Supplementary Materials for

Human population dynamics and Yersinia pestis in ancient northeast Asia


*Corresponding author. Email: anders.goetherstrom@arklab.su.se (A.G.); jan.stora@ofl.su.se (J.S.); gulsahkilinc@hacettepe.edu.tr (G.M.K.)

DOI: 10.1126/sciadv.eabc4587

The PDF file includes:

Supplementary Text
Figs. S1 to S32
Tables S1 to S24
Legends for data files S1 to S5
References

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/7/2/eabc4587/DC1)

Data files S1 to S5
Supplementary Text

Archaeological background

General archaeological information about the northeast Asia

The archaeological finds from the Pleistocene, corresponding to the Middle to the Upper Palaeolithic periods, in the northeast Asia are rather scarce. This has been explained by the volatile climatic conditions and overall fluctuations and changes of the environment affecting the conditions for the human groups: the Ice Age cold period, the late Pleistocene temperature fluctuations and the post-tectonic activity of the Baikal rift zone. During the period corresponding to the Palaeolithic period, Baikal became a deep lake while the eastern terraces of the Baikal basins were uplifted and the formation of the Baikal outlet into the Angara River 60,000 years ago occurred (70). There is a correlation between these events and the scarcity of the archaeological sites in the Pleistocene. The Khaiyrgas cave is one of the few notable sites providing the opportunity to view the collections of artefacts from several periods, starting from the Upper Palaeolithic.

A different scenario was seen in the Holocene populations of northeast Asia, represented by a diversity of cultural complexes, with a continuous presence of human occupation in some of the ecologically favourable areas up until today. The general chronological outline for the human groups close to Baikal during the first half of the Holocene is characterized by two main complexes in the Early Mesolithic-Middle Neolithic and in the Late Neolithic-Bronze Age associated with the Kitoi and the Glazkovo cultural complexes, respectively. On the territory of the Baikal region (Angara, Upper Lena, Baikal, Vitim), the late Mesolithic burial complexes were identified - related to the Khinsky and Schukin burial traditions; Early Neolithic - the Chinese tradition of burials; Late Neolithic — Isakovskaya, Serovskaya, Late-Sarovskaya for Priolkhonya, “burial traditions” for the Upper Lena “archaic”.

A Mesolithic human occupation has been discovered in both the Cis-Baikal and the Trans-Baikal areas. Some aspects of the Mesolithic cultures in northeast Asia are considered as pan-Eurasian, such as the use of red ochre. Red ocher is found in burial complexes of the Early and Late Neolithic of Pribaikalye. Regional differences are seen in the broad variety of local lithic traditions, use of different materials (for example nephrite) as well as composite tools, fishery-associated artefacts
and, finally, the early use of pottery. The complexity of the Mesolithic/Early Neolithic cultures of the northeast Asia is seen in the variation of the material used by people to make lithic tools and exceptional in the case is the Dzhylinda burial included in the present study (sample: irk00x, Dzyhlinda-1). The variation of the cultures in the early Holocene may be explained by the specialization of regional groups as means of adaptation to different ecological environments, which vary from alpine and forested areas to lowland steppe and tundra in northeast Asia (71). The Late Neolithic-Bronze Age complex of cultures (Serovo, Isakovo and Galzkovo) show more homogenous expressions and characteristics than the Mesolithic and Early Neolithic cultural-complexes, but with a large variation in pottery styles.

There is a period between the two main cultural complexes mentioned above, during which archaeological finds in the Cis-Baikal area are few. This “hiatus” has been in focus of discussions and different views and interpretations have been presented, suggesting that it may have lasted anywhere between 100 and a 1,000 years. Other “hiatus” periods have most often been associated to unfavourable ecological conditions (72).

The development of the human populations and the movements of groups are complicated to follow and interpret due to the complexity of the archaeological record. Often the burial sites have several cultural layers associated to different cultural manifestations and the specific chronological circumstances are not always fully understood. Exemplary in this case is the Cyclodrom/Lokomotiv, the large necropolis in Irkutsk, which contained burials from the Mesolithic and up to the Bronze Age. Regarding the movements of people during the first part of the Holocene, there has been a general consensus of directions of mobility and migrations, which are linked to the appearance and dispersal of Late Neolithic and Bronze Age culture complexes through the areas near Baikal and the Lena river valley. The most often identified directions of mobility in these periods have been from west to east and from south to north.

Archaeological contexts of the studied individuals

In this study, we studied 40 samples from 5 distinct administrative regions in Russian Federation including: Trans-Baikal, Cis-Baikal, Yakutia, Krasnoyarsk Krai and Amur Oblast. Chronologically, the samples cover a time period between the Late Upper Paleolithic and the
Medieval era. The majority of individuals in our study are dated to the Neolithic and Bronze Ages. A Late Upper Paleolithic individual (yak025, Khaiyrgas-1) included in this study was recovered in the Khaiyrgas Cave in Yakutia (18). Archaeological information for 38 of the 40 samples (Table S1) sequenced here was provided in (18). In this study, we report 2 additional individuals including cta016 which is from the Trans-Baikal region and dated to 8,345 to 8,185 BP and Matta-1 (N2a) which is from the Yakutia region and dated to 6,845 to 6,665 BP. We provide archaeological information for these two samples below.

