Master's Thesis of Biological Sciences

The genetic history of ancient Siberian populations

고대 시베리아 집단의 유전학적 역사

February 2023

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Haechan Gill

The genetic history of ancient Siberian populations

Advised by Professor Choongwon Jeong

Submitting a master's thesis of
Biological Sciences

February 2023

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Haechan Gill

Confirming the master's thesis written by
Haechan Gill
February 2023

Chair	Lim, Young Woon	(Seal)
Vice Chair	Jeong, Choongwon	(Seal)
Examiner	Shin, Seunggwan	(Seal)

Abstract

Siberia is a vast geographic region spanning North Asia from Ural Mountain to the Pacific Ocean that is covered with tundra and taiga forests. Although Siberia has been sparsely populated, the history of the Siberian populations is complicated and interesting. This complication is because multiple migrations and admixtures have occurred in Siberia, and ancient genomes have contributed to revealing these events. Indeed, Eastern and Western Eurasians mixed frequently in Siberia, and also the ancestors of Native Americans were formed and migrated from Siberia to America. However, since ancient Siberian genomes have been analyzed focusing on the relationship with Native Americans for recent years, a comprehensive understanding of the relationship between ancient Siberian populations is lacking. In order to solve this problem, published genomes were analyzed comprehensively to reconstruct the detailed genetic history of ancient Siberian populations and their impact on other populations. As a result, before the Middle Holocene began, Siberian populations were already divided into population located in Lake Baikal and population located in Yakutia region. Genetic exchanges between the two populations occurred during the Neolithic period, consistent with changes in genetic profiles and cultures in each region. Then, the impact of the Yakutia population was explored. They were found to influence the genetic makeup of present-day Uralic-speaking people, Nganasan, Selkup people, and north-eastern Europeans. Furthermore, the Yakutia population also affected both Yenisei-speaking Kets and Na-

Dene-speaking Athabaskans in North America, suggesting that the Yenisei and Na-Dene languages shared common Siberian ancestors. This result will be used as an essential example for studying a correlation between linguistic similarity and genetic similarity and used for reconstructing the formation of North Eurasians and their migration to America in detail.

Keyword: Siberia, Archaeogenetics, Middle Holocene, Human population genetics, Uralic-language, Dene-Yenisei language, Yakutia
Student Number : 2021-23420

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. General introduction to Archaeogenetics

As DNA sequencing and Polymerase Chain Reaction (PCR) technologies advance, ancient DNA can be used as molecular data in population genetics (Pääbo et al., 1989). Existing population genetic and evolutionary theories could be proved and applied to practical research using ancient genome data (Patterson et al., 2012). Archaeogenetics is a study that reconstructs demographic histories such as migration, admixture, and domestication by analyzing increasing ancient genome data (Kim et al., 2018). In particular, studies of human evolution have received much attention and significantly impacted the understanding of human history and finding the causes of genetic diseases (Bouwman & Rühli, 2016). Archaeogenetics will be studied more actively and developed further in the future.

1.2. Archaeogenetic studies of early Siberian populations

Siberia is a vast geographical region spanning north Asia that is covered with tundra and taiga forests (Cocks & Torsvik, 2007). Since Siberia has unsuitable vegetation for crop cultivation, thus hunting, gathering, and nomadic pastoralism were common in this region (Krupnik, 2000; Weber & Bettinger, 2010). However,

although Siberia has been populated sparsely, the history of the Siberian population is quite complicated and interesting. This complication is because multiple migrations and admixtures have occurred in Siberia, and ancient genomes have revealed demographic events over the past several years (Raghavan et al., 2014; Sikora et al., 2019).

According to the oldest Siberian genome from Ust'-Ishim in western Siberia (Ust'-Ishim from here on), Siberia was occupied by anatomically modern humans at least 45 thousand years ago (kya) (Fu et al., 2014). Since Ust'-Ishim is genetically symmetric with Western Eurasian and Eastern Eurasian, a common Eurasian ancestor occupied Siberia first. After 10,000 years, two Upper Paleolithic 31,000-years-old individuals genetically close to Western Eurasian were found in the Yana river (Yana_RHS from here on) in northern Siberia (Sikora et al., 2019). Interestingly, while the major ancestry of Yana_RHS is Western Eurasian, they also have minor Eastern Eurasian ancestry represented by a 34,000-year-old individual from Salkhit Valley in northeastern Mongolia, suggesting a pretty ancient admixture between Western and Eastern Eurasian (Massilani et al., 2020). Therefore, the Upper Paleolithic Siberian ancestors had a distinctive genetic composition, including Western Eurasian and Eastern Eurasian ancestry (Figure 1A).

As the Last Glacial Maximum began 33,000 years before present (BP), northern Siberians migrated south to avoid cold climates (Sikora et al., 2019). Consequently, descendants of Yana_RHS, a 24,000-year-old individual from Mal'ta (MA1) and two 18,000-year-old individuals from Afontova Gora (AG2, AG3)

appeared in south-central Siberia during LGM (Fu et al., 2016; Raghavan et al., 2014). These individuals had similar genetic profiles that were named "Ancient North Eurasian" (ANE) and found to affect western Siberians, Europeans as well as Native Americans (Fu et al., 2016). More specifically, the ANE population mixed with Native American ancestors who diverged from east Asian, and after that, Native American ancestors migrated from Siberia to America by crossing the Bering Land Bridge (Fu et al., 2014) (Figure 1B). Interestingly, Native American ancestry remained in Siberia. A 14,000-year-old individual from the Ust-Kyakhta-3 site in the south of Lake Baikal (UKY) and a 9,800-year-old individual from the Kolyma river in northern Siberia (Kolyma_M) were suggested to represent close relatives of Native American and referred as "Ancient Paleo Siberian" (APS) (Sikora et al., 2019; Yu et al., 2020) (Figure 1B).

