution between man and the environment in Eastern Siberia, mainly in Yakutia or Sakha Republic, during the last five centuries. It involved three French partners (CNRS, Université Paul Sabatier Toulouse III, as well as the Université de Strasbourg until 201X, and then Université de Paris-2015), as well as the Russian Foundation for Basic Research, the Northeastern Federal University in Yakutsk, as well as the Krasnoyarsk State Medical University. The COSIE International Associated Laboratory was co-led by Prof. Crubézy (Université Paul Sabatier Toulouse III) and Prof. Anatoly Alexeev who served as the rector of the Northeastern Federal University (NEFU) in Yakutsk in 2011, and as the director of the interdisciplinary laboratory entitled "Evolution of Nature and Humans in the North". The laboratory focused on the long-term interaction between humans and their environment, particularly over the past four centuries, when the arrival of Yakut and later European populations gave rise to epidemic diseases, some of which are re-emerging today. In addition to joint archaeological excavations, paleogenetics was proposed as a method to better understand the history of human settlement, archaeological sites, and the evolution of infectious diseases. Knowledge transfer between French and Yakut partners was facilitated through workshops and the exchange of professors, researchers, and students. Official collaborative agreements between Université Paul Sabatier Toulouse III and NEFU were signed from 2011, and renewed twice for a five-year period, to facilitate exchange between both institutions.

In 2015, the collaborative framework was extended to laboratories specifically focused on Indigenous populations. This involved Dr. Anatoly Nikolaev, NEFU rector at the time, as well as Dr. Lilia Alekseeva (NEFU Museum of Archaeology and Ethnography), Dr. Sardana Fedorova (NEFU Molecular Biology Research Laboratory of the Institute of Natural Sciences), Prof. Olga Melnichuk (NEFU Institute of Modern Languages and Regional Studies), Dr. Maxim Cheprasov (NEFU Laboratory-Museum of the Mammoth at the Institute of Applied Ecology), and Dr. Natalya Popova (Institute for Humanities Research and Indigenous Studies of the North, under the Siberian Branch of the Russian Academy of Sciences). This collaboration later evolved to include the same partners in an international research project.

1.3.3 Fieldwork, Ethics committee and custom clearance

Field archaeological excavations were led by Yakut colleagues both before (2004-2011) and after (2011-2019) the collaboration was established, in collaboration with the institutions mentioned above for the latter. In June 2012, ethical authorization for conducting ancient DNA analyses was granted by the Local Committee for Biomedical Ethics of the Federal State Budgetary Institution, which is referred to as the “Yakut Scientific Center for Comprehensive Medical Problems” of the Siberian Branch of the Russian Academy of Medical Sciences. The authorization aimed “to collect information on the evolution of bacteria or viruses identified for the studied period, as well as to study the influence of socio-ecological factors on movement, adaptation, and dynamics of ancient populations” (Excerpt from the Minutes, dated June 20, 2012, No. 31).

Subsequently, scientific material, including archaeological remains, was issued customs clearance certificates by the rector of Yakutsk University. These materials were transported to France by the company Genomed, under customs resolution number RU/2019/05-41350/19.”

POINT 4.4:

1b) The community outreach is important also because the communities might be faced with a lot of questions regarding the absence of impact of Russian conquest. That can be even generalised for other communities that were under Russian and other colonial power. It would be good to discuss this issue in this section also and possibly use the literature to describe the impact of Russian conquest historically more clearly in general. This is also important to fully understand the impact of this study and the relationship of biomolecular methods and historical evidence. At the moment, the manuscript suggests that because of the lack of genetic impact, Yakutia was a middle ground, a region that was not conquered demographically or militarily - it should be heavily noted that there might be strong effects on livelihoods and the populations as a whole even without the genetic impact.

ANSWER 4.4:

We totally agree with the reviewer that nuance in this matter is critical to avoid suggesting that this conquest, and by extension any other colonization process, did not have any impact on the daily lives of people. In fact, historical records show that Yakut communities were severely impacted by epidemics. For example, it is established that the Yakuts were struck by smallpox epidemics in 1652, and again in 1681, 1691, and 1695, which caused terrible losses. However, as early as 1652, many Yakut groups fled the areas around the Russian forts to settle in the Vilyuy region (Tokarev, 1953). The people were able to migrate to other regions thanks to their equestrian way of life. Yakut nobility quickly cooperated with Russian authorities and obtained territorial autonomy, which laid the foundation for the current Republic of Saha-Yakutia. By contrast, settled hunter-gatherers and fishermen, such as the Itelmen (Kamchadal) and the Yukaghir, could not easily flee. As a result, they suffered repeated mortality rates, estimated around 40-60%. Most of the Yukaghir territory was depopulated and soon occupied by the Yakut (Forsyth, 1992). In general, in Siberia during the Russian rule, Turco-Mongolian populations (Buryat, Tuvan, Yakut, Khakas) were less impacted by epidemics than Paleo-Asiatic populations, and experienced positive demographic growth. It is likely that isolated Paleo-Asiatic groups were immunologically naïve, in contrast with Turco-Mongolian pastoralists who had ancient contact with Chinese and Western societies through Silk Road trade networks.

We have now developed the main text, especially the discussion section to provide example of how the Russian conquest has impacted the livelihoods of local communities. The fact that their way of life, and social organization was largely resilient to Russian colonization, and that the Yakut people did not disappear and eventually expanded in the region, should not undermine the heavy biological, cultural and demographic toll documented in other regions of the world.

This is reflected in the main text on page 7, lines 352-363:

“Additionally, the harsh environment, often lethal to European settlers and unsuitable for large-scale cereal agriculture, hindered the development of demographically-thriving Russian colonies⁵⁵ (Supplementary Information Section 1.1 and Fig. S1_1). Consequently, the Russian expansion in Yakutia relied heavily on local populations, representing a form of colonialism distinct from the settler colonialism of the Americas, which proceeded through Indigenous displacement and eradication⁵⁶. It also diverged from the indirect rule characteristic of colonial India, where colonial authority was maintained through local leaders without administering the territory. In contrast, Russian rule actively controlled taxation, and granted citizenship to local Yakut elites, integrating them as full subjects of the empire.

Despite the absence of significant genetic impact, Russian conquest profoundly affected Yakut livelihoods and demographics. Russian contact introduced numerous infectious agents, including the smallpox identified here and *Mycobacterium tuberculosis*¹³, with devastating demographic effects on immunologically naïve Indigenous populations¹³.”

2) I am not persuaded by the statements about social practices.

POINT 4.5:

2a) There are 4 cases of close relatives and 2 of these contain a case that is an exception to patrilocality. The patrilocality hence has to be more established, e.g. by exploring differences between IBD sharing between and within sites between females and males explicitly and exploring if it is significantly different (uniparental markers are not sufficient to claim patrilocality). Patrilocality is also potentially questioned by the findings of important female individuals in the society. If anthropological/ethnological studies are used as a basis for the patrilocality statement, they and this assumption need to be clearly described and discussed.

ANSWER 4.5:

Our analyses now include Identity-By-Descent sharing comparison between sex and locality. We restricted the comparison between males and females to adult individuals, because subadults might represent an obvious confounding factor when assessing the sex-specific patterns of mobility and connectedness (children

indeed are too young to move accordingly to the local rule; for example, the young daughter of a woman will live together with her mother, even if the society is patrilocal). We first compared the sum of IBD segment >12 cM shared between females and males within the same region. We performed the same measurement between regions. We observe that there is a significant difference in long-IBD fragments between men and women buried in different region (Wilcoxon test, Benjamini-Hochberg correction, p -value<0.0001). There is also a trend for men from the same region to share more long IBD segments with men from different regions, although this remains only marginally significant after correction for multiple testing (Wilcoxon test, Benjamini-Hochberg correction, p -value=0.07).

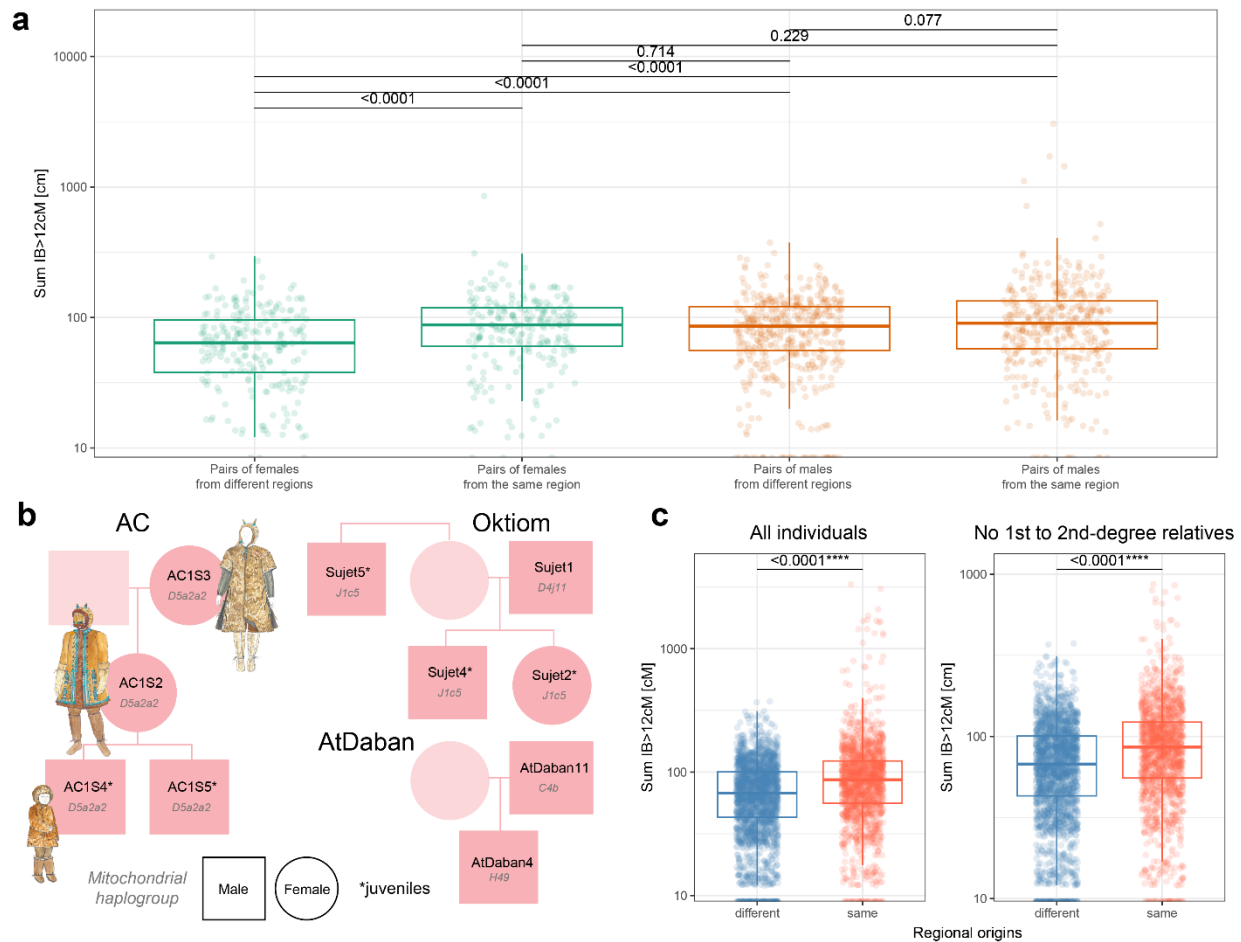
We now discuss these result in the main text (page 5, line 253-258):

“Historical and ethnographic sources describe Yakut communities structured around patrilineal and patrilocal clans, with strict exogamy⁴⁸. Mitochondrial diversity was high across Stages 1–4, while the number of Y-chromosomal haplotypes remained markedly limited (Fig. 4a). Furthermore, the proportion of long IBD (Identity-By-Descent) segments (>12 cM) in pairs of same-sex adults buried in different region was significantly greater amongst males than females (Wilcoxon test, p -value<0.0001; Extended Data Fig. 5a). These findings are consistent with the expectations of patrilocal clans practicing exogamy. ”

Combined, our analyses support patrilocality, in line with ethnographic reports. Ancient Yakut marriage rituals are indeed known to have celebrated the departure of the bride from her parents’ home and her establishment in her husband’s family (Vitashevskii, 1910; Serochevskii, 1993). As a rule, the new family settled on the territory of the husband’s father, which corresponds to a patrilocal residence pattern. Of note, exceptions to this pattern existed, as we found two cases of individuals buried alongside relatives from their mother’s family (the mother’s mother in one case, the mother’s brother in the other, which is more indicative of a matrilocality residence. This is reflected in our statement (Page 5, lines 260-263):

“While these findings suggest non-strict patrilocal practices, the funerary context of the latter two sites is unique across Yakutia(Duchesne et al., 2025). It shows individuals deposited simultaneously, which is indicative of a catastrophic event, such as the smallpox outbreak that decimated people buried at Arbre Chamanique.”

The results of the patrilocality tests described above are presented in a new panel of Extended Data Figure 5 (panel a), copied below:



Extended Data Figure 5: Examples of genealogical reconstructions and IBD sharing.

a) Sum of IBD segments longer than 12 cM (log-scaled) shared between pairs of adult individuals of the same sex from the same region, or from different regions. The p -value of the Wilcoxon test, with Benjamini-Hochberg correction for multiple testing, is shown. **b)** Genealogy of the individuals from AC, AtDaban and Okiom burials. **c)** Sum of IBD segments longer than 12 cM shared between pairs of individuals from the same region (red), or from different regions (blue). Y-axis is log-scaled. The p -value of the t-test is shown if significant. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over 75th or under 25th percentiles, respectively.

POINT 4.6:

2b) The exogamy is considered to be relatively local as not many connections are found on close relatedness between the regions - however given the time and space, how likely is it to find such relatives? Also, IBD sharing of all individuals within regions is considered higher within a region than between regions and a test that the individuals vary from panmixia is used to support this to showcase that there are preferential unions and contacts between close neighbours. This could be stated for almost any population from any species as there are clear patterns of genetic differentiation and relatedness being mainly driven by geography (isolation by distance). If the claim is that this population is special in this as having an increased rate of within region connectivity when compared to other populations (as this is then connected in the text to the statement that there is a high connection to the land of their ancestors), it needs to be substantiated: at least a departure from a panmictic population simulated in space should be shown.

ANSWER 4.6:

We agree with the reviewer that our findings depict a Yakut population following the most common patterns of isolation-by-distance found across the world. The reason why we decided to carry out those analyses is that the dominant archaeological model (Duchesne 2020) hypothesized significant rates of migration across regions to maintain the striking homogeneity of the material culture excavated over 31,000 square kilometers, with burials separated by 1,500+ kilometers. Here, we not only identify significant cultural differences across the main four regions, but we also demonstrate closer genetic ties within regions than between (in line with what found in the majority of human groups). Therefore, we find no support in the dominant archaeological model. We have reflected the above in the main text, as follows (page 6, lines 278-281):

“Combined, these results suggest social practices conducive to preferential unions, contacts and microbial exchanges intra-regionally, between close neighbors. This contrasts with the prevailing archaeological model, which posited that substantial migration between regions played a key role in maintaining striking similarities in the material culture (Duchesne et al., 2020).”

POINT 4.7:

3) A clear presentation of historical debate and various theories regarding Yakut origins should be also included in the SI. While it is well written in the main text, upon looking through the literature, I got the impression that the question is even more complex and that the evidence (oral, linguistic, cultural, historical) does not align so clearly with the presented results as it can appear from the main text. Also while reading the main text, I was left wondering what evidence does and what does not support the results (e.g. slightly enigmatic mentions of oral histories). Showcasing how debated this question is (even in the main text) would strengthen the impact of this study.

ANSWER 4.7:

We thank the reviewer for their suggestion. We have now developed a specific section of the Supplementary Information, entitled “*1.1 Background information on Yakutia, the Yakuts and the Russian conquest*”. This section provides a relevant summary of the underlying geographic area, linguistic diversity, and is followed by individual sections presenting common narratives about the origins of the Yakut people as depicted in local myths, or presented by early explorers and travelers, or the Soviet school of archaeology. A more objective description of the archaeological record then follows, before the section closes with a summary of the main findings of previous (archaeo)genetic work.

POINT 4.8:

4) From the archeological description, it is clear that the Christian graves have been excavated not systematically but only accidentally when (usually child) solitary graves have been unearthed under the assumption that these are from predating periods. It should be discussed if there are any issues related to this bias in the dataset: these for-their-time unusually buried individuals might come from particular subgroups and might not be representative of the whole population and their oral microbiome might be also affected by their ‘special’ status and by the fact that they were children (though their oral microbiomes are likely linked to those of their close relatives and their methylation patterns are likely not affected). It would be good to discuss in the discussion if there are any other potential technical/sampling biases that could be linked to the absence of chronological patterns in the dataset (the most surprising finding in my assessment).

ANSWER 4.8:

See ANSWER 2.2.

5) Use of reference data.

POINT 4.9:

5a) In the main text, only AADR paper is cited and the authors of papers where the data are actually from are not acknowledged for the data they produced.