Archaeological periods do not have heuristic meaning, since there are no specific criteria for distinguishing these periods as in European context while analyzing materials from Asia. For example, pottery appears in the Late Paleolithic and Mesolithic; Paleolithic traditions of lithic technologies exist in the Neolithic. Still, archaeological periods for the analyzed individuals have been determined according to a general chronological frame of the north Asian archaeological cultural complexes (73–80) (Table S1). In brief, genome sequence data from Yakutia include a Late Upper Palaeolithic (c.16,900-16,450 BP) individual (0.3X genome coverage); two Early/Middle Neolithic (c.6,850-6,190 BP) individuals (0.03X to 0.05X); six Late Neolithic (c.4,780-3,080 BP) individuals (0.2X to 8.9X); and an Early Iron Age (c.2,750-2,490 BP) individual (0.07X); from Trans-Baikal encompasses two Mesolithic (c.8,515-8,185 BP) individuals (1.1X to 4.7X), five Neolithic (c.7,460-5,940 BP) individuals (0.1X to 3.8X); and a Bronze Age (c.3,210-3,000 BP) individual (0.1X); from Cis-Baikal consists of two Mesolithic (c.8,980-7,855 BP) individuals (2.17X to 3.49X); nine Neolithic (c.6,100-4,600 BP) individuals (0.4X to 14.5X); eight Bronze Age (c.4,530-3,700 BP) individuals (0.1X to 8.7X); and a Medieval (c.675-560 BP) individual (0.5X); from Krasnoyarsk Krai encompasses a Bronze Age (c.4,280-4,085 BP) individual (13.6X); and from Amur Oblast covers an Iron Age (c.1,345-1,270 BP) individual (0.7X) (fig. S1, table S1). Below we present a short description of the archaeological contexts of the studied individuals:

**Cis-Baikal group**

Of the studied 40 individuals, 20 were from the Cis-Baikal area encompassing Angara River basin, Lena River basin and Lake Baikal region.
Angara river group (10 individuals)
The sites of Cyclodrome (Lokomotiv) (irk051, Cyclodrome-1) and Glazkovo (irk036) both belong to the Glazkovo necropolis in Irkutsk (81, 82). The sites were excavated by several researchers during the 20th century (83–85). The Glazkovo necropolis in Irkutsk contains burials belonging to both the Kitoi and the Glazkovo cultural-complexes (81). The individual from Glazkovo necropolis (irk036) belongs to the Glazkovo culture, which dates to Early Bronze Age. The Suhaja Pad’ Buret’ site, burial 2 (irk025) is from the Early Bronze Age (81–84). The Ust’-Dolgoe (Ust’-Dolgaya) site, burial 3 (irk022) is also a Glazkovo site, based on the burial inventory (81). Chastaja (Chastye) Padi (irk033) is both a burial and a settlement site. The burials belong to Glazkovo culture while the settlement site had a layer with fragments of pottery with net-imprint (81). Anosovo-1 (irk050) was excavated from an Eneolithic Glazkovo site (81). Podostrozhnoe is a term used for a group of sites, including Kirpichnyj Saraj (irk034), Gorodische site N 1 (irk040) and Podostrozhnoe site N 3 (irk057). Burial 3 from the Zarubino site (irk032) is a Medieval burial.

Lena river group (7 individuals)
The sites at Zvjozdochka (irk061) were excavated by Okladnikov in 1930 at the left bank of Lena river. These sites were classified as Glazkovo culture sites. The sites at Korkino (irk030), Shishkino N1 (irk068) and Makarovo (irk017) are also classified as Glazkovo culture sites. The Silinskij burial site (irk008) lies on the left bank of Lena River. Popovskij Lug is a collective name used for a number of Neolithic-Bronze Age sites (86). In 2010 an archaeological expedition at Popovskij Lug II unearthed a single male burial (irk007, Popovskij-1). The burial is the oldest earliest documented burial in Cis-Baikal with ceramic as grave goods (87) and attributed to the early Neolithic of Baikal region based on the morphology of the burial equipment and the presence of ceramics. Makrushino (burial 26, mak026) is a burial site containing material from the Early Neolithic to the Bronze Age in the Kachug District (88).

Baikal Lake (3 individuals)
The Mys Uyuga burial 2 (individual 1- irk071) is a double burial situated on the peninsula of Mys Uyuga (89, 90). The Shamanka 2 site (individual 1, irk076) is a site situated on the southern side of the cape Shamanka (78). The Sokhter site, complex 1 contains a single burial (burial 5, irk075)
excavated in 1994 (91). This burial is a *Shumilikhinskaja* burial type, and dated to the Neolithic/Early Bronze Age (78).

Trans-Baikal group (8 individuals)
The analysed burials in the Trans-Baikal group are all located east of the Lake Baikal. The sites are: Izvestkovaja-1 (burial 1-brn008 and burial 2-brn001), Nozhlyj lake burial site-2 (burial 6, skeleton 3, brn012 and one individual from a burial without a number, brn002), Okoshki-1 (irk078), Dvorcy-Dacha burial (brn003) and Dzhylinda (irk00X, Dzhylinda-1). Chronologically the samples are spread from c.8,514 and and 5,940 cal BP (92). The Okoshki-1 Burial 23 (irk078) is dated to the Bronze Age (93, 94). We include one new individual from the area in this study (cta016). It is from one of the few known Stone Age burial grounds in Trans-Baikal region. The site is located 7 kilometers off the Duroy village, on the western side of the Bol’shaja Kanga - 1 mountain (Большая Канга-1) by the Argun river. From 65 burials two were considered to be Stone Age burials, based on for example appearance of red ochre by the remains of the individual. A tooth from mandible of the individual 1 was taken for analyses. The bone tools and remaining inventory from this part of Trans-Baikal suggests a hunter/fisher-gatherer lifestyle (95).