1.3. Development of diverse Siberian cultures during the Middle Holocene

At the beginning of the Holocene (10,000 years BP), two distinctive populations already existed in Siberia: ANE and APS. However, "Ancient North Asian" (ANA), the East Asian population represented by hunter-gatherers from Devil's Cave in Russia Far East (DevilsCave_N) (Siska et al., 2017) and Amur river basin (AR_EN) (Ning et al., 2020) migrated north to Siberia during the Middle Holocene (7,000 years BP to 5,000 BP) (Kılınç et al., 2021). As a result, genomic aspects of the Middle Holocene Siberia

became complicated, and diverse cultures appeared during this period (Figure 1C).

Lake Baikal and Yakutia, the largest republic of Russia in the Far East, are the representative archaeological sites where artifacts and ancient genomes from the Middle Holocene have been discovered (de Barros Damgaard et al., 2018; K ı l ı nç et al., 2021; Sikora et al., 2019; Yu et al., 2020). According to the previous study, ANA ancestry was first transferred to the east Baikal region, which is close to the Amur river (K ı l ı nç et al., 2021). Then, ANA ancestry spread to the west Baikal and Yakutia regions during the Neolithic period. As a result, the preexisting APS and ANE ancestry were largely replaced by ANA ancestry in both Baikal and Yakutia regions.

Before the introduction of ANA ancestry in Yakutia, the Middle Neolithic individual in Yakutia were closer to APS and belonged to Belkachi culture, descendants of Syalakh culture (Slobodin, 2019). Unlike the Middle Neolithic culture, following Late Neolithic Yakutia culture was affected by the migration of ANA ancestry during the Neolithic period (K ı l ı nç et al., 2021). Late Neolithic Yakutia culture, Ymyiakhtakh spread as far as Chukotka Peninsula and contributed largely to present-day Far East populations (Flegontov, Alt ı n ı ş ı k, et al., 2016; Stepanov et al., 2012).

Similarly, Early Neolithic west Baikal individuals, Kitoi people, were closer to ANA like Ymyiakhtakh, which implies the introduction of ANA ancestry in the west Baikal region (de Barros Damgaard et al., 2018). While no significant genetic changes occurred in the Yakutia region after the migration of ANA ancestry, ANE ancestry resurged in the west Baikal region during the Early

Bronze Age (de Barros Damgaard et al., 2018; K ı l ı nç et al., 2021; Yu et al., 2020). During this period, cultural change also occurred from Kitoi culture to Serevo-Glazkovo culture in the west Baikal region (Weber et al., 2002). The reintroduction of distal ANE ancestry has explained this resurgence of ANE ancestry in previous studies (de Barros Damgaard et al., 2018; K ı l ı nç et al., 2021; Yu et al., 2020).

However, it is still being determined where distal ANE ancestors came from and whether the reintroduction of ANE ancestry occurred. Moreover, although the genetic makeup of Yakutia and Lake Baikal was adequately explained independently, spatiotemporal analysis of whole Middle Holocene populations still needs to be improved. For instance, archaeological evidence supported the migration of Kitoi people from the west Baikal region to the Yakutia region (Kuzmin & Bellwood, 2015). However, the genetic relationship between Kitoi and Belkachi culture has never been studied. Thus, a comprehensive understanding of the relationships between Middle Holocene Siberian populations is necessary.

1.4. Shared Siberian ancestry among Uralic language family

Present-day Siberians speak diverse languages, including Uralic, Yukaghir, Turkic, Tungusic, and Yenisei (Abondolo, 2015; Khabtagaeva, 2019). Among them, Uralic-language is a widespread language in North Eurasia that is spoken by present-day north

Siberian such as Nganasan, Nenets, and Selkup, as well as by north-eastern Europeans such as Finns, Saami, and Hungarian (Lamnidis et al., 2018). Interestingly, north-eastern Europeans and northern Siberians share linguistic similarities despite the long physical distance and geographical barrier of the Ural Mountains. Moreover, not only linguistic similarity between them but the genetic makeup of north-eastern Europeans was also shaped by the migration of Siberian ancestors (Jeong et al., 2019). For instance, 3,500-year-old early Metal Age individuals from Bolshoy Oleni Ostrov in the Kola Peninsula were found to possess Siberian ancestry and transferred Siberian ancestry to present-day Uralic-speaking north-eastern European (Lamnidis et al., 2018).

However, it has yet to be revealed how the Siberian ancestors influenced north-eastern Europeans. In the previous study, Siberian ancestry was represented by Nganasan, present-day Uralic-speaking people (Lamnidis et al., 2018). In other words, the ancestors of Nganasan and when the Siberian ancestors of north-eastern Europeans migrated are still uncertain. Moreover, since Nganasan is an eastern representative of other present-day north Eurasian such as Selkup, finding Siberian ancestors of Nganasan is also necessary to understand the genetic makeup of north Eurasians (Jeong et al., 2019).

1.5. Dene-Yenisei language hypothesis and its relevance with Paleo Eskimo

As mentioned above, the Yenisei language is one of the

language families in Siberia. However, the Yenisei language is only spoken by Ket people, and other Yenisei-speaking populations are extinct (Georg, 2008; Vajda, 2013). In addition, since the Yenisei language is distinctive from other languages spoken in Siberia, the origin of the Yenisei language has been controversial (Vajda, 2019; Vovin, 2000). One of the hypotheses for the origin of the Yenisei language is about the shared origins of the Yenisei language and Na-Dene language, which Athabaskans speak in north America (Vajda, 2019). Not only linguistic evidence but genetic evidence also has been suggested. For instance, the major Y haplogroup of Ket is the Q1 subclade (90%) which is common in Native Americans but rare in Siberian except for Selkup people (60%) (Pinotti et al., 2019). In addition, autosomal analysis unraveled the genetic connection between Ket people and Paleo-Eskimo, represented by the Saqqaq culture who influenced Athabaskan (Flegontov et al., 2019; Flegontov, Changmai, et al., 2016). In summary, the Ket people and Athabaskans share common ancestors through Paleo-Eskimos. However, whether Ket people were influenced by back migration of Paleo-Eskimo ancestry is still uncertain. The existence of common Siberian ancestors of Ket and Paleo-Eskimo can also explain this similarity.

