Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia

Graphical Abstract

Highlights
- An Upper Paleolithic Siberian shows a deep link with the First Peoples of the Americas
- A 10,000-year continuum of Ancient North Eurasian ancestry in the Lake Baikal region
- The Neolithic to Bronze Age population formation occurred through prolonged local admixture
- Long-range human and Y. pestis mobility across Eurasia during the Early Bronze Age

Authors
He Yu, Maria A. Spyrou, Marina Karapetian, ..., Cosimo Posth, Choongwon Jeong, Johannes Krause

Correspondence
posth@shh.mpg.de (C.P.), jeong@shh.mpg.de (C.J.), krause@shh.mpg.de (J.K.)

In Brief
Genome-wide ancient DNA data from the Lake Baikal and its surroundings, comprising a time transect of 10,000 years from the Upper Paleolithic to the Early Bronze Age, reveals deeply divergent ancestry that links Upper Paleolithic Siberians and the First Peoples of the Americas and delineates the complex transition between Early Neolithic and Early Bronze Age populations in Siberia.

Yu et al., 2020, Cell 181, 1–14
June 11, 2020 © 2020 Elsevier Inc.
https://doi.org/10.1016/j.cell.2020.04.037
Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia

He Yu,1 Maria A. Spyrou,1 Marina Karapetian,2 Svetlana Shnaider,3 Rita Radzewičiūtė,1 Kathrin Nägele,1 Gunnar U. Neumann,1 Sandra Penske,1 Jana Zech,4 Mary Lucas,4 Petrus LeRoux,5 Patrick Roberts,4 Galina Pavlenok,3 Alexandra Buzhilova,2 Cosimo Posth,1,6,* Choongwon Jeong,1,7,* and Johannes Krause1,8,*

1Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena 07745, Germany
2Research Institute and Museum of Anthropology, Moscow State University, Moscow 125009, Russia
3Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia
4Department of Archaeology, Max Planck Institute for the Science of Human History, Jena 07745, Germany
5Department of Geological Sciences, University of Cape Town, Rondebosch 7701, South Africa
6Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen 72070, Germany
7School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea
8Lead Contact
*Correspondence: posth@shh.mpg.de (C.P.), jeong@shh.mpg.de (C.J.), krause@shh.mpg.de (J.K.)

SUMMARY

Modern humans have inhabited the Lake Baikal region since the Upper Paleolithic, though the precise history of its peoples over this long time span is still largely unknown. Here, we report genome-wide data from 19 Upper Paleolithic to Early Bronze Age individuals from this Siberian region. An Upper Paleolithic genome shows a direct link with the First Americans by sharing the admixed ancestry that gave rise to all non-Arctic Native Americans. We also demonstrate the formation of Early Neolithic and Bronze Age Baikal populations as the result of prolonged admixture throughout the eighth to sixth millennium BP. Moreover, we detect genetic interactions with western Eurasian steppe populations and reconstruct <i>Yersinia pestis</i> genomes from two Early Bronze Age individuals without western Eurasian ancestry. Overall, our study demonstrates the most deeply divergent connection between Upper Paleolithic Siberians and the First Americans and reveals human and pathogen mobility across Eurasia during the Bronze Age.

INTRODUCTION

The Lake Baikal region in Siberia has been inhabited by modern humans since the Upper Paleolithic and has a rich archaeological record (Katzenberg and Weber, 1999; Weber, 1995). In the past 5 years, ancient genomic studies have revealed multiple genetic turnovers and admixture events in this region. The 24,000-year-old individual (MA1) from the Mal’ta site represents an ancestry referred to as “Ancient North Eurasian (ANE),” which was widespread across Siberia during the Paleolithic (Fu et al., 2016; Raghavan et al., 2014a; Sikora et al., 2019) and that contributed to the genetic profile of a vast number of present-day Eurasian populations as well as Native Americans (Haak et al., 2015; Lazaridis et al., 2014, 2016; Raghavan et al., 2015). ANE ancestry was suggested to have been largely replaced in the Lake Baikal region during the Early Neolithic by a gene pool related to present-day eastern Siberia (Damgaard et al., 2018a). It is also suggested that a limited resurgence of ANE ancestry by the Early Bronze Age (Damgaard et al., 2018a).

Siberia has also been proposed as a source for multiple waves of dispersals into the Americas, the first of which was shown to be driven by a founding population estimated to have formed around 25,000–20,000 years before the present (BP) (Raghavan et al., 2015). The so-called Ancient Beringian ancestry represented by a 11,500-year-old Alaskan individual (USR1) was shown to be part of this founding population, estimated to have split from other Native Americans around 23,000 BP (Moreno-Mayar et al., 2018). In addition, the recently published 9,800-year-old Kolyma genome from northeastern Siberia was suggested to represent the closest relative to Native American populations outside of the Americas (Sikora et al., 2019). Moreover, the Paleo-Eskimo ancestry represented by a 4,000-year-old Saqqaq individual from Greenland was also estimated to have split from north-eastern Siberian groups and migrated to Arctic America around 6,000–5,000 BP (Flegontov et al., 2019; Raghavan et al., 2014b; Rasmussen et al., 2010). Although these waves of migration are generally linked to ancient Siberian populations, their origins in the context of the Siberian genetic history remain poorly understood. Further studies of the Siberian population history using ancient genomes are, therefore, critical for the better understanding of the formation of Native American populations.
Furthermore, the Neolithic to Bronze Age transition in Eurasia was marked by complex cultural and genetic changes facilitated by extensive population movements, though their impact in the Lake Baikal region is still unclear. Looking to the west, the Early Bronze Age groups from the Pontic-Caspian steppe associated with the Yamnaya complex spread both east and west along with their distinct genetic profile often referred to as “Steppe ancestry” (Allentoft et al., 2015; Haak et al., 2015). The eastward expansion of this group is considered to be associated with the Early Age Afanasievo culture. However, the later Middle Bronze Age Okunevo-related population from the central steppe as well as the Late Bronze Age Khövsgöl-related population from the eastern steppe harbor only a limited proportion of Steppe ancestry (Jeong et al., 2018, 2019). Therefore, the effect of steppe migrations in eastern Eurasia, particularly the interactions of Bronze Age Baikal hunter-gatherers with the contemporaneous and geographically proximal Afanasievo population, is still largely unexplored.

In this study, we report 19 newly sequenced ancienthunter-gatherers from the Lake Baikal and its surrounding regions, spanning from the Upper Paleolithic to the Early Bronze Age. Their analyses alongside published data reveal the most deeply divergent ancestry that link Upper Paleolithic Siberians and the First Peoples of the Americas, and more clearly delineate the complex transition between Early Neolithic and Early Bronze Age populations in the Lake Baikal region. We also provide both human and pathogen genomic evidence demonstrating the influence of western Eurasian steppe populations in this region during the Early Bronze Age and discuss the genetic contribution of Lake Baikal hunter-gatherers to Siberian populations through time.

RESULTS

Ancient DNA Sequencing

We generated genome-wide genotype data from 19 ancient humans, including one Upper Paleolithic individual (dated to 14,050–13,770 BP, see Oriova, 1995; Pavlenok et al., 2019), four Early Neolithic individuals (7,320–6,500 BP), and 14 Late Neolithic to Early Bronze Age (LNBA) individuals (4,830–3,570 BP) from a total of 10 archaeological sites (Figure 1; Table S1). The radiocarbon date offsets caused by the local freshwater reservoir were estimated using the carbon and nitrogen isotopic values as described in previous studies on the same region (see STAR Methods; Schulting et al., 2014, 2015). We built single- and double-stranded DNA libraries from teeth or petrous portions of the temporal bone for the studied individuals, and shotgun sequencing revealed high levels of DNA preservation with endogenous DNA contents ranging from 0.12% to 50.54% (Table S1). Subsequently, libraries were enriched for human DNA by SNP-capture targeting a set of 1.24 million variable sites (Fu et al., 2015) and sequenced to mean coverage ranging from 0.04X to 2.07X. Pseudo-haploid genotypes were called on the targeted SNPs by randomly sampling a single allele at each position, with 34k to 886k SNPs covered by our samples. Additionally, we performed deep shotgun sequencing on eight individuals with high endogenous DNA levels (12%–51%), to achieve genomic coverage that ranged from 0.1X to 1.9X and refined their diploid genotypes by genotype likelihood-based imputation, resulting in 386k to 518k SNPs overlapping with the Human Origins dataset (Table S1).

We determined genetic sex by comparing the coverage on the sex chromosomes with the autosomal chromosomes, which revealed four females and 15 males. All individuals revealed low modern human DNA contamination at the mitochondrial level as well as through an estimation of X chromosomal heterozygosity on male individuals, except for KAG001 that showed 9.6% nuclear contamination (Table S1). No kin relationship was found among these individuals. We finally intersected our genotypes with SNPs on the Affymetrix Human Origins array (Lazaridis et al., 2014) and combined with published genotype data from 3,014 present-day worldwide individuals and 453 ancient individuals for population genomic analysis (Table S1).

Population Structure

We first performed principal-component analysis (PCA) to understand the genetic background of the studied individuals, against modern Eurasian and Native American populations, and projected selected ancient individuals onto the PCs calculated with modern ones (Figure 1C). Most of the Lake Baikal individuals occupied the space on a “ANE-NEA” cline running between “Northeast Asian” (NEA) ancestry represented by Neolithic hunter-gatherers from the Devil’s Gate in the Russian Far East (Sikora et al., 2019; Siska et al., 2017), and the ANE ancestry represented by Upper Paleolithic Siberian individuals MA1, AfontovaGora 2 (AG2), and AfontovaGora 3 (AG3) (Fu et al., 2016; Raghavan et al., 2014a), which was first described by Damgaard et al. (2018a). Our newly sequenced Upper Paleolithic genome from the Ust-Kyakhta-3 site (UKY) just south to the Lake Baikal is placed close to the Mesolithic northeastern Siberian Kolyma individual (Sikora et al., 2019) and is shifted toward Native American populations compared to the rest of the ancient Baikal individuals along PC2. All four Early Neolithic individuals cluster with published Early Neolithic groups from the same region (Shamanka_EN, Lokomotiv_EN, UstBelaya_EN) (Damgaard et al., 2018a; Flegontov et al., 2019) designated as the “Baikal_EN” population. The LNBA individuals were divided into four groups. The major “Baikal_LNBA” group included 10 individuals and clustered with published Late Neolithic to Bronze Age Baikal populations (Shamanka_EBA, Kurma_EBA, UstIlda_EBA, UstIlda_LN, UstBelaya_BA). These individuals were positioned in PCA closer to ANE-related individuals compared with the Early Neolithic individuals from the same region, as well as closer to the Paleo-Eskimo Saqqaq individual (Rasmussen et al., 2010). Another two individuals (GLZ001 and GLZ002) from the Glazkovskoe predmestie site, unlike the third individual from the same archaeological site (GLZ003), seemed shifted from the main cluster and showed closer genetic affinity to the Devil’s Gate and Early Neolithic Baikal individuals. One of the six individuals from the Kachug site (KPT005) was substantially displaced from the Baikal_LNBA group toward western Eurasians along PC1, not along the ANE-NEA cline but toward later Bronze Age populations, suggesting a potential introgression of the Steppe-related ancestry. Finally, an Early Bronze Age individual (BZK002) from the Bazaika site in the Yenisei River region further to the west of the Lake Baikal was significantly displaced.
toward ANE-related individuals and located close to published Bronze Age individuals associated to the Okunevo culture (Damgaard et al., 2018a).

Population clustering with ADMIXTURE based on worldwide populations also showed a similar clustering pattern. When selecting a K value of 16 (see STAR Methods), the published and newly sequenced individuals belonging to main Early Neolithic to Bronze Age Baikal groups all showed genetic profiles composed of a mixture of three major components that were mostly enriched in ANE-related individuals, northeast

See also Figure S1 and Table S1.
Asians, and central Siberians represented by the Uralic-speaking Nganasan population (Figure 1D). The ANE and central Siberian ancestries were both of higher proportion in most LNBA Baikal individuals than in the Early Neolithic ones, while GLZ001 and GLZ002 showed higher NEA ancestry, similar to the Early Neolithic population. The BZK002 individual presented a profile similar to the published Okunevo group (Damgaard et al., 2018b), with a much larger ANE component compared to other Lake Baikal individuals. The KPT005 individual also displayed a substantial contribution derived from European “Western Hunter-Gatherer” (WHG) ancestry, likely acquired through gene flow from the west.

We estimated the runs of homozygosity (ROH) of selected individuals together with published Baikal individuals (Table S1) and did not identify an inbreeding signal in any individual. The Kolyma individual showed significantly more ROH compared with other individuals, suggesting a smaller population size in Mesolithic northeastern Siberia (Figure S1). The sharing of in-identity-by-descent (IBD) segments between individuals suggested a close relationship between UKY and Kolyma, supporting our analyses based on genome-wide SNP data, and also revealed that Early Neolithic and LNBA Baikal individuals shared genetic affinity with each other as well as with the older UKY and Kolyma genomes (Figure S1).