ANSWER 4.9:

See ANSWER 2.7. We have decided to only cite the relevant AADR publication in order to respect the limits in the number of citations authorized for a publication in Nature. We have, however, now cited in the Supplementary Information Section 2.10 every single original publication contributing the AADR version 5 used as a comparative panel. This lists as follow (pages 35-36, lines 316-321):

“PCA was based on pseudo-haploid genotype calls for all the individuals presented in this study and carried out using smartPCA from EIGENSOFT (Patterson et al., 2006) version 7.2.170, projecting 913 ancient Eurasian and American individuals(Rasmussen et al., 2010, 2014; I. Lazaridis et al., 2013; Raghavan et al., 2014; Skoglund, Malmstrom, et al., 2014; Allentoft et al., 2015; Jones et al., 2015; Fu et al., 2015, 2016; Mathieson et al., 2015, 2018; I. P. Lazaridis et al., 2016; Omrak et al., 2016; Hofmanová et al., 2016; Jeong et al., 2016, 2020; Broushaki et al., 2016; Kılınç et al., 2016, 2021; Lipson et al., 2017; Pinhasi et al., 2017; Yang et al., 2017, 2020; Sikora et al., 2017, 2019; Olalde et al., 2018; Günther et al., 2018; P. de B. Damgaard et al., 2018; de Barros Damgaard et al., 2018; McColl et al., 2018; Mittnik et al., 2018; Flegontov et al., 2019; Järve et al., 2019; Ning et al., 2019, 2020; Narasimhan et al., 2019; Wang et al., 2019, 2021; Feldman et al., 2019; Yu et al., 2020; Mao et al., 2021; Yaka et al., 2021; Zhang et al., 2021) and 106 ancient Yakut individuals (coverage>0.02-fold) onto the principal components (PCs) obtained from in 2,761 Eurasian modern individuals(Meyer et al., 2012; Patterson et al., 2012; I. Lazaridis et al., 2013; Prüfer et al., 2014; Raghavan et al., 2014, 2015; Rasmussen et al., 2014; The 1000 Genomes Project Consortium et al., 2015; Skoglund et al., 2015, 2016; I. P. Lazaridis et al., 2016; Mallick et al., 2016; de Barros Damgaard et al., 2018; McColl et al., 2018; Lamnidis et al., 2018; Flegontov et al., 2019; Jeong et al., 2019; Mao et al., 2021; Wang et al., 2021).”

POINT 4.10:

Also, in the main text the authors use abbreviations (like the enigmatic Baikal_sib and Baikal_his) that are not described there and the same goes for Yakut_IA. Since these are crucial datasets for modelling the population studied, they should be clearly acknowledged and it should be described to the reader where and when these samples come from (in the main text at least shortly, in the SI the logic for their grouping should be described).

ANSWER 4.10:

See ANSWER 2.12. The precise lists of individuals forming Baikal_sib and Baikal_his, as well as any other reference group used in our analyses, are provided in the Supplementary Information Section 2.12 and Material and Methods, and the ADXMITURE profiles of these two groups are displayed in Fig. 2b. (see ANSWER 1.16).

POINT 4.11:

5b) Also: for example, ‘Mongolia_Dornod_LateMedieval’ (as in AADR) was published in Jeong et al. 2020 though this paper is not cited for this data in the main text and in the SI another paper (Kilinc et al. 2019) has been cited for this data instead (probably something happening to the citations).

ANSWER 4.11:

We thank the reviewer for spotting this mistake. The reference list has now been corrected accordingly (and for other typos that were previously overlooked; see ANSWER 4.9).

POINT 4.12:

5c) I am also not happy with the interchangeable use of `Baikal_his` and `Baikal_sib` in the main text. I understood after reading the methods and SI but when reading the body of the text, I thought it was a typo. I do not understand why not to use both datasets even if the interval for DATES estimate for one group was wide (which did not look so wide to me when looking at the SI table).

ANSWER 4.12:

We agree with the reviewer and have now mentioned the qpAdm analyses using both population sources in the main text and explained the use of `Baikal_Sib` for DATES in the main text too. We also realized that the wrong data for the Standard errors for the `Baikal_his` models were copy-paste in the SI Table, and we rectified the values in Supplementary Table 1h. The true values show a difference of almost 2 generations in standard errors between `Baikal_his` models and `Baikal_sib`, which justifies presenting the `Baikal_sib` model preferentially. Also, the higher number of individuals constituting `Baikal_sib` is also a motivation to prefer models with this group. We now also display the ADMIXTURE profile of these two group in Fig. 2b. (see ANSWER 1.16). The corresponding results and methods were edited accordingly:

For the mention of both qpAdm models (page 4, lines 169-173):

“Except for rare outliers (Supplementary Table 1f), the genetic makeup of all other Stage 1–4 Yakuts could be modelled as a two-ways admixture between populations inhabiting Yakutia in the Iron Age (Yakutia_IA: 62-84%) and the Baikal region from the 13th to 15th centuries CE (`Baikal_his`, 16-38% and `Baikal_sib` 40-79% with Yakutia_IA: 21-60%; Supplementary Table 1f; Supplementary Information Sections 2.12 and Fig. S2_9).”

For the DATES results (page 4, 188-191):

“The stable and homogeneous genetic composition of historical and modern Yakuts helped refine existing models of their origins. DATES³³ estimated the admixture time between the two primary ancestry sources between ~1100 and ~1250 CE (using Yakutia_IA+Nganasan and `Baikal_sib`; see Supplementary Information Section 2.12 for the rationale; Fig. 2d, Supplementary Table 1h).”

The method section (page 13, lines 670-675):

“We further applied DATES³³, using both the pseudo-haploid and imputed dataset, to 2-ways models to estimate the time of the admixture event between the local ancestry source (Yakutia_IA+Nganasan) and Baikal populations (Supplementary Table 1h; Supplementary Information Section 2.12). As the confidence intervals using the `Baikal_sib` source were more restrained (Supplementary Table 1h), and `Baikal_sib` covered more individuals, the analyses including the `Baikal_sib` source were preferred (Fig. 3d).”

POINT 4.13:

5d) Additionally, the Yakut_AI group (described also only in the main text and with the same issues with the citations in SI as above) consists only of two individuals. Except for Admixture analysis, it is not justified that they should be grouped. As such it is not clear if these samples were used individually, the same results would be obtained and it is not discussed how the low number of samples might affect the results.

ANSWER 4.13:

See ANSWERS 2.9 and 1.16. The two Iron Age individuals (Mokp and yak030) from the region show similar genetic profiles (Fig. 2b) and similar PCA placement (Fig. 2a). We, however, agree that even though our work doubled the number of individuals genetically characterized at the genome-wide level in the region, a number of 2 individuals is unlikely to reflect the whole genetic landscape present in such a vast region. We, thus, have nuanced some of our statements accordingly (page 6, lines 318-320):

“The precise contribution of the local population requires further investigation, as current models rely on a limited number of Iron Age individuals, likely failing to fully capture the genetic diversity across the vast territory of Yakutia at the time of admixture.”

We’ve also edited the Method section (page 12, lines 647-650):

“qpAdm models were aimed at testing whether the Yakut genomic makeup was compatible with a 2-ways admixture from a local Siberian background (Yakutia_IA, N=2, comprising Mokp and yak03041, as they showed similar genetic profile and PCA placements), and another source, potentially from the Baikal region”.

POINT 4.13:

5e) Related to this: Given the lack of possible sources, more distal modelling of the studied individuals would help to confirm that there are no differences between them - from Supplementary Table 1B it seems only simple combinations of Russian, Yakutia_IA, Baikal_his and Baikal_sib has been attempted. Finestructure shows on an intercontinental scale some slight differences between individuals (e.g. Oktom1Sujet, Lepse1, AtDaban4) so I am wondering if there is some unexplored diversity (though there is a lack of modern and ancient reference data so it might be hard to explore).

ANSWER 4.13:

We had in fact run many more models than those originally presented in the manuscript previously submitted. These were omitted as an attempt to only focus on the most relevant analyses. The results of such models are now presented in Supplemental Supplementary Table 1g, and largely support our previous findings. Briefly, we first attempted to pinpoint the best source to represent the Southern Siberian/Baikal/Central Asia origin. We used qpADM and the “rotating” method from Harney and colleagues (2021) to test every model of 2-ways admixture between Yakutia_IA (as described in our paper) and one of these populations: Russia_AngaraRiver_Medieval, Modern Buryat, Mongolia_Khuvsgul_MLBALatEMedieval and Mongolia_Sukhbaatar_Xiongnu. All models involving Yakutia_IA and another population from the considered set yielded similar results: they could fit the data, with feasible admixture proportions, for the majority of the individuals, apart from a small fraction of outliers (we noticed that a larger proportion of models were infeasible or rejected when considering the modern Buryat as a source). This indicates that the Yakut genomic makeup can largely be modelled from a local Iron Age source plus other local sources from Siberia and/or Mongolia. We then tested whether adding another source from more western origins could also fit the data, using 3-ways models admixture rotating model while still rotating through the Baikal possible sources. The western sources considered were between Russian, Polish, Bulgarian, Czech (accounting for Slavic speaking populations), Adygei, Abkhasian, Chechen, Lezgin and North Ossetian (accounting for the North Caucasus), Mansi (to represent Uralian speaking populations) and Altaian (Turko-Mongolic speaking populations). A vast majority of such models could not fit the data. Those fitting the data already showed good model fit with 2-ways models devoid of western sources and indicated rather relatively limited fractions of western ancestry (eg. all 2-ways models were accepted for individual AC1S5, while only 3 3-ways models, out of 44 tested, were accepted; they indicated western ancestry sources contributing 0.3%-2.7% ancestry; this is also true for AtDaban4, although more 3-ways models were accepted (30), yet with western ancestry contribution of 0.1%-3.5%). These extra analyses are now presented in the Supplementary Information, section 2.12, as follows (page 40, lines 368-388): “We initially sought to identify the most suitable source population to represent the Baikal_his source. To achieve this, we applied qpAdm on the 1240k data set and the “rotating” method outlined by Harney et al. (2021), testing various two-way admixture models between Yakutia_IA (N=2, comprising Mokp and yak03041) and one of the following populations: Russia_AngaraRiver_Medieval.SG (Kılınç et al., 2021) (N=1), Mongolia_Sukhbaatar_Xiongnu (N=1) and Mongolia_Khuvsgul_MLBALateMedieval (Jeong et al., 2020) individuals (N=2) and modern Buryat (Buryat.SG (Raghavan et al., 2015; de Barros Damgaard et al., 2018), N=4) (Supplementary Table 1G). Our results indicated that all models incorporating Yakutia_IA and another population from this set produced comparable outcomes. These models successfully fit the data and yielded plausible admixture proportions for most individuals, aside from a small subset of outliers. However, we observed a higher rate of infeasible or rejected models when using modern Buryat as a source (Supplementary

Table 1G). To maximize statistical power, we then used all the individuals from these groups except the modern Buryats in the Baikal_his group for the final qpAdm models. Next, we investigated whether incorporating an additional western source would further improve model fit. We tested three-way admixture models, on the HO data set (to include more modern populations) while continuing to rotate through the Baikal-related sources. The western sources considered included Russian, Polish, Bulgarian, and Czech (representing Slavic-speaking populations); Adygei, Abkhazian, Chechen, Lezgin, and North Ossetian (representing the North Caucasus); Mansi (as a representative of Uralic-speaking populations); and Altaian (as a proxy for Turko-Mongolic-speaking populations). The majority of these three-way models failed to fit the data (Supplementary Table 1g). In cases where they did, the two-way models without western sources already provided a strong fit, suggesting that western ancestry played only a minor role (Supplementary Table 1g). No other western sources outperformed the Russian group so we kept it as the western source for the final models.”

6) Some statements are hard to evaluate given how the data has been presented:

POINT 4.14:

6a) A claim is made that relatives within the sites are buried close together but the cemetery plan is not shown (or referenced if I did not miss it). This can be also statistically evaluated by actually comparing GIS coordinates of the related individuals to unrelated individuals.

ANSWER 4.14:

We apologize for the confusion. We tested kinship patterns within and between regions, but not within sites. However, the kinship networks (Fig. 4a) reconstructed indeed identified close familial ties within multiple burials (eg Arbre Chamanique, Oktiom, and Ieralakh), or between individuals buried in different graves located in the same archaeological site (eg AtDaban, Oulakh, TouoraUrek). The number of such individuals remains, however, generally limited, except for AtDaban where 10 individuals from the same site were excavated. All related individuals were males, while the only female (AtDaban6) was found unrelated, in line with the expectation of woman exogamy. In these conditions, it appears practically impossible to analyze whether those individuals buried next to each other are more genetically related than those located further away in the same site, as requested. This is why this analysis was not originally carried out. However, we have added the GPS coordinates of each and every individual analyzed (Supplementary Table 1a), and provided maps of 2 sites in which ≥ 2 subjects were excavated in the companion article that was submitted to Zenodo (*Beyond the Permafrost: Excavated Tombs of Eastern Siberia. Archaeological data from the French-Sakha missions* <https://doi.org/10.5281/zenodo.15017495>). This document also provide a synthetic description of 114 graves, and showing photographs of the associated material. See ANSWERS 2.3 and 2.5.

POINT 4.15:

6b) Relationships like father-son or grand-mother, daughter and grand-children are discussed but the genetic relatedness is showcased as a network and not as pedigrees and these relationships are not readable from that figure (it would from pedigrees) - this should be changed so the major relationships that were resolved are presented.

ANSWER 4.15:

We did display two possible pedigrees in Extended Figure 4A. These figures are still provided in the revised version of our work, and have been extended to another additional pedigree. For clarity, we have, however, decided to show the results discussed through networks, instead of pedigrees, given the multiple pedigrees compatible with the data presented (except for those three pedigrees). See also ANSWER 4.5 for the new figure and caption.

POINT 4.16:

7) It is a bit of detective work to figure out how many samples have been used for what analysis. It is also

unclear how many and what individuals with what types of datasets have been generated. Please show that clearly in a figure (e.g. as an extension of Fig 1c) and describe the numbers clearly in the text. The absence of this leads to questions about how representative the assessment of methylation of 21 individuals between the periods was, if the same individuals have been compared or how much of the comparison relies on different sets of individuals.

ANSWER 4.16:

The requested numbers are now provided on the respective figures, and specified in the relevant captions. We have also prepared a file, deposited as “Source data”, providing the exact numbers underlying 13 figure panels.

POINT 4.17:

Minor issues:

1) In some cases (e.g. line 102), ‘Yakut’ is not capitalised.

ANSWER 4.17:

Fixed, here, and throughout all documents.

POINT 4.18:

2) It is not always clear in the methods if imputed or not imputed data was used (eg. PCA, hapROH), please clarify.

ANSWER 4.18:

We apologize for the lack of clarity and have scrutinized the Material and Methods, as well as the Supplementary Information, to provide all necessary details. These can be summarized as follows:

*HAPROH: “Effective population sizes for each stage and region were estimated on the 1240k SNPs *pseudo-haploid* Yakut panel” (page 11, lines 585-587)

*ADMIXTURE: “Unsupervised ADMIXTURE (version 1.3.0) analyses were carried out to estimate the proportions of genetic ancestries present in Yakuts (coverage $\geq$ 0.03-fold, *pseudo-haploid*), using autosomal positions as part of the 1240K HO panel, and a total of 3,639 Eurasian and American individuals.” (page 12, lines 611-613)

D-statistics: “Calculations were carried out on the pseudo-haploid 1240K data set” (page 12, lines 637-638)

qpAdm: “Admixture models for ancient Yakut individuals (coverage $>$ 0.1) were assessed using the pseudo-haploid 1240K data set and qpAdm” (page 12, lines 646-647)

DATES: “We further applied DATES, using both the pseudo-haploid and imputed dataset, to 2-ways models to estimate the time of the admixture” (page 13, lines 670-671)

PCA: “PCA was based on pseudo-haploid genotype calls for all the individuals presented in this study” and “. A second PCA was carried out to validate our imputation pipeline by confirming similar projections for imputed genotype data and pseudo-haploid data (Fig. S2_6).” (page 11, lines 600 and page 12, lines 606-607)

POINT 4.19:

3) Unclear how the individuals used for IBD were filtered, usually coverage or n. of SNPs is used and the

results are then filtered by the length and number of IBD segments.