A recent study reported genetic similarity between Paleo-Eskimo and a Middle Neolithic individual from Yakutia (K ı l ı nç et al., 2021). However, whether Middle Neolithic Yakutia ancestry can replace Paleo-Eskimo ancestry in Athabaskan or Ket people has never been tested. Therefore, studying Middle Neolithic Yakutia ancestry and its spread is necessary for unraveling the origin of the Yenisei and Na-Dene language.

1.6. Purpose of the study

This study will reveal the demographic history of Middle Holocene Siberian populations. Since genetic interactions between Yakutia and Lake Baikal cultures have been understudied, the first purpose will be to dissect two populations and confirm if there was a genetic exchange between the two regions. Then, the influence of Yakutia ancestry will be explored on worldwide populations to explain shared Siberian ancestry among present-day Uralic-speaking populations. Finally, the Yakutia genome will be compared with the Paleo-Eskimo genome and tested whether Yakutia ancestry was transferred to Na-Dene-speaking Athabaskans. This result will shed light on the origins of the Yenisei language. The proposed three primary purposes of this study are presented in Figure 1C.

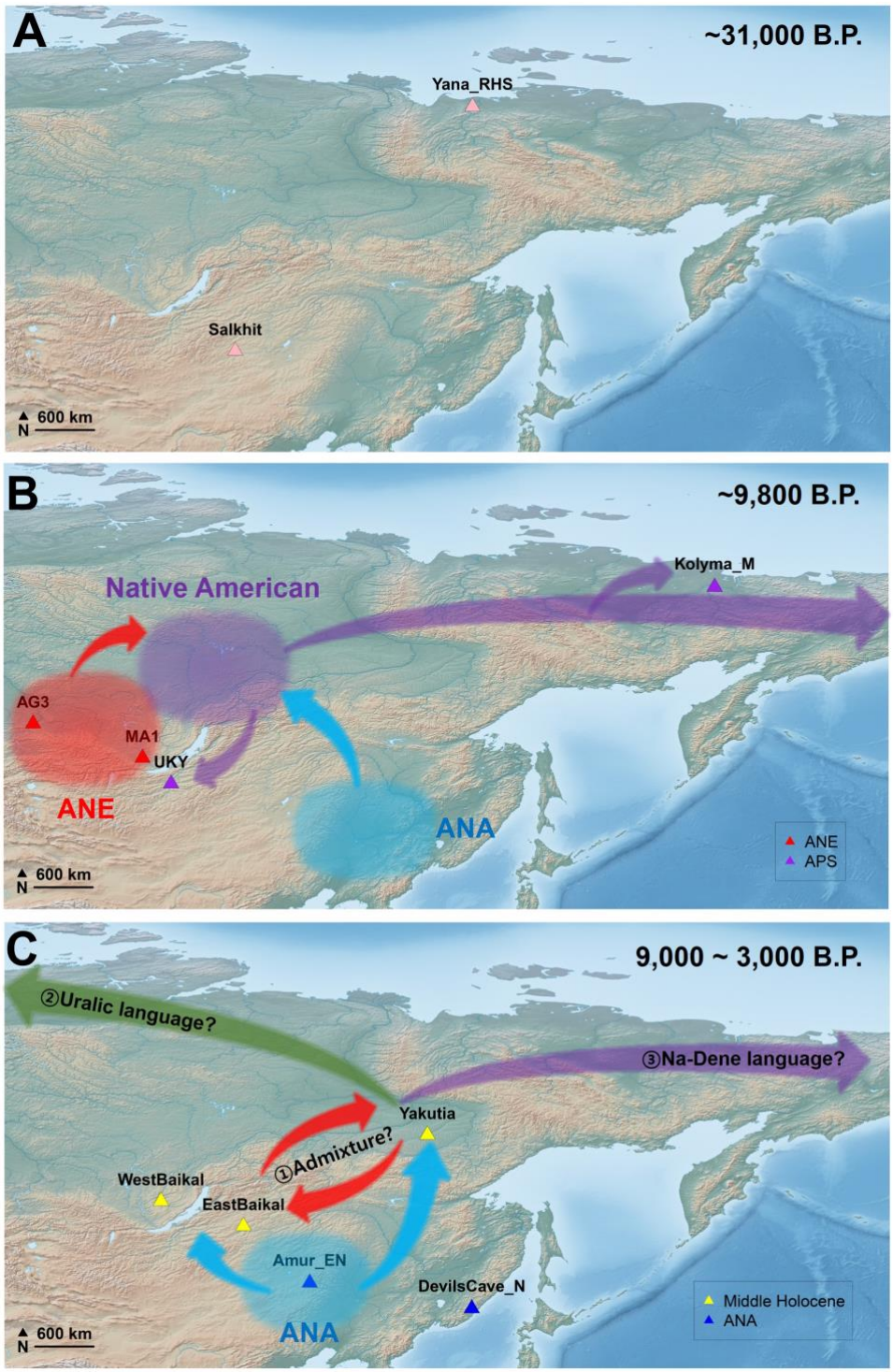

Figure 1. The sampling sites of the key ancient Siberian genomes and major demographic events. (A) The sampling sites of Upper Paleolithic

Siberian ancient genomes, Yana_RHS, and Eastern Eurasian ancient genomes, Salkhit, are presented. (B) The sampling sites of Ancient North Eurasian (ANE) and Ancient Paleo Siberian (APS) are presented. According to previous studies, the ancestors of Native Americans diverged from the East Asian population, Ancient North Asian (ANA), and then met the ANE population. When migrating to America, Native American ancestry was transferred to the APS population. The process of a series of admixture and migration is presented on the map. (C) The sampling sites of Middle Holocene Siberian populations and ANA populations are presented. The migration of ANA ancestry to Siberia is presented, and three major purposes of this study are presented. The base map data is downloaded from Natural Earth (https://www.naturalearthdata.com/downloads/), and the coordinates of each sample are from the AADR annotation file v52.2 (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data).