Yakutia (Sakha Republic) group (10 individuals)
The samples from Yakutia located geographically in three areas: Southwestern Yakutia (Khaiyrgas Cave at the Lena river (yak025, Khaiyrgas-1)), central Yakutia (samples=yak030 (N3a), N4a1, N4b2, N5a (Onnyos-1)) and northeastern Yakutia (yak021-024). The following sites are included in the present study; *Khaiyrgas cave* (yak025, Khaiyrgas-1), *Onnyos* (N5a, Onnyos-1), *Kyordyughen* (Kyordyughen 2, N4a1; Kyordyughen 1, burial 1, N4b2), *Kamenka 2* (individual 1, Kamenka-1, yak022; individual 2, Kamenka-2, yak023; and individual 3, yak024, Kamenka-3); Pomazkino (yak021), and Dyupsya (yak030), (8, 91). The Onnyos burial (N5a, Onnyos-1) belongs to the Belkachi culture complex (96). The Kyordyughen site is dated to the Late Neolithic (97, 98). Kyordyughen 1 is a double burial and we include both individuals (N4b2, Ind 1) and (N4a1, Ind 2) (99). The Kamenka-II burial contained three juvenile individuals (96, 100) (individual 1- yak022, individual 2-yak023 and individual 3-yak024) (101). From the Pomazkino-III site we include burial 2 (yak021) (97). The Dyupsya burial (yak030, labelled as N3a in (18) is dated to the Early Iron Age (102, 103). We include one previously unpublished individual from the area (N2a,
Matta-1). It is from the Matta burial which was found on the terrace of the Matta Lake which is located in Megino-Kangalassky ulus (district), 26 km east of the Lena River. The skeleton belonged to a female and had bones from a metatarsal or metacarpal from a species of the genus Lepus instead of the left hand of the individual (I04). The fragments taken for the analyses are bones from left foot: metatarsale 1, os cuneiforme 2 (I05).

Krasnoyarsk Krai (1 individual)
The region of Krasnoyarsk Krai is represented by burial 1 (kra001) from the Nefteprovod-2 site.

Amur Oblast (1 individual)
The Blagoveschensky District of Amur Oblast is represented by burial 2 (bla001) from the Oktyabr’skoe site.

Radiocarbon dating and stable isotope analysis
All samples except yak024 (Kamenka-3) were radiocarbon dated using AMS radiocarbon dating (table S1) (I8). Kamenka-3 could not be dated due too poor quality of the sample but it was found in the same burial with yak022 (Kamenka-1) and yak023 (Kamenka-2). Calibration of the conventional radiocarbon ages was performed using Oxcal version 4.3.2 (I06), and the IntCal13 atmospheric curve (I07). Calibrated results (BCE) were reported with 95% confidence intervals (2σ) (figure S1, table S1). The isotopic values for a few samples exhibit elevated δ15N-values (>+16) (table S1) (70). A few of these values were measured on deciduous teeth and probably exhibit an effect of weaning. Although some level of freshwater reservoir effect (FRE) might affect the age estimations this does not have major consequences for the present study (70).

Evaluation of biological relatedness
In Trans-Baikal group, we found that brn002 was a second or third degree relative of brn012. In Cis-Baikal group, irk022 was a 3rd-5th degree relative of irk025. Within the Yakutia group, yak021, yak022 (Kamenka-1), yak023 (Kamenka-2) and yak024 (Kamenka-3) were biologically related at different degrees from first to third-fifth degree (Table S1). We excluded from the frequency-based analyses one individual from each pair of related individuals with a k0 values < 0.9. The excluded individuals are brn002, irk022, yak024 (Kamenka-3), yak023 (Kamenka-2),
yak021 (Table S1). Note that the high genetic diversity within the ancient samples due to their wide spatial and chronological differences may produce an underestimation of the relatedness degree between some of the studied individuals.

Uniparental markers

Y haplogroups
All individuals were accredited either to Y macro-haplogroup Q or N. The one possible exception was individual irk032 (probable carrier of Hg I*), which was dated to the Medieval period. In short, all of the oldest (Mesolithic and Neolithic) and most eastern (Trans-Baikal and Yakutia) individuals were carriers of N1c1a1 lineage, while Bronze Age males from Cis-Baikal region were all carriers of Q1a2a and its variants. The individual from Cis-Baikal region (irk007, Popovskij-1) may have been a carrier of N1c2b2 lineage associated with present-day Khanty and Komi populations of western Siberia (14). This pattern is in line with earlier findings suggesting both lineages were widespread among individuals from Neolithic Lokomotiv and Shamanka cultures around Lake Baikal (16) (Table S18).

Mitochondrial haplogroups
All individuals carried non-African mitochondrial macrohaplogroups of M, N and R (18) (Table S1). Individuals from Trans-Baikal group carried east Eurasian haplogroups of A, D, C, G and their sub-haplogroups. Cis-Baikal group carried A, C, D, F and their sub-haplogroups and also one Cis-Baikal individual carried R1b mitochondrial haplogroup. Similarly this R1b haplogroup was also found in Khaiyrgas-1 (yak025), the Late Upper Palaeolithic subadult female from Yakutia. Other Yakutia individuals carried the A, C, D and F haplogroups (Table S1) (18).

Model-based clustering for modern populations
We run the ADMIXTURE on modern samples and projected the ancient samples on this. At this point, to decide which K explains the observed pattern better, we examined each K and present the results for all Ks (fig. S3, table S4).

In brief, at K=2, modern populations separated into African (dark-red) and non-African (navy) components; ancient Siberians carried non-African component. At K=3, the new component
separated present-day east Asia & Native America populations (dark-green) and west Eurasia populations (navy); ancient Siberians carried mostly the east Asia component (dark-green) and to a lesser degree the west Eurasia component (navy). At K=4, the new component separated present-day east Asia (dark-green) and Native America (purple) populations; ancient Siberians carried slightly more east Asia (dark-green), and to a lesser degree Native America (purple) and west Eurasia (navy) components. At K=7, the new component separated present-day southeast Asia (brown) and east Asia (dark-green) populations; ancient Siberians carried this component at low levels compared to east Asia (dark-green) component. At K=8, the new component (yellow) maximized in present-day Far East populations including Itelmen and Koryak (see fig. S4 for geographical locations of populations); ancient Siberians carried almost equal amount of Far East (yellow), east Asia (dark-green), and west Eurasia (navy) components and less amount of Native America (purple) and southeast Asia (brown) components. At K=9, the new component (light-blue) split up the west Eurasia populations into north (navy) and south (light-blue) components and also separated the west European Mesolithic hunter-gatherers and Neolithic farmers; this component present in very low levels in ancient Siberians. At K=13, Siberian ancestry split up into Far East (pink-orange) and north Siberian (yellow) components; this new Far East (pink-orange) component was present in ancient Siberian genomes in low amount compared to north Siberian (yellow) component. Starting from K=14 to K=16 the observed pattern continued, Therefore, we present the results for K=14 in (Fig. 2 and fig. S4).