ANSWER 4.19:

We have now detailed the methodology underlying the analyses carried out with ancIBD (Ringbauer et al., 2023). As this software requires non-filtered VCF-formatted data, and performs an internal filtering procedure, we used the non-filtered VCF-formatted output of GLIMPSE2 as input for ancIBD. Analyses were run with default parameters, as these are optimized for ancient DNA data. ancIBD only retains IBD block with a minimal SNP density of 220 SNPs per cM. We then categorized IBD blocks according to their size (ie ≥ 8 cM, ≥ 12 cM, ≥ 16 cM, and ≥ 20 cM) to calculate the total number of such blocks, and the genome fraction encompassing these categories (in Mb). Extended Figure 5 panel a and c only show the genome fraction encompassing the size category of blocks ≥ 12 cM. This procedure is now integrated into the Supplementary Information, as follows (page 9, lines 270-278):

“Finally, Identify-By-Descent (IBD) content were calculated using ancIBD (Ringbauer et al., 2023). As this software requires non-filtered VCF-formatted data, and performs an internal filtering procedure, we used the non-filtered VCF-formatted output of GLIMPSE2 (Rubinacci et al., 2021) as input for AncIBD. As recommended, the Yakut data set was downsampled to the 1,240,000 SNPs from the 1240k dataset, for which ancIBD is optimized, and IBD sharing was screened for every pair with default settings (Ringbauer et al., 2023). Analyses were run with default parameters, as pre-optimized for ancient DNA data. AncIBD only retains IBD block with a minimal SNP density of 220 SNP per cM. We then categorized IBD blocks according to their size (ie ≥ 8 cM, ≥ 12 cM, ≥ 16 cM, and ≥ 20 cM) to calculate the total number of such blocks, and the genome fraction encompassing these categories (in Mb).”

POINT 4.20:

4) Line 672: Manually corrected alignments. Is that an established practice? It does not sound good to me.

ANSWER 4.20:

Alignments are indeed generally visually inspected, and manually corrected, to avoid misalignments around indels. However, we repeated all the analyses using the uncorrected MAFFT alignments and obtained a very similar tree topology, with comparable node support. The trees obtained with and without manual alignment correction are shown below, for comparison. We now mention this analysis in the Method section (page 15, line 774-776) :

“A tree was also generated using the exact same procedure as described above but without the manual correction of the modern genomes alignment (Supplementary Information Section 2.18 and Fig. S2_21).”

We have also added the following sentence to the Supplementary Information to mention those analyses and results, as well as to precise where the manual correction happens (page 27, lines 632-644):

“The multifasta sequence data comprising a total of the 45 modern viral genomes were further aligned using MAFFT (Katoh & Standley, 2013), with default parameters (--auto), visually checked, and manually corrected wherever appropriate. Gaps were added to the six ancient samples according to the gaps in the reference genome after the alignment procedure, and all fasta were merged to form a multifasta sequence of 52 viral genomes. Positions in which at least 50% of the sequences were covered were retained for Maximum Likelihood (ML) reconstruction in IQtree 1.6.12 (Nguyen et al., 2015). (Fig. 3e) The best substitution model (-m MFP) was identified as TVM+F+ASC+G4, and node support was estimated from 1,000 ultrafast bootstrap (Hoang et al., 2018) replicates (UFBoot), with each bootstrap tree optimized using a hill-climbing nearest neighbor interchange (NNI) search based on bootstrap alignment (-bb 1000 -bnni). To ensure reproducibility, we also performed the exact same procedure ML tree reconstruction as above but based on the direct output of MAFFT (Katoh & Standley, 2013) without visual checking and correcting of the modern viral genomes alignment, with the ancient genomes added as described above. The resulting tree yielded the same topology for the well supported nodes (**Error! Reference source not found.**)”

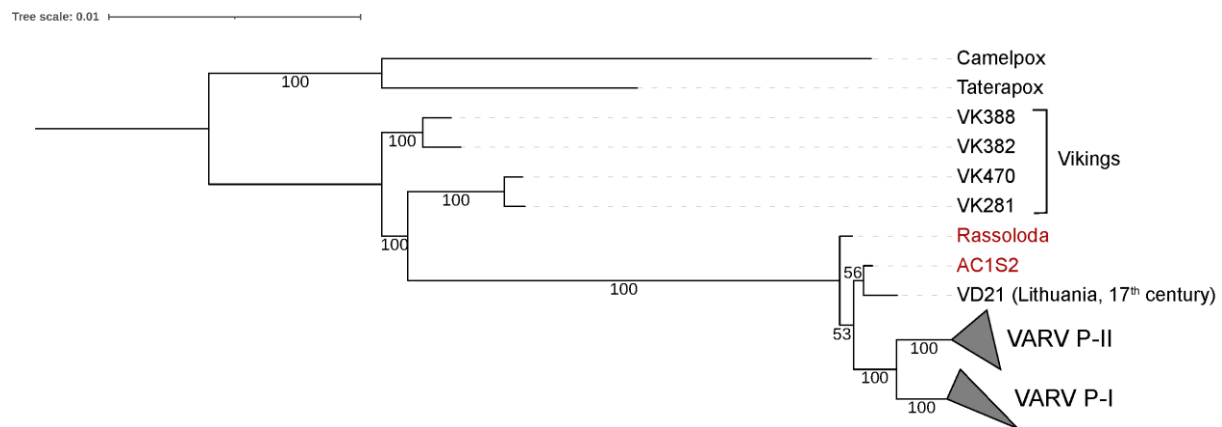


Figure S2_21. Maximum Likelihood phylogenetic tree of the two Yakut smallpox strains, based on raw MAFFT alignment. Node supports are based on 1,000 ultrafast bootstraps.

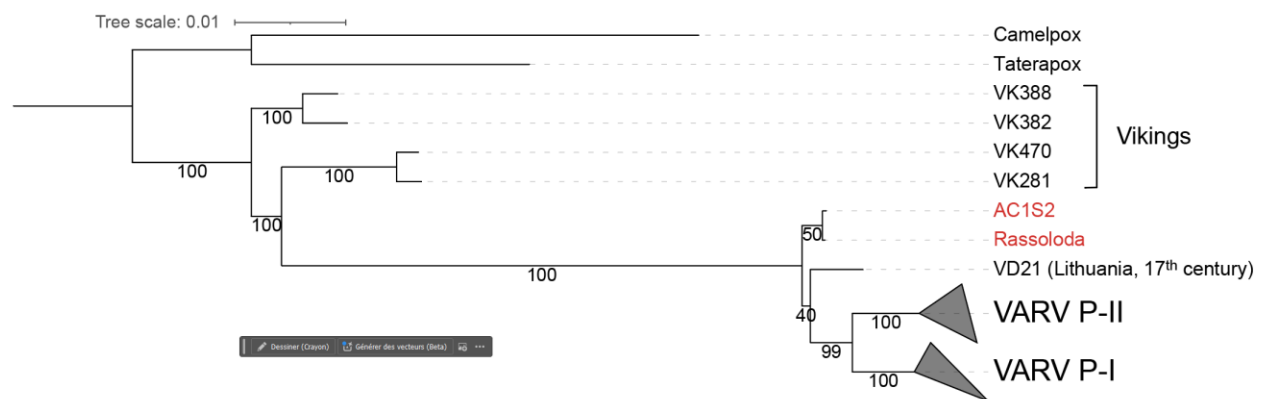


Figure 3.e) Maximum Likelihood phylogenetic tree of the two Yakut smallpox strains. Node supports are based on 1,000 ultrafast bootstraps. Boxplots represent the 25%, 50% and 75% quantiles, with upper and lower whiskers showing values within the 1.5 interquartile range. The line is the mean value for each species at each stage

POINT 4.21:

5) Soft tissue samples: What were they used for? Any results from them worth mentioning? Did I miss that? It is methodologically interesting and a lot of space is dedicated to their description so it is a bit anticlimactic.

ANSWER 4.21:

The newly prepared Supplemental Supplementary Table 1b breaks down DNA libraries per tissue type. Soft tissues were analyzed for only one individual, AC1num2, and consisted of lung and muscle. The lung tissue turned out positive for smallpox (see also ANSWER 1.9). Otherwise, soft tissues showed limited endogenous content, and provided extremely limited sequence data aligning against the human genome, resulting in 0.0000587- 0.0003046-fold coverage, depending on the DNA library considered. Most of the human reads characterized for this individual were generated from the four DNA libraries prepared on bone DNA extracts. The relatively limited DNA preservation observed in soft tissues match previous work on woolly mammoths (Seguin-Orlando et al., 2015). As a result, hard tissues, mostly petrosal bones and teeth, which are known to maximize DNA preservation (Gamba et al., 2014; P. B. Damgaard et al., 2015), were preferentially analyzed. The section presenting the identification of smallpox DNA in three individuals has been edited to describe the various types of tissues showing positive hits, as follows (page 5, lines 240-243): “Sequence data were derived from permafrost-preserved lung and bone tissues, as well as teeth and dental calculi (Supplementary Table 1p), albeit only limited portions of the smallpox genome could be characterized (average depth-of-coverage=0.12 and 0.15-fold for AC1S2 and Rassoloda, respectively).”

POINT 4.22:

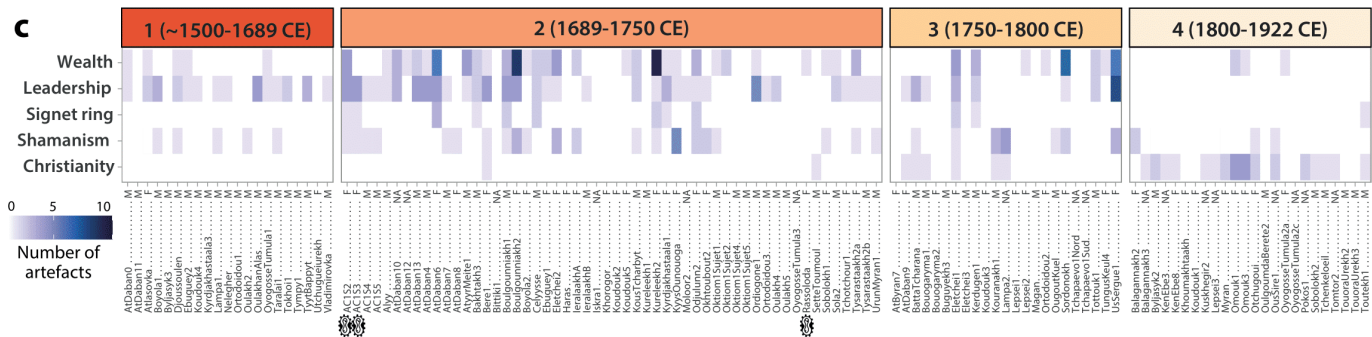
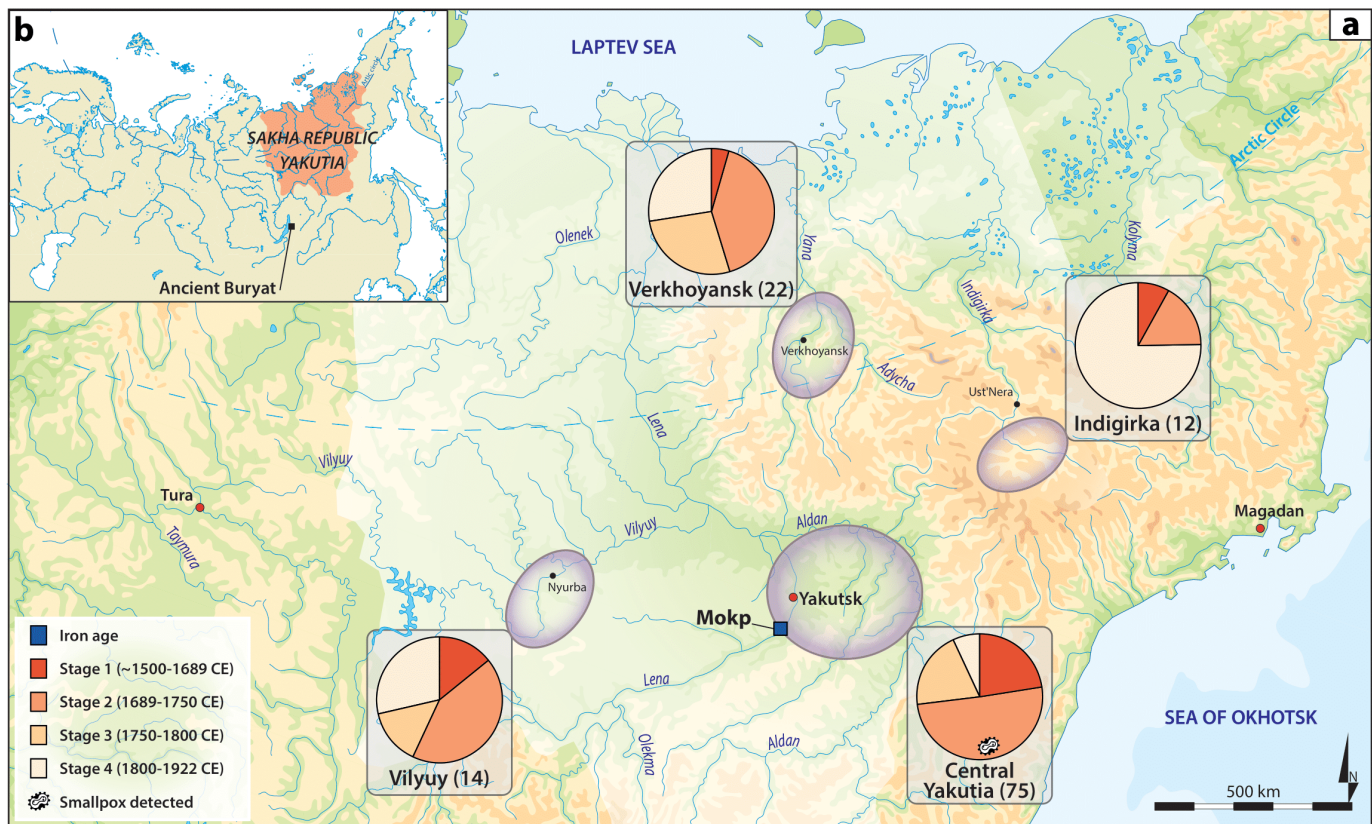
6) Fig 1a - unclear what are the colours of the regions, no legend and very similar to other colours on the plot

ANSWER 4.22:

The colors outlining the four main regions investigated were selected for showing no overlapping with the other colors used elsewhere in the manuscript, and for maximizing contrast. They were, thus, only meant to facilitate the identification of the regions on Figure 1A. We have now revised the corresponding figure as shown below to enhance readability:

Figure 1 (next page): Study area and archaeological material.

a) Location of archaeological remains. Biological material sampled from 123 individuals spread across 113 archaeological sites. Those sites located in very close areas were grouped in pie charts for improving readability, with colours reflecting time stages. The three individuals positive for smallpox are indicated with a graphical representation of the virus. **b)** Map of Eurasia showing Yakutia and the burial location of one 17th century individual from Buryatia (black rectangle). **c)** Cultural heatmap summarizing the number of archeological artefacts tracing wealth, leadership, shamanism and Christianity (as defined in the Supplementary Information Section 2.20) for Stages 1-4. The M and F suffixes appended to individual names indicate genetic males and females, respectively. **d)** Typical Yakut artefacts reflecting wealth, leadership, shamanism and Christianity. **a:** Cauldrons, copper alloy with horse meat offering in a *kytia* (wooden pot) (Individual Alyy, a1), or iron-made (Ebuguey2, a2); **b:** Signet ring (Celysse, b1+ b 2; AC1, b3+ b4; Kerdugen, b5; Boulgouniakh1 and Boulgouniakh2, b6+ b7; Toutekh, b8); **c:** Ring (Kureleekh); **d:** Solar disk (Sordonokh); **e:** Bronze torque necklace, twisted with counterweight composed of pearls and metal ornaments (Eletchei2, e1), or flat with counterweight made of silver coins (Sordonokh, e2); **f:** Earrings (Lepsei2, f1; Kyrdjakhastaala, f2; Bere1, f3); **g:** *Batilla*, sword with a long wooden handle (AtDaban12, g1; AtDaban11, g2); **h:** Copper alloy bracelet (Kureleekh); **i:** *Choron*, carved wooden monopod or tripod vase (Kerdugen, i1; Bakhtakh3, containing an offering of dairy product, i2); **σ:** Mammoth ivory comb (Eletchei1); **k:** Pipes (wooden bowl k1; Bere, with its copper alloy bowl and wooden stem, k2; AC1, mammoth ivory bowl, k3); **λ:** Carved wooden spoon (Tottouk1); **m:** Iron knife (IeralakhA); **n:** Iron curb chain for the harness (Boulgounniakh3); **o:** Wooden saddle covered with decorative metal plate (Eletchei1); **p:** Iron stirrup (Boulgounniakh1).