2. Material and Methods

2.1. Genotype data preparation

All genotype data used in this study are from Allen Ancient DNA Resource (AADR) v52.2, uniformly curated genotypes for thousands of ancient and present-day individuals (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data", version 52.2). There are two SNP datasets: 1240K, a set of 1,233,013 SNP sites in the human genome, and Human Origins (HO), a subset of 1240K merged with present-day human data typed on the Human Origins array with 597,573 sites (Consortium, 2015). The HO dataset was used only for analysis including present-day populations, and the 1240K dataset was used for the rest of the analysis.

Since group-based methods assume independence between individuals (i.e., each genome is treated as an independent sample from a population), close relatives must be excluded for further analysis. To exclude close relatives, pairwise mismatch rates (PMR) between every pair of genomes were calculated by dividing the number of mismatched genotypes by the number of sites covered in both genomes (Kennett et al., 2017). Then, kinship coefficients were calculated by PMR, and one extra pair of duplicates which was not reported was found. For each first-degree pair or duplicate, one of the individuals with low sequencing coverage was removed for further analysis. In addition, outliers

reported in previous studies were also removed. Samples used in this study are listed in Appendix 1 and Appendix 2.

2.2. Principal Component Analysis

Since genotype matrix data is high-dimensional, dimension reduction methods were frequently used for data visualization. Principal Component Analysis (PCA), one of the dimension reduction methods, was performed with present-day individuals in the HO dataset using *smartpca* v18140 from EIGENSOFT version 8.0.0, and the 'lsqrproject: YES' option was used to project ancient samples on precalculated principal components (Patterson et al., 2006). The first PCA was performed with present-day Eurasian and Native American populations, and the second PCA was performed with present-day Eurasian. Present-day populations used in PCA are listed in Appendix 2.

2.3. F-statistics

Although PCA is practical and intuitive for describing the data pattern, quantitative analysis and demographic hypothesis tests are impossible with PCA. Thus, quantitative analysis and demographic hypothesis are tested by F-statistics in this study. F-statistics is a measurement of shared genetic drift between populations and is calculated by the allele frequencies of populations (Patterson et al., 2012). F-statistics between two different populations (referred to as "F_2 ") is estimated by the mean of the square of allele frequency

difference and can be interpreted as a branch length subtending two populations in the phylogenetic tree (Peter, 2016).

F-statistics can also be calculated for three and four populations, but the purpose of each test is different (Patterson et al., 2012). F_3 statistics is defined as the mean product of the allele frequency difference of the other two populations against one fixed population. In other words, for fixed population A and the other two populations B and C, F_3 (A; B, C) is defined as the mean of $(a-c)(a-b)$ (a, b, and c represent allele frequency of A, B, and C). The meaning of F_3 statistics depends on what the fixed group is. If a fixed population, A is an outgroup of B and C, then F_3 (A; B, C), referred to as *outgroup*-F_3 , measures shared genetic drift between B and C against A (Raghavan et al., 2014). Otherwise, F_3 (A; B, C) can be used for testing admixture because it can be negative only if A has intermediate allele frequencies due to admixture between B and C (Durand et al., 2011). The final form of F-statistics, F_4 (A, B; C, D) (A is usually outgroup), is defined as the mean of $(a-b)(c-d)$ and used for testing the cladality of proposed tree topology (Durand et al., 2011). If no asymmetric gene flow has occurred between B and C or B and D (i.e., tree topology (A, (B, (C, D))) is correct), then the expected F_4 (A, B; C, D) is zero. Otherwise, F_4 (A, B; C, D) can be negative or positive when the proposed tree topology is wrong or needs admixture. A significantly positive F_4 value implies asymmetric gene flow between B and C, and a negative value implies asymmetric gene flow between B and D.

In this study, F-statistics were calculated by *qp3pop* and *f4* functions from the R library ADMIXTOOLS2 2.0.0. ver 2.0.0.

(https://github.com/uqrmaie1/admixtools, publication pending). Central African population, Mbuti was used as an outgroup to calculate *outgroup*$-F_3$ and F_4 statistics in the form of F_4 (Mbuti, worldwide; target1, target2). *Outgroup*$-F_3$ was used for measuring shared genetic drift to infer proximal sources or descendants of the target population, and F_4 statistics were used for testing symmetricity between targets or searching additional admixture sources. Worldwide populations used in F-statistics are listed in Appendix 2.

2.4. Admixture modeling analysis

As an application of F-statistics, *the qpWave* method was developed to infer how many independent genetic streams are necessary for a set of target populations against a set of outgroups (Reich et al., 2012). *qpWave* is based on multiple F_4 statistics and infers the rank (the number of independent column or row vectors) of the matrix composed with F_4 statistics. For example, if all F_4 statistics of the form F_4 (target1, target2; outgroup1, outgroup2) are zero, then the rank of *the* F_4 matrix is also zero. Zero rank of the matrix implies that targets are symmetric to each other against all outgroup populations, and only one genetic stream from the outgroup is enough to explain all target populations. Likewise, if the rank of F_4 matrix is r, then at least $(r+1)$ independent streams from outgroups are necessary to explain target populations.

qpAdm is also an application of F-statistics and can be interpreted as a particular case of *qpWave* (Lazaridis et al.,

2016). *qpAdm* supposes that the target population is a mixture of source populations. If the admixture is correct, one of *the* F_4 vectors can be expressed as a linear combination of other F_4 vectors, and then the rank of *the* F_4 matrix must be the number of sources. *qpAdm* compares whether the full-rank model (assumes no admixture and has the most parameters) is significantly better than the admixture model (has fewer parameters than the full-rank model). If the full-rank model is significantly better (i.e., the P-value is lower than the cutoff, 0.05), then it means the admixture model is infeasible. Otherwise, the admixture model is feasible, and admixture proportions can be estimated.