**Analysis for Yersinia pestis (Y. pestis)**

*Y. pestis* is recently diverged from a less virulent species *Y. pseudotuberculosis* (108). Therefore, one of the main methodological problems to identify *Y. pestis* sequences in archaeological samples is the sequence similarity of these two closely related species. To this end, a hybrid genome combining *Y. pestis* and *Y. pseudotuberculosis* reference genomes was created in (39). In this study, aDNA reads were aligned to this hybrid genome and the reads that were mapped to *Y. pestis* CO92 reference genome were extracted. To judge the taxonomical origin of *Y. pestis* sequences, edit distance distribution, length distribution, deamination patterns, and breadth of coverage information were used. Edit distance is the number of substitutions between the DNA sequence and its respective reference genome portion. It implies the similarity between the reference genome and the aligned read and it could be used as a marker to judge the quality of the alignment. Most
of the reads should be perfectly aligned to the reference genome (that means 0 edit distance). Deamination patterns at the 5’ and 3’ ends of the DNA sequences prove the authenticity of the sequences. Right skewed length distribution (with a mean around 50bp) shows that the DNA sequences are fragmented thus supporting the authenticity. The genome coverage and a homogenous genome-wide distribution of the alignments is an important factor to prove authenticity and presence of the microbe (109). Generally, reads could stack in multi-locus repetitive regions that share the sequence homology between bacterial species. Those areas results with a peak in the coverage and causes misinterpretation. So, a careful examination of the read distribution is also important.

In this study we analyzed 205 sequence libraries from 40 different samples to detect the \textit{Y. pestis} presence. In Anosovo-1 (irk050), Kamenka-1 (yak022) and Kamenka-2 (yak023) we detected notable amounts of aDNA reads associated with \textit{Y. pestis}. Specifically, Our 9,523 aDNA reads in Anosovo-1, 2,480 aDNA reads in Kamenka-1, and 4,176 aDNA reads in Kamenka-2 were assigned to \textit{Y. pestis} CO92 reference genome (Table S20). The Kamenka-2 from Yakutia in eastern Siberia is to our knowledge the most northeastern findings of \textit{Y. pestis} to this day (fig. S23).

\textbf{Identification of \textit{Y. pestis} aDNA sequences in Anosovo-1}

After the mapping to the hybrid genome and filtering processes, we found that 9,395 aDNA reads specifically mapped to \textit{Y. pestis} genome in the Anosovo-1 sample. After removing the optical PCR duplicates, 3,289 of aDNA reads were used for the further analysis. We observed that 1650 aDNA reads aligned to CO92 reference genome with 0 edit distance followed by 991 aDNA reads with an edit distance 1, 363 aDNA reads with 2 edit distance, and 166 aDNA reads with an edit distance 3 (fig. S24A). Also the fragment length distribution was skewed with a mean of 54.1657 nucleotide (fig. S24B). Nucleotide misincorporation patterns showed the authentic deamination patterns at the end of the DNA sequences (fig. S24C). The overall coverage was 1.85% for the whole genome. \textit{Y. pestis} CO92 genome has one chromosome and three plasmids so it is possible to check the coverage for each of them. Table S21 shows the statistics for the individual sequences in \textit{Y. pestis} genome. Results show that we covered %3.66 of the \textit{Y. pestis} CO92 chromosome (table S21). Reads were also homogeneously distributed among the chromosomes except pPCP1 plasmid (fig.
These plots should be interpreted with caution since it is not possible to make X axis wide enough to accommodate all alignments. Still, reads were distributed among the genome homogeneously. Also, we checked the alignments using IGV. We did not find any spurious mappings. We observed that more than 90% of the aDNA reads aligned to *Y. pestis* genome with less than 3 edit distance.

**Identification of *Y. pestis* aDNA sequences in Kamenka-2 (yak023)**

After mapping to the hybrid genome and filtering, we found that 4,176 aDNA reads were specifically mapped to *Y. pestis* genome in Kamenka-2. After removing 26% of the clonal sequences, 3,070 of aDNA reads were used for the further analysis. We observed that 1,784 aDNA reads aligned to CO92 reference genome with 0 edit distance followed by 848 aDNA reads with an edit distance 1, 251 aDNA reads with 2 edit distance, and 67 aDNA reads with an edit distance 3 (fig. S26A). Also the fragment length distribution was skewed with a mean 51.36 nucleotide length (fig. S26B). Nucleotide misincorporation patterns showed the authentic deamination patterns at the end of the DNA sequences (fig. S26C). The overall coverage was 1.65% for the whole genome. We checked the coverage for one chromosome and three plasmids of *Y. pestis*. Table S22 shows the statistics for the individual sequences in *Y. pestis* genome. Results show that we covered %3.16 of the *Y. pestis* CO92 chromosome (table S22). However we could not find aDNA reads mapped to plasmid pCP1. Then we analysed the genome coverage (fig. S27). We observed that more than 90% of the aDNA reads aligned to *Y. pestis* genome with less than 3 edit distance. Also we observed a homogenous distribution in coverage.