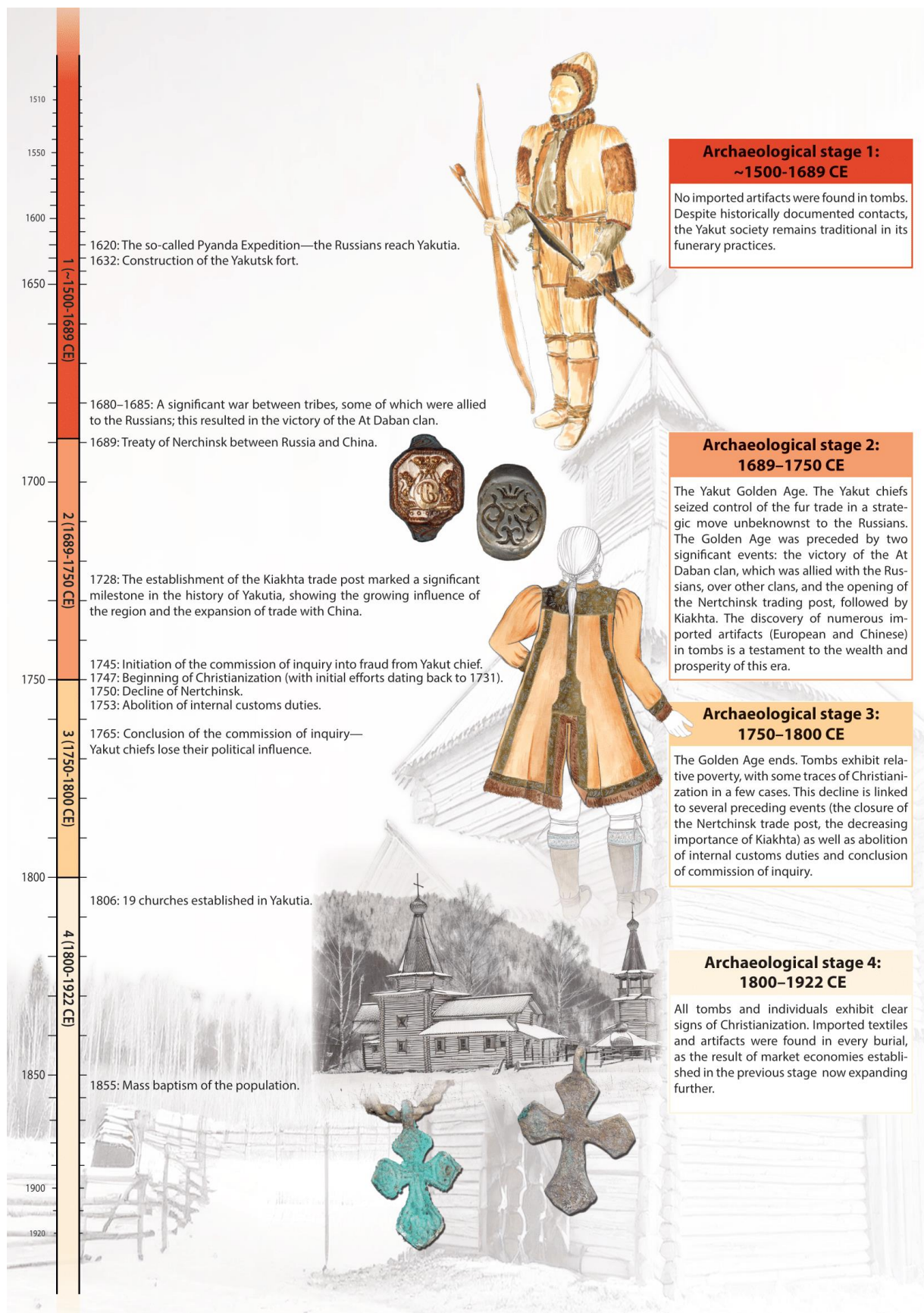
POINT 4.23:

7) Fig 1c: mark Russian conquest and potentially other important historical events mentioned in the SI

ANSWER 4.23:

We followed the reviewer's request, and have added Russian conquest. We have, however, not added other important historical events, such as epidemics, because the subjects present in each Archaeological Stage cannot be sorted chronologically (ie we don't have yearly precision, except a few within Archaeological Stage 1). The size of each box representing the different Archaeological Stages is a function of the number of subjects, not time. Therefore, precisely locating any event with reference to a specific subject appears impossible. However, we have prepared an additional figure (Extended Fig. 1) summarizing some of the milestones marking the history of Yakutia. The revised figure shows as follows:

Extended Data Figure 1 (next page): Main historical events punctuating the four main Archaeological Stages of Yakutia. More extensive discussions about the chronology of the various archaeological contexts can be found in Crubézy and Nikolaeva, 2017 (Crubézy & Nikolaeva, 2017b).



POINT 4.24:

8) Fig 1c: caption - refer to where in SI described the variables (Wealth, Leadership..)

ANSWER 4.24:

The description of these variables was provided in the Supplementary Information, however, with no direct reference to Figure 1c. This has now been corrected, see ANSWER 4.22 for the caption of Figure 1.

POINT 4.25:

9) Fig2a: use the same scale for both axis

ANSWER 4.25:

This is now fixed. See ANSWER 1.16.

References cited in the point-by-point answer to reviewers.

- Alexander DH, Novembre J, and Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19(9):1655–1664 [https://doi.org/10.1101/gr.094052.109]
- Allentoft ME, Sikora M, Sjögren KG et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–172 [https://doi.org/10.1038/nature14507]
- Anderson D (2000) *Identity and Ecology in Arctic Siberia: The Number One Reindeer Brigade* Oxford University Press
- Anderson D, Arzyutov D, and Alymov S (2019) *Life Histories of Etnos Theory in Russia and Beyond* Open Book Publishers
- Antezack A, Etchecopar-Etchart D, La Scola B et al. (2023) New putative periodontopathogens and periodontal health-associated species: A systematic review and meta-analysis. *Journal of Periodontal Research* 58(5):893–906 [https://doi.org/10.1111/JRE.13173]
- Bergström A, McCarthy SA, Hui R et al. (2020) Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367(6484) [https://doi.org/10.1126/science.aay5012]
- Biagini P, Thèves C, Balaesque P et al. (2012) Variola Virus in a 300-Year-Old Siberian Mummy. *New England Journal of Medicine* 376(21):2057–2059 [https://doi.org/10.1056/NEJMc1212281]
- Blanco-Míguez A, Beghini F, Cumbo F et al. (2023) Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology* 2023 41:11 41(11):1633–1644 [https://doi.org/10.1038/s41587-023-01688-w]
- Briggs AW, Stenzel U, Johnson PLF et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* 104(37):14616–14621 [https://doi.org/10.1073/pnas.0704665104]
- Briggs AW, Stenzel U, Meyer M et al. (2009) Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research* 38(6):e87 [https://doi.org/10.1093/NAR/GKP1163]
- Broushaki F, Thomas MG, Link V et al. (2016) Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353(6298):499–503 [https://doi.org/10.1126/science.aaf7943]

- Chintalapati M, Patterson N, and Moorjani P (2022) The spatiotemporal patterns of major human admixture events during the European Holocene. *ELife* 11 [https://doi.org/10.7554/ELIFE.77625]
- Crosby A (2009) *Ecological Imperialism: The Biological Expansion of Europe 900-1900* (second edition) Cambridge University Press
- Crubézy E, Duchesne S, Razafindrazaka H et al. (2021) Sucrose Is Not the Whole Story: Risk Factors and Oral Health at the Contact (Yakutia, Siberia-16th/19th). In *Biology* (Vol. 10, Issue 10) [https://doi.org/10.3390/biology10100974]
- Crubézy E, Melnichuk O, and Alexeev A (2020) Archaeology genetics and history 15 years of research in Yakutia (2002 – 2017). *Vestnik Arkheologii, Antropologii i Etnografii* 4(1):110–119 [https://doi.org/https://doi.org/10.20874/2071-0437-2020-51-4-11]
- Crubézy E and Nikolaeva D (2017a) Vainqueurs ou vaincus ? Odile Jacob
- Crubézy E and Nikolaeva D (2017b) Vainqueurs ou vaincus? [Winners or losers?] Odile Jacob
- Dabernat H, Thèves C, Bouakaze C et al. (2014) Tuberculosis epidemiology and selection in an autochthonous Siberian population from the 16th-19th century. *PLoS ONE* 9(2) [https://doi.org/10.1371/journal.pone.0089877]
- Damgaard P de B, Marchi N, Rasmussen S et al. (2018) 137 ancient human genomes from across the Eurasian steppes. *Nature* 557(7705):369–374 [https://doi.org/10.1038/s41586-018-0094-2]
- Damgaard PB, Margaryan A, Schroeder H et al. (2015) Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports* 5(1):11184 [https://doi.org/10.1038/srep11184]
- de Barros Damgaard P, Martiniano R, Kamm J et al. (2018) The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360(6396):eaar7711 [https://doi.org/10.1126/science.aar7711]
- De Sales A (1981) La Relation forgeron-chaman chez les Yakoutes de Sibérie. *L'Homme* 21(4):35–61
- Diamond JM (1997) *Guns, germs, and steel : the fates of human societies* (1st ed.) W.W. Norton & Co https://search.library.wisc.edu/catalog/9910007689602121
- Duchesne S, Bravina R, Popov V et al. (2020) Frozen graves of Yakutia, a chronological sequence. *Vestnik Arkheologii, Antropologii i Etnografii* 4(51):120–130 [https://doi.org/10.20874/2071-0437-2020-51-4-11]
- Duchesne S, Romanova L, Gérard P et al. (2025) Beyond the Permafrost: Excavated Tombs of Eastern Siberia Archaeological data from the French-Sakha missions Zenodo
- Eisenhofer R, Kanzawa-Kiriyama H, Shinoda KI et al. (2020) Investigating the demographic history of Japan using ancient oral microbiota. *Philosophical Transactions of the Royal Society B* 375(1812) [https://doi.org/10.1098/RSTB.2019.0578]
- Fedorova S and Barashkova N (2022) Issledovaniia populiatsionnoi genetiki naseleniia Respubliki Sakha (Iakutiia)[Population-Genetic Studies of the Population of the Sakha Republic (Yakutia)] Nauka
- Feldman M, Fernández-Domínguez E, Reynolds L et al. (2019) Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nature Communications* 10(1):1218 [https://doi.org/10.1038/s41467-019-09209-7]
- Fellows Yates JA, Andrades Valtueña A, Vågene ÅJ et al. (2021) Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. *Scientific Data* 2021 8:1 8(1):1–8 [https://doi.org/10.1038/s41597-021-00816-y]
- Fernandes DM, Sirak KA, Ringbauer H et al. (2020) A genetic history of the pre-contact Caribbean. *Nature* 2020 590:7844 590(7844):103–110 [https://doi.org/10.1038/s41586-020-03053-2]

- Ferret C (2010) Ferret, C. (2010). Une civilisation du cheval. Les usages de l'équidé, de la steppe à la taïga [A civilization of the horse. The uses of the equid, from the steppe to the taiga] (Belin)
- Fischer RH (1943) The russian fur trades, 1500-1700. University of California Press.
- Flegontov P, Altınışık NE, Changmai P et al. (2019) Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570(7760):236–240
[<https://doi.org/10.1038/s41586-019-1251-y>]
- Forsyth J (1992) A history of the peoples of Siberia. Russia's north Asian colony 1581–1990. Cambridge University Press
- Fox R (1967) Kinship and marriage : an anthropological perspective Cambridge University Press
- Fu Q, Hajdinjak M, Moldovan OT et al. (2015) An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524(7564):216–219
[<https://doi.org/10.1038/NATURE14558>]
- Fu Q, Posth C, Hajdinjak M et al. (2016) The genetic history of Ice Age Europe. *Nature* 534(7606):200–205 [<https://doi.org/10.1038/nature17993>]
- Gamba C, Jones ER, Teasdale MD et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* 2014 5:1 5(1):1–9
[<https://doi.org/10.1038/ncomms6257>]
- Günther T, Malmström H, Svensson EM et al. (2018) Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biology* 16(1):e2003703 [<https://doi.org/10.1371/JOURNAL.PBIO.2003703>]
- Hall V, Collins MD, Lawson PA et al. (2005) *Actinomyces dentalis* sp. nov., from a human dental abscess. *International Journal of Systematic and Evolutionary Microbiology* 55(1):427–431
[<https://doi.org/10.1099/IJS.0.63376-0/CITE/REFWORKS>]
- Harney E, Patterson N, Reich D et al. (2021) Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* 217(4)
[<https://doi.org/10.1093/genetics/iyaa045>]
- Hoang DT, Chernomor O, Von Haeseler A et al. (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35(2):518–522
[<https://doi.org/10.1093/MOLBEV/MSX281>]
- Hofmanová Z, Kreutzer S, Hellenthal G et al. (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* 113(25):6886–6891
[https://doi.org/10.1073/PNAS.1523951113/SUPPL_FILE/PNAS.1523951113.SD03.XLSX]
- Humphrey C and Sneath D (1999) The End of Nomadism? Society, State and the Environment in Inner Asia. Duke University Press.
- Ivanov V (1974) Istoricheskoe i etnograficheskoe issledovanie Iakutii XVII–XVIII vv. [Historical and Ethnographic Study of Yakutia in the 17th–18th Centuries] Nauka
- Järve M, Saag L, Scheib CL et al. (2019) Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* 29(14):2430–2441 [<https://doi.org/10.1016/j.cub.2019.06.019>]
- Jeong C, Balanovsky O, Lukianova E et al. (2019) The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution* 3(6):966–976 [<https://doi.org/10.1038/s41559-019-0878-2>]
- Jeong C, Ozga AT, Witonsky DB et al. (2016) Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proceedings of the National Academy of Sciences* 113(27):7485–7490 [<https://doi.org/10.1073/pnas.1520844113>]

- Jeong C, Wang K, Wilkin S et al. (2020) A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* 183(4):890-904.e29 [https://doi.org/10.1016/j.cell.2020.10.015]
- Jochelson W (1926) The Yukaghir and the Yukaghirized Tungus. In F. Boas (Ed.), *Memoir of the American Museum of Natural History*; [v. 9] E.J. Brill, G.E Stechert
http://hdl.handle.net/2246/26
- Jochelson W (1933) The Yakut: Vol. XXXIII Par The American Museum of Natural History
- Jones ER, Gonzalez-Fortes G, Connell S et al. (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications* 6:1–8 [https://doi.org/10.1038/ncomms9912]
- Jónsson H, Ginolhac A, Schubert M et al. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684 [https://doi.org/10.1093/bioinformatics/btt193]
- Kabanovitch R (1952) Iz istorii bor'by s tuberkulezom [From the History of the Fight Against Tuberculosis] Izdatel'stvo Akademii meditsinskikh nauk SSSR
- Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772–780 [https://doi.org/10.1093/MOLBEV/MST010]
- Kazarina A, Petersone-gordina E, Kimsis J et al. (2021) The Postmedieval Latvian Oral Microbiome in the Context of Modern Dental Calculus and Modern Dental Plaque Microbial Profiles. *Genes* 2021, Vol. 12, Page 309 12(2):309 [https://doi.org/10.3390/GENES12020309]
- Khazanov AM (1994) Nomads and the Outside World. (2nd ed.) University of Wisconsin Press
- Kılınç GM, Kashuba N, Koptekin D et al. (2021) Human population dynamics and *Yersinia pestis* in ancient northeast Asia. *Science Advances* 7(2):1–14 [https://doi.org/10.1126/sciadv.abc4587]
- Kılınç GM, Omrak A, Özer F et al. (2016) The Demographic Development of the First Farmers in Anatolia. *Current Biology* 26(19):2659–2666 [https://doi.org/10.1016/j.cub.2016.07.057]
- Kononova S, Vinokurova D, Barashkov NA et al. (2021) The attitude of young people in the city of Yakutsk to DNA-testing. *International Journal of Circumpolar Health* 80(1):1973697 [https://doi.org/10.1080/22423982.2021.1973697]
- Lamnidis TC, Majander K, Jeong C et al. (2018) Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications* 9(1):5018 [https://doi.org/10.1038/s41467-018-07483-5]
- Lazaridis I, Patterson N, Mitnik A et al. (2013) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413 [https://doi.org/10.1038/nature13673]
- Lazaridis IP, Nadel D, Rollefson G et al. (2016) Genomic insights into the origin of farming in the ancient Near East. *Nature* 536(7617):419–424 [https://doi.org/10.1038/nature19310]
- Levi-Strauss C (1969) *The Elementary Structures of Kinship* Beacon Press
- Lipson M, Szécsényi-Nagy A, Mallick S et al. (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551(7680):368–372 [https://doi.org/10.1038/nature24476]
- Macé M, Richeval C, Alcouffe A et al. (2021) At the Origins of Tobacco-Smoking and Tea Consumption in a Virgin Population (Yakutia, 1650-1900 A.D.): Comparison of Pharmacological, Histological, Economic and Cultural Data. *Biology* 10(12):1–26 [https://doi.org/10.3390/biology10121271]
- Mallick S, Li H, Lipson M et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206 [https://doi.org/10.1038/nature18964]