Qpwave and *qpadm* function from R library ADMIXTOOLS2 2.0.0. (https://github.com/uqrmaie1/admixtools, publication pending) were used for admixture modeling analysis. Following populations were used as a base outgroup set for both *qpWave* and *qpAdm* analysis: present-day central Africa population Mbuti (n=5), Southeast Asian Ami (n=2), Native American Mixe (n=3), South Asian Onge (n=2) (Mallick et al., 2016), Neolithic Iranians from the Ganj Dareh site Iran_N (n=8) (Lazaridis et al., 2016), Epipaleolithic European Villabruna (n=1) (Fu et al., 2016), Anatolian Neolithic Anatolia_N (n=23) (Mathieson et al., 2015), and Neolithic southern Russia West_Siberia_N (n=3) (Narasimhan et al., 2019). In addition, when multiple admixture models were feasible, *qpAdm* rotating approach, which systematically shifts candidates from target to outgroup, was used to find the best proximal source (Harney et al., 2021).

2.5. Graph-based analysis

In order to test component-wise admixture models comprehensively, graph-based analysis was implemented by using *the qpgraph* function from the R library ADMIXTOOLS2 2.0.0. (https://github.com/uqrmaie1/admixtools, publication pending). *Qpgraph* automatically fits the branch length, which is measured by *F*-statistics and estimates admixture proportions of a given graph topology (Patterson et al., 2012). Since *qpgraph* does not search other graph topologies, it differs from automated-graph search programs like treeMix (Pickrell & Pritchard, 2012). However, a fully-automated graph search algorithm based on *qpgraph*, *find_graph* function has been developed recently in ADMIXTOOLS2. Although the graph is likely to be over-fitted and fitted to the local optimum by the *find_graph* function, consensus features between fitted graphs are informative.

In this study, the *qpgraph* function was used to construct the admixture graph of Middle Holocene Siberian populations manually. Before graph fitting, F_2 statistics between all pairs of target populations are calculated by *extract_f2* function in ADMIXTOOLS2 with 'max_miss=0' option, the same as 'allsnps: NO' in the previous version. The number of within SNP sets was 182,628. Mbuti population was also used as an outgroup in this analysis, and the following populations were used for distal representatives: MA1 for ANE; WHG for Mesolithic hunter-gatherers from Europe (Mathieson et al., 2018); USR1 for Native Americans (Moreno-Mayar et al., 2018); EastBaikal_N for ANA. Then, Middle Holocene populations were systematically added by following orders: irk030,

Dzhylinda-1, WestBaikal_EN, Yakutia_MN, Saqqaq, WestBaikal_LNBA, and Yakutia_LN.

Then, Lake Baikal populations and Yakutia populations were replaced with Metal Age north-eastern European (Russia_Bolshoy) (Lamnidis et al., 2018) and Comb Ceramic Complex hunter-gatherers (Estonia_MN_CCC) (Saag et al., 2017) to find the best admixture graph explaining the migration of Siberian ancestry to north-eastern Europe. In this case, the number of within SNP sets was 151,255.

3. Results

The result section will be divided into two parts because of quantity and complexity. The first part (3.1) delineates a general description of data preparation and visualization. Then, the spatiotemporal demographic history of Middle Holocene Siberian populations will be tested by population genetic analysis. The second part (3.2) explains the impact of Middle Holocene Siberian populations on other regions and linguistic families.

3.1.1. Data preparation and archaeological background

Ancient and present-day genomes data used in this study are included in the Allen Ancient DNA Resource (AADR, version 52.2) dataset (Method). First, three well-known distal sources of Middle Holocene Siberian were prepared: Ancient North Eurasian (ANE) represented by Upper Paleolithic genomes, Afontova Gora3 (AG3) and Mal'ta1 (MA1) (Fu et al., 2016; Raghavan et al., 2014); Ancient North Asian (ANA) represented by early Neolithic hunter-gatherers from Devil's Gate Cave (DevilsCave_N) and Amur river (AR_EN) (Ning et al., 2020; Siska et al., 2017); Ancient Paleo Siberian (APS) represented by Paleolithic genome from Lake Baikal (UKY) and Mesolithic genome from Kolyma river (Kolyma_M) (Sikora et al., 2019; Yu et al., 2020) (Figure 2).

Then, the Middle Holocene Siberian genomes were prepared.

Middle Holocene Siberian genomes are mainly from Lake Baikal, Yakutia, and Kolyma river region, and the number of genomes from Middle Holocene is relatively larger than that of genomes from Pleistocene (de Barros Damgaard et al., 2018; K ı l ı nç et al., 2021; Sikora et al., 2019; Yu et al., 2020). Since these genomes were reported from multiple different studies, relatedness needs to be checked. Thus, the pairwise mismatch rate (PMR) between every pair of genomes was calculated to infer the kinship coefficient (Kennett, Douglas J. et al. 2017) (Methods, Appendix 1). As a result, an extra pair of duplicates not reported in AADR was found. Then, duplicates, first-degree relatives, and outliers reported in previous studies were removed for further analysis.

Finally, one ancient population and three present-day populations were selected as target groups to explore the influence in north-eastern Europe (Russia_Bolshoy) (Lamnidis et al., 2018), Uralic-speaking (Nganasan, Selkup) and Yenisei-speaking (Ket) populations. The sampling sites of key samples are presented in Figure 2.