**Identification of *Y. pestis* aDNA sequences in Kamenka-1 (yak022)**

After mapping to the hybrid genome and filtering process, we found that 2,480 aDNA reads were specifically mapped to *Y. pestis* genome in Kamenka-1 sample. After removing 34% of the clonal sequences, 1,640 of aDNA reads were used for the further analysis. We observed that 1,403 aDNA reads aligned to CO92 reference genome with 0 edit distance followed by 148 aDNA reads with an edit distance 1, 28 aDNA reads with edit distance of 2, and 6 aDNA reads with an edit distance of 3 (fig. S28A). Also the fragment length distribution was right-skewed with a mean 46.4207 nucleotide length (fig. S28B). Nucleotide misincorporation patterns did not show the authentic deamination patterns at the end of the DNA sequences (fig. S28C). The overall coverage
was 0.79% for the whole genome. We checked the coverage for one chromosome and three plasmids of Y. pestis. Table S23 shows the statistics for the individual sequences in Y. pestis genome. Results show that we covered %1.56 of the Y. pestis CO92 chromosome (table S23). In Kamenka-1 sample, we could not find aDNA reads mapped to plasmid pCP1. Then we analysed genome coverage (fig. S29). We observed that more than 90% of the aDNA reads aligned to Y. pestis genome with less than 3 edit distance. Therefore, we could not validate the presence of Y. pestis in this individual.

Overall, in our analysis, deamination rate in Y. pestis reads is relatively higher than the rate in human reads (table S24) for all there samples.

Functional SNPs

Accelerator phenotype

N-acetyltransferase 2 gene (NAT2) is involved in the metabolism of a wide variety of foreign chemical substance ( xenobiotics) (110). The metabolism can be either fast, intermediate or slow and the status has been linked to differential response to various drugs and carcinogens (111). The variation in worldwide distribution of different NAT2 phenotypes suggests that the distribution of the phenotypes correlates with the type of subsistence practice. It has been noted that hunter-gatherers often have high frequencies of the fast phenotype (>77%), while pastoralists and agriculturalists have reduced levels of the phenotype (51-54.9%) (112) . The favorable selection of the slow phenotype in modern populations, which is higher in west Eurasians and Africans, compared to hunter-gatherers, Native Americans and east Asians was hypothesized to be related to dietary changes accompanying the transition to agriculture (113, 114). Indeed, in modern Asian populations the distribution of acetylation phenotype has been linked to subsistence strategy with slow phenotype more common among populations of sedentary agriculturalists than the nomadic populations (115).

In order to investigate the status of acetylator phenotype in ancient Siberian individuals we predict the NAT2 phenotype from seven informative SNPs (rs1801279, rs1801280, rs1799930, rs1799931, rs1041983, rs1799929, rs1208) using an online NAT2 prediction tool - NAT2PRED (http://nat2pred.rit.albany.edu) (116). To confirm the status we further use a tag SNP (rs1495741)
and/or a 2-SNP panel (rs1041983 and rs1801280) (117) which has been suggested to outperform the single tag SNP prediction in non-Caucasian populations (118), but may result in overrepresentation of the intermediate phenotype.

We have used the combination of all three methods to predict the acetylation status in 22 individuals sequenced to genomic coverage above 1X (Table S22). We compare the results obtained from both damage repaired (Table S19) and non-damage repaired libraries (Table S19) using the latter for finale predictions due to low impact of observed damages coupled with much higher coverage (especially in irk034) and what follows, increased probability of identification of heterozygotes. We only accept consensus predictions whenever all three prediction tools yielded results and when at least two of the predicted phenotypes matched. Using those criteria we were able to make phenotype predictions in 15 out of 22 tested individuals. We find that a single Mesolithic individual was a carrier of a rapid phenotype, and so were five Neolithic individuals. Only one Neolithic individual was a carrier of an intermediated phenotype while Bronze Age individuals were either intermediate (N=5) or slow acetylators (N=3) (Table S19). Our data suggests that intermediate phenotypes were uncommon in Siberia during the Mesolithic and Neolithic with significant frequency increase during the Bronze Age. Thus, our results are in line with hypothesis suggesting selective pressures associated with transition to agriculture and major change of lifestyle may have shaped present-day distribution of the NAT2 phenotype in Asia.

**Diet-related SNPs**

The ability to digest sugars found in milk into the adulthood, so called lactase persistence, is often regarded as an important cultural adaptation in humans associated with dairy animal domestication. We have tested 22 individuals sequenced to average genome coverage exceeding 1X for the two SNPs most commonly associated with the trait: LCTa (rs4988235 - "C/T(-13910)", (119) and LCTb (rs182549 - "G/A(-22018)"; (120) as well as other variants associated with lactase persistence which could be more common in individuals of non-European descent: rs41380347 - "C/T(-13915)" (East Asia), rs41525747 - "C/G(-13907)" (East/North Africa and Middle East) and rs145946881 - "G/C-14010" (East Africa) (121). However, we find no evidence of lactase persistence in the 22 tested ancient Siberians.
Another variant, potentially associated with the spread of agriculture is the derived status of ADH1B*47His which among other associations has been linked to alcohol metabolism in East Asians and the so-called ‘Asian blush’. It has been suggested that prevalence of derived ADH1B*47His could be explained by the increased alcohol consumption in connection to Neolithic rice domestication (114). None of the tested individuals was a carrier of the derived (A) allele at position rs1229984. Additionally, the lack of derived alleles (A) at the ALDH2 rs671 suggests absence of Asian flush and reduced hangovers in all 22 individuals. Thus, present-day distribution of ADH1B*47His and characteristic distribution of alcohol metabolism associated variants among Asian populations was shaped in more recent times.

Other phenotypic traits
We have tested all individuals for a number of SNPs associated with various phenotypic traits such as AGT gene, associated with increased risk of hypertension (122); KITLG gene, associated with UV protective tanning in lighter skinned individuals (123); CYP3A4 and CYP3A5 associated with certain cancers and drug metabolism (124, 125); EDAR gene, associated with incisor and hair shape as well as Asian ancestry (126), ABCC1 gene associated with various phenotypic traits as well as Asian ancestry (127), and finally, C282Y and H63D variant in HFE gene associated with hereditary haemochromatosis.