**Upper Paleolithic Baikal Ancestry Links with Non-Arctic Native Americans**

In the population structure analysis, we found the Upper Paleolithic UKY individual to be closely related with the northeastern Siberian Kolyma individual. This is further validated by outgroup $f_2$ statistics (Figure 2A) where, similarly to Kolyma, UKY showed close genetic affinity with Native American and Beringian populations (Figure 2A). $F_3$ statistics in the form of $f_{3}(\text{Mbuti}, X; \text{Kolyma}, \text{UKY})$ revealed that Kolyma is more closely related to populations from northeastern Siberia and North America compared with UKY (Figure S2). We further applied $f_2$ statistics to explore the relationship of UKY and Kolyma with Native Americans and USR1 that was described as an outgroup to all non-Arctic Native Americans (Moreno-Mayar et al., 2018). Both UKY and Kolyma were symmetrically related with non-Arctic Native Americans and USR1, while USR1 shared significantly more genetic affinity with Native American populations compared to UKY and Kolyma (Figure S2; Table S2).

We also investigated their genetic composition using qpAdm modeling (Haak et al., 2015) and found that both UKY and Kolyma possessed a similar level of ANE contribution, around 30%, when modeled as two-way mixture of Devil’s Gate (representing NEA ancestry) and AG3 (representing ANE ancestry) (Table S3). Noticeably, this model did not fit well for both UKY ($p = 1.45E-03$) and Kolyma ($p = 3.98E-08$), as the Native American Karitiana population showed extra affinity with the tested individuals compared to the fitted model (Table S3). This observation suggests that UKY and Kolyma shared a certain degree of genetic drift with Native American populations that occurred after the ancestors of Native Americans diverged from ANE and NEA ancestries.

We further explored the relationships among UKY, Kolyma, and ancient Native American groups using the graphic-based qpGraph modeling (Patterson et al., 2012; Reich et al., 2009). We found that both UKY and Kolyma could be modeled as mixture between a northeast Asian lineage and a sister lineage giving rise to distinct ancient Siberian populations. This basal Native American group experienced multiple genetic contacts with northeast Asian populations giving rise to distinct ancient Siberian populations.

**Complex Transition between the Early Neolithic and Bronze Age in the Lake Baikal Region**

A previous study described the transition between Early Neolithic and Bronze Age populations from the Lake Baikal region as the result of a discrete admixture event of ANE ancestry into the local gene pool (Damgaard et al., 2018a). In this study, we combined the newly sequenced Baikal_EN and Baikal_LNBA individuals with published data from the same time period (Figure 1C) and analyze these two combined datasets, Baikal_EN_all ($n = 19$) and Baikal_LNBA_all ($n = 34$), to better elucidate the genetic transition that occurred in this region. Prior to analyzing the combined groups, we confirmed the similarity between the new individuals and published groups using outgroup $f_2$ statistics. Both Baikal_EN and Baikal_LNBA groups showed the highest genetic affinity with published Early Neolithic and LNBA Baikal populations, respectively (Table S4).

From outgroup $f_2$ statistics of the combined groups, we found both of the Baikal Early Neolithic and LNBA groups to be sharing high genetic affinity with ancient and modern northeast Asian and Siberian populations (Figure S3). The LNBA Baikal population also showed a high genetic affinity with the Paleo-Eskimo Saqqaq individual. Compared to their NEA proxy, they both carried extra genetic affinity with ANE-related populations while the LNBA population more so than the Early Neolithic population, as shown by $f_4$ statistics (Figure S3). These results revealed the existence of ANE-related ancestry in the Early Neolithic population and, at the same time, validated the previous finding that an extra ANE ancestry gene flow is responsible for the genetic shift between Early Neolithic and Bronze Age Baikal populations.

We further applied qpAdm modeling to quantify the proportion of ANE-related ancestry in Early Neolithic and LNBA Baikal populations, Saqqaq and Nganasan. Using Devil’s Gate as the NEA proxy, the Upper Paleolithic UKY was found to be a better fit than Kolyma as the ANE-related proxy for both Baikal populations,
Figure 2. Genetic Affinity between Upper Paleolithic UKY, Kolyma, and Native Americans

(A) Genetic affinity between UKY and worldwide population assessed by $f_3$ (Mbuti; X, UKY). The sampling location of UKY is shown with a green triangle. The 10 test populations with highest $f_3$ values are shown in diamonds and other populations in circles.

(B) Graphic model of the relationship among UKY, Kolyma, and Native American populations. We first find the best fitted model with only UKY or only Kolyma as described in Figure S2 and then add Kolyma on the selected model with UKY and choose the best model based on maximum $f$-statistics $Z$ scores and final scores reported for each model. The lineages related with Native American population are colored orange, and the northeast Asian-related lineages are colored red. See also Figure S2 and Table S2.
AG3 provided a good fit for the Early Neolithic population (Table S5). Using Devil’s Gate and AG3 as the two proxies of NEA and ANE ancestries, respectively, we estimated the ANE-related ancestry increasing from 14.3% in the Early Neolithic Baikal population to 22.7% in the LNBA population (Figure 3A; Table S3). Of note, the northeastern Siberian Kolyma individual could work as a sufficient ANE proxy for Saqqaq, as described in the study where this genome was first reported (Sikora et al., 2019) but did not provide a good fit for the Baikal populations and Nganasan. This suggests that the Baikal hunter-gatherer and Nganasan populations are more likely to have formed in central or southern Siberia while Paleo-Eskimo ancestry could have emerged in either central or northeastern Siberia.

Furthermore, the program DATES was used to date the admixture events between ANE and NEA ancestries in the Baikal population based on the decay of ancestry covariance (Moorjani and Patterson, 2018). We detected a recent admixture signal in the Early Neolithic population, estimated to around 21 generations ago, while the admixture signal in LNBA population was dated to 71 generations ago, although this group harbored significantly more ANE-related ancestry (Table S3). When considering the average radiocarbon date of each population and the standard errors of their admixture dates, we identified contiguous intervals for the admixture events that spanned ~8,500–6,000 BP, considering a generation time of 29 years (Figure 3B; Figure S4; Table S1; Table S5). Assuming a dating offset of 400–500 years due to freshwater reservoir effect estimated for the newly reported individuals, the admixture timings ranged between ~8,000 and 5,500 BP. This suggests that both Baikal populations could have been formed through an extended admixture process between local groups and northeast Asian-related populations. The Early Neolithic groups were thus found to have experienced a prolonged admixture process, in contrast to the discrete and rather abrupt event suggested earlier (Damgaard et al., 2018a). This admixture, however, did not continue substantially in the Late Neolithic and Bronze Age, as suggested by the older admixture date for the LNBA population (Figure 3B) and the relatively larger genetic variation among Early Neolithic individuals compared to the homogeneous LNBA cluster, as shown in the PCA plot (Figure 1C).

High Mobility in Bronze Age Siberia Revealed by Genetic Outliers
Among the LNBA Baikal individuals, we identified three outlier individuals, which showed distinct genetic backgrounds from the major group. The genetic profile of the KPT005 individual suggested its affinity to western Eurasian populations inferred from PCA and ADMIXTURE results (Figure 1). Using qpAdm modeling, we found that this individual could indeed be modeled as the mixture between the LNBA Baikal and multiple Bronze Age western steppe populations, with the Steppe ancestry contribution ranging between 42% and 48% (Figure 4A; Table S5). Considering geography and radiocarbon dates, we suggest the most likely candidate in our dataset for the source of...
admixtures to be the Yamnaya-related Afanasievo population from nearby Altai-Sayan region.

Additionally, we also identified two Early Bronze Age Baikal individuals (GLZ001 and GLZ002) showing significantly lower ANE admixture levels compared to the major Baikal_LNBA group (Figure 1C). We estimated the ANE contribution in these two individuals to be around 10%, using Devil’s Gate and AG3 as proxies of NEA and ANE ancestries, respectively (Figure 4A; Table S3). Furthermore, these two individuals both carried Y chromosome haplogroup C2b1, which was predominant in Early Neolithic Baikal populations (Table S1), while the other LNBA individuals carried exclusively Q1a haplogroup. This could indicate either a relic population with ANE ancestry proportion similar to, or even less than, the Early Neolithic population preserved until the Bronze Age, or immigrants from surrounding regions carrying more NEA ancestry.

The mobility of the Baikal individuals during their lifetimes was thus investigated through the application of strontium isotope analysis to the enamel of eleven teeth samples (Bentley et al., 2004; Price et al., 2002). We found that ten out of the eleven individuals all had values within the bioavailable 87Sr/86Sr range (0.708–0.711) comprehensively reported for the Baikal region /C24 (0.709–0.712) (Haverkort et al., 2010). By contrast, GLZ001 had a 87Sr/86Sr value >0.713 that suggested a non-local origin for this individual, thus investigated through the application of strontium isotope analysis to the enamel of eleven teeth samples (Bentley et al., 2004; Price et al., 2002). We found that ten out of the eleven individuals all had values within the bioavailable 87Sr/86Sr range (0.708–0.711) comprehensively reported for the Baikal region /C24 (0.709–0.712) (Haverkort et al., 2010). By contrast, GLZ001 had a 87Sr/86Sr value >0.713 that suggested a non-local origin for this individual, followed by movement to the Angara River Valley region after the period of formation of the second molar (>7 years of age).
(Alexander Bentley, 2006). Based on existing bioavailable strontium data, sources of origin plausibly include areas to the south and east of Lake Baikal on Archean and Proterozoic geology (Haverkort et al., 2010). GLZ002 appears instead “local” to the Angara River Valley in terms of $^{87}$Sr/$^{86}$Sr, despite the genetic similarity to individual GLZ001. Though a speculation, it is possible that GLZ002 could have migrated to the region during childhood, prior to 7 years of age and, hence, prior to the formation of the second molar, or from a region with similar $^{87}$Sr/$^{86}$Sr range to the Angara River Valley.

Identification of Y. pestis Infections among Genetic Outlier Individuals

Recent studies have reported evidence of Yersinia pestis infections in humans across multiple regions of Europe and central Asia between the Middle Neolithic and Bronze Age (~4,900–3,500 BP) (Andrades Valtueña et al., 2017; Rascovan et al., 2019; Rasmussen et al., 2015b; Spyrou et al., 2018). To date, the majority of identified strains group on an extinct phylogenetic lineage previously designated as the LNBA lineage (Andrades Valtueña et al., 2017). Such data have been interpreted alongside human population genetic frameworks, suggesting the spread of this lineage via Yamnaya-related human migrations across Eurasia during the Late Neolithic and Early Bronze Age (Andrades Valtueña et al., 2017).

To investigate the presence of pathogen DNA signatures among individuals from the Lake Baikal region, we screened all analyzed specimens (n = 19) using the pipeline HOPS (Hüblicher et al., 2019). Our metagenomic analysis of shotgun sequenced data revealed evidence of ancient Y. pestis DNA in two Early Bronze Age individuals (GLZ001 and GLZ002) (Figure S5), which were dated to 4,556 and 4,430 BP (median dates), respectively, after accounting for the freshwater reservoir effect present in this region (Table S1). Subsequent whole-Y. pestis-genome enrichment yielded a 7.2- and 12.8-fold genomic coverage for GLZ001 and GLZ002, respectively (Table S6). We built a maximum likelihood phylogeny in order to compare the newly reconstructed genomes with previously published ancient and modern Y. pestis isolates (Figure 4B). GLZ001 and GLZ002 grouped on the previously described LNBA Y. pestis lineage (Andrades Valtueña et al., 2017; Rasmussen et al., 2015b), together with 4,800- to 3,500-year-old isolates from across Eurasia (Figure 4B). Specifically, they appeared most closely related (distance $d = 17–19$ SNPs) to a genome from the Baltic Sea region (Kunila2, 4,520–4,290 BP), retrieved from an individual culturally affiliated with the Corded Ware complex (Figure 4B; Andrades Valtueña et al., 2017). Such a result is intriguing since, based on human genetic analyses, individuals GLZ001 and GLZ002 appear as genetic outliers compared to other analyzed Bronze Age individuals from the same region. In addition, they lack Yamnaya-related Steppe ancestry (Figure 1; Figure 4A), common to all published plaque victims with Y. pestis genomes grouping on the LNBA phylogenetic lineage (Andrades Valtueña et al., 2017; Rasmussen et al., 2015b).

To further characterize the Y. pestis genomic contents of GLZ001 and GLZ002, we computed the coverage across previously defined virulence-associated and evolutionary-determinant genes (Demeure et al., 2019; Zhou and Yang, 2009). Our analysis revealed similar patterns to those previously observed in LNBA strains, which are consistent with a decreased efficiency in flea-mediated transmission (Figure S5). These include the absence of ymt on the pMT1 plasmid and the presence of active PDE-2, PDE-3, rcsA, and ureD gene variants (Figures S5 and S6; Andrades Valtueña et al., 2017; Rasmussen et al., 2015b; Sun et al., 2014). Additionally, we identified here the absence of ympf1.66c (Figure S5), a putative helicase considered to be involved in the bacterium’s initial within-macrophage survival and subsequent bubonic disease progression (Pradel et al., 2014).