- Mallick S, Micco A, Mah M et al. (2024) The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data* 2024 11:1 11(1):1–10
[https://doi.org/10.1038/s41597-024-03031-7]
- Mallick S and Reich D (2023) The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes, V5 (V8 ed.) Harvard Dataverse
[https://doi.org/doi/10.7910/DVN/FFIDCW]
- Mao X, Zhang H, Qiao S et al. (2021) The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* 184(12):3256–3266.e13
[https://doi.org/10.1016/j.cell.2021.04.040]
- Margaryan A, Hansen HB, Rasmussen S et al. (2018) Ancient pathogen DNA in human teeth and petrous bones. *Ecology and Evolution* 8(6):3534–3542
[https://doi.org/https://doi.org/10.1002/ece3.3924]
- Mathieson I, Alpaslan-Roodenberg S, Posth C et al. (2018) The genomic history of southeastern Europe. *Nature* 555(7695):197–203 [https://doi.org/10.1038/nature25778]
- Mathieson I, Lazaridis I, Rohland N et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503 [https://doi.org/10.1038/nature16152]
- McColl H, Racimo F, Vinner L et al. (2018) The prehistoric peopling of Southeast Asia. *Science* 361(6397):88–92 [https://doi.org/10.1126/science.aat3628]
- Meyer M, Kircher M, Gansauge M-T et al. (2012) A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338(6104):222–226
[https://doi.org/10.1126/science.1224344]
- Middendorff AF (1878) *Sibir' i severnye Yakuty* [Siberia and the Northern Yakuts] Tipografiya Imperatorskoi Akademii Nauk
- Mitnik A, Wang C-C, Pfengle S et al. (2018) The genetic prehistory of the Baltic Sea region. *Nature Communications* 9(1):442 [https://doi.org/10.1038/s41467-018-02825-9]
- Modi A, Attolini D, Zaro V et al. (2023) Combined metagenomic and archaeobotanical analyses on human dental calculus: A cross-section of lifestyle conditions in a Copper Age population of central Italy. *Quaternary International* 653–654:69–81
[https://doi.org/10.1016/J.QUAINT.2021.12.003]
- Narasimhan VM, Patterson N, Moorjani P et al. (2019) The formation of human populations in South and Central Asia. *Science* 365(6457):eaat7487 [https://doi.org/10.1126/science.aat7487]
- Nguyen LT, Schmidt HA, Von Haeseler A et al. (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32(1):268–274 [https://doi.org/10.1093/MOLBEV/MSU300]
- Ning C, Li T, Wang K et al. (2020) Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nature Communications* 2020 11:1 11(1):1–9
[https://doi.org/10.1038/s41467-020-16557-2]
- Ning C, Wang CC, Gao S et al. (2019) Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan. *Current Biology* 29(15):2526–2532.e4
[https://doi.org/10.1016/J.CUB.2019.06.044/ATTACHMENT/6613EC96-18A0-42FF-81A6-A70A7D54DDAF/MMC4.PDF]
- Okladnikov A (1955) *Yakutiya do prisoyedineniya k Russkomu gosudarstvu* // *Istoriya Yakutskoy ASSR* [Yakutia Before Its Incorporation into the Russian State // History of the Yakut ASSR]
- Okladnikov A (1970) *Iakuty: Proiskhozhdenie naroda i ego kul'tury* [The Yakuts: The Origin of the People and Their Culture] Nauka

- Olalde I, Brace S, Allentoft ME et al. (2018) The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555(7695):190–196 [https://doi.org/10.1038/nature25738]
- Omrak A, Günther T, Valdiosera C et al. (2016) Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* 26(2):270–275 [https://doi.org/10.1016/j.cub.2015.12.019]
- Orlando L, Allaby R, Skoglund P et al. (2021) Ancient DNA analysis. *Nature Reviews Methods Primers* 2021 1:1 1(1):1–26 [https://doi.org/10.1038/s43586-020-00011-0]
- Ottoni C, Borić D, Cheronet O et al. (2021) Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. *Proceedings of the National Academy of Sciences* 118(32) [https://doi.org/10.1073/pnas.2102116118]
- Pakendorf B (2007) Contact in the prehistory of the Sakha (Yakuts):Linguistic and genetic perspectives *Lot*
- Patterson N, Moorjani P, Luo Y et al. (2012) Ancient Admixture in Human History. *Genetics* 192(3):1065–1093 [https://doi.org/10.1534/genetics.112.145037]
- Patterson N, Price AL, and Reich D (2006) Population Structure and Eigenanalysis. *PLoS Genetics* 2(12):e190 [https://doi.org/10.1371/journal.pgen.0020190]
- Pedersen JS, Valen E, Velazquez AMV et al. (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research* 24(3):454–466 [https://doi.org/10.1101/GR.163592.113]
- Pinhasi R, Zarina G, Moiseyev V et al. (2017) The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology* 27(4):576–582 [https://doi.org/10.1016/j.cub.2016.12.060]
- Popov AA (1964) Yakuty: Opyt etnograficheskogo issledovaniia [The Yakuts: An Ethnographic Study] Izdatel'stvo Akademii Nauk SSSR
- Prüfer K, Racimo F, Patterson N et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49 [https://doi.org/10.1038/nature12886]
- Raghavan M, Skoglund P, Graf KE et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91 [https://doi.org/10.1038/nature12736]
- Raghavan M, Steinrücken M, Harris K et al. (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250) [https://doi.org/10.1126/SCIENCE.AAB3884/SUPPL_FILE/RAGHAVAN-SM.PDF]
- Rasmussen M, Anzick SL, Waters MR et al. (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 2014 506:7487 506(7487):225–229 [https://doi.org/10.1038/nature13025]
- Rasmussen M, Li Y, Lindgreen S et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762 [https://doi.org/10.1038/nature08835]
- Ringbauer H, Huang Y, Akbari A et al. (2023) Accurate detection of identity-by-descent segments in human ancient DNA. *Nature Genetics* 2023 56:1 56(1):143–151 [https://doi.org/10.1038/s41588-023-01582-w]
- Rohland N, Harney E, Mallick S et al. (2015) Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1660):20130624 [https://doi.org/10.1098/rstb.2013.0624]
- Romanova L, Balter V, Simon L et al. (2019) Diet of autochthonous populations in Yakutia using isotopic , ethnographic , historical and archaeological data. *Journal of Archaeological Science: Reports* 28(August):102022 [https://doi.org/10.1016/j.jasrep.2019.102022]

- Rubinacci S, Ribeiro DM, Hofmeister RJ et al. (2021) Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics* 53(1):120–126 [https://doi.org/10.1038/s41588-020-00756-0]
- Seguin-Orlando A, Gamba C, Sarkissian C Der et al. (2015) Pros and cons of methylation-based enrichment methods for ancient DNA. *Scientific Reports* 5 [https://doi.org/10.1038/SREP11826]
- Serochevskii VL (1993) *Âkuty. Opyt ètnograficheskogo issledovaniâ* [Yakuts: An Attempt at an Ethnographic Study], (second edi)
- Seroshevsky VL (1993) *Yakuti. Opit etnograficheskogo issledovaniâ* [The Yakuts. The Experience of the Ethnographical Research] ROSSPEN
- Sikora M, Pitulko V V., Sousa VC et al. (2019) The population history of northeastern Siberia since the Pleistocene. *Nature* 570(7760):182–188 [https://doi.org/10.1038/s41586-019-1279-z]
- Sikora M, Seguin-Orlando A, Sousa VC et al. (2017) Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* 358(6363):659–662 [https://doi.org/10.1126/science.aao1807]
- Skoglund P, Mallick S, Bortolini MC et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature* 525:7567 525(7567):104–108 [https://doi.org/10.1038/nature14895]
- Skoglund P, Malmstrom H, Omrak A et al. (2014) Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* 344(6185):747–750 [https://doi.org/10.1126/science.1253448]
- Skoglund P, Northoff BH, Shunkov M V. et al. (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences* 111(6) [https://doi.org/10.1073/pnas.1318934111]
- Skoglund P, Posth C, Sirak K et al. (2016) Genomic insights into the peopling of the Southwest Pacific. *Nature* 538:7626 538(7626):510–513 [https://doi.org/10.1038/nature19844]
- Socransky SS, Haffajee AD, Cugini MA et al. (1998) Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology* 25(2):134–144 [https://doi.org/10.1111/J.1600-051X.1998.TB02419.X]
- The 1000 Genomes Project Consortium, Auton A, Abecasis GR et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74 [https://doi.org/10.1038/nature15393]
- Tikhonov D, Gurkan C, Demirdov K et al. (2019) On the origins of the Sakhas' paternal lineages: Reconciliation of population genetic / ancient DNA data, archaeological findings and historical narratives. *Siberian Research* 1(1):91–111 [https://doi.org/10.33384/26587270.2019.01.004e]
- Tokarev S (1953) *Istoriia Iakutii v XVII veke* [History of Yakutia in the 17th Century].
- Truong DT, Tett A, Pasoli E et al. (2017) Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Research* 27(4):626–638 [https://doi.org/10.1101/GR.216242.116]
- Van Dyke TE, Bartold PM, and Reynolds EC (2020) The Nexus Between Periodontal Inflammation and Dysbiosis. *Frontiers in Immunology* 11:530286 [https://doi.org/10.3389/FIMMU.2020.00511/BIBTEX]
- Velsko IM, Fellows Yates JA, Aron F et al. (2019) Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* 7(1):1–20 [https://doi.org/10.1186/S40168-019-0717-3/FIGURES/9]
- Velsko IM, Semerau L, Inskip SA et al. (2022) Ancient dental calculus preserves signatures of biofilm succession and interindividual variation independent of dental pathology. *PNAS Nexus* 1(4):1–14 [https://doi.org/10.1093/PNASNEXUS/PGAC148]

- Vitashevskii NA (1910) Brak i rodstvo u iakutov [Marriage and Kinship among the Yakuts] Tipografiia Imperatorskoi Akademii Nauk
- Wang C-C, Reinhold S, Kalmykov A et al. (2019) Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nature Communications* 10(1):590 [https://doi.org/10.1038/s41467-018-08220-8]
- Wang C-C, Yeh H-Y, Popov AN et al. (2021) Genomic insights into the formation of human populations in East Asia. *Nature* 591(7850):413–419 [https://doi.org/10.1038/s41586-021-03336-2]
- Weyrich LS, Duchene S, Soubrier J et al. (2017) Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 2017 544:7650 544(7650):357–361 [https://doi.org/10.1038/nature21674]
- White R (2010) *The Middle Ground. Indians, Empires, and Republics in the Great Lakes Region, 1650–1815* Cambridge University Press
- Yaka R, Mapelli I, Kaptan D et al. (2021) Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Current Biology* 31:1–14 [https://doi.org/10.1016/j.cub.2021.03.050]
- Yang MA, Fan X, Sun B et al. (2020) Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369(6501):282–288 [https://doi.org/10.1126/SCIENCE.ABA0909/SUPPL_FILE/PAPV2.PDF]
- Yang MA, Gao X, Theunert C et al. (2017) 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Current Biology* 27(20):3202–3208.e9 [https://doi.org/10.1016/j.cub.2017.09.030]
- Yates JAF, Velsko IM, Aron F et al. (2021) The evolution and changing ecology of the African hominid oral microbiome. *PNAS* 118(20):1–11 [https://doi.org/10.1073/pnas.2021655118]
- Yu H, Spyrou MA, Karapetian M et al. (2020) Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. *Cell* 181(6):1232–1245.e20 [https://doi.org/10.1016/j.cell.2020.04.037]
- Zerjal T, Dashnyam B, Pandya A et al. (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *American Journal of Human Genetics* 60(5):1174–1183
- Zhang F, Ning C, Scott A et al. (2021) The genomic origins of the Bronze Age Tarim Basin mummies. *Nature* November 2020:0–3 [https://doi.org/10.1038/s41586-021-04052-7]
- Zlojutro M, Tarskaia LA, Sorensen M et al. (2009) Coalescent simulations of Yakut mtDNA variation suggest small founding population. *American Journal of Physical Anthropology* 139(4):474–482 [https://doi.org/10.1002/ajpa.21003]
- Zvénigorosky V, Duchesne S, Romanova L et al. (2020) The genetic legacy of legendary and historical Siberian chieftains. *Communications Biology* 3(1):1–9 [https://doi.org/10.1038/s42003-020-01307-3]

Appendix 1.

Poster of the exhibition organized at the NEFU Archaeological Museum between March 2nd and 20th 2023.

Музей археологии и этнографии СВФУ
Высшая школа мастерства "Уран Иис" ИЯКН СВФУ
Институт гуманитарных исследований и проблем малочисленных народов Севера СО РАН
Союз художников Якутии
НПО "Өбүгэ суола / Путь предков"
Объединение мастериц "Сунтаар ситимэ" Сунтарского улуса

Творческая выставка
объединения мастериц
"Сунтаар ситимэ"
Сунтарского улуса

СҮКТҮҮ

СВАДЕБНАЯ
ОДЕЖДА XIX ВЕКА



Открытие 2 марта в 15:00
Сроки выставки: 2 - 20 марта 2023 г.
Музей археологии и этнографии СВФУ,
Кулаковского, 48, правое крыло
Справки по телефону 49-68-41

+0

Appendix 2.

Poster of the exhibition organized as part of the CNRS 80th anniversary, and presented at the Historical Park Russia – My history (Yakutsk), from September 1st 2019 to the end of the year 2019.



**РОССИЯ
МОЯ ИСТОРИЯ**

**LIBERTÉ • ÉGALITÉ • FRATERNITÉ
REPUBLIQUE FRANÇAISE
MINISTÈRE
DES CULTURES ET DES
AFFAIRES ÉTRANGÈRES**

**Université Fédérale
Sakha (Yakutia)**

**INSTITUT POLAIRE
FRANÇOIS VICTOR**

SB RAS

1887

**UNIVERSITÉ
TOULOUSE III
PAUL SABATIER**

AMIS

&C

ВЫСТАВОЧНЫЙ ПРОЕКТ

**К ПРОПХОЖУЕНИЮ
ЯКУТОВ:**

**ФОТОВЫСТАВКА
АРХЕОЛОГА П. ЖЕРАРА**

**ВЫСТАВКА МАТЕРИАЛОВ
ФРАНКО-РОССИЙСКИХ
АРХЕОЛОГИЧЕСКИХ
РАСКОПОК
с 2002 по 2019 гг.**

**ПУБЛИЧНАЯ
ЛЕКЦИЯ С ДИСКУССИЕЙ**

**ОТКРЫТИЕ И ЛЕКЦИЯ
1 сентября 2019 г.
в 15.00**

ВХОД СВОБОДНЫЙ

Благодарим за помощь в организации выставки:
Исторический парк «Россия – моя история», Россия
Институт гуманитарных исследований и проблем малочисленных народов Севера СО РАН, Россия
Институт зарубежной филологии и регионоведения СВФУ, Россия
Музей археологии, этнографии и высшей школы СВФУ, Россия
Национальный центр научных исследований (CNRS), Франция
Французский полярный институт Поль Эмиль Виктор (IPEV), Франция
Университет Тулузы 3 им. П. Сабатье (UPS), Франция
Национальный институт превентивной археологии (INRAP), Франция

Appendix 3.

Photographs of the inauguration and exhibition titled *"At the Origins of the Yakuts: Franco-Russian Archaeological Excavations 2002-2019"*, presented in Yakutsk at the *"Russia—My History"* Museum. The exhibition, a grand display of one hundred major artifacts unearthed during our excavations, was a sight to behold. Before the inauguration, with Iakut colleagues, we held a public conference attended by over 200 people. The exhibition ran from September 2019 to December 31, 2019.



Appendix 4.

Field photographs illustrating our interactions with local communities. Reviewers can pay particular attention to the detailed and informative explanations provided to local populations on archaeological sites and in official village buildings, enriching their understanding of the excavations conducted on their territories.



Reviewer #1

Overall, Crubezy et al. have done a good job of addressing my comments on the previous version of the manuscript. I am restricting my comments to the oral microbiome, pathogen screening, and phylogenetic analyses, as the other analyses are beyond my expertise. To me, the revised manuscript seems to be much improved. I have a couple of minor concerns which should be addressed.

POINT 1.1:

Are there any cases where StrainPhlAn identified closely related strains within the same metagenome for the six bacterial species? How are such instances of multiple strains present within the same individual being accounted for when building the bacterial phylogenetic trees? We often see such examples of multiple strains in oral metagenome data and their presence will impact the placement of those samples in the tree.

RESPONSE 1.1:

By default, and proceeding one sample at a time, the methodology implemented in StrainPhlAn retains only those species-specific marker genes showing a breadth-of-coverage of at least 80%, considering nucleotide positions covered by a minimum of eight high-quality read alignments, and for which the dominant allele is found at a frequency above 80%. Positions below this threshold are masked and, thus, are not used for phylogenetic reconstruction. Therefore, the method is tailored to detect the presence of a single dominant strain at most per metagenome. It follows that there are no instances in which we detected multiple strains with single individuals. For transparency, we have developed the corresponding Methods section accordingly (SI, Section 2.15, lines 560-564, pages 49):

“StrainPhlAn4 was run with default parameters to extract species-specific MetaPhlAn4 markers per extract. We removed any sample with less than 80% of the markers covered, and kept all markers showing a breadth-of-coverage of more than 80%, considering nucleotide positions covered by a minimum of eight high-quality read alignments, and for which the dominant allele is found at a frequency above 80%.”

POINT 1.2:

For the six bacterial phylogenies, please specify how many clade-specific genes each tree is based on, how many data-points i.e. number of sites in each alignment, and percentage of missing data allowed.