We only find variation in AGT, KITLG, EDAR, ABCC1 and CYP3A5 genes. While the different variants and incidence of heterozygotes in AGT and KITLG is spread out among various groups and time periods and likely represents population variation, it seems that the increase in heterozygosity in ABCC1 and EDAR genes is mostly restricted to Neolithic and Bronze Age individuals from Cis-Baikal group. On EDAR we observed increased occurrence of (A) on otherwise widespread (G) variant associated with straight hair and shovel shaped incisors (note that the phenotypic representation of heterozygotes would be the same), while four individuals were heterozygous on ABCC1 gene, otherwise all tested individuals were homozygotes for variant T associated with Asian ancestry. Thus, it seems that on some phenotype encoding SNPs we detect signal of western admixture in Cis-Baikal during the Neolithic and Bronze Age.

Blood type
AB0 and Rh blood group system are a set of genetic markers which has been used in the study of human genetic variation for decades and presents distinct geographic distribution \cite{110, 128}. In order to investigate blood group variation in our dataset we followed a strategy for blood group prediction adopted in \cite{129} and checked 22 individuals at five SNPs within the ABO gene region: rs8176719 (G/-), rs7853989 (C;C/G;G), rs8176746 (C;C/A;A), rs8176747 (G;G/C;C) and rs590787 (C;C/T;T) \cite{130}

It was possible to make broad blood group inferences in 19 of the tested individuals of which four were phenotypically carriers of blood group ARh+, while remaining 15 individuals were probable carriers of 0Rh+. Both phenotypes are common in present-day Asia, especially 0Rh+, however seeing that we only identify ARh+ in Mesolithic and Neolithic individuals in our dataset and not among more numerous Bronze Age samples could potentially suggest that ARh+ was more widespread in earlier periods and declined in frequency sometime during or after the Neolithic.
Fig. S1

**Radiocarbon dates of the studied samples.** Samples excluded from the comparison are yak025 (Khaiyrgas-1), irk032, yak030, and bla001. yak024 (Kamenka-3) could not be dated due to too poor quality of the sample but it was found in the same burial with yak022 (Kamenka-1) and yak023 (Kamenka-2). Full data was presented in Table S1.
Fig. S2

**Outgroup $f_3$-statistics for Khaiyrgas-1.** Plot showing the $f_3$-statistics in the form of $f_3(Khaiyrgas-1, Test\_Population, Yoruba)$ plotted with 2 standard errors (se). Test population is one of the 194 present-day world populations from Human Origins dataset (see table S3).
Fig. S3

ADMIXTURE analysis for K=2 to K=16. Final results for each K which was obtained by running CLUMPP on 30 independent runs per K was plotted in R. Color codes which were used for plotting are presented in table S4. R library "RColorBrewer" was used to create that color vector (table S4).
Fig. S4

ADMIXTURE result for modern populations. (a) ADMIXTURE analysis result for K=14 ancestral clusters for present-day world populations from Human Origins data set. (b) Approximate geographical positions corresponding to each of the present-day populations plotted in the fig. S4a (table S4).
**Fig. S5.**

**Admixture graph fitting using Treemix.** (a) A maximum likelihood tree for a model which allows for zero migration and (b) Pairwise residual matrix in which large values imply a poor fit for the indicated population pairs.
Fig. S6

$F_r$-statistics for $f_2$(Yoruba, Khaiyrgas-1; TestPop, Kolyma_M). Results were plotted with two se intervals for the tree like topology of $f_2$(Yoruba, Khaiyrgas-1; TestPop, Kolyma_M) in which TestPop corresponds to one of the ancient groups on the left y-axis (table S6).
**Fig. S7**

$F_r$-statistics for $f_d(Yoruba, Kolyma\_M; TestPop, Khaiyrgas-I)$. Result were plotted with two se intervals for the tree like topology of $f_d(Yoruba, Kolyma\_M; TestPop, Khaiyrgas-I)$ in which TestPop$_x$ corresponds to one of the ancient groups on the left y-axis (table S6).
Fig. S8

$F_4$-statistics for $f_4$(Yoruba, Anzick-1; Khaiyrgas-1, Kolyma_M) and $f_4$(Yoruba, Saqqaq; Khaiyrgas-1, Kolyma_M). Result plotted with two se intervals for the tree like topologies in the form of $f_4$(Yoruba, Anzick-1; Khaiyrgas-1, Kolyma_M) and $f_4$(Yoruba, Saqqaq; Khaiyrgas-1, Kolyma_M).
Fig. S9

$F_r$-statistics to formally test grouping of individuals on PCA. Tests were run for tree-like topology of $f4(Yoruba, ind_{1..n}group_x; ind_{y1..ngroup_y, ind_{x1..n}group_x})$. Each individual shared more alleles with individuals from the same group compared to the individuals from other groups ($Z>|3|$).
Fig. S10

Outgroup $f_3$-statistics for $\text{Yakutia}_{\text{Lena}}_{6850-6190}\text{ BP}$. Plot showing the $f_3$-statistics in the form of $f_3(\text{Yakutia}_{\text{Lena}}_{6850-6190}\text{ BP}, \text{Test Population}, \text{Yoruba})$ plotted with two se. Test population is one of the 194 present-day world populations from Human Origins dataset (table S7).
Fig. S11

**Outgroup $f_3$-statistics for Yakutia_Lena_Kolyma_4780_2490_BP.** Plot showing the $f_3$-statistics in the form of $f_3(Yakutia_Lena_Kolyma_4780_2490_BP, Test\ _Population, \ Yoruba)$ plotted with two se. Test population is one of the 194 present-day world populations from Human Origins dataset (table S8).
Northeast Asia- and Native America-related genetic components. A subset of ADMIXTURE analysis results for K=14 ancestral clusters for Yakutia individuals compared to Native American populations (Anzick1, purple component) and Northeast Asia populations (Nganasan, yellow component).
Fig. S13