Moreover, to investigate whether the radiocarbon date correction using stable isotope values could provide a reliable estimation of the freshwater reservoir offsets in GLZ001 and GLZ002, we explored an additional dating approach for these specimens based on the Y. pestis molecular phylogeny. For this, we first assessed the temporal phylogenetic signal across the LNBA lineage using previously published radiocarbon dates as tip calibrations (Andrades Valtueña et al., 2017; Rasmussen et al., 2015b). The resulting correlation coefficient ($r$) and $R^2$ values were 0.88 and 0.78, respectively (Figure S5), which supported a linear relationship between root-to-tip genetic distance and specimen age for this dataset (including a modern 0.PE2 genome as outgroup). As such, we used the Bayesian framework BEAST (Drummond and Rambaut, 2007) to independently re-estimate the tip dates of GLZ001 and GLZ002 and tested the coalescent constant size and coalescent skyline tree priors in combination with strict and lognormal relaxed clock models (see STAR Methods). All runs produced overlapping estimates supporting median ages of ~4,400 BP for both genomes (95% highest posterior density [HPD] range across all models: 4,115–4,590 BP) (Figure 4C; Table S6). As such, our analysis supports the estimated offset in the radiocarbon date estimates of both Glazkovskoe predmestie individuals, caused by the freshwater reservoir effect described previously for the Lake Baikal region (Schulting et al., 2014; Weber et al., 2016).

Taken together, these results provide evidence for the presence of human Y. pestis infections in the Lake Baikal region during the Early Bronze Age (Figure 4B). To our knowledge, this is the easternmost appearance of Y. pestis strains associated with the LNBA lineage, despite the lack of Steppe-related ancestry in both affected individuals.

Genetic Influence on the Okunevo Culture

The Okunevo is a Bronze Age culture of the central steppe, geographically located west of the Lake Baikal region (Figure 1A). Previous studies have suggested a genetic relationship between Okunevo and ancient Lake Baikal individuals (Damgaard et al., 2018b; Jeong et al., 2019). In this study, we detected a high genetic affinity between the Early Bronze Age BZK002 genome from the region where the Okunevo culture was subsequently practiced and Okunevo-related individuals (Damgaard et al., 2018a), as revealed by PCA and outgroup $f_2$ statistics (Figure 1C; Table S4). The BZK002 genome could be modeled with qpAdm as a two-way admixture between LNBA Baikal and AG3 or Botai-like populations, which are central Eurasian populations without any Steppe ancestry (Figure 5A). BZK002 also performed as a good proxy for the populations that mixed with Yamnaya/
Afanasievo-related groups to form the ancestry observed in Okunevo-associated individuals (Table S3). Moreover, we verified the modeling of Okunevo as a three-way admixture among the LNBA Baikal group, AG3 or Botai-like individuals, and Steppe-related populations (Figures 5A; Table S3).

Finally, we used DATES to estimate the timing of admixture events among those three ancestries in the Okunevo-related population, as well as to estimate the admixture timing of LNBA Baikal and Botai ancestries in the BZK002 individual. The estimated dates of these admixture events fell into two distinct time periods. Using LNBA Baikal and Botai-like populations as sources, the admixture events in Okunevo and BZK002 were dated to 42.9 and 23.4 generations ago, respectively, corresponding to an overlapping range of 6,000–5,000 BP, considering the specimens ages. The estimations including Yamnaya-related populations as one of the sources in Okunevo were instead both younger, around 17 generations, corresponding to 1,500–4,500 BP (Figure 5B; Figure S4; Table S5). Based on the dating results, we propose a scenario where the formation of the Okunevo-related gene pool resulted from an initial admixture of Botai-like and LNBA Baikal ancestries, followed by a gene flow of Yamnaya-related ancestry via its eastward expansion during the fifth millennium BP. BZK002 was directly radiocarbon dated to around 4,700 BP, predating the published Okunevo-related individuals by 200–800 years and overlapping with the estimated time range of the Steppe-ancestry admixture. Moreover, the Baikal-Botai admixture for BZK002 falls within the same period estimated for the Okunevo population, suggesting that this individual could represent an intermediate status during the formation of the Okunevo-associated genetic profile.

DISCUSSION

In this study, we demonstrated that multiple population turnovers took place in the Lake Baikal region from the Upper Paleolithic to the Bronze Age periods as a dynamic variation of ANE and NEA ancestry proportions. However, instead of the suggested complete replacement of the ANE-related ancestry by an East Asian gene pool (Damgaard et al., 2018a), we show that the ancestry first described in the 24,000-year-old MA1 individual survived in the region throughout the Upper Paleolithic until the Bronze Age, as revealed in the genome of the 14,000-year-old UKY individual as well as in the Early Neolithic population. Furthermore, the genetic transition from Early Neolithic to Bronze Age Baikal groups could be well explained by a prolonged gene flow between NEA- and ANE-related ancestries throughout the eighth to the sixth millennium BP (Figure 5B), and the UKY genome was attested to be a better local source for the ANE ancestry than a possible southward expansion of Kolyma-related ancestry (Sikora et al., 2019). In contrast, the Kolyma genome represents a good proxy for the Paleo-Eskimo Saqqaq genetic profile, supporting the view that the Paleo-Eskimo ancestry emerged around northeastern Siberia (Flegontov et al., 2019), though we are unable to rule out the scenario that this ancestry formed in central or southern Siberia from an UKY-related gene pool.

Noticeably, we detect a strong genetic connection between Upper Paleolithic Siberians from the Lake Baikal region with non-Arctic Native Americans. According to our demographic modeling, Native Americans, the 14,000-year-old southern Siberian UKY individual and the 9,800-year-old northeastern Siberian Kolyma individual were all descendants, at least in part, of the same admixed population that carried both ANE and NEA ancestries. This population was likely widespread across Siberia during the Upper Paleolithic and experienced frequent genetic contacts with northeast Asian-related populations, which resulted in varying proportions of ANE/NEA ancestry through different areas and time periods. Considering the earlier population split of UKY compared with Kolyma and the Alaskan USR1 genome from the region throughout the Upper Paleolithic and experienced frequent genetic contacts with northeast Asian-related populations, which resulted in varying proportions of ANE/NEA ancestry through different areas and time periods. Considering the earlier population split of UKY compared with Kolyma and the Alaskan USR1 genome from the lineage leading to the rest of Native American populations, our result challenges the hypothesis that this basal Native American ancestry formed in northeastern Siberia (Sikora et al., 2019).

Further genetic evidence from Upper Paleolithic Siberian groups will be necessary to describe where and when exactly the ancestral gene pool of Native Americans came together.

Moreover, we provided evidence for the genetic contact between southern Siberia and the western steppe during the Early Bronze Age, based on both human and pathogen DNA data, which suggests a high mobility across Eurasia. The genetic influence of Yamnaya-related populations in the Lake Baikal region is evident by the presence of Steppe ancestry in individual KPT005 (Figure 4A). In addition, previous studies have
suggested that such migrations likely also facilitated the spread of the Y. pestis LNBA lineage across Eurasia (Andrades Valtüena et al., 2017), as all individuals associated with this lineage carry Steppe ancestry (Allentoft et al., 2015; Andrades Valtüena et al., 2017; Haak et al., 2015; Mathieson et al., 2018; Mittnik et al., 2018, 2019; Wang et al., 2019). In this study, the two individuals shown to be infected with Yersinia pestis (GLZ001 and GLZ002) did not display genetic evidence of Yamnaya-related Steppe ancestry in their genomes but had substantially more NEA ancestry than all other Bronze Age individuals from the same and other surrounding sites (Figure 1). Moreover, one of the two individuals (GLZ001) showed a non-local signal in strontium isotope analysis, supporting a homeland outside the Angara River Valley. Instead, based on the Y. pestis phylogeny, the newly reconstructed Baikal genomes were found to be genetically closest related to a strain from the Baltic region in northeastern Europe (Figure 4B), isolated from an individual associated with the Corded Ware complex (Andrades Valtüena et al., 2017; Mittnik et al., 2018). Although our current resolution is insufficient for inferring patterns of pathogen transmission between western Eurasia and southern Siberia, the presented data may rather be indicative of a century- or decade-long process that lead to the bacterium’s long-distance spread. Nevertheless, the phylogenetic topology of both Lake Baikal genomes within the LNBA lineage most parsimoniously suggests that Y. pestis spread into this region within the context of Yamnaya-related steppe expansions during the fifth millennium BP (Andrades Valtüena et al., 2017). Importantly, our results provide prime evidence that strains of the LNBA Y. pestis lineage also affected individuals that were not genetically impacted by such migrations.

In conclusion, our study describes dynamic changes in the population structure of the Lake Baikal region and reveals a widespread occurrence of the genetic ancestry that gave rise to the First People of the Americas in Upper Paleolithic Siberia. In addition, we provide evidence of high human mobility across the Eurasian steppe revealing new insights about the spread of Y. pestis during the fifth millennium BP.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Archaeological information
- **METHOD DETAILS**
  - Radiocarbon dating and calibration
  - Stable isotope analysis
  - Strontium isotope analysis
  - Ancient DNA processing
  - Human genome enrichment and sequencing
  - Y. pestis genome enrichment

**QUANTIFICATION AND STATISTICAL ANALYSIS**

- Genotyping and dataset preparation for analysis
- Population genetics analysis
- Admixture modeling with qpAdm
- Demographic modeling with qpGraph
- Admixture dating with DATES
- Pathogen DNA screening with HOPS
- Read processing, SNP calling and phylogenetic analysis of Yersinia pestis
- BEAST v1.8 tip dating of GLZ001 and GLZ002
- Assessment of Y. pestis gene profiles

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2020.04.037.

**ACKNOWLEDGMENTS**

We thank the wet laboratory and computational teams at MPI-SHH. We are grateful to Matthias Meyer and Sarah Nagel from the Max Planck Institute for Evolutionary Anthropology in Leipzig for generating the single-stranded library of UKY. We thank Stephan Schiffels for his suggestions and comments and Aida Andrades Valtüena for computational support. G.P. was supported by the Russian Foundation for Basic Research (Project No. 19-59-44010). The human and pathogen genomics research were funded by the Max Planck Society and the European Research Council ERC-CoG 771234 PALEoRIDER.

**AUTHOR CONTRIBUTIONS**


**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: November 18, 2019
Revised: March 16, 2020
Accepted: April 21, 2020
Published: May 20, 2020

**REFERENCES**


Oalaide, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Zhan, Y., Genet, M.-L. 373 s.


R Development Core Team (2008). Computational Many-Particle Physics (Springer Berlin Heidelberg).


Vlast'rud'a (1928). 21 oktyabrya. 2746, 6.

Vlast'rud'a (1929). 28 iyulya. 17, 4.


### REAGENT or RESOURCE SOURCE IDENTIFIER

<table>
<thead>
<tr>
<th>Biological Samples</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>ANG001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>BZK002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>GLZ001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>GLZ002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>GLZ003</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>IU0001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KAG001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KAG002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT003</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT004</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT005</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>KPT006</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>STB001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>STB002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>ZPL001</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>ZPL002</td>
</tr>
<tr>
<td>Ancient skeletal element</td>
<td>This study</td>
<td>UKY001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chemicals, Peptides, and Recombinant Proteins</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D1000 ScreenTapes</td>
<td>Agilent Technologies</td>
<td>Cat# 5067-5582</td>
</tr>
<tr>
<td>D1000 Reagents</td>
<td>Agilent Technologies</td>
<td>Cat# 5067-5583</td>
</tr>
<tr>
<td>PfuTurbo Cx Hotstart DNA Polymerase</td>
<td>Agilent Technologies</td>
<td>Cat# 600412</td>
</tr>
<tr>
<td>Herculase II Fusion DNA Polymerase</td>
<td>Agilent Technologies</td>
<td>Cat# 600679</td>
</tr>
<tr>
<td>1x Tris-EDTA pH 8.0</td>
<td>AppliChem</td>
<td>Cat# A8569,0500</td>
</tr>
<tr>
<td>Sodiumhydroxide Pellets</td>
<td>Fisher Scientific</td>
<td>Cat# 10306200</td>
</tr>
<tr>
<td>Sera-Mag Magnetic Speed-beads</td>
<td>GE LifeScience</td>
<td>Cat# 6515210500250</td>
</tr>
<tr>
<td>Carboxylate-Modified (1 mm, 3EDAC/PA5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 M EDTA pH 8.0</td>
<td>Life Technologies</td>
<td>Cat# AM9261</td>
</tr>
<tr>
<td>10x Buffer Tango</td>
<td>Life Technologies</td>
<td>Cat# BY5</td>
</tr>
<tr>
<td>GeneRuler Ultra Low Range DNA Ladder</td>
<td>Life Technologies</td>
<td>Cat# SM1211</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Merck</td>
<td>Cat# 1070222511</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Merck</td>
<td>Cat# 1009832511</td>
</tr>
<tr>
<td>USER enzyme</td>
<td>New England Biolabs</td>
<td>Cat# M5505</td>
</tr>
<tr>
<td>Uracil Glycosylase inhibitor (UGI)</td>
<td>New England Biolabs</td>
<td>Cat# M0281</td>
</tr>
<tr>
<td>Bst DNA Polymerase2.0, large frag.</td>
<td>New England Biolabs</td>
<td>Cat# M0537</td>
</tr>
<tr>
<td>BSA 20mg/mL</td>
<td>New England Biolabs</td>
<td>Cat# B9000</td>
</tr>
<tr>
<td>T4 Polynucleotide Kinase</td>
<td>New England Biolabs</td>
<td>Cat# M0201</td>
</tr>
<tr>
<td>T4 DNA Polymerase</td>
<td>New England Biolabs</td>
<td>Cat# M0203</td>
</tr>
<tr>
<td>PEG-8000</td>
<td>Promega</td>
<td>Cat# V3011</td>
</tr>
<tr>
<td>20% SDS</td>
<td>Serva</td>
<td>Cat# 39575.01</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Sigma Aldrich</td>
<td>Cat# P2308</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guanidine hydrochloride</td>
<td>Sigma Aldrich</td>
<td>Cat# G3272</td>
</tr>
<tr>
<td>3M Sodium Acetate (pH 5.2)</td>
<td>Sigma Aldrich</td>
<td>Cat# S7899</td>
</tr>
<tr>
<td>Water</td>
<td>Sigma Aldrich</td>
<td>Cat# 34877</td>
</tr>
<tr>
<td>Tween-20</td>
<td>Sigma Aldrich</td>
<td>Cat# P9416</td>
</tr>
<tr>
<td>5M NaCl</td>
<td>Sigma Aldrich</td>
<td>Cat# S5150</td>
</tr>
<tr>
<td>Denhardt’s solution</td>
<td>Sigma Aldrich</td>
<td>Cat# D9905</td>
</tr>
<tr>
<td>ATP 100 mM</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# R0441</td>
</tr>
<tr>
<td>1 M Tris-HCl pH 8.0</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 15568025</td>
</tr>
<tr>
<td>dNTP Mix</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# R1121</td>
</tr>
<tr>
<td>SSC Buffer (20x)</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# AM9770</td>
</tr>
<tr>
<td>GeneAmp 10x PCR Gold Buffer</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 4379874</td>
</tr>
<tr>
<td>Dynabeads MyOne Streptavidin T1</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 65602</td>
</tr>
<tr>
<td>Salmon sperm DNA</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 15632-011</td>
</tr>
<tr>
<td>Human Cot-I DNA</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 15279011</td>
</tr>
<tr>
<td>0.5M HCl</td>
<td>Carl Roth</td>
<td>Cat# 9277.1</td>
</tr>
<tr>
<td>HNO₃</td>
<td>Merck</td>
<td>Cat# 1.00456.2500</td>
</tr>
</tbody>
</table>