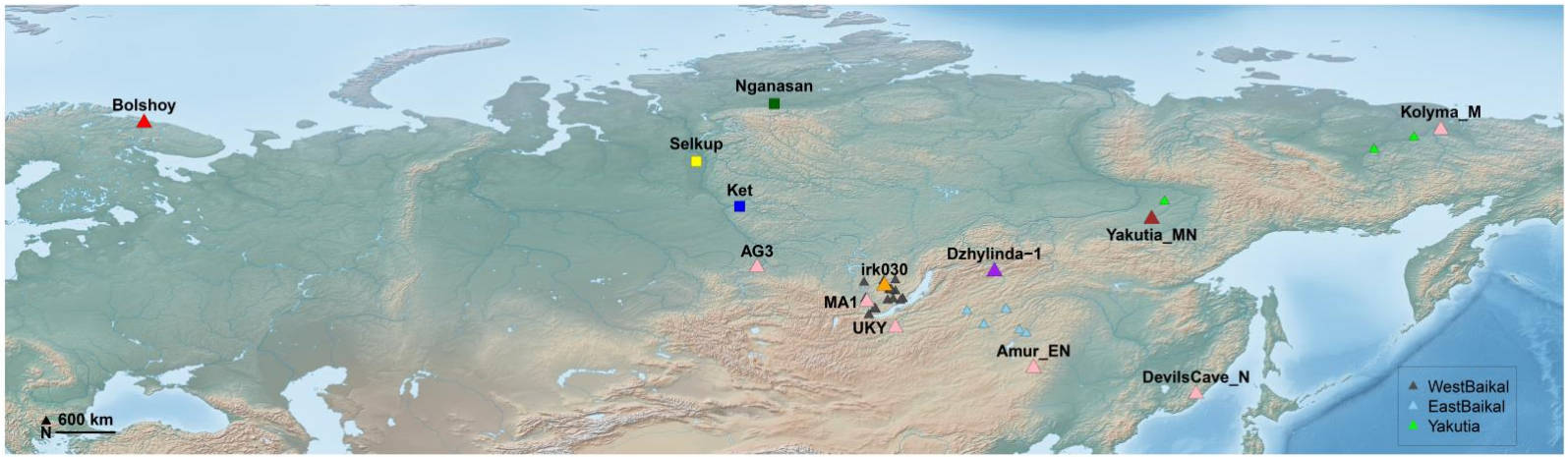

Figure 2. The sampling sites of the key ancient and present-day genomes used in this study. The sampling sites of ancient and present-day genomes used in this study are presented. Triangles represent ancient genomes, and squares represent present-day genomes. Distal source populations (AG3, MA1 for ANE; UKY, Kolyma_M for APS; Amur_EN, DevilsCave_N for ANA), which are used for explaining the formation of Middle Holocene Siberians, are represented by light pink triangles. Among Middle Holocene Siberians, key individuals, irk030, Dzhylinda-1, and Yakutia_MN, are marked separately, and the rest of the Middle Holocene populations are marked based on geographic locations. A red triangle represents the north-eastern European target from Bolshoy Oleni Ostrov (Bolshoy). Three present-day Siberian targets, Nganasan, Selkup, and Ket, are represented by green, yellow, and blue squares, respectively. The base map data is downloaded from Natural Earth (https://www.naturalearthdata.com/downloads/), and the coordinates of each sample are from the AADR annotation file v52.2 (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data).

3.1.2. Population structure of ancient Siberian

Principal component analysis (PCA) was performed with present-day Eurasian and American populations, and ancient genomes were projected on the calculated PCs to explore the population structure of studied genomes. As reported in previous studies, PC1 separated individuals from west to east, and PC2 separated individuals from Eurasians to Americans (Jeong et al., 2019). Almost Middle Holocene genomes from Lake Baikal and Yakutia were located on "ANE-ANA" cline stretching between ANE represented by AG3 and ANA represented by DevilsCave_N (Figure 3). Exceptionally, two Lake Baikal samples, Dzhylinda-1 and irk030, were not on the ANE-ANA cline and shifted toward Native American populations represented by USR1 (Moreno-Mayar et al., 2018) along the PC2 axis like APS such as UKY and Kolyma_M.

Except for irk030 and Dzhylinda-1, the rest of the Middle Holocene Siberian genomes were clustered by period and location. Each of the clusters was named as follows: Early Neolithic West Baikal (WestBaikal_EN); Late Neolithic to Bronze Age West Baikal (WestBaikal_LNBA); Neolithic East Baikal (EastBaikal_N); Middle Neolithic Yakutia (Yakutia_MN); Late Neolithic Yakutia (Yakutia_LN); These clusters were divided into two groups according to affinity against ANE and ANA: ANE-like Siberian, WestBaikal_LNBA and Yakutia_MN that were shifted toward ANE; ANA-like Siberian, WestBaikal_EN, Yakutia_LN and EastBaikal_N that were shifted toward ANA (Figure 3). Interestingly, the frequency of the Y-haplogroup is almost fixed in each subgroup.

For example, 26 of 27 Y-haplogroups of ANE-like Siberians are Q, and 12 of 16 Y-haplogroups of males in ANA-like Siberians are N (Appendix 1).

Next, PCA was performed with only present-day Eurasian populations (Figure 4). In this result, PC2 separated the Asian population from north to south. Yakutia_LN and WestBaikal_EN were split, and Yakutia_LN was shifted toward Nganasan population along PC2. This shift suggested that ancient individuals in Yakutia shared more genetic drift with Nganasan than that other ancient Siberians.