$F_4$-statistics for $f_4$(Yoruba, Devil’s Cave; YakutiaPopX, YakutiaPopY). Result plotted with two standard se intervals for the tree like topologies in the form of $f_4$(Yoruba, Devil’s Cave; YakutiaPopX, YakutiaPopY) in which YakutiaPopX, YakutiaPopY corresponds to one of the ancient Yakutia groups (table S9).
Fig. S14

$F_4$-statistics for $f_4(\text{Yoruba, Anzick-1}; \text{YakutiaPopX, YakutiaPopY})$. Result plotted with two se intervals for the tree like topologies in the form of $f_4(\text{Yoruba, Anzick-1}; \text{YakutiaPopX}, \text{YakutiaPopY})$ which YakutiaPopX, YakutiaPopY corresponds to one of the ancient Yakutia groups. (table S9).
Fig. S15

F_4-statistics for $f_4$(Yoruba, Yakutia_Lena_6850_6190_B; PopX, Saqqaq). Result plotted with two se intervals for the tree like topology in the form of $f_4$(Yoruba, Yakutia_Lena_6850_6190_BP; PopX, Saqqaq) where PopX corresponds to one of the ancient groups (table S10).
Fig. S16

$F_{st}$-statistics for $f_4$(Yoruba, Saqqaq; PopX, Yakutia_Lena_6850_6190_BP). Result plotted with two se intervals for the tree like topology in the form of $f_4$(Yoruba, Saqqaq; PopX, Yakutia_Lena_6850_6190_BP) where PopX corresponds to one of the ancient groups (table S10).
Outgroup $f_3$-statistics for Dzyhlinda-1. Plot showing the $f_3$-statistics in the form of $f_3(Dzyhlinda-1, Test\_Population, Yoruba)$ plotted with two se. Test population is one of the 194 present-day world populations from Human Origins dataset (table S12).
Fig. S18

$F_r$-statistics for $f_4$(Yoruba, Dzyhinda-1; PopX, Yakutia_Lena_6850_6190_BP). Result plotted with two se intervals for the tree like topology in the form of $f_4$(Yoruba, Dzyhinda-1; PopX, Yakutia_Lena_6850_6190_BP) where PopX corresponds to one of the ancient groups (table S13).
Fig. S19

$F_{st}$-statistics for $f_4$(Yoruba, Yakutia_Lena_6850_6190_BP; PopX, Dzyhinda-1). Result plotted with two se intervals for the tree like topology in the form of $f_4$(Yoruba, Yakutia_Lena_6850_6190_BP;PopX, Dzyhinda-1) where PopX corresponds to one of the ancient groups (table S13).
Fig. S20

Outgroup $f_3$-statistics for Trans-Baikal_8345_3000_BP. Plot showing the $f_3$-statistics in the form of $f_3$(Trans-Baikal_8345_3000_BP;Test_Population,Yoruba) plotted with two se. Test population is one of the 194 present-day world populations from Human Origins dataset (table S14).
Fig. S21

Outgroup \( f_3 \)-statistics for \( bla001 \). Plot showing the \( f_3 \)-statistics in the form of \( f_3(bla001, Test\_Population, Yoruba) \) plotted with two standard errors (se). Test population is one of the 194 present-day world populations from Human Origins dataset (table S16).
Fig. S22

Distribution of long Runs of Homozygosity.
Fig. S23

Overview of the analyzed individuals with *Y. pestis*. Radiocarbon dates are compiled from (37). Rise509 and Rise511 contains the oldest *Y. pestis* genomes in Asia (found in Afanasievo Gora cultural site). Dates are shown in BCE.
Fig. S24

Summary for *Y. pestis* analysis in Anosovo-1. (a) The edit distance (less than 10 are shown in the plot), (b) length distribution observed from the reads that are aligned to *Y. pestis* CO92 reference genome in the Anosovo-1, (c) Nucleotide misincorporation patterns.
Fig. S25

The coverage distribution among the sequences in *Y. pestis* genome from the Anosovo-1.

Please note that this plot is to demonstrate the distribution of the coverage among the genomes. Because of the limited space on the X axis, the coverage could be misinterpreted.
**Fig. S26**

**Summary for *Y. pestis* analysis in Kamenka-2.** (a) The edit distance (less than 10 are shown in the plot), (b) length distribution observed from the reads that are aligned to *Y. pestis* CO92 reference genome in the Kamenka-2, (c) Nucleotide misincorporation patterns.
Fig. S27

The coverage distribution among the sequences in Y. pestis genome from the Kamenka-2.

Please note that this plot is to demonstrate the distribution of the coverage among the genomes. Because of the limited space on the X axis, the coverage could be misinterpreted.
**Fig. S28**

**Summary for Y. pestis analysis in Kamenka-1.** (a) The edit distance (less than 10 are shown in the plot), (b) length distribution observed from the reads that are aligned to *Y. pestis* CO92 reference genome in the Kamenka-1, (c) Nucleotide misincorporation patterns.
Fig. S29

The coverage distribution among the sequences in *Y. pestis* sequences from Kamenka-1. Please note that this plot is to demonstrate the distribution of the coverage among the references. Because of the limited space on the X-axis, the coverage could be misinterpreted.
Comparison of deamination rates and read length distribution between *Y. pestis* and human sequences of Ansanovo-1. (a) Deamination patterns of *Y. pestis* reads, (b) Read length distribution of *Y. pestis* reads, (c) Deamination patterns of human reads, (d) Read length distribution of human reads.
Comparison of deamination rates and read length distribution between *Y. pestis* and human sequences of Kamenka-2. (a) Deamination patterns of *Y. pestis* reads, (b) Read length distribution of *Y. pestis* reads, (c) Deamination patterns of human reads, (d) Read length distribution of human reads.
Fig. S32

**Comparison of deamination rates and read length distribution between Y. pestis and human sequences of Kamenka-1.** (a) Deamination patterns of Y. pestis reads, (b) Read length distribution of Y. pestis reads, (c) Deamination patterns of human reads, (d) Read length distribution of human reads.
Table S20.