Critical Commercial Assays

<table>
<thead>
<tr>
<th>Critical Commercial Assays</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Pure Viral Nucleic Acid Large Volume Kit</td>
<td>Roche</td>
<td>Cat# 5114403001</td>
</tr>
<tr>
<td>HiSeq 4000 SBS Kit (50/75 cycles)</td>
<td>Illumina</td>
<td>Cat# FC-410-1001/2</td>
</tr>
<tr>
<td>DyNaMo Flash SYBR Green qPCR Kit</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# F415L</td>
</tr>
<tr>
<td>MinElute PCR Purification Kit</td>
<td>QIAGEN</td>
<td>Cat# 28006</td>
</tr>
<tr>
<td>Quick Ligation Kit</td>
<td>New England Biolabs</td>
<td>Cat# M2200L</td>
</tr>
<tr>
<td>Oligo aCGH/Chip-on-Chip Hybridization Kit</td>
<td>Agilent Technologies</td>
<td>Cat# 5188-5220</td>
</tr>
</tbody>
</table>

Deposited Data

<table>
<thead>
<tr>
<th>Deposited Data</th>
<th>ENA: PRJEB37007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw and analyzed data (European nucleotide archive)</td>
<td>This study</td>
</tr>
<tr>
<td>1240K Genotype data (Edmond Data Repository of Max Planck Society)</td>
<td>This study</td>
</tr>
</tbody>
</table>

Software and Algorithms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EAGER 1.92.55</td>
<td>Schubert et al., 2016</td>
<td><a href="https://github.com/MikkelSchubert/adapterremoval">https://github.com/MikkelSchubert/adapterremoval</a></td>
</tr>
<tr>
<td>AdapterRemoval 2.2.0</td>
<td>Li and Durbin, 2009</td>
<td><a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a></td>
</tr>
<tr>
<td>BWA 0.7.12</td>
<td>Peltzer et al., 2016</td>
<td><a href="https://github.com/apeltzer/DeDup">https://github.com/apeltzer/DeDup</a></td>
</tr>
<tr>
<td>DeDup 0.12.1</td>
<td>Jónsson et al., 2013</td>
<td><a href="https://github.com/ginolhac/mapDamage">https://github.com/ginolhac/mapDamage</a></td>
</tr>
<tr>
<td>bamUtil 1.0.13</td>
<td><a href="https://github.com/statgen/bamUtil">https://github.com/statgen/bamUtil</a></td>
<td><a href="https://github.com/statgen/bamUtil">https://github.com/statgen/bamUtil</a></td>
</tr>
<tr>
<td>CircularMapper</td>
<td>Peltzer et al., 2016</td>
<td><a href="https://github.com/apeltzer/CircularMapper">https://github.com/apeltzer/CircularMapper</a></td>
</tr>
<tr>
<td>Schmutzi</td>
<td>Renaud et al., 2015</td>
<td><a href="https://github.com/grenaud/schmutzi">https://github.com/grenaud/schmutzi</a></td>
</tr>
<tr>
<td>SAMtools 1.3</td>
<td>Li et al., 2009</td>
<td><a href="http://www.htslib.org/doc/samtools.html">http://www.htslib.org/doc/samtools.html</a></td>
</tr>
<tr>
<td>GATK 3.5</td>
<td>DePristo et al., 2011</td>
<td><a href="https://gatk.broadinstitute.org/hc/en-us">https://gatk.broadinstitute.org/hc/en-us</a></td>
</tr>
<tr>
<td>Genelmp 1.4</td>
<td>Spiliopoulou et al., 2017</td>
<td><a href="https://pm2.phs.ed.ac.uk/genelmp/">https://pm2.phs.ed.ac.uk/genelmp/</a></td>
</tr>
<tr>
<td>SHAPEIT v2.r837</td>
<td>Delaneau et al., 2013</td>
<td><a href="https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html">https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html</a></td>
</tr>
<tr>
<td>yHaplo</td>
<td>Poznik, 2016</td>
<td><a href="https://github.com/23andMe/yhaplo">https://github.com/23andMe/yhaplo</a></td>
</tr>
<tr>
<td>ANGSD 0.910</td>
<td>Korneliussen et al., 2014</td>
<td><a href="http://www.popgen.dk/angsd/index.php/ANGSD">http://www.popgen.dk/angsd/index.php/ANGSD</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johannes Krause (krause@shh.mpg.de).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
Aligned reads to the human reference genome and mtDNA from the 19 newly reported ancient individuals and the Y. pestis capture sequencing reads of two individuals are available at the ENA archive under the accession number PRJEB37007. Haploid genotype data for the 1240K panel are available in the eigenstrat format at the Edmond Data Repository of Max Planck Society (https://edmond.mpdl.mpg.de/imeji/collection/z9C4DN1vFlqsmbHO).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological information
In this study we collected 7 petrous bones and 13 teeth from 19 individuals excavated from 10 different sites, including one Ust-Kyakhta-3 site from the south of Lake Baikal and the other nine sites in cis-Baikal region.

Sample from the Western Trans-Baikal
Ust-Kyakhta-3 Site.

- UKY001, tooth fragment (Ust'Kyakhta-3 1976 Excavation area 1 Layer 1 Square B-3)

The stratified Ust-Kyakhta-3 site is located on the right bank of the Selenga River in the vicinity of Ust-Kyakhta village in the Kyakhtinski Region of the Republic of Buryatia (Russian Federation). The site was discovered in 1947 by an archaeological team led by Academician A. P. Okladnikov. Initial excavations at Ust-Kyakhta-3 were carried out in 1976 and 1978, revealing two culture-bearing horizons. More than 40,000 stone artifacts were recovered; abundant faunal remains were also collected (Okladnikov, 1977, 1979). Archaeological work undertaken at the site in 2012 (Pavlenok, 2015a) yielded two human tooth fragments among archaeological materials collected in the course of sieving the sediments of Layer 1 (Pavlenok et al., 2019).

Excavations at Ust-Kyakhta-3 in 2012 provided additional information on the stratigraphy of the site. The stratigraphic column was subdivided into 12 lithological strata. Strata from 1 to 5 represent diluvial and aeolian sediments. The lower segment of the column, including Strata 6 through 12, was formed by flood-plain alluvium. Archaeological materials were noted in association with two
lithological strata. Archaeological Layer 1 is enclosed within lithological Stratum 9, comprising a humus-containing fine-grained sandy loam from 4 to 12 cm thick. Archaeological Layer 2 is associated with lithological Stratum 11, composed of light brown sandy loam from 15 to 20 cm thick.

Cultural remains associated with Ust-Kyakhta-3 Layer 1 were accumulated during the terminal Pleistocene with associated dates of 14,165 ± 236 – 14,183 ± 234 cal BP. Archaeological materials from Layer 2 yielded similar dates ranging from 13,758 ± 138 – 14,326 ± 266 cal BP (calibration was carried out following http://www.calpal-online.de/) (Pavlenok et al., 2019).

The lithic industry is characterized by flaking technology of typical wedge-shaped cores aimed at the production of micro-blades; blades were produced through flaking flat and sub-prismatic cores. The archaeological assemblage from Layer 1 is larger and contains a greater variety of various typologically and technologically significant artifacts. All the major inferences concerning stone reduction at Ust-Kyakhta-3 are based on the collection from Layer 1. The materials from Layer 2 do not contradict these inferences. One specific feature of this culture is a special strategy of core reduction. Upon exhaustion of the flaking surface, some flat cores were modified into blanks for wedge-shaped cores together with pebbles and thick spalls. Two modification strategies were reconstructed (Pavlenok, 2015b; Tashak, 2000). Analyses of the tool kit have shown that the targets of primary reduction; blades and micro-blades; were primarily selected for secondary working. The tool kit is dominated by knives, side- and end-scrapers and specific Ust-Kyakhta points (Tashak, 2005); burins fashioned mostly on blades are less numerous. Borerst made on micro-blades (Pavlenok, 2015a) constitute a characteristic series in the tool kit.

In addition to lithic artifacts, bone implements were also recovered, including points, awls, needles, hafts of composite tools with one or two cutting edges, fishhooks (Pavlenok, 2014; Tashak, 2005), and a series of ostrich eggshell beads (Zotkina et al., 2018).

Analytical data from the Ust-Kyakhta-3 archaeological collection provide solid evidence for their attribution to the Selenga Terminal Paleolithic culture (Pavlenok, 2015a). The origin of this culture is associated with older, autochthonous complexes in Trans-Baikalia.

**Individuals from the cis-Baikal region**

These human remains are currently housed in the Research Institute and Museum of Anthropology (RIMA) of the Lomonosov Moscow State University (MSU). Most of these remains were collected by Alexey Okladnikov during his first few independent archaeological excavations as an employee of the Irkutsk museum of regional studies between 1926 and 1929 in the Verkholenski and Irkutski uyezds of the Irkutsk Governorate. Now these territories form several districts of the East Siberian region. The works were regularly reported by the local press, informing general public not only about the archaeological excavations, but also about the establishment of a special department in the Irkutsk museum of regional studies to store the excavation materials, as well as the establishment of the District committee for the protection of the Paleo-Ethnological heritage of the region (Vlast’truda, 1928).

In 1926 Okladnikov documented about ten Late Stone Age dwelling sites in the Verkholenski uyezd, and he also discovered a Neolithic cemetery nearby lishino village (Okladnikov, 1926). During 1927 an archaeological exploration took place in Kachugsky District along the Lena river tributaries, including Biryulka, Anga, Manzurka and some other rivers. As a result, several Neolithic dwelling sites were discovered as well as a cemetery in Khaptsagai area near Bayraki village on the right bank of the Manzurka river (Okladnikov, 1928, 1929). One month prior to this, local press distributed an information about excavations in the vicinities of the Stepno-Baltaiskii ulus, by Okladnikov and Kveiko, resulting in the discovery of human remains, stone scrapers, polished green nephrite knife, bone ornaments and more (Vlast’truda, 1929). The same year Okladnikov conducted archaeological excavations near Zapleskino village in the Zhigalovsky District. The resulting findings were also transferred to the Irkutsk museum of regional studies.

In 1929 most of the anthropological materials from Okladnikov’s excavations were transferred to the Anthropological museum of the Lomonosov Moscow State University.

Through the work of several generations of Russian specialists (N.I. Vitkovsky, M.P. Ovchinnikov, G.F. Debets, A.P. Okladnikov, M.M. Gerasimov, L.P., L.P. Chlobystin and others), characteristic archaeological cultures of the Neolithic and the Early Bronze Age of the Cisbaikalia were defined. Over the past decades, the chronology and periodization of the Neolithic and the Bronze Age of the Cisbaikalia were refined (Bazalijskij, 2012; Mamonova and Sulerzhickij, 1989, 2008; Weber et al., 2002, 2006). As a result, two clusters of the Early Neolithic (EN) sites and a multicomponent cluster of the Late Neolithic and Early Bronze Age (LN-EBA) sites were defined. The first Early Neolithic group includes sites of the Kitoi mortuary tradition, represented by large well-known cemeteries in South Angara region of Southern Baikal, such as the Shambanka II cemetery. The second cluster of the Early Neolithic sites is located at adjacent regions in the Upper Lena River, Olkhon and Western Transbaikal area, different from the Kitoi group mainly in artifact composition and mortuary tradition. This cluster is represented by several small cemeteries and few burials (Movsesian et al., 2015). The EN groups practiced hunting, fishing and sealing, with large and unevenly distributed population, differential mobility and substantial social differentiation. The LN-EBA (Serovo-Isakovo-Glazkovo) cultures are represented by various local variants spread over large territory. Similar with EN population, they had formal cemeteries, practiced hunting, fishing and sealing, but are represented by larger and evenly distributed populations that had moderate mobility and social differentiation (Webber and Bettinger, 2010).