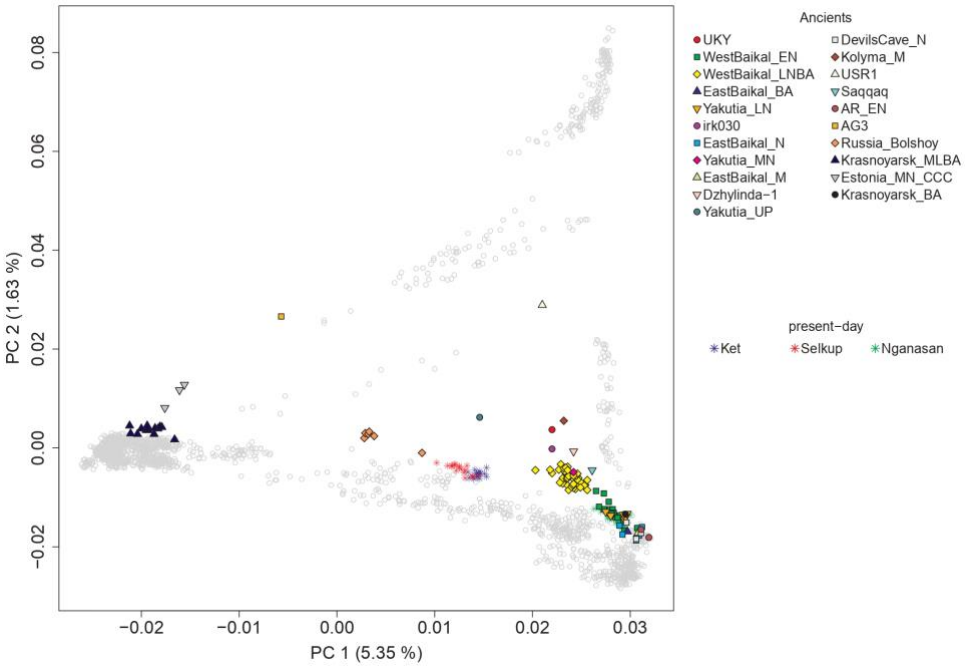

Figure 3. Principal component analysis performed with present-day Eurasian and American individuals. Principal component analysis is performed with present-day Eurasian and American individuals, and each present-day sample is placed on principal component 1 and 2 coordinates by grey circle. Key ancient genomes and present-day genomes are projected on pre-calculated principal components and labeled.

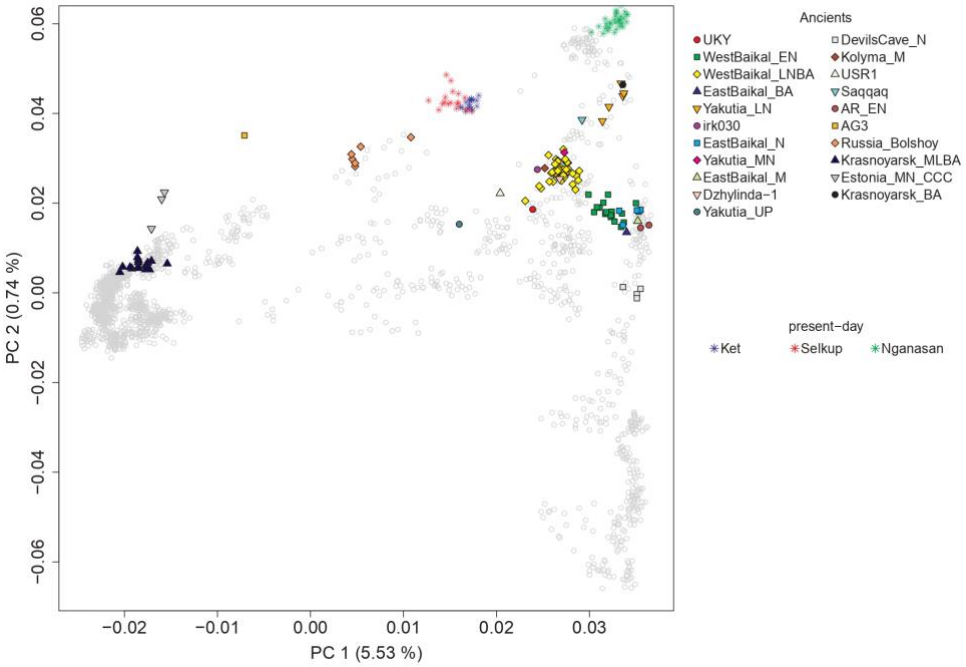

Figure 4. Principal component analysis performed with present-day Eurasian individuals. Principal component analysis is performed with present-day Eurasian individuals only, and each present-day sample is placed on principal component 1 and 2 coordinates by grey circle. Key ancient genomes and present-day genomes are projected on pre-calculated principal components and labeled.

3.1.3. Proximal source of ANE and ANA ancestry for Middle Holocene Siberia

The previous study explained Middle Holocene Siberians by admixtures between APS and ANA populations (Yu et al., 2020). However, the time between APS and the Middle Holocene Siberian is over 2,000 years. It means reconstructing high-resolution dynamics of Middle Holocene populations with distal APS sources is impossible. Therefore, proximal sources for the Middle Holocene Siberian were searched. Dzhylinda-1 and irk030 were assumed to be proximal sources of APS ancestry in Middle Holocene Siberian populations and referred to as APS-like Siberians because they were shifted toward APS on PCA results (Figure 3).