Alignment statistics of the aligned *Y. pestis* reads in different samples. BF: Before read length and quality filtering, AF: After read length and quality filtering, BD: Before duplication removal, AD: After duplication removal.

<table>
<thead>
<tr>
<th>Sample</th>
<th>NO_reads_BF</th>
<th>NO_reads_AF</th>
<th>Mean_length_BD</th>
<th>Mean_coverage_BD</th>
<th>Clonality</th>
<th>NO_reads_AD</th>
<th>Mean_length_AD</th>
<th>Mean_coverage_AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>irk050</td>
<td>9523</td>
<td>9395</td>
<td>54.23</td>
<td>0.05</td>
<td>64.99</td>
<td>3289</td>
<td>54.17</td>
<td>0.02</td>
</tr>
<tr>
<td>yak022</td>
<td>2503</td>
<td>2480</td>
<td>46.57</td>
<td>0.01</td>
<td>33.87</td>
<td>1640</td>
<td>46.42</td>
<td>0.01</td>
</tr>
<tr>
<td>yak023</td>
<td>4227</td>
<td>4176</td>
<td>51.25</td>
<td>0.00</td>
<td>26.48</td>
<td>3070</td>
<td>51.37</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table S21.

The coverage statistics for the sequences in *Y. pestis* genome in the Anosovo-1. Coverage describes the percent covered regions at least one time (breadth of coverage more than 1X).

<table>
<thead>
<tr>
<th>NCBI_ID</th>
<th>Length</th>
<th>Coverage</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_003131.1</td>
<td>70305</td>
<td>3.41</td>
<td>plasmid pCD1</td>
</tr>
<tr>
<td>NC_003132.1</td>
<td>9612</td>
<td>0.52</td>
<td>plasmid pCP1</td>
</tr>
<tr>
<td>NC_003134.1</td>
<td>96210</td>
<td>0.97</td>
<td>plasmid pMT1</td>
</tr>
<tr>
<td>NC_003143.1</td>
<td>4653728</td>
<td>3.66</td>
<td>CO92 chromosome</td>
</tr>
</tbody>
</table>

Table S22.

The coverage statistics for the sequences in *Y. pestis* genome in Kamenka-2. Coverage describes the percent covered regions at least one time (percent coverage at least 1X).
<table>
<thead>
<tr>
<th>NCBI_ID</th>
<th>Length</th>
<th>Coverage</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_003131.1</td>
<td>70305</td>
<td>3.60</td>
<td>plasmid pCD1</td>
</tr>
<tr>
<td>NC_003134.1</td>
<td>96210</td>
<td>2.13</td>
<td>plasmid pMT1</td>
</tr>
<tr>
<td>NC_003143.1</td>
<td>4653728</td>
<td>3.16</td>
<td>CO92 chromosome</td>
</tr>
</tbody>
</table>

**Table S23.**

The coverage statistics for the sequences in *Y. pestis* genome in Kamenka-1. Coverage describes the percent covered regions at least one time (breadth of coverage more than one x).

<table>
<thead>
<tr>
<th>-</th>
<th>Length</th>
<th>Coverage</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC_003131.1</td>
<td>70305</td>
<td>1.58</td>
<td>plasmid pCD1</td>
</tr>
<tr>
<td>NC_003134.1</td>
<td>96210</td>
<td>1.02</td>
<td>plasmid pMT1</td>
</tr>
<tr>
<td>NC_003143.1</td>
<td>4653728</td>
<td>1.56</td>
<td>CO92 chromosome</td>
</tr>
</tbody>
</table>

**Table S24.**

Comparison of single stranded deamination rates of *Y. pestis* and human DNA reads. Single stranded deamination rates were calculated using mapDamage software.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Single strand deamination rate for <em>Y. pestis</em> reads</th>
<th>Single strand deamination rate for human reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>irk050</td>
<td>0.51 +/- 0.05</td>
<td>0.38 +/- 0.001</td>
</tr>
<tr>
<td>yak022</td>
<td>0.15 +/- 0.06</td>
<td>0.038 +/- 0</td>
</tr>
<tr>
<td>yak023</td>
<td>0.31 +/- 0.05</td>
<td>0.25 +/- 0.001</td>
</tr>
</tbody>
</table>
REFERENCES AND NOTES


83. V. I. Bazaliiskii, E. V. Menshagin, V. V. Golbert, V. N. Prokopijev, in Problems of Archaeology and Ethnography of Siberia (Irkutsk State University, 1982), pp. 68–69.


92. A. P. Okladnikov, I. I. Kirillov, South-East Transbaikalia in the Stone Age and Early Bronze Age (Nauka, 1980).


94. A. V. Kharinskii, E. V. Kovychev, N. N. Kradin, S. G. Bocharov, A. G. Sittikov, in Dialogue of Urban and Steppe Cultures in the Eurasian Range. Historical Geography of Golden Horde:

95. N. G. Dyatchina, The description of bone and horn articles found at Priargun Sites of Late Neolithic-Bronze Age. Gumanitarnyj vektor 30, 47–59 (2012).


104. N. P. Sentekjaeva, in *Archeology* (Novosibirsk State University, 2015), pp. 11–17.

105. N. Kashuba, thesis, Uppsala University, Disciplinary Domain of Humanities and Social Sciences, Faculty of Arts, Department of Archaeology and Ancient History, Archaeology (2017).