Earlier craniometric studies by Russian anthropologist Debets suggested that the EN and the LN-EBA populations were genetically distinct (Debets, 1930, 1948). Mamonova shown that while differences between Serovo and Glazkovo cranioanatomical samples were minor, Kitoi crania differed substantially from both of these samples, and the extent of the difference between Kitoi and Serovo crania was greater than between Kitoi and Glazkovo crania (Mamonova, 1973). It is likely that Serovo and Glazkovo series represent single cranioanatomical complex with local variants. Overall, this is the same cranioanatomical complex present in Angara (Alekseev and Mamonova, 1979).
Recent craniometric studies as well as studies of non-metric cranial traits suggest that the Early Neolithic Kitoi groups and the Late Neolithic Serovo groups belong to different gene pools. However, at the same time similarities were found between the Kitoi group from Angara region and chronologically distant Bronze Age Glazkovo groups from Angara, Upper Lena and Transbaikal territories. It is hypothesized that the morphological diversity of Glazkovo groups could be due to their mixture with the descendants of the Kitoi groups who left the Cisbaikalia and later migrated back from the neighboring territories. It is assumed that the population of the Transbaikal, which exhibited extremely stable anthropological profile from the Early Neolithic to the Bronze Age time period, penetrated Cisbaikalia during the Glazkovo stage and influenced the formation of the Early Bronze Age population’s gene pool (Movsesian et al., 2014, 2015).

Five skulls (RIMA number 4620 and 4571) were donated to the Anthropological museum of the MSU in 1885 by Siberian archaeologist I.T. Savenkov. In 1884 he conducted archaeological research in the villages: Ladeya, Nyasha, Bazaikha and Sobakino (Krasnoyarsk district).

Angara Site.
- ANG001, tooth M3 lower right (RIMA number 4620)

Savenkov found this cranium while excavating ancient graves on the banks of Angara River. No archaeological context is available.

Bazaikha Site.
- BZK002, tooth M2 lower right (RIMA number 4571)

Savenkov found two skulls near the river Bazaikha on the right banks of Yenisei river in Krasnoyarsk district. No archaeological context is available. According to Okladnikov (1950), the skeletal remains may be attributed to the Bronze Age Karasuk culture between late second to early first millennium BC, based on the archaeological findings described by Savenkov.

Glazkovskoe Predmestie Site.
- GLZ001, tooth M2 upper left, (RIMA number 4616),
- GLZ002, tooth M2 upper right, (RIMA number 4617),
- GLZ003, tooth M1 lower right, (RIMA number 4618)

Archaeological and anthropological findings were from the Glazkovo outskirts of Irkutsk city found during excavations for the foundation of a new infant asylum. Construction works were conducted under the lead of V.P. Sukachov from the summer of 1887. An ancient burial was discovered in the autumn of the same year. N.I. Vitkovski, an employee of the museum of the East Siberian division of the Russian Geographical society, was called in upon discovery. This researcher was known for his excavations on the Kitoi cemetery between 1880-1881 that gave the name to the Early Neolithic culture of the region. By the time Vitkovski arrived at the site, construction workers had destroyed the burials. He collected six crania and a few artifacts, including two rings made from white nephrite. According to several eyewitnesses, the interred individuals were in the sitting position, facing East. Under Vitkovski’s supervision, excavation continued and an undamaged burial was discovered. The skeleton was in flexed position with the legs bent at the knees. Two perforated pendants from cleaved wild boar tusk were found on the cranium. The findings were transported to the museum of the Russian Geographical society’s division (Debets, 1930). Among these, three crania were later transferred to the MSU and became part of its osteological collections.

Iushino Site.
- IU0001, part of left os temporale (pars petrosa), (RIMA number 8311)

In 1926 Okladnikov discovered a Neolithic burial ground near Iushino village (Okladnikov, 1926). Human skeletal remains and artifacts from three burials were transferred to the Irkutsk museum. Based on these, the remains are attributed to the Early Neolithic period of the Kitoi culture (Alekseev and Mamonova, 1979). The cranium from the Grave 3 was later transferred to the MSU.

Kachug Site.
- KAG001, part of left os temporale (pars petrosa), (RIMA number 8307)
- KAG002, tooth M1 lower left, (RIMA number 8308)

The Kachug burial ground was located on the right bank of the Lena river, above the Kachug village. Currently, the exact location of this burial ground cannot be identified. Here in 1927 three burials were excavated, dated to the Eneolithic time (Okladnikov, 1955). Skeletal remains from the grave 2 (# 8307) and grave 3 (# 8308) were transferred to the Anthropological museum. Grave 2 (# 8307) contained polished nephrite ring and a black stone disc (Okladnikov, 1955). From grave 3 (# 8308) Okladnikov described the following artifacts: a small copper or bronze knife “firmly attached to the massive horn handle,” “this knife is flat, with double edged blade, its protruding part has triangular outline,” and two copper tubes (Okladnikov, 1955). Items found in these burials have analogies in well-documented and dated burial complexes of Angara, Upper Lena and Oikhon areas, attributed to the Glazkovo culture of the Late Neolithic – Early Bronze Age (Alekseev and Mamonova, 1979; Zubkov, 2010).

Khaptsagai Site.
- KPT001, part of left os temporale (pars petrosa), (RIMA number 8313),
- KPT002, part of left os temporale (pars petrosa), (RIMA number 8300),
- KPT003, M2 upper right, (RIMA number 8302),


In 1927 an archaeological exploration in Kachug area took place. As a result, several Neolithic dwelling sites were discovered and a burial ground in Khaptsagay area near Bayraki village, on the right bank of Manzurka river (Okladnikov, 1928). Twenty ancient burials were excavated. Skeletal remains from six graves were transferred to the Anthropological museum: from grave 7 (inventory #8300), grave 12 (#8304), grave 13 (#8305), grave 14 (#8302), grave 19 (#8301) and from a looted grave (#8313). Some of the individuals were in a flexed position within the interment, under a stone vault. Numerous artifacts were unearthed such as white nephrite rings, green nephrite knives, stone arrow heads and spear points, stone chisel, awls and stone needles in ornamented needle cases (Okladnikov, 1929). According to the researchers, human skeletal remains are dated to the Glazkovo stage of the Late Neolithic based on the artifacts (Alekseev and Mamonova, 1979).

Zhigalovo Site.

- STB001, part of right os temporale (pars petrosa), (RIMA number 8315)

This cranium has an abbreviation “Zh. 1” written on it, which likely stands for Zhigalovo, grave 1, where the cranium was presumably discovered. Based on the finding date (year 1929) this may be a cranium from a disintegrated grave, discovered by Okladnikov on the Lena river bank on the opposite side of the Zhigalovo village, the district center. The grave was severely destroyed due to displacement and erosion of clay layers of the ancient terrace. Here, stone bow plates attributed to Late Neolithic Serovo culture were discovered. This was the first finding of the bow remnants from this period (Okladnikov, 1950).

Stepno-Baltaiskii ulus Site.

- STB002, part of left os temporale (pars petrosa), (RIMA number 8314)

Okladnikov and Kveiko found Stone Age burials in the vicinity of the Stepno-Baltaiskii ulus. Stone scrapers, a polished green jade knife, bone ornaments and patterned sown elk shoulder blades were found within these burials (Vlast’truda, 1929). The skeletal remains (the cranium and postcranial skeleton) from the grave 1 were transferred to the Anthropological museum of the MSU. According to the researcher, this skeleton may belong to the Early Neolithic period of the Kitoi culture (Alekseev and Mamonova, 1979).

Zapleskino Site.

- ZPL001, part of left os temporale (pars petrosa), (RIMA number 8309),
- ZPL002, M2 lower right, (RIMA number 8310)

These remains are from burials in the vicinity of Zapleskino village in the Zhigalovsky District. Okladnikov transferred two crania to the Anthropological museum, including crania from grave 1 (#8309) and grave 4 (#8310). These findings are dated to the Late Neolithic – Early Bronze Age stage of the Glazkovo culture (Alekseev and Mamonova, 1979).

METHOD DETAILS

Radiocarbon dating and calibration

Except for UKY001, all the other 19 samples were analyzed by accelerator mass spectrometry (AMS) at Manheim (MAMS). For UKY001 we used the three published 14C dates of charcoal and bone materials from the same layer reported as references of the sample age (Orlova, 1995; Pavlenok et al., 2019), and combined them in calibration. All the radiocarbon dates were calibrated using the dataset IntCal 13 (Reimer et al., 2013). Prior to conversion to a calendar age it was important to determine whether there were any dietary offsets that could influence the calibrated ages, as freshwater reservoir effects, including those from the Lake Baikal region, are known to make radiocarbon ages appear as much as > 1,000 years in some systems (Nomokonova et al., 2013; Ramsey et al., 2014; Schulting et al., 2014, 2015). We used stable carbon (δ13C) and nitrogen (δ15N) measurements of bone collagen from the same individuals, as well as published regression equations to estimate the offset between conventional and terrestrial radiocarbon dates (Schulting et al., 2014, 2015). The estimated offsets are given in Table S1.

Stable isotope analysis

Background

δ13C variability in terrestrial ecosystems is primarily driven by plants that variously utilize two dominant photosynthetic pathways, C3 and C4 (Smith and Epstein, 1971). C3 plants, including virtually all trees, shrubs, and temperate grasses, and domesticates such as wheat and barley, have values from -24 to -36‰ (global mean -26.5‰). C4 plants, including domesticated grasses such as millet, have values ranging from c. -9 to -17‰ (global mean -12‰) (Smith and Epstein, 1971). These non-overlapping distinctions are reflected in the tissues of consumers (Ambrose and Norr, 1993). δ15N varies with trophic level, and δ15N trophic shifts of +2-6‰ being documented in terrestrial ecosystems (Deniro and Epstein, 1981; Sealy et al., 1987). The long length of aquatic foodchains, leads to distinctively high δ15N in consumers (Minagawa and Wada, 1984). However, while marine consumers have both high δ13C and δ15N, δ13C is less predictable in freshwater ecosystems necessitating detailed baseline reconstruction (Katzenberg and Weber, 1999; Kiyashko et al., 1998). δ13C and δ15N analysis of human bone collagen primarily determines the isotopic values of the protein input...
to the diet, with a much more minor contribution of lipid and carbohydrate sources (Ambrose and Norr, 1993). This means that the $\delta^{13}C$ and $\delta^{15}N$ values of bone and dentine collagen will be heavily affected by foods that are high in protein (Ambrose and Norr, 1993). In addition, tooth dentine forming during a period of breastfeeding (e.g., permanent M1) will be elevated in $\delta^{15}N$ to appear a whole trophic level higher (Fuller et al., 2003). The same may also apply to certain portions of the skull including the petrus (Jørkov et al., 2009).

**Stable isotope analysis**

We sampled the petrous bones of seven individuals and tooth dentine of 12 individuals (Table S1) for stable carbon and nitrogen isotope analysis in order to determine differences in diet and the potential impact of the reservoir effect on the ages of individuals analyzed.

We obtained c. 0.5 g of dentine or bone powder using a handheld Dremel drill. All specimens were first cleaned using abrasion by sandblaster. We extracted collagen from the resulting powder using a modified Longin methodology (Longin, 1971). Samples were demineralised by immersion in 0.5M HCl for 1-7 days. Once demineralisation was complete, samples were rinsed three times with ultra-pure H$_2$O. The residue was gelatinized in pH3 HCl at 70 °C for 48 hours and Ezee-filters were used to remove the insoluble residues from the soluble collagen solution (Brock et al., 2013). Samples were lyophilized for 48hrs. 1.0 mg of the resulting purified collagen was weighed in duplicates into tin capsules for analysis.

The $\delta^{13}C$ and $\delta^{15}N$ ratios of the weighed-out bone collagen were measured using a Thermo Scientific Flash 2000 Elemental Analyzer coupled to a Thermo Delta V Advantage Isotope Ratio Mass Spectrometer at the Stable Isotope Laboratory of the Department of Archaeology, Max Planck Institute for the Science of Human History, Jena. Values are reported as the ratio of the heavier isotope to the lighter isotope ($^{13}C/^{12}C$ or $^{15}N/^{14}N$) as $\delta$ values in parts per mill (‰) relative to international standards, VPDB for $\delta^{13}C$ and atmospheric N2 (AIR) for $\delta^{15}N$. The results reported for the samples were calibrated against international standards of (IAEA-CH-6: $\delta^{13}C = -10.80 \pm 0.47$ ‰, IAEA-N2: $\delta^{15}N = 20.3 \pm 0.2$ ‰, and USGS40: $\delta^{13}C = -26.38 \pm 0.042$ ‰, $\delta^{15}N = 4.5 \pm 0.1$ ‰). Machine error was determined using repeat runs of a laboratory standard (fish gelatin: $\delta^{13}C = -15.1 \pm 0.1$ ‰, $\delta^{15}N = -14.3$ ‰). Based on replicate analyses machine error over the course of a year is ± 0.2‰ for $\delta^{13}C$ and ± 0.2‰ for $\delta^{15}N$. Overall measurement precision was studied through the measurement of repeats of fish gelatin ($n = 80$, ± 0.2‰ for $\delta^{13}C$ and ± 0.2‰ for $\delta^{15}N$).