RESPONSE 1.2:

We have reported in the captions of each corresponding figure the information requested, including: the substitution model used; the number of biomarkers underlying each bacterial phylogenetic tree, as well as; the number of sites in each alignment, with missingness rates (average, and range). See Supplementary Information, section 2.15, pages 50-54:

Figure S2.1 (Next page). Strain-level analysis of species *Treponema denticola* and *Porphyromonas gingivalis*.

A. The Maximum Likelihood (ML) phylogenetic tree for *T. denticola* was reconstructed using IQtree 1.6.12¹²³, with the TVMe+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 24 Yakut individuals, plus the sequence of the reference genome. In total, the consensus sequence alignment included 126 StrainPhlAn biomarkers, spanning 1,209 nucleotide positions, with an average individual missingness of 20.68% (range=15.14-34.90%). **B.** The ML phylogenetic tree for *P. gingivalis* was reconstructed using IQtree 1.6.12¹²³, with the TVM+F+ASC+G4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 9 ancient Yakut individuals, supplemented with the data from one historical human, plus the sequence of the reference genome. In total, the consensus sequence alignment included 52 StrainPhlAn biomarkers, spanning 686 nucleotide positions, with an average individual missingness of 11.46% (range=1.02- 23.91%). Trees are rooted at mid-points. Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶.

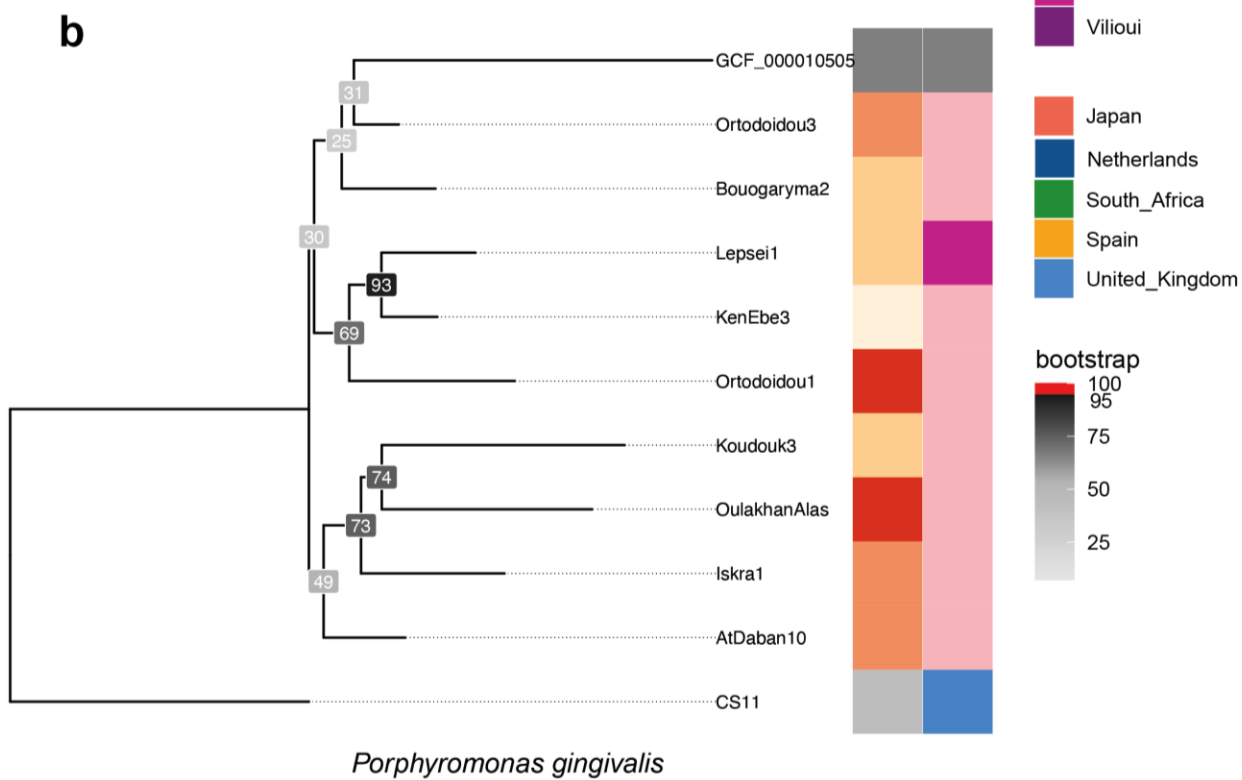
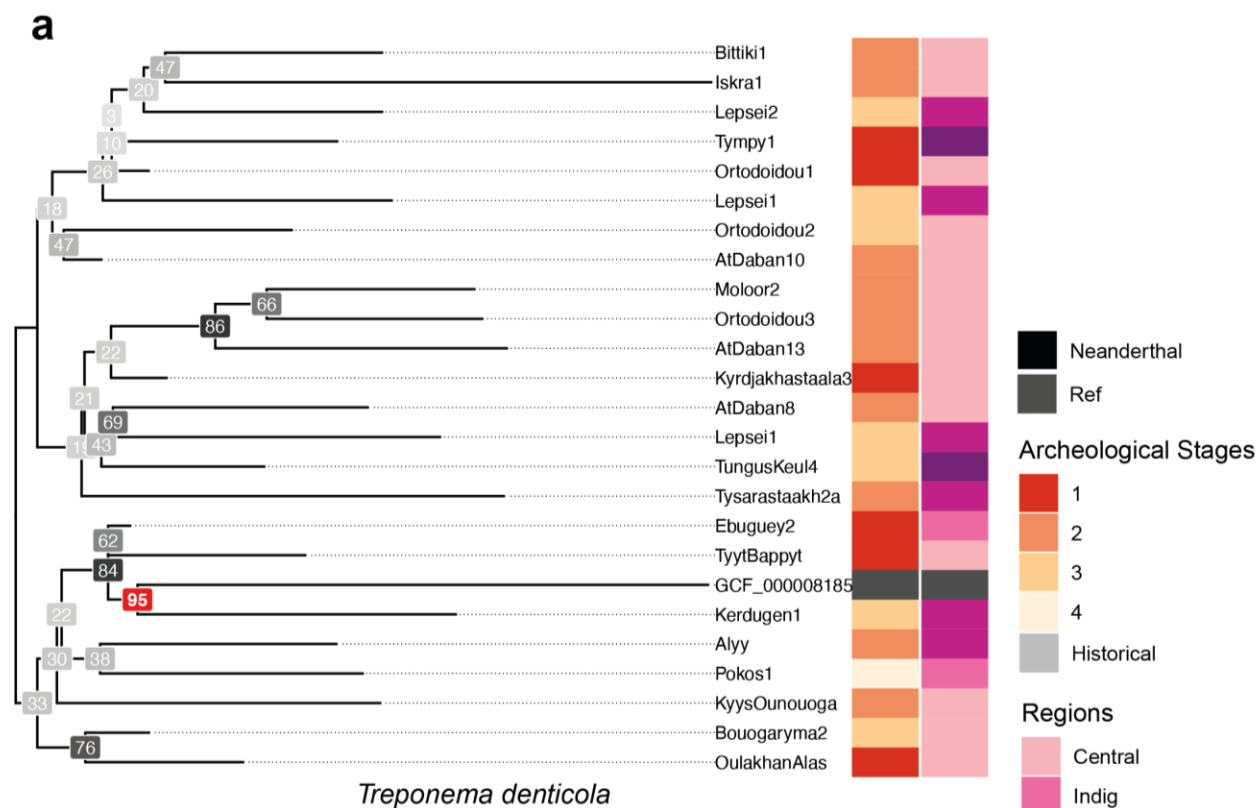




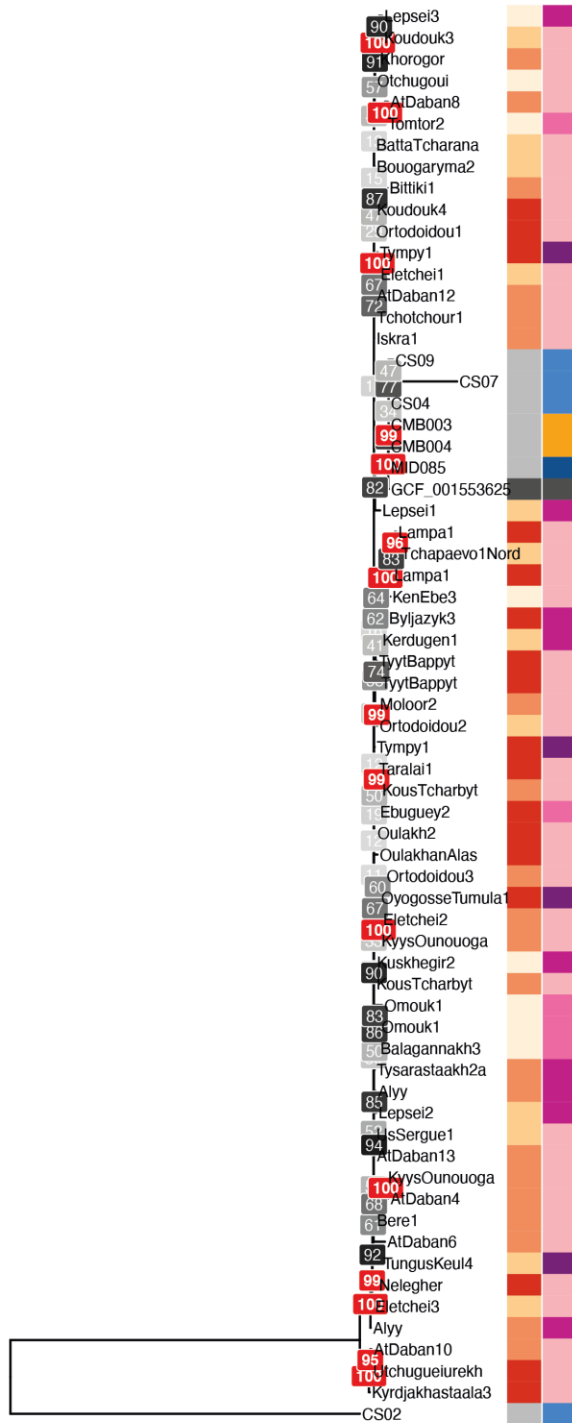
Figure S2_2 (Previous page). Strain-level analysis of *Actinomyces dentalis* and *Desulfobulbus oralis*.

A. The ML phylogenetic trees for *A. dentalis* was reconstructed using IQtree 1.6.12¹²³, with the TIM+F+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 84 ancient Yakut extracts, supplemented with the data from three Neanderthal outgroups and 19 historical humans, plus the sequence of the reference genome. In total, the consensus sequence alignment included 193 StrainPhlAn biomarkers, spanning 1,656 nucleotide positions, with an average individual missingness of 29.08% (range=7.61%- 36.59%). **B.** The Maximum Likelihood phylogenetic tree for *D. oralis* was reconstructed using IQtree 1.6.12¹²³, with the TVM+F+ASC+R4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 73 ancient Yakut extracts, supplemented with the data from six Neanderthal outgroups and 22 historical humans, plus the sequence of the reference genome. In total, the consensus sequence alignment included 196 StrainPhlAn biomarkers, spanning 4,340 nucleotide positions, with average individual missingness of 6.91% (range=1.47- 34.15%). Trees are rooted using the Neanderthal samples as outgroups. Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶. Individuals appearing more than once in the phylogeny correspond to individuals for which different archaeological materials were analysed, such as tooth roots and dental calculus.

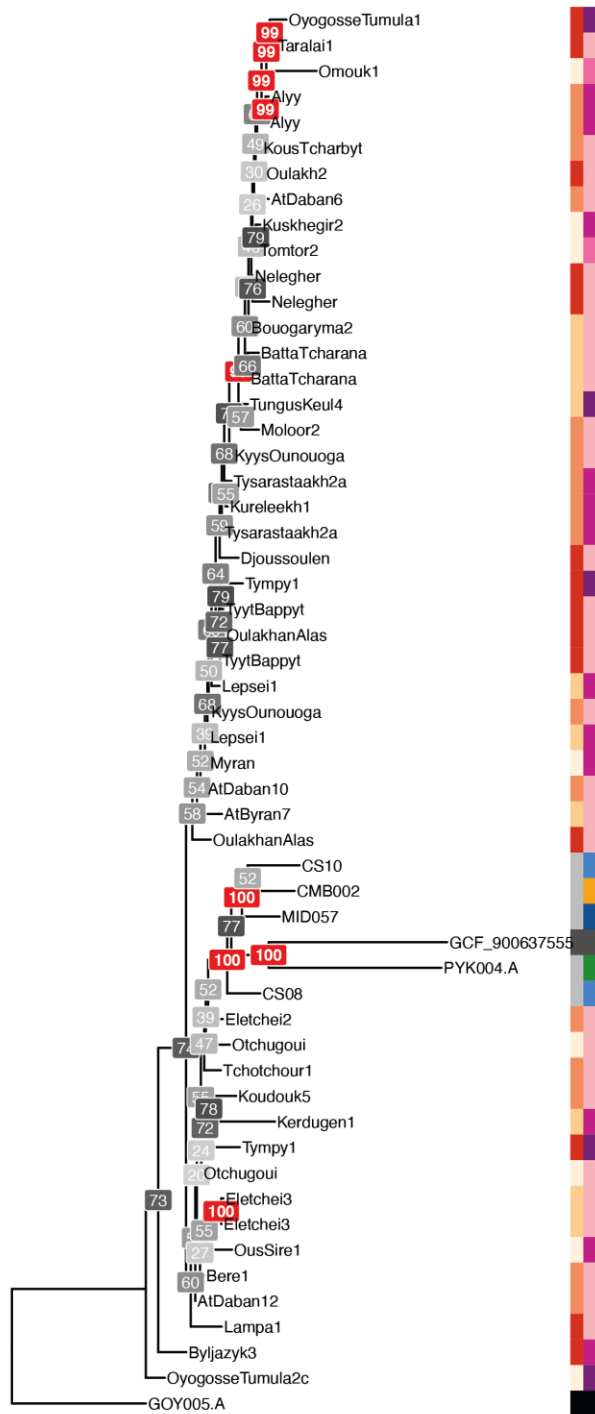
Figure S2_3 (Next page). Strain-level analysis of *Desulfomicrobium orale* and *Lautropia mirabilis*.

A. The Maximum Likelihood phylogenetic tree of *D. orale* was reconstructed using IQtree 1.6.12¹²³, with the TIM+F+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 58 ancient Yakut extracts, supplemented with the data from seven historical humans, plus the sequence of the reference genome. The tree is rooted at mid-point. In total, the consensus sequence alignment included 195 StrainPhlAn biomarkers, spanning 5,048 nucleotide positions, with an average individual missingness of 3.72% (range=0.00- 20.90 %). **B.** The Maximum Likelihood phylogenetic tree for *L. mirabilis* was reconstructed using IQtree 1.6.12¹²³, with the TVM+F+ASC+R4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 48 ancient Yakut extracts, supplemented with the data from one Neanderthal outgroup and six historical humans, plus the sequence of the reference genome. The tree is rooted using the Neanderthal sample as outgroup. In total, the consensus sequence alignment included 186 StrainPhlAn biomarkers, spanning 6,916 nucleotide positions, with an average individual missingness of 14.82% (range=0.10- 34.69%). Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶. Individuals appearing more than once in the phylogeny correspond to individuals for which different archaeological materials were analysed, such as tooth roots and dental calculus.

a

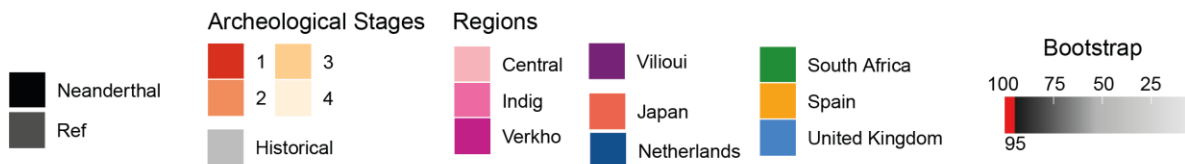


b



Desulfomicrobium orale

Lautropia mirabilis



POINT 1.3:

The circular trees shown in figures S2 15-17 are a bit difficult to interpret. I recommend showing these as regular bracket-style trees. For some of the trees, not many of the branches seem to have more than 95% bootstrap support, which is the recommended threshold when using ultrafast bootstrapping. This should be mentioned while interpreting the results.

RESPONSE 1.3:

We have followed the reviewer's recommendation and we now show non-circular trees on Supplementary Figures S2_15-17. We have also edited the figures to highlight those nodes showing minimal bootstrap support of 95%, which helps readers focus their attention on those most reliable nodes. We, however, note that PhyloPhlAn phylogenetic strain analyses are commonly based on non-bootstrapped phylogenies (cf. Asnicar, F., Thomas, A.M., Beghini, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11, 2500 (2020). <https://doi.org/10.1038/s41467-020-16366-7>). Therefore, the following warning was added to the corresponding figure captions:

Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶.

See Response 1.2 for the figures.

Reviewer #2

I am happy with the extensive changes made in response to my comments and the thorough explanations provided. It is especially nice to see the maps of historical Indigenous and Russian settlements as they help to clarify the demographic situation (Figure S1_1), and to see the very extensive collection of primary archaeological data posted at Zenodo. This a great resource for future researchers!