As with APS, irk030 and Dzhylinda-1 were expected to share Native American ancestry. In order to validate this assumption quantitatively, *qpAdm* analysis was performed to test whether a three-way admixture: ANE + ANA + Native American is also feasible for irk030 and Dzhylinda-1. As a result, irk030 and Dzhylinda-1 were adequately modeled as a three-way admixture with similar ancestry proportion to UKY and Kolyma_M (Figure 5, Table 1). The same admixture model was tested for Middle Holocene populations, and all Middle Holocene populations showed diminished Native American ancestry compared to the former four individuals (Table 1). Furthermore, *qpWave* was implemented to infer how many independent genetic streams were necessary for grouped populations, and only one stream was enough for (APS, irk030) and (APS, Dzhylinda-1) (Table 2). In other words, irk030

and Dzhylinda−1 can be explained by only genetic drift from APS without additional admixture. Meanwhile, an additional genetic stream was necessary when Dzhylinda−1 and irk030 were grouped (Table 2).

Next, *outgroup*−F_3 statistics were calculated to confirm genetic continuity from Dzhylinda−1/irk030 to the Middle Holocene Siberian (Figure 6). Then, irk030 shared the most genetic drift with WestBaikal_LNBA, and Dzhylinda−1 shared the most genetic drift with Yakutia_MN. This result mirrored the geographical location of each sample. Indeed, irk030 was found in the West Baikal region as WestBaikal_LNBA individuals, and Dzhylinda−1 was found in the northeast Baikal region, which is close to the Yakutia region (Figure 2). Then, *qpWave* analysis further confirmed that only one stream was enough for (Dzhylinda−1, Yakutia_MN) and at least two independent genetic streams were necessary to explain (irk030, WestBaikal_LNBA) (Table 2). Interestingly, Dzhylinda−1 showed close genetic affinity to not only Yakutia_MN but also Paleo−Eskimo Saqqaq. All combinations of subgroups among Yakutia_MN, Saqqaq, and Dzhylinda−1 were tested by *qpWave*, and the best topology was (Dzhylinda−1, (Yakutia_MN, Saqqaq)) (Table 2).

Likewise, EastBaikal_N was assumed as the proximal ANA source because EastBaikal_N is close to AR_EN, one of the ANA populations, both geographically and genetically (Figure 2, Figure 3). *QpWave* analysis confirmed that no extra genetic stream was necessary for (AR_EN, EastBaikal_N), however at least two genetic streams were necessary for EastBaikal_N and other ANA populations such as DevilsCave_N, Yumin (Table 3). Therefore, proximal ANA ancestry was introduced from the Amur river basin

to the East Baikal region. *Outgroup*−F_3 results also supported close genetic affinity between ANA−like Siberian and EastBaikal_N (Fig. 7). However, both Yakutia_LN and WestBaikal_EN were found to require extra genetic stream against EastBaikal_N by *qpWave* analysis (Table 3).

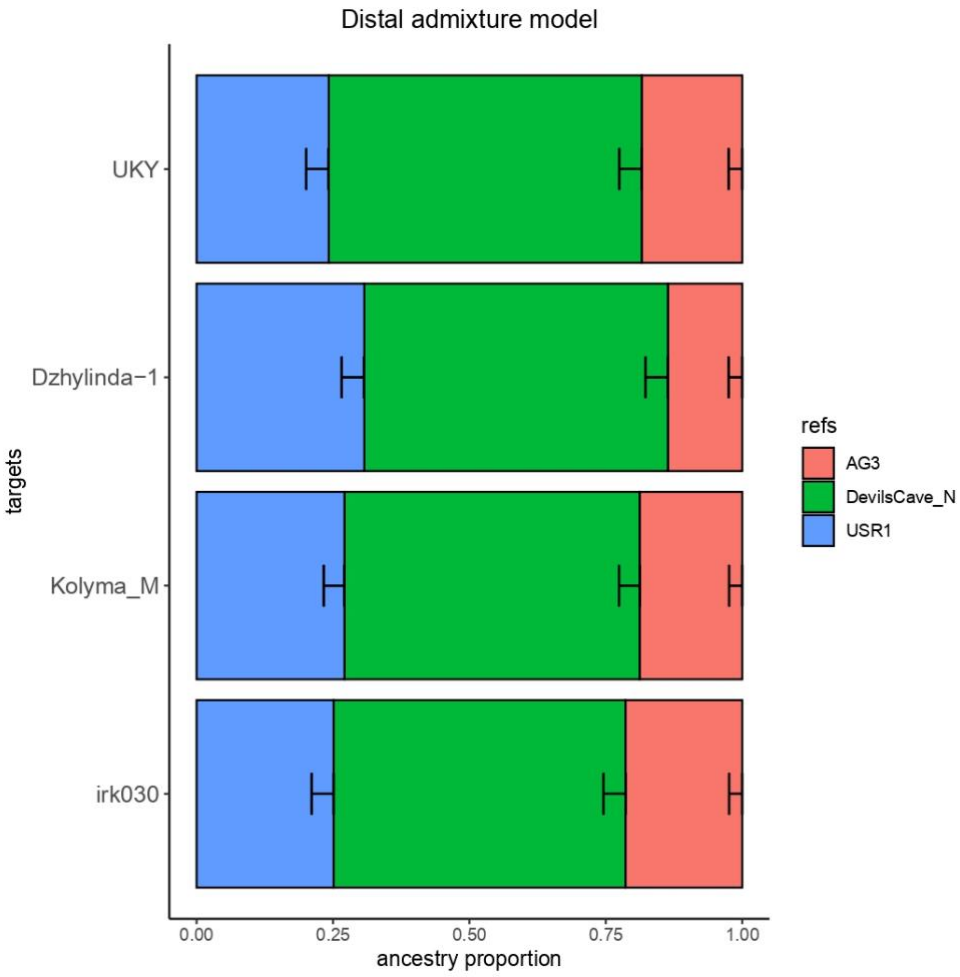

Figure 5. Distal *qpAdm* modeling of APS individuals and APS-like Middle Holocene Siberians. The results of three-way admixture model: AG3 (ANE) + DevilsCave_N (ANA) + USR1 (Native American) are presented by stacked-bar plot. Horizontal bars represent the standard error of ancestry proportion calculated by the 5cM block jackknifing