The resulting $\delta^{13}C$ and $\delta^{15}N$ values were used to independently estimate the potential impacts on the ‘reservoir effect’ of radiocarbon ages from the skeletons caused by the consumption of freshwater fish in the Lena River and Angara River regions. There has been a long history of research into the ‘reservoir’ effects of prehistoric individuals living in this region based on the paired radiocarbon dating of human bone and associated animal remains from the same graves, and correlations have been found between the extent of the offset in radiocarbon ($^{14}C$) years and human stable isotope values (Ramsey et al., 2014; Schulting et al., 2014, 2015). Here, we used established equations from Schulting et al. (2014, 2015) to estimate the radiocarbon offset based on our data. For individuals from the Lena River we have used Schulting et al.’s (2015, Table 3 pp. 586) “Full data set, $\delta^{13}C$” equation for the Early Neolithic samples and the “EBA sites, $\delta^{13}C$ and $\delta^{15}N$” equation for Early Bronze Age samples. For individuals from the Angara River we have used Schulting et al.’s “SW Baikal/Angara” Equation (2014, Table 4 pp. 998) that is based on their treating analyzed graves separately. Due to the fact that we are analyzing petrous bones and dentine, however, it should be noted that our reservoir estimates are best treated as ‘maximum’ estimates given the potential impact of the ‘weaning effect’ on $\delta^{13}C$ and $\delta^{15}N$, particularly on First Molars (Fuller et al., 2003) and even perhaps petrous bones (Jørkov et al., 2009).

The $\delta^{13}C$ and $\delta^{15}N$ results of the samples analyzed are shown in Table S1. The data is broadly consistent with individuals living in a predominantly C3 environment and consuming significant amounts of freshwater resources (see also Katzenberg et al., 2010). There is a distinction between individuals excavated along the Lena and Angara Rivers, with the latter having higher $\delta^{13}C$ and $\delta^{15}N$ suggesting increasing amounts of freshwater resources in the diets of these individuals (Table S1). Katzenberg et al. (2010) reported Late Neolithic and Bronze Age human $\delta^{13}C$ of between −20 and −19 ‰ for the Lena and Angara Rivers, with $\delta^{15}N$ of c. 10.0 ‰. Our human data ($\delta^{13}C = c. −20 \text{ to } −16$ ‰, $\delta^{15}N$ c. 10 to 16 ‰) is consistent with interpretations of a C3-dominated diet, though with significantly varying, and often greater, contributions of freshwater fish to the diet relative to Katzenberg et al.’s dataset for the same region. While some of this difference is likely a product of the fact that one of the individuals with the highest $\delta^{13}C$ and $\delta^{15}N$ is dated to the Early Neolithic at Angara (where more fish consumption was also noted by Katzenberg et al., 2010), and another individual with highest $\delta^{13}C$ and $\delta^{15}N$ is represented by an M1 that is likely influenced by the breastfeeding effect, it is clear that even on a cross-tissue basis, freshwater resources may have resulted in radiocarbon offsets in a number of these individuals (Table S1). The calculated offsets are shown in Table S1, with individuals from the Lena River having offsets between 161 and 618 years (with many centering around 300 years) and those from the Angara River having offsets between 246 and 485 years (with 3 of the 4 individuals centering around 400 years) (Table S1).

**Stable isotope analysis**

**Background**

The use of strontium isotope ratios ($^{87}Sr/^{86}Sr$) to determine human mobility rests on the basic concept that rocks of different ages and compositions have distinctive values that do not fractionate from the bedrock into the biosphere (including the tissues of any individual measured) (Graustein and Armstrong, 1983; Montgomery, 2010). Older rocks have relatively higher $^{87}Sr/^{86}Sr$ when compared to younger rocks, while different mineral contents (with different $^{87}Sr/^{86}Sr$ ratios) leads to characteristic $^{87}Sr/^{86}Sr$ based on rock ‘type’ (Capo et al., 1998). For example limestones have lower Rb/Sr ratios than continental granites (Montgomery, 2010). While there is a
lack of mass-dependent fractionation from the bedrock into the foodchain, bedrock $^{87}\text{Sr/}^{86}\text{Sr}$ does not always correlate with bioavailable $^{87}\text{Sr/}^{86}\text{Sr}$. Differential weathering of rocks with different $^{87}\text{Sr/}^{86}\text{Sr}$ ratios, as well as the geographic variability of hydrology and aeolian transport, lead to variations in the $^{87}\text{Sr/}^{86}\text{Sr}$ of soils and plants overlying a given distribution of rocks (Montgomery, 2010).

From plants, $^{86}\text{Sr}$ actively replaces $^{40}\text{Ca}$ in consumer tissues as part of the process of nutrient uptake and excretion (Montgomery, 2010; Rokita et al., 1993), with the amount of strontium in consumer tissues believed to be directly reflective of that available from the diet (and environment) (Montgomery, 2010; Price et al., 1986). In the context of tooth enamel from permanent teeth, $^{87}\text{Sr/}^{86}\text{Sr}$ ratios will be incorporated during the process of mineralization which will vary depending on the tooth sampled anywhere between ten weeks before birth and 16 years of age (Hillson, 1996). By sampling teeth that form during different periods of life, and comparing the $^{87}\text{Sr/}^{86}\text{Sr}$ values of these teeth to bioavailable $^{87}\text{Sr/}^{86}\text{Sr}$ values in the burial locale of an individual, it is possible to determine whether someone grew up in a different location to that in which they were buried, thus determining potential migration during the life of an individual (Bentley et al., 2003; Montgomery, 2010). Robustness and detail of the bioavailable baseline created is important as the $^{87}\text{Sr/}^{86}\text{Sr}$ of an individual will be the combined consequence of all food consumed and all water drunk, as well as the extent of the range over which it obtains food and drink (significant in the context of trade in food or presence of long rivers) (Montgomery, 2010), which can complicate interpretation.

**Strontium isotope analysis**

We sampled a total of 11 human teeth (Table S1) from the Angara River Valley (GLZ001, GLZ002, GLZ003, ANG001) and the broader Baikal region (KAG002, KPT002, KPT006, ZPL001) for strontium isotope analysis. The teeth sampled were permanent 1st, 2nd and 3rd molars, which represent different periods of life (0-3 years, 3-7 years and adult respectively as per Alexander Bentley, 2006), to see if there were any distinctions between their $^{87}\text{Sr/}^{86}\text{Sr}$ ratios during childhood and the bioavailable $^{87}\text{Sr/}^{86}\text{Sr}$ of their burial location. We used the comprehensive dataset of Haverkort et al. (2010) that includes modern and archaeological fauna from terrestrial and aquatic settings as an indication of the local baseline.

For each tooth, adhering sediments were removed and teeth were cleaned ultrasonically in ultra-pure water (deionized water). Teeth were then air-dried overnight and a groove was made along the buccal edge of the tooth, in order to average the signal for the period of enamel formation (Alexander Bentley, 2006), using a Dremel drill at the Stable Isotope Laboratory of the Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany.

Drill powdered samples of 20 mg were then shipped to the clean laboratory in the Department of Geological Sciences at the University of Cape Town. Samples were dissolved in 2 mL 65% 2B HNO$_3$ in a closed Teflon beaker and placed for one hour on a hotplate at 140°C. The samples were then dried down and re-dissolved in 1.5 mL 2 M HNO$_3$. Strontium separation chemistry followed methods discussed in Pin et al. (1994). After separation, the solutions for each sample were dried, dissolved in 2 mL 0.2% HNO$_3$ and diluted to 200 ppb Sr concentrations for strontium isotope analysis. Radiogenic $^{87}\text{Sr/}^{86}\text{Sr}$ ratios were measured using a Nu Instruments NuPlasma HR MC-ICP-MS in the Department of Geological Sciences at the University of Cape Town. Sample values were corrected for instrumental mass fractionation using an $^{88}\text{Sr/}^{86}\text{Sr}$ ratio of 0.1194 (Nier, 1938) and isobaric $^{87}\text{Rb}$ interference using the measured $^{86}\text{Rb}$ signal and natural Rb isotope ratios. All data presented are referenced to bracketing analyses of NIST SRM987 ($^{87}\text{Sr/}^{86}\text{Sr}$ reference value of 0.710255). Results for repeat analyses of an in-house carbonate standard (NM95) processed and measured with the batches of unknown samples in this study gave an $^{87}\text{Sr/}^{86}\text{Sr}$ ratio of 0.708885 (2 s.d. = 0.000010; n = 414) in agreement with long-term results for this in-house standard having an average $^{87}\text{Sr/}^{86}\text{Sr}$ ratio of 0.708911 (2 s.d. = 0.000040; n = 414).

**Ancient DNA processing**

All samples were processed in dedicated laboratories at the Max Planck Institute for the Science of Human History in Jena, Germany, except the single-stranded library of sample UKY001.B that was generated at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. Bone powder for DNA extraction was obtained from petrous bones by drilling the densest osseous matter around the coeclea, and from teeth by cutting at the junction between the root and crown, sampling the dental pulp. The sample UKY001 was already fragmented and two fragments were milled using the MM200 mixer mill (Retsch). The resulting bone powder was split in three equal aliquots. For detailed information on the analyzed samples, their archaeological context and radiocarbon age see Table S1 and STAR Methods.

DNA from the 19 ancient individuals was extracted following established protocols (Dabney et al., 2013). The sample UKY001.B was pre-treated with a washing step to reduce surface contamination (Korlević et al., 2015). A negative and cave bear positive controls were included. To release DNA from 30–100 mg of bone powder, a solution of 900 mL EDTA, 75 mL H2O and 25 mL Proteinase K was added. In a rotator, samples were digested for at least 16 h at 37°C, for UKY001.B this was followed by an additional hour at 56°C (Rohland and Hofreiter, 2007). The suspension was then centrifuged and transferred into a binding buffer as previously described (Dabney et al., 2013). To bind DNA, silica columns for high volumes (High Pure Viral Nucleic Acid Large Volume Kit; Roche) were used. After two washing steps using the manufacturer’s wash buffer, DNA was eluted in TET (10 mM Tris, 1 mM EDTA and 0.05% Tween) in two steps for a final volume of 100 μL.

Double-stranded DNA libraries were built from 25 μL of DNA extract in the presence of uracil DNA glycosylase, following a single-stranded ‘UDG-half’ library preparation to reduce, but not eliminate, the amount of deamination-induced damage toward the ends of ancient DNA (aDNA) fragments (Rohland et al., 2015). Negative and positive controls were carried alongside each experiment. Libraries were quantified using the IS7 and IS8 primers (Meyer and Kircher, 2010) in a quantification assay using a DyNAmo SYBP Green qPCR Kit (Thermo Fisher Scientific) on the LightCycler 480 (Roche). Each aDNA library was double indexed (Kircher et al.,...
3

quantification (IS7/IS8) so that each indexing reaction does not exceed 2 of indexing primers (each primer containing a unique 8bp identifier), and was split into multiple PCR reactions based on the initial libraries using a qPCR (Roche). For a double-indexing step (Kircher et al., 2012), each library was assigned a unique combination published protocol (Meyer and Kircher, 2010). Subsequently, the primer combination IS7/IS8 was used to quantify the resulting library and subsequent DNA repair (Briggs et al., 2010). The rest of the library preparation steps were proceeded with according to a used an initial uracil-DNA-glycosylase (UDG) and endonuclease VIII treatment (USER enzyme, New England Biolabs) for uracil excision with 50

reads were mapped to human reference mitochondrial sequence GenBank: NC_012920.1 with BWA 0.7.12 aln/samse algorithm (Li and Durbin, 2009), duplications were removed by DeDup 0.12.1 (https://github.com/apeltzer/DeDup) and damage patterns of each library were checked with mapDamage 2.0.6 (Jonsson et al., 2013). Then we masked 2bp from both ends of the reads from double-stranded libraries with trimBam in bamUtil 1.0.13 (https://github.com/statgen/bamUtil) to remove the damaged sites.

The mitochondrial capture sequencing reads were cleaned by AdapterRemoval 2.2.0 to remove the adapters. Then the cleaned reads were mapped to human reference mitochondrial sequence GenBank: NC_012920.1 with BWA 0.7.12 aln/samse algorithm and realigned with CircularMapper (Peltzer et al., 2016). After removing duplication with DeDup, the consensus sequences were generated by Schmutzi (Reanaud et al., 2015).