I agree that all genetic analyses for present-day Yakuts (published and unpublished ones, cited in answer 2.2) demonstrate very low levels of Russian ancestry, supporting the authors' claim that almost no admixture happened between ancient Yakuts and Russian colonizers (at least outside of the Russian settlements in the south of Yakutia, which were not excavated). This, and the much more detailed supplemental section on the sampling strategy, addresses my main concern (point 2.2).

POINT 2.1:

However, I'm still unconvinced that the preferred qpAdm models for the admixture between the migrating Yakuts and the local Iron Age population of Yakutia are robust. Namely, I am puzzled by the rather low fraction of trans-Baikal ancestry estimated, 16-38% (point 2.12), and the high fraction of indigenous IA ancestry (62-84%). The relevant preprint which I referred to is now published (Flegontova et al. 2025, Performance of qpAdm-based screens for genetic admixture on graph-shaped histories and stepping stone landscapes. doi: 10.1093/genetics/iyaf047), and I'd like to highlight few points below.

RESPONSE 2.1:

The work from Flegontova et al., to which the reviewer has contributed, represents an important contribution to understanding limitations of qpADM modeling, especially regarding the importance of spatial and temporal sampling design and population histories deviating from simple scenarios involving discrete admixture pulses. These authors reveal serious limitations in cases of spatially, continuous landscapes, such as those resulting following stepping-stone migration and isolation by distance, including mis-leading-confidence in wrong population models and admixture proportions bias. Nonetheless, we would like to stress that other analyses than qpADM, such as ADMIXTURE, are consistent with a predominant fraction of local (Indigenous IA)

ancestry (57-78%) in ancient and modern Yakut individuals (Fig. 2b). This contributes to build confidence in the ancestry fractions estimated by qpADM.

POINT 2.2:

1. Crucial proxy source groups used in the analysis (Yakutia_IA and Baikal_his) are very small, comprising just 2 or 4 individuals (answer 2.12), which limits the accuracy of admixture fraction estimates. I also see that other alternative source populations were very small (answer 4.13): Russia_AngaraRiver_Medieval.SG (Kılınç et al., 2021) (N=1), Mongolia_Sukhbaatar_Xiongnu (N=1) and Mongolia_Khuvsgul_MLBALateMedieval (Jeong et al., 2020) individuals (N=2) and modern Buryat (Buryat.SG (Raghavan et al., 2015; de Barros Damgaard et al., 2018), N=4). I suggest increasing the samples size for the proxy sources, if possible.

RESPONSE 2.2:

We have scrutinized the latest archive of ancient human genetic variation compiled through the Allen Ancient DNA Resource (v62.0, see <https://dataverse.harvard.edu/file.xhtml?fileId=10537416&version=9.0>). This archive includes genomic data for 9 individuals from eastern Mongolia dating from the Xiongnu period onwards, all coming from a single publication (Lee et al. 2024. Medieval genomes from eastern Mongolia share a stable genetic profile over a millennium. Hum Popul Genet Genom 4(1):0004. <https://doi.org/10.47248/hpgg2404010004>). The Xiongnu individuals analyzed, however, showed genetic profiles with inflated East Asian ancestry, reflecting the established composite confederation nature of this Empire (Lee et al. 2024). Consequently, the sample size of our proxy sources could not be extended. To reflect this limitation, we have now updated the main text as follows (page 6, lines 328-330):

“Current models rely on a limited number of genetically characterized individuals used as population sources, which likely do not capture the full extent of genetic diversity across the vast territory of Yakutia at the time of admixture.”

POINT 2.3:

2. The "right" population set (answer 2.17) includes no group even remotely close to the region/period of interest and to the most plausible ancestry sources, which may also contribute to biased admixture fraction estimates (even if their SEs reported by qpAdm are low). The "right" (or reference) set includes: an African group (N=4), a Southeast Asian group (N=15), a Native American group (N=6), an Upper Paleolithic Siberian (N=1), a Neolithic Levantine group (N=4), a Neolithic Anatolian farmer group (N=16), a Paleolithic individual from West Europe (N=1), and a Paleolithic individual from the East European Plain (N=1). In our study of qpAdm performance on admixture-graph-shaped histories and stepping stone landscapes (Flegontova et al. 2025), we conclude that qpAdm results are much more robust when sampling is dense: when all the groups participating in the analysis in any role are sampled from a relatively small region such as West Eurasia or East Eurasia, rather than just few highly non-contemporaneous groups around the world. I suggest that the authors replicate their qpAdm analysis on several different sets of "right" populations (larger and more regionally and chronologically-focused sets); that would make the results much more robust!

RESPONSE 2.3:

We thank the reviewer for his suggestion, which helped strengthen our previous conclusions. We have repeated our analyses using two alternative sets of 'right' (outgroup) populations, representing Eurasian ancestry, and more precisely East Eurasian ancestry. The first set, referred to as "set 1", includes all reference populations from our original right set ("set 0"), with the addition of WSHG (N=2), EEHG (N=5), Iran_GanjDareh_N (N=8), Russia_Ekven_OldBeringSea (N=16), Russia_BA_Okunevo.SG (N=14), and China_Upper_YR_LN (N=6). The second set, "set 2", is identical to set 1, excepting that individuals from Mbuti, Mixe, Villabruna, Ami, and Natufian populations were removed. We then attempted to model the two models previously obtained, i.e., Yakutia_IA + Baikal_his and Yakutia_IA + Baikal_sib, using each of the three sets of 'right' populations. We applied the feasibility criteria proposed by Flegontova et al. (2025), especially that (1) estimated admixture proportions ± 2 SE are within the (0, 1) interval, and (2) p-values are ≥ 0.01 , to identify those most reliable models. We found that a majority of the ancient Yakut individuals could be successfully modelled as Yakutia_IA + Baikal_his, with 'set 0' leading to the acceptance of 96.8% of the models

tested, versus 59.1% and 69.9% for ‘set 1’ and ‘set 2’, respectively. Individuals successfully modeled in more than a single set also showed consistent admixture proportions (see Table R2.3). Using Baikal_sib instead of Baikal_his as a proxy for the Baikal source returned a reduced proportion of models passing acceptance criteria and lower proportions from the Baikal source, as previously reported for ‘set 0’. These results largely confirm our initial modelling, with variation in local ancestry depending on the Baikal source, and demonstrate that the genomic makeup of a majority of the individuals conform to a simple population history in which some exogenous population best modelled as Baikal_his admixed with local Indigenous to form the historical Yakut gene pool.

‘Right’ (outgroup) set	Baikal source	Percentage of accepted models (N=93)	Range of Yakutia IA admixture coefficient
Set 0	his	96,77	49-84%
Set 0	sib	74,19	21-60%
Set 1	his	59,14	55-78%
Set 1	sib	49,46	17-44%
Set 2	his	69,89	48-77%
Set 2	sib	41,94	18-45%

Table R2.3. Summary of qpADM accepted when modelling N=93 historical Yakut individuals as a two-way admixture between Yakutia IA and Baikal sources. Three different ‘right’ (outgroup) populations sets were considered (set 0, set 1, and set 2), as well as two different Baikal sources (Baikal_his and Baikal_sib). Models were accepted if admixture coefficients ± 2 SE are within the (0, 1) interval, and when showing p -values ≥ 0.01 . Set 0 corresponds to our initial set of ‘right’ populations, while ‘set 1’ and ‘set 2’ represent larger, and more regionally and chronologically focused, alternative sets.

We have updated Table S1f, including the results of the new qpAdm models shown above. See also Response 2.5 for the updated text in the Supplementary Materials. We have also nuanced our discussion about the local contribution, to better highlight the range found with different Baikal sources and reflect the reviewer’s concerns (page 6, lines 323-332):

“This timeline, along with the broad connection with the Baikal region, concurs with oral traditions about their origins¹. However, the persistence of a substantial local genetic component contrasts with cultural and linguistic evidence pointing to a predominant central Asian Turkic-Mongol influence³. We note, however, that the estimated contribution from the local population varies depending on the Baikal source considered, underscoring the need for denser sampling across the region and time period. Current models rely on population sources defined by a limited number of genetically characterized individuals, which likely do not capture the full extent of genetic diversity across the vast territory of Yakutia at the time of admixture. Despite these limitations, our results suggest that the Yakut history represents a fusion of a local population with a migrating group, both genetically and culturally.”

POINT 2.4:

3. I agree that the ADMIXTURE results supporting the high IA ancestry estimates for Yakust are suggestive (answer 2.4), but we showed that both ADMIXTURE and qpAdm often detect ancestry from the same false sources (see Suppl. text 2 in Flegontova et al. 2025). All analyses based on allele frequencies are somehow correlated (PCA, f-statistics, qpAdm, ADMIXTURE) and should be interpreted with caution if they are not supported by other data types such as autosomal haplotypes.

RESPONSE 2.4:

We agree with the reviewer that finding consistency between analyses based on allele frequencies and haplotypes can strengthen the conclusions. This is precisely why our study design included both types of analyses. We note

that the total length of haplotype blocks donated by Mokp (Yakutia IA) is significantly larger than that donated by modern Mongolian individuals (Wilcoxon test, p -value=1.156e-13), as shown in the figure below, as well as on Extended Data Figure 3c. This aligns with the dominant IA ancestry found in the genomic makeup of historical Yakut individuals, on the basis of analyses relying on allelic frequencies.

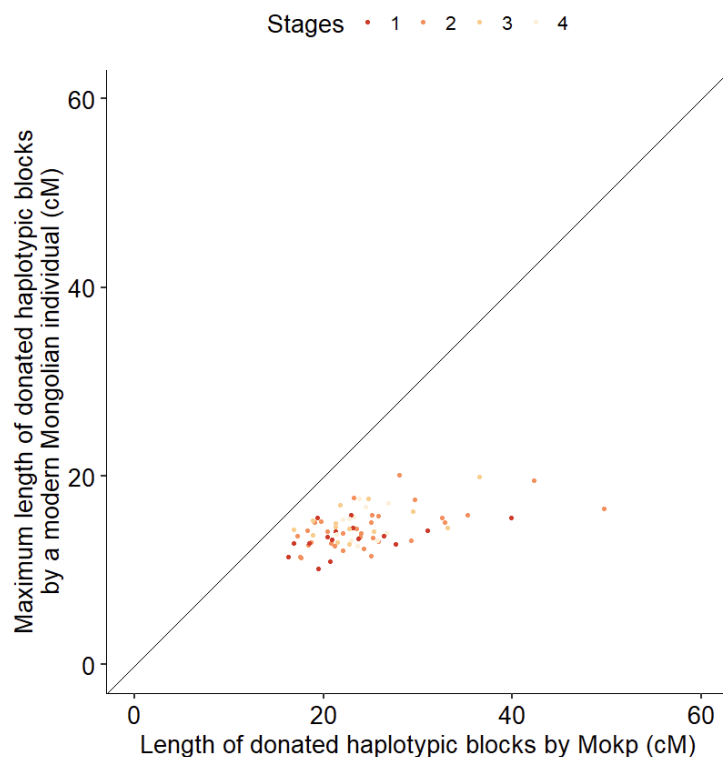


Figure R2.4: Total length of haplotype blocks donated by Mokp (IA Yakutia) and modern Mongolian individuals in each historical Yakut genome. The total length of haplotype blocks donated by modern Mongolian individuals represents the maximal estimates identified in a subset of N=6 individuals (HGDP01225, HGDP01232, HGDP01231, HGDP01228, HGDP01227, HGDP01223).

POINT 2.5:

4. Answer 2.4 "All admixture proportion estimates obtained with qpAdm were within the $[0 - 1]$ interval, whereas estimates outside these bounds are expected under the Isolation-By-Distance model from Harney et al. (Harney et al., 2021)."

This is not true since Harney et al. (2021) explored unrealistic one-dimensional stepping stone landscapes and only briefly, while Flegontova et al. focused on two-dimensional landscapes sampled in random and systematic ways. On 2D landscapes, estimated admixture fractions between 0 and 1 are typical for symmetric source arrangement around the target, but are also relatively common for non-optimal sources (those far away from the target in the geographic and genetic space) and for highly asymmetric source arrangements, which drives high false discovery rates typical for qpAdm-based admixture screens (Flegontova et al. 2025). In general, the qpAdm method is very noisy (especially p-values, but admixture fraction estimates too) and should be applied cautiously, especially if the vast parameter space of potential "right" or rotating deme sets is explored just a tiny bit (see my recommendation above).

RESPONSE 2.5:

We agree with the reviewer that our original statement needs to be nuanced, now that the study from Flegontova et al. has been published. We have, thus, added the following paragraph to reflect their findings (Supplementary Information, section 2.12, lines 410-426, page 40-41):

"Models were accepted if admixture proportion estimates ± 2 SE are within the $(0, 1)$ interval, and when p -values ≥ 0.01 . Estimates outside the $[0, 1]$ interval are frequently expected under one-dimensional Isolation-By-Distance models, such as those presented by Harney et al.¹⁰⁶. However, recent work from Flegontova et al. (2025)¹⁰⁷ has revealed that estimates within the $(0, 1)$ interval can be retrieved under more realistic scenarios of Isolation-By-Distance involving two-dimensional landscapes, even for non-

optimal sources. To reduce the risk of false positive model matches, we repeated qpAdm modeling using alternative sets of 'right' populations, each representing regionally and chronologically focused groups of interest. This approach was the only one practically feasible given that the sample size of the different sources could not be extended from data previously published, and that all detectable ancient graves within radii of just a few dozen kilometres have been systematically excavated by MAFSO in regions such as Verkhoyansk. The first set ('set 1') expanded our original reference set ('set 0') by including additional groups representing East Eurasian and West Siberian ancestries, including: WSHG (N=2), EEHG (N=5), Iran_GanjDareh_N (N=8), Russia_Ekven_OldBeringSea (N=16), Russia_BA_Okunovo.SG (N=14), and China_Upper_YR_LN (N=6). The second set ('set 2') was derived from 'set 1', but excluded Mbuti, Mixe, Villabruna, Ami, and Natufian, in order to reduce potential bias deriving from the presence of deep or divergent ancestries. Both sets yielded results consistent with those obtained using our initial set ('set 0'). In particular, the genomic makeup of the majority of historical Yakut individuals could be successfully modelled as a two-way admixture between Yakutia_LA and Baikal_his sources, generally with greater admixture proportions from the local Indigenous population (Yakutia_LA)."

POINT 2.6:

5. On the other hand, I do appreciate that no extensive exploration of hundreds of alternative qpAdm models was attempted by the authors who instead chose to focus on the most historically plausible models (answer 4.13). High-throughput model exploration is dangerous since prestudy odds are very low, and false discovery rates of such protocols often reach 90% (Flegontova et al. 2025).

RESPONSE 2.6:

We agree with the reviewer, and were glad to see that best practices recommended by Flegontova et al. (2025) included to focus only on a small number of qpAdm models that are biologically plausible and grounded in prior knowledge to avoid exploring a huge number of speculative or poorly justified combinations of sources. This strategy indeed increases the probability that the models tested include the true history, hence, that the statistical method used has a better chance of identifying the right one.

Reviewer #4

Responses: Referee #4 (Remarks to the Author):

I am happy with the improved manuscript. Only: With more information, I am a bit less sure about the ethical oversight (see my comment below to answer 4.3).

POINT 4.1:

While improved, I would also still suggest toning down the language regarding social structure and mention the caveats: given the vast spaces and regions, the sampling is very sparse and there are no sizable sites/cemeteries that would allow study complete communities. It is thus hard to maintain that patrilocality is strongly present in the data and the differences within/between regions metrics could still contain mobility between regions. The findings are interesting and impactful even if it is admitted that more research is needed to elucidate these issues (answers 4.5 and 4.6).

RESPONSE 4.1:

We have followed the reviewer's recommendations and have edited the main text accordingly. It now reads as follows:

- *Marital practices generally maintained low consanguinity, with the exception of one female bearing the latest markers of traditional shamanism, who was the daughter of second-degree relatives. (Summary, lines 54-56)*
- *Further research, incorporating genetic analyses of large-scale funerary sites that represent entire communities, alongside denser sampling across broader geographic regions, is needed to assess the prevalence of patrilocal residence patterns. In the present data set, exogamy... (Results, The past social life of Yakut communities (lines 271-274)*
- *Finally, the significant association between genetic distances and regional affiliation, but not with indicators of leadership, suggests that political power was not concentrated in a single clan. (Results, last paragraph, lines 301-303)*

- *Although further research is required to assess how common this practice was, social norms allowed for close-kin unions, as shown by the UsSerguei female, identified as the daughter of second-degree relatives. (Discussion, lines 341-343)*

POINT 4.2:

REACTION OF R4 TO ANSWER 4.1: I am glad these efforts have been made and I am glad that they are now featured in the SI. I understand the need not to over-self-promote (and it can be also stated in the SI that this was a result of the review process if the authors so wish) but I do think that this new addition also serves as a good example for studies where such efforts were not made (and should be made) potentially raising standards in archeological/archeogenetic studies in general. It also highlights the amount of efforts of local researchers that, while having limited resources, often spearheaded them.