Y. pestis genome enrichment

Samples GLZ001 and GLZ002 that were putatively positive for Y. pestis were also converted into Illumina double-stranded libraries with 50 μL of input DNA extract. As post-mortem aDNA cytosine deamination can influence read mapping and SNP identification, we used an initial uracil-DNA-glycosylase (UDG) and endonuclease VIII treatment (USER enzyme, New England Biolabs) for uracil excision and subsequent DNA repair (Briggs et al., 2010). The rest of the library preparation steps were proceeded with according to a published protocol (Meyer and Kircher, 2010). Subsequently, the primer combination IS7/IS8 was used to quantify the resulting libraries using a qPCR (Roche). For a double-indexing step (Kircher et al., 2012), each library was assigned a unique combination of indexing primers (each primer containing a unique 8bp identifier), and was split into multiple PCR reactions based on the initial quantification (IS7/IS8) so that each indexing reaction does not exceed 2 × 10^10 input DNA copies. The indexing PCR reaction was carried out for 10-cycles using the enzyme PfU Turbo Cx Hotstart DNA Polymerase (Agilent). Reaction products were cleaned up using the MinElute DNA purification kit (QIAGEN), eluted in TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20), and quantified using a qPCR (Roche) with the IS5/IS6 primer combination (Meyer and Kircher, 2010). Moreover, indexed libraries were amplified using the enzyme Herculase II Fusion DNA Polymerase (Agilent) to achieve 1–2 μg of total DNA in 7 μL, which would be used as input for whole-genome Y. pestis enrichment. PCR products were purified with the MinElute DNA purification kit (QIAGEN) and eluted in EBT (QIAGEN EB, 0.05% Tween20). In-solution Y. pestis captures of both full-UDG and partially-UDG-treated libraries were performed using a published protocol (Fu et al., 2013), adapted for Y. pestis as described previously (Andrades Valtueña et al., 2017).

QUANTIFICATION AND STATISTICAL ANALYSIS

Genotyping and dataset preparation for analysis

The cleaned reads with base quality and mapping quality over 30 were pooled up with mpileup in SAMtools 1.3 (Li et al., 2009). For double-stranded libraries, we called pseudo-haploid genotypes with pileupCaller 1.2.2 (https://github.com/stschiff/sequenceTools) under random haploid calling mode. For the single-stranded library UKY001_B0102, we called pseudo-haploid genotypes with pileupCaller 1.4.0.3 under single-strand mode using both shotgun and capture sequencing reads, which ignores forward reads at C/T polymorphisms and reverse reads at G/A polymorphisms. Then we combined our genotypes with published ancient individuals (Allentoft et al., 2015; Damgaard et al., 2018b, 2018a; Flegontov et al., 2019; Jeong et al., 2018, 2019; Jones et al., 2015; Lazaridis et al., 2016; Mathieson et al., 2015; Moreno-Mayar et al., 2018; Olalde et al., 2018; Rasmussen et al., 2010, Article 2012) in 1–4 parallel 100 μ L reactions using PfuTurbo DNA Polymerase (Agilent). The indexed products for each library were pooled, purified over MinElute columns (QIAGEN), eluted in 50 μ L TET and again quantified using the 1010 input DNA copies. The indexing PCR reaction was set up with 100 μ L of the purified product were amplified in multiple 100 μ L reactions using Herculase II Fusion DNA Polymerase (Agilent) following the manufacturer’s specifications with 0.3 μ M of each primer IS5 and IS6 primers. After another MinElute purification, the product was quantified using the Agilent 2100 Bioanalyzer DNA 1000 chip. An equimolar pool of all libraries was then prepared for shotgun sequencing on Illumina Hiseq4000 platform using 75bp single-end reads for screening. The single-stranded DNA library for sample UKY001.B was built from 30 μ L of DNA extract in the absence of uracil DNA glycosylase (non-UDG library) followed by double indexing, using an automated version of the protocols described in Gansauge et al. (2017) and Kircher et al. (2012) on a liquid handling system (Agilent Technologies Bravo NGS Workstation).

Human genome enrichment and screening

Both double-stranded UDG-half and single-stranded non-UDG-treated libraries were further amplified with IS5/IS6 primers to reach a concentration of 200–400 ng μ L–1 as measured on a NanoDrop spectrophotometer (Thermo Fisher Scientific). Mitochondrial DNA capture (Fu et al., 2013) was performed on screened double-stranded libraries, after shotgun sequencing, showed the presence of aDNA, highlighted by the typical C-to-T and G-to-A substitution pattern toward 5’ and 3’ molecule ends, respectively (Table S1). Eight of the libraries with percentages of human DNA over 10% were deep shotgun sequenced on an Illumina HiSeq 4000 instrument with 75 pair-end-run cycles using the manufacturer’s protocol. Furthermore, samples with a percentage of human DNA in shotgun data around 0.1% or greater were enriched for a set of 1,237,207 targeted SNPs (1240K capture) across the human genome (Fu et al., 2015). The enriched DNA product was sequenced on an Illumina HiSeq 4000 instrument with 75 single-end-run cycles using the manufacturer’s protocol.

The de-multiplexed capture sequencing reads were cleaned and mapped to human reference genome hs37d5 using EAGER pipeline 1.92.55 (Peltzer et al., 2016). Within the pipeline, the adapters were removed by AdapterRemoval 2.2.0 (Schubert et al., 2016), reads were mapped with BWA 0.7.12 aln/samse algorithm (Li and Durbin, 2009), duplications were removed by DeDup 0.12.1 (https://github.com/apeltzer/DeDup) and damage patterns of each library were checked with mapDamage 2.0.6 (Jonsson et al., 2013). Then we masked 2bp from both ends of the reads from double-stranded libraries with trimBam in bamUtil 1.0.13 (https://github.com/statgen/bamUtil) to remove the damaged sites.

The mitochondrial capture sequencing reads were cleaned by AdapterRemoval 2.2.0 to remove the adapters. Then the cleaned reads were mapped to human reference mitochondrial sequence GenBank: NC_012920.1 with BWA 0.7.12 aln/samse algorithm and realigned with CircularMapper (Peltzer et al., 2016). After removing duplication with DeDup, the consensus sequences were generated by Schmutzi (Reanaud et al., 2015).

Y. pestis genome enrichment

Samples GLZ001 and GLZ002 that were putatively positive for Y. pestis were also converted into Illumina double-stranded libraries with 50 μL of input DNA extract. As post-mortem aDNA cytosine deamination can influence read mapping and SNP identification, we used an initial uracil-DNA-glycosylase (UDG) and endonuclease VIII treatment (USER enzyme, New England Biolabs) for uracil excision and subsequent DNA repair (Briggs et al., 2010). The rest of the library preparation steps were proceeded with according to a published protocol (Meyer and Kircher, 2010). Subsequently, the primer combination IS7/IS8 was used to quantify the resulting libraries using a qPCR (Roche). For a double-indexing step (Kircher et al., 2012), each library was assigned a unique combination of indexing primers (each primer containing a unique 8bp identifier), and was split into multiple PCR reactions based on the initial quantification (IS7/IS8) so that each indexing reaction does not exceed 2 × 10^10 input DNA copies. The indexing PCR reaction was carried out for 10-cycles using the enzyme PfTurbo Cx Hotstart DNA Polymerase (Agilent). Reaction products were cleaned up using the MinElute DNA purification kit (QIAGEN), eluted in TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20), and quantified using a qPCR (Roche) with the IS5/IS6 primer combination (Meyer and Kircher, 2010). Moreover, indexed libraries were amplified using the enzyme Herculase II Fusion DNA Polymerase (Agilent) to achieve 1–2 μg of total DNA in 7 μL, which would be used as input for whole-genome Y. pestis enrichment. PCR products were purified with the MinElute DNA purification kit (QIAGEN) and eluted in EBT (QIAGEN EB, 0.05% Tween20). In-solution Y. pestis captures of both full-UDG and partially-UDG-treated libraries were performed using a published protocol (Fu et al., 2013), adapted for Y. pestis as described previously (Andrades Valtueña et al., 2017).
Individuals were detected using Refined IBD algorithm in BEAGLE v4.1 (Browning and Browning, 2013), under the default parameters. The shared IBD segments between individuals are divided into four groups, namely UKY, Kolyma, Baikal EN and Baikal LNBA (Table S1). The ROH was detected using qpDstat calculated by distinguish the Yamnaya-related and Sintashta-related ancestries. Before running American Karitiana (n = 16), to distinguish ancestries across Eurasian continent. For the modeling of KPT005 with possible West Natufian (n = 6), Neolithic Iranian farmer Iran_N (n = 5), modern South Asian Onge (n = 11), Southeast Asian Ami (n = 10) and Native American modern populations (Figure 1C; Table S1). We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.

We identified the ROH and shared IBD segments in the 24 shotgun sequenced individuals and 81 individuals from several modern Siberian populations (Buryat, Even, Evenk_FarEast, Evenk_Transbaikal, Ngasan, see Table S1). The unsupervised population clustering pattern was estimated by ADMIXTURE 1.3.0 (Alexander et al., 2009). We included 3488 individuals in the ADMIXTURE analysis, including 2986 worldwide modern individuals, 483 selected ancient individuals and 19 individuals newly genotyped in this study. Before ADMIXTURE analysis, the dataset was filtered to 120,599 SNPs with PLINK 1.90 (Purcell et al., 2007) to remove SNPs with minor allele frequency lower than 0.01 (--maf 0.01) and linked SNPs with r^2 > 0.2 (--indep-pairwise 200 25 0.2). The ADMIXTURE analysis was carried out with population number (K) from 2 to 20, five duplications for each K value and we chose K = 16 as it gave the lowest cross-validation (CV) error.
Demographic modeling with qpGraph

We used qpGraph 6450 to reconstruct a demographic model of studied populations. For the resolution of graphic modeling, we applied “useallsnps” mode to correct for low-coverage sample and carried out this part of analysis using 1240K SNP dataset. Starting from (Mbuti, (AG3, Onge)), we first iteratively added the Devil’s Gate, USR1, Ancient Southwestern Ontario (ASO, n = 6) and Early San Nicolas (ESN, n = 17) to build a skeleton graph based on previous publications (Posth et al., 2018), with USR1, ASO and ESN forming a single group which is the mixture of AG3 and Devil’s Gate, and USR1 outgroup to ASO and ESN.

Then we added UKY and Kolyma independently to this graph, considering them as either sister lineage of existing node, or two-way admixture of the existing lineages. For each population, we tested 91 different models and chose the best fitted model based on the worst Z scores and the final likelihood scores reported by each model. After obtaining the best fitted model with UKY or Kolyma (Figure S2), we further tested 340 different models by adding Kolyma onto the best UKY model, or vice versa, to find the best fitted graph including both of these individuals.

Admixtue dating with DATES

We applied the ancestry covariance pattern-based DATES 753 program (Moorjani and Patterson, 2018) to estimate the time of admixture events in Lake Baikal and Okunevo populations. For the Baikal populations, we used a pooled population ANE_pool (n = 15, including MA1, AG2, AG3, EHG, Botai, West_Siberia_N) to represent the ANE-related ancestry, and population EA_pool (n = 39, including Devil’s Gate, Ulchi and Nanai) to represent the NEA ancestry. For the Okunevo population, the three ancestries were represented by Baikal_LNBA_all (n = 34), Yamnaya_pool (n = 43, including Yamnaya_Samara, Yamnaya_Karagash, Yamnaya_Kalmkia and Afanasievo) and Botai_pool (n = 8, including Botai and West_Siberia_N), respectively. For the BZK002 individual, we used the same Baikal_LNBA_all and Botai_pool to represent the two ancestries. The bin size for covariance calculation was 0.1cM and exponential fitting started at d ≥ 0.5cM.

Pathogen DNA screening with HOPS

The Megan Alignment Tool (MALT version 0.4.0) (Vågene et al., 2018) was used within the pathogen screening pipeline HOPS (Hübener et al., 2019), for the assessment of pathogen DNA in all Lake Baikal specimens analyzed in this study. Pre-processed shotgun NGS reads were used as input for MALT, against a custom database of all NCBI RefSeq bacterial and viral assemblies marked as complete, as well as a selection of eukaryotic pathogen genomes and the human GRCh38 reference (RefSeq, November 2017). Genomes with description keywords such as “unknown” were removed from the database, retaining a total of 15,361 entries. MALT was run with an 85 minimum percentage identity filter (–minPercentIdentity), 1 for the minimum support filter (–minSupport), a top percentage parameter of 1 (–topPercent) and a semi-global alignment mode, with all other parameters remaining as default. The MALT output was filtered with MALTEXtract (within HOPS) based on a predefined list of bacterial, viral and eukaryotic pathogen candidates, where assigned reads were evaluated based on different parameters such as their coverage distributions, their edit distance and their aDNA damage profiles.

Read processing, SNP calling and phylogenetic analysis of Yersinia pestis

All captured libraries were sequenced on an Illumina HiSeq4000 (2x76+8+8 cycles). Processing of de-multiplexed NGS reads was performed within the automated pipeline EAGER v1.92.55 (Peltzer et al., 2016). In brief, AdapterRemoval v2 (Schubert et al., 2016) was used for removing Illumina adapters and for read filtering based on sequencing quality (minimum base quality of 20) and length (≥30 bp). Moreover, for partially-UDG-treated libraries, both ends of NGS reads were trimmed to remove remaining deaminated sites, and the resulting files were concatenated with pre-processed read data from full-UDG-treated libraries. Furthermore, BWA (Li and Durbin, 2009) was used for read mapping against the Y. pestis chromosomal reference genome (GenBank: NC_003143.1), using a –n parameter of 0.1 and -l 32. The identical parameters were also used for mapping against the plasmid reference sequences (GenBank: NC_003131.1, GenBank: NC_003134.1 and GenBank: NC_003132.1), where nucleotide positions 3,000 – 4,200 were masked in pPCP1 as they were previously shown to display homology with an expression vector present in laboratory reagents (Schuenemann et al., 2011). Subsequently, SAMtools was used to remove reads with mapping quality lower than 37 and MarkDuplicates was used to remove PCR duplicates (http://broadinstitute.github.io/picard/).

SNPs were called on the chromosomal Y. pestis mapped reads using the UnifiedGenotyper in GATK (DePristo et al., 2011). GLZ001 and GLZ002 were analyzed alongside a dataset of modern Y. pestis genomes (n = 233) (Cui et al., 2013; Eroshenko et al., 2017; Kisilchikina et al., 2015, 2018a, 2018b; Kutreyev et al., 2018; Morelli et al., 2010; Zhgenti et al., 2015) (as listed in (Keller et al., 2019)). In addition, we included previously published historical Y. pestis genomes with ≥ 10-fold coverage (Bos et al., 2011, 2016; Feldman et al., 2016; Keller et al., 2019; Namouchi et al., 2018; Syprou et al., 2016, 2019), and all previously published Late Neolithic and Bronze Age genomes with ≥ 3-fold coverage (Andrades Valtueña et al., 2017; Rasmussen et al., 2015b; Syprou et al., 2018). Geographic isolation locations of genomes within the comparative dataset are as follows: China (CHN), United States of America (USA), Madagascar (MDG), India (IND), Myanmar (MMN), Congo (COG), Uganda (UGA), Mongolia (MNG), Nepal (NPL), Iran (IRN), Kazakhstan (KAZ), Kyrgyzstan (KGR), Armenia (ARM), Georgia (GEO), Azerbaijan (AZE), Uzbekistan (UZB), Turkmenistan (TKM), Tajikistan (TJK), Russia (RUS), unspecified regions of the Former Soviet Union (FSU), Switzerland (CHE), Germany (DEU), Spain (ESP), France (FRA), United Kingdom (GBR), Netherlands (NLD), Norway (NOR), Lithuania (LTU), Croatia (CRO), Estonia (EST).
SNPs were called using the ‘EMIT_ALL_SITES’ option in GATK, which generated a call for every position in the chromosomal reference genome (CO92). Subsequently, a custom java tool, MultiVCFAnalyzer v0.85 (https://github.com/alexherbig/MultiVCFAnalyzer), was used to produce a SNP table of variant positions across the entire dataset, excluding previously defined non-core regions, as well as mRNAs, tRNAs and tmRNAs (Cui et al., 2013; Morelli et al., 2010). Variants were retained based on the following conditions: homozygous SNP positions were called with a minimum coverage of 3-fold and a minimum genotyping quality of 30. For heterozygous positions, SNPs were retained when at least 90% of reads supported the call. When none of the conditions were met, an “N” would be inserted in the respective genomic position. A total of 7,052 SNPs were identified in the present dataset. The SNP alignment produced by MultiVCFAnalyzer v0.85 was used as input for a maximum likelihood (ML) phylogeny with 99% partial deletion filter, which retained 4,368 SNP positions. ML trees were built using the program RAxML (version 8.2.9) (Stamatakis, 2014), under the Generalized Time Reversible (GTR) substitution model (Tavaré, 1986) (four gamma rate categories). 1,000 bootstrap replicates were used to determine node support.

**BEAST v1.8 tip dating of GLZ001 and GLZ002**

The program TempEst v1.5 (http://tree.bio.ed.ac.uk/software/tempest/) was used to assess of the temporal signal across previously published genomes and calibration points on the Y. pestis LNBA lineage (Andrades Valtuerna et al., 2017; Rasmussen et al., 2015b), including the modern isolate 0.PE.2 Pestoides F (Garcia et al., 2007) as outgroup. All overlapping variant positions across this set of genomes were used for the analysis, retaining a total of 374 SNPs in the dataset. The resulting correlation coefficient (r) was 0.88 and R² was 0.78 when including the 0.PE2 outgroup and 0.98 and 0.97, respectively, for the LNBA lineage alone. Such result suggests a strong temporal signal in the dataset.

Subsequently, we used BEAST v1.8 (Drummond and Rambaut, 2007) to estimate the tip dates of GLZ001 and GLZ002, using published radiocarbon date ranges across the LNBA lineage as uniform priors (in years before the present) as follows: RISE509 (uniform prior: 4836–4625, tip date: 4729), RK1.001 (uniform prior: 4828–4622, tip date: 4720), GEN72 (uniform prior: 4833–4592, tip date: 4721), Gyvakarai1 (uniform prior: 4571–4422, tip date: 4485), Kunila2 (uniform prior: 4524–4290, tip date: 4427), 1343Untal (uniform prior: 4346–4098, tip date: 4203), 6Post (uniform prior: 3957–3832, tip date: 3873), RT5 (uniform prior: 3868–3704, tip date: 3789), RISE505 (uniform prior: 3694–3575, tip date: 3635). The tip date of the 0.PE2 outgroup was set to 0 (present). The input date range for GLZ001 and GLZ002 was constrained to be between 5000 and 3500 years before the present, which spans the entire temporal range of the LNBA lineage, and the starting value tip date was set to 4250 (midpoint). The graphical user interface BEAUti v1.8 (Drummond and Rambaut, 2007) was used to set up multiple runs using the coalescent constant size and coalescent skyline tree priors in combination with both a strict and a lognormal relaxed clock model. In addition, a GTR substitution model with four gamma rate categories was used for all runs. All set-ups were run in BEAST v1.8 (Drummond and Rambaut, 2007) using two independent chains of 25,000,000 states each. After run completion, chains were combined using LogCombiner (Drummond and Rambaut, 2007) with 10% burn-in and were then inspected in Tracer v1.6 to ensure run convergence (http://tree.bio.ed.ac.uk/software/tracer/) with all ESS values being > 200. The resulting mean tip dates and HPD95% intervals produced for GLZ001 and GLZ002 across all runs were viewed and analyzed in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

**Assessment of Y. pestis gene profiles**

In order to assess the gene profiles of GLZ001 and GLZ002, we computed the coverage across of 150 previously defined virulence-associated and evolutionary-determinant Y. pestis genes (Demeure et al., 2019; Zhou and Yang, 2008). For this, the two newly reconstructed genomes were comparatively assessed alongside previously published representatives of the LNBA lineage (RISE509, RK1001, GEN72, Gyvakarai1, Kunila2, 1343Untal, 6Postilionstrasse, RISE505), a Bronze Age isolate showing signatures of flea adaptation (RT5), historical Y. pestis genomes from the first and second plague pandemics (Altenerding 2148, London ES 8124/B2911/11972), representatives of modern isolates (0.PE2 Pestoides F, 0.PE4 Microtus 91001, 1.ORI CO92), as well as a Y. pseudotuberculosis strain (IP31953). Raw pre-processed reads from all aforementioned datasets were mapped against the CO92 reference chromosome and plasmids using BWA (-n 0.1, -l 32), without the subsequent use of a mapping quality filter. The coverage across the defined gene regions was calculated using BEDTools (Quinlan and Hall, 2010). The computed coverages were plotted in the form of a heatmap using the ggplot2 package within R version 3.2.1 (R Development Core Team, 2008; Wickham, 2016). In addition, previously defined substitutions or InDels within functionally relevant genes were manually inspected in GLZ001 and GLZ002 using the program IGV (Thorvaldsdóttir et al., 2013). These were: a non-synonymous substitution defining the pla gene variant involved in within-host bacterial dissemination (Zimblert et al., 2015), pseudogenization mutations in genes PDE2, PDE3, ureD and a 30bp duplication in rcsA associated with Y. pestis flea-colonization (Sun et al., 2014), as well as a pseudogenization mutation in fliD associated with immune evasion (Minnich and Rohde, 2007).
Figure S1. Population Size and Relatedness of Lake Baikal Populations Revealed by ROH and IBD Segments, Related to Figure 1 and Table S1

This figure summarizes the accumulative ROH length detected in each individual (row 1), shared IBD segment length of individuals within population (row 2), and shared IBD segment length of UKY, Baikal_EN and Baikal_LNBA individuals with other population (row 3-5), respectively. The long segments (> 8Mb) and short segments (< 2Mb) are also summarized separately.
Figure S2. Relationship between UKY, Kolyma, and Modern-Day Populations Based on $f_4$ Statistics and qpGraph Modeling, Related to Figure 2
(A) This figure shows the different genetic affinities between UKY, Kolyma with worldwide population, assessed by $f_{4}(Mbuti, X; Kolyma, UKY)$. The test populations with significant $f_4$ values ($|Z| > 3$) are shown in diamonds and other populations in circles.
(B) This figure shows the graphic modeling of UKY (left) and Kolyma (right) on the skeleton graph including Mbuti, AG3, Onge, Devil’s Gate, USR1, ASO and ESN described in STAR Methods. The best fitted model for each individual is selected based on the maximum $f$-statistics $Z$ scores and final scores reported for each model.
The outgroup $f_3$-statistics in the form of (A) $f_3$(Mbuti, X; Baikal_EN_all) and (B) $f_3$(Mbuti, X; Baikal_LNBA_all) are applied to measure the genetic affinity of Early Neolithic and LNBA Baikal individuals with worldwide population. The ten population with highest $f_3$ are shown in diamonds. Then (C) $f_4$(Mbuti, X; Devil’s Gate, Baikal_EN_all) and (D) $f_4$(Mbuti, X; Baikal_EN_all, Baikal_LNBA_all) are used to show the genetic difference between NEA ancestry, Early Neolithic Baikal population and LNBA Baikal population.
Figure S4. Dating of the Admixture Events in Baikal, Okunevo Population, and the BZK002 Individual, Related to Figures 3 and 5 and Table S5
This figure shows the DATES estimation of (A) time of admixture events in Early Neolithic and LNBA Baikal population and (B) time of admixture events in Okunevo population and BZK002 with different ancestor pairs. The red cross dots show the weighted ancestry covariance in different genetic distances, and the green curves show the exponential fitting starting at 0.5 cM. Details of the results are listed in Table S5.
Figure S5. Overview of Y. pestis Screening, Gene Content, and Temporal Signal, Related to Figures 4 and S6 and Tables S1 and S6

(A) Y. pestis screening results for samples GLZ001 and GLZ002. The histograms were calculated in HOPS and represent the edit distance of GLZ001 and GLZ002 MALT assigned reads to their closest matching reference within the Y. pseudotuberculosis complex. In addition, aDNA damage patterns were generated using MapDamage2.0, after BWA mapping of all reads against the Y. pestis reference genome CO92.

(B) Presence/absence analysis of virulence-associated and evolutionary-determinant genes across Y. pestis genomes. The genomic profiles of GLZ001 and GLZ002 were investigated by calculating the coverage across virulence-associated and evolutionary-determinant genes, in comparison to those from previously published ancient strains from the LNBA period (RISE509, RK1001, GEN72, Gyvakarai1, Kunila2, 1343UnTal85, 6Postillionstrasse, RISE505), the Late Bronze Age (RT5), as well as from the medieval and early modern periods (Altenerding 2148 and London BD 8124/8291/11972). In addition, representatives of modern lineages were also included for comparative purposes. The investigated genes were located on the Y. pestis chromosome and the pMT1, pPCP1 and pCD1 plasmids. The heatmap was constructed in R version 3.2.1 using the ggplot2 package.

(C) Calculating the temporal signal within the Y. pestis LNBA lineage. Regressions of root-to-tip genetic distance against specimen age were calculated using TempEst v1.5 for all published LNBA genomes using a maximum likelihood phylogeny of 374 overlapping variant positions for the analysis. The input ages were as follows: RISE509 (4729 BP), RK1.001 (4720 BP), GEN72 (4721 BP), Gyvakarai1 (4485 BP), Kunila2 (4427 BP), 1343UnTal (4203 BP), 6Post (3873 BP) and RISE505 (3635 BP). The left plot represents a linear regression of root-to-tip genetic distance against specimen age including a modern Y. pestis genome as outgroup (0.PE2 Pestoides F), whereas the right panel represents the regression including only previously published ancient genomes within the LNBA lineage (without the outgroup).
Figure S6. Visual Inspection of Functionally Informative \( Y. \) pestis SNPs in Virulence-Associated and Evolutionary-Determinant Genes, Related to Figures 4 and S5

Screenshots created in IGV (Thorvaldsdóttir et al., 2013) for a virtualisation of virulence-associated genes affected by substitutions or InDels. Visualized genomic positions are as follows: (A) presence of a “T” at CO92 pPCP1 plasmid position 7500 suggests ancestral \( \text{pil} \) variant; (B) lack of “T” insertion at CO92 chromosomal position 1892659 suggest active \( \text{flhD} \) gene variant; (C) Presence of “G” at chromosomal \( Y. \) pseudotuberculosis IP32953 position 3944534 suggests ancestral (active) PDE-3 variant; (D) presence of “T” at chromosomal \( Y. \) pseudotuberculosis IP32953 position 3944166 suggests derived (pseudogenised) PDE-3-pe’ variant; (E) lack of “T” insertion at CO92 chromosomal position 1434044 suggests active PDE-2 variant; (F) lack of “G” insertion at CO92 chromosomal position 2997296 suggests active \( \text{ureD} \) variant; (G) imprecise mappability of reads spanning a 30bp duplication within \( \text{rcsA} \) gene in GLZ001 and GLZ002, mapped against the \( \text{rcsA} \) [y1741] gene in KIM10), compared to the published ancient strain RTS showing genetic evidence of flea adaptation (Spyrou et al., 2018); (H) exclusion of one 30bp duplicate from a constructed \( \text{rcsA} \) reference restores read mappability within this region in GLZ001 and GLZ002, suggesting an absence of the duplication in those genomes.