RESPONSE 4.2:

We completely agree with the reviewer and are pleased that, with the additional information that s/he previously requested, our study can serve as a model for future research that meaningfully engages local communities and supports the decolonization of research practices.

POINT 4.3:

REACTION OF R4 TO POINT 4.2: Great, see my reaction above. Though the actual discussions of genetic results that worried me are a bit hidden within a large scale archaeological outreach and I would suggest to separate them and highlight them specifically.

RESPONSE 4.3:

We have restructured the corresponding section of the Supplementary Information to indicate activities aimed at circling back our final results to local communities. We shall insist that the covid19 pandemics, and the current Russo-Ukrainian war, have limited scope for in-person activities involving western scientists since 2020. To further demonstrate that informing even the most remote Yakut communities of our findings remains paramount to our approach, one of our Yakut co-authors, Prof Rosalia Bravina (Institute for Humanities Research and Indigenous Studies of the North, Yakutsk), decided to write a new article in the leading local newspaper ("Sakha sire"). This article was published on May 31st, and can be accessed freely online at the following link: <https://sakha-sire.ru/arheologiya-uonna-genetika-noruottar-ikki-ardylarynaagy-byrajyakka/>. The article offers a detailed account of all research activities carried out since excavations began in 2004, concluding with the following statement:

"Although the archaeological expedition conducted as part of this international project was suspended due to the pandemic, scientists from both countries continued to investigate the archaeology and genetics of these populations in the laboratory. There is no doubt that this project is of vital importance not only for understanding the history of the Sakha people but also for broader studies on human settlement in the Far North."

POINT 4.4:

REACTION OF R4 TO ANSWER 4.3: I disagree that 'no genetic relatedness could be expected' (even when it might be unlikely given how many modern and ancient DNA samples were analysed) and I am surprised the authors think so. I personally know people living (and researching) today that have their grandparent born in 1850s and with IBD fragments, the detection of relatedness is well possible beyond 4th degree. So it does not seem to be at all impossible that individuals dated to stage 4 (1800-1922) and potentially even earlier stages could be found to have direct links to living people. Given the newly-stated facts in the SI, I agree that legally everything seems correct. But one could discuss public availability of the recent data related to a historically-exploited minority in a situation where one cannot be sure about governmental or private company misuse of private data when this is only loosely supported by an ethical permission dating to 2012, quite earlier before any large-scale autosomal aDNA data existed and many issues regarding DNA and its privacy have not yet been established in genetics (as is also evidenced in the wording of the 2012 permission that mentions bacteria and viruses and 'influence of socio-ecological factors on movement, adaptation, and

dynamics of ancient populations’ but does not touch on the issues of human DNA directly). I have seen decisions from ethical commissions at European/African institutions in recent years for issues related to recently-dated aDNA samples that prefer to protect this data (EGA – data is available but upon request and does not allow automatic use for example forensic or genealogical companies and searches) and I would like to ask that the authors consider getting an updated opinion from the ethical commission from a French institution (as getting it from both countries might be difficult and time-intensive and importantly, it might be good to avoid an impression of double-standards and ethics dumping).

RESPONSE 4.4:

We thank the reviewer for emphasizing this critical point. As the primary investigators of this study have previously focused mainly on non-human animals, we shall admit that they did not initially give sufficient consideration to the potential misuse of human genetic data once publicly released. Following established protocols from their work on horses, we originally submitted the data to the European Nucleotide Archive (ENA) to ensure transparency and reproducibility. In hindsight, this was a mistake and we sincerely thank the reviewer for their insistence on this matter. We have now withdrawn all data from ENA (following the ticket addressed to ENA on June 1st 2025, ‘ENA METAGENOME #814436’, we have received confirmation on July 28th that the study initially submitted was now suspended). We then transferred the corresponding data to the European Genome-phenome Archive (EGA, Accession Nb. EGAS0001002458), in line with the reviewer’s recommendation. This applies to all our ancient human samples, ensuring that access is now governed by an ethics committee and restricted to non-commercial research purposes, thereby minimizing any risk of individual identification or tracking.

REACTION OF R4 to ANSWER 4.4: Thank you for updating the text, as far as I am concerned, this is fine (though some other reviewers from history will be more knowledgeable in this regard and raised similar issues).

POINT 4.5:

REACTION OF R4 to ANSWER 4.5: I appreciate the increased precision in the text and the additional tests. I would still mention the caveats of low sample size, sparse sampling, archaeological biases (more men buried at a site, more men likely to be determined as related at a site → driving the signal). So I would be careful about the sweeping statements of how much in line the findings are with the expectations as the significance of the tests is low and as mentioned previously because within the 3 pedigrees resolved (extended fig. 5), there can be actually a case made for 2 possible matrilineal situations (not only the mother who is also a daughter in AC but also in Oktiom – a woman that has a brother at the site) and not even one patrilineal case (father with adult son might be a son prior to marriage). I do not argue that the society was not patrilineal. I argue that this could be argued from genetics: this is mainly a historical statement to which genetics finds exceptions rather than confirms it really. If the historical records spoke about matrilineality, I am sure that then matrilineality would be also claimed as in line with the evidence due to the matrilineal cases. Genetics has history of over-interpreting social practices, for example from mtDNA/Ychrom patterns (despite the issues with effective population size). I am not suggesting to delete these statements about social practices but I would suggest being clear about the limitations and minimize statements like ‘Social organization remained mainly structured around patrilineal clans showing strong ties to the land of their ancestors’ as that then sounds as inferred from genetics.

RESPONSE 4.5:

Agreed. See RESPONSE 4.1. Moreover, the sentence ‘*Social organization remained mainly structured around patrilineal clans showing strong ties to the land of their ancestors*’ was removed from the Summary, which now concludes with (lines 54-56):

Marital practices generally maintained low consanguinity, with the exception of one female bearing the latest markers of traditional shamanism, who was the daughter of second-degree relatives.

We have also nuanced the statement made in the last sentence of the first paragraph of the section about ‘*The past social life of Yakut communities*’, as follows (lines 258-260):

These findings are consistent with the expectations of patrilocal clans practicing exogamy, though the predominance of males in the sampled assemblages may enhance the detection of genetic relatedness among males.

POINT 4.6:

REACTION OF R4 to ANSWER 4.6: It still feels a bit overinterpreted, I would be very careful about statements about social structure giving the sampling. For example, as I have mentioned, how likely it would be to find such relatives between the regions given the space and time covered? How much migration and mobility there could be that would not be visible? I would say quite a lot. This observation should be stated but I would be clear about the increased precision in the text and the additional tests. More work in future with further samples and other sampling strategies might answer these questions

RESPONSE 4.6:

We have rephrased the sentence as suggested, emphasizing the limitations of our present sample set, as follows (lines 285-289):

Combined, these results suggest that unions, interactions, and microbial exchanges occurred primarily within regional boundaries. Although the current sample size and geographic range are limited relative to the vast expanse of Yakutia, this pattern contrasts with the prevailing archaeological interpretation, which proposed, based on the same assemblages, that extensive migration between regions was responsible for the striking similarities in the material culture¹⁷.

RESPONSE TO ANSWER 4.9: OK.

RESPONSE TO ANSWER 4.10: OK.

RESPONSE TO ANSWER 4.11: OK..

RESPONSE TO ANSWER 4.12: OK..

RESPONSE TO ANSWER 4.13.1: OK..

RESPONSE TO ANSWER 4.13.2: Thank you for the addition.

POINT 4.7:

RESPONSE to ANSWER 4.14: OK, if the intra-site analysis was not the goal, I would suggest rewording the statement in the discussion: ‘Additionally, kinship analysis revealed that first- and second-degree relatives were primarily buried in geographically close locations.’ as this to me rather suggests that comparison of related and unrelated individuals within a site was made. Thank you for sharing the detailed archaeological information with the community, this is a great addition and allows further usability of the results of this study.

RESPONSE 4.7:

We have rephrased the suggested sentence as follows (lines 337-340):

Additionally, kinship analysis revealed that while first- and second-degree relatives could be buried in the same archaeological site, they were primarily found in geographically close locations, within a maximum of 159 kilometers from one another (median distance=45m).

RESPONSE to ANSWER 4.15: I think both presentations are helpful and great to have, thanks.

RESPONSE to ANSWER 4.16: OK.

RESPONSE to ANSWER 4.17: OK.

RESPONSE TO ANSWER 4.18: OK.

POINT 4.8:

RESPONSE TO ANSWER 4.19: Actually there is an additional filter recommended by Ringbauer et al. 2023 that can be expressed in SNPs or in coverage. Namely it is 0.25x coverage for WGS data (as stated in the paper but it is likely better for most samples and also for IBD fragments longer than 12cm that were used here). It is unclear how low quality samples have been used in the analysis (e.g. I would have checked if such samples are the ones that show extreme numbers of IBD fragments shared, e.g. Extended fig. 5 for pairs of males from the same regions).

RESPONSE 4.8:

We only used in IBD analysis samples that have been imputed. As we input samples with a coverage of at least 0.35X, we did respect the filter from Ringbauer et al. 2023 and even exceeded it. We updated the Method section to make this point explicit (page 11):

As recommended, the Yakut data set was down-sampled to 1,240,000 SNPs, for which ancIBD is optimized, and IBD sharing was screened for every pair of imputed individuals (coverage ≥ 0.35 -fold) with default settings⁹⁴.

RESPONSE TO ANSWER 4.20: OK, I find the analysis reproduced without the manual alignment convincing.

POINT 4.9:

RESPONSE TO ANSWER 4.21: OK. I would actually mention these observations in the SI but I leave that to the authors' discretion.

RESPONSE 4.9:

We now have added the information shared in our answer to the SI (Section 2.3, page 28-29):

"Soft tissues were analyzed for only one individual, AC1num2, and consisted of lung and muscle. They showed limited endogenous content, and provided extremely limited sequence data aligning against the human genome, resulting in 0.0000587-0.0003046-fold coverage, depending on the DNA library considered. The relatively limited DNA preservation observed in soft tissues match previous work on woolly mammoths²⁶. As a result, hard tissues, mostly petrosal bones and teeth, which are known to maximize DNA preservation^{11,12}, were preferentially analyzed."

RESPONSE TO ANSWER 4.22: OK

RESPONSE TO ANSWER 4.23: OK

RESPONSE TO ANSWER 4.24: Yes, only reference was needed. Thanks.

RESPONSE TO ANSWER 4.25: OK

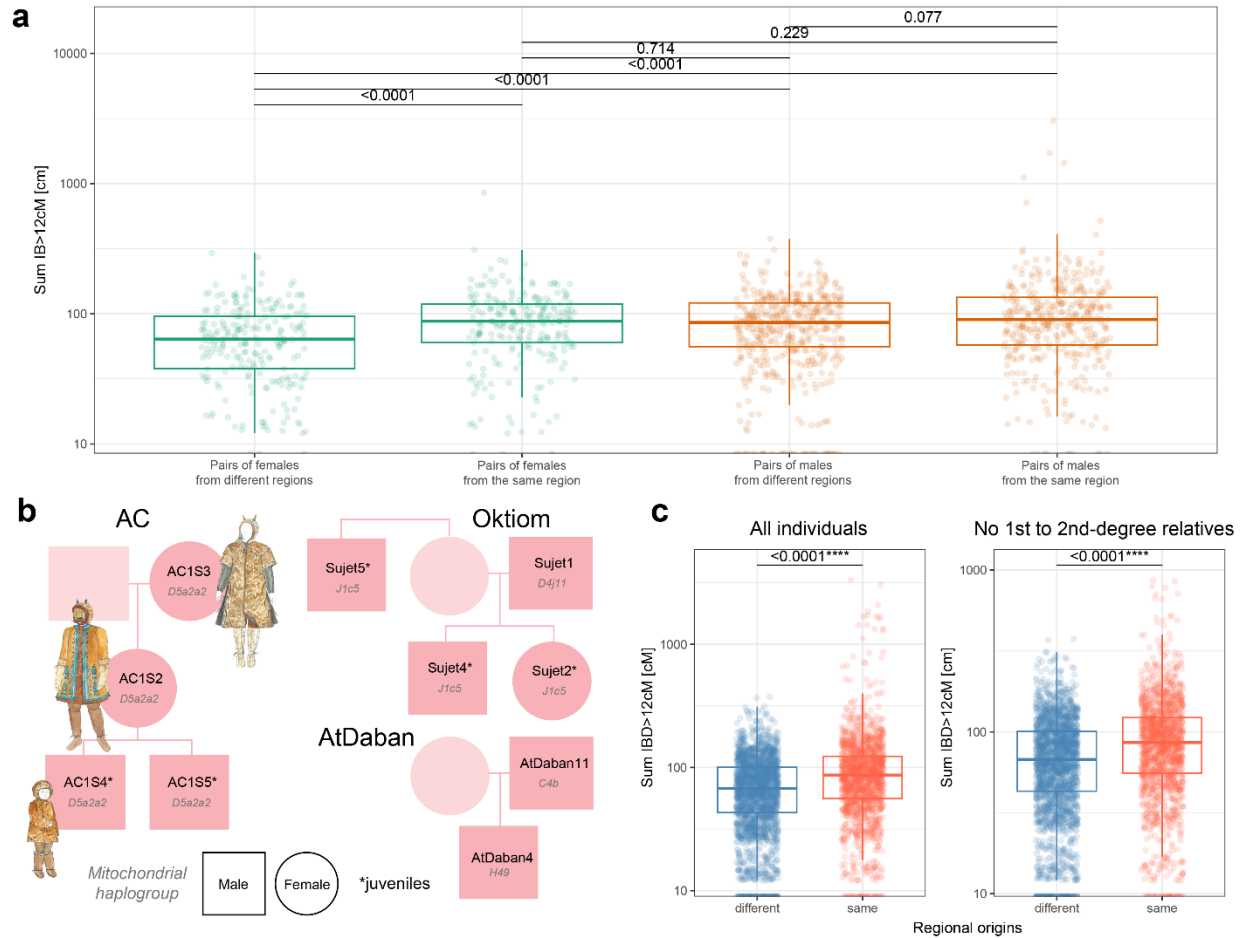
POINT 4.10:

New small point:

Axis label in extended Fig. 5 is missing D in IBD.

RESPONSE 4.10:

We apologize for overlooking this. The corrected corresponding figure is pasted below:



I was asked to add reactions to the comments of reviewer 3:

RESPONSE of R4 to ANSWER 3.1

I found this new addition rather too positive for Russian settlers. Indeed the other cases of colonialism might have been harsher in some ways though one could question the need to relativize one injustice with another. Also, I am not keen on the wording from the perspective of Russia ('hindered the development of demographically-*thriving* Russian colonies' - this could be reworded from the perspective of Yakuts, for example population replacement of Yakuts or similar). Otherwise ok as far as I can judge.

RESPONSE 3.1:

It was not our intention to make our statement too positive for Russian settlers, nor to relativize injustice. We have now rephrased the paragraph as followed (page 7, lines 364-373):

Additionally, the harsh environment, often lethal to European settlers and unsuitable for large-scale cereal agriculture, hindered the development of large Russian settlements⁵⁵ (Supplementary Information Section 1.1 and Fig. S1_1). Consequently, the Russian expansion in Yakutia relied heavily on local populations, representing a form of colonialism distinct from the settler colonialism of the Americas, which proceeded through Indigenous displacement and eradication⁵⁶. It also diverged from the indirect rule characteristic of colonial India, where colonial authority was maintained through local leaders without administering the territory. In contrast, Russian authorities imposed their tax systems on Yakut communities and co-opted Yakut elites by granting citizenship and incorporating these leaders into the imperial structure. This fundamentally transformed the traditional political and social organization, making Yakuts the subjects of a foreign empire.

RESPONSE of R4 to ANSWER 3.2

I understand but I suggest that this section is referred to in the main text - not the section only but the fact that there is a terminological issue (e.g. “for discussion of the terminology of ethnonyms, see Section..”) so it is clear to the readers that it is not so simple.

RESPONSE 3.2:

To make the terminological issue transparent to the reader, we now mention it in the main text (page 2, lines 68-69):

This contrasts with the reindeer-based economies of neighboring Indigenous groups⁵ (for a discussion of ethnonym terminology see Supplementary Information Section 1.1.3).

RESPONSE of R4 to ANSWER 3.3

OK I think, though the idea of the reviewer 3 seems to be also that alcohol could get to Yakut communities via trade with other Steppe people not only through European/Chinese/Mongolian route.

RESPONSE 3.3:

We agree with what reviewer 3 suggested, and as a result, edited our statement accordingly. It originally referred to ‘alcohol’ in general. It is now more specific, and restricted to vodka.

RESPONSE of R4 to ANSWER 3.4

As stated in my comments, I found this sentence overinterpreted but this comment (exogamy) seemed to have been addressed.

RESPONSE 3.4:

We updated the main text to tone down our conclusions on exogamy and kinship. See RESPONSE 4.5, 4.6 and 4.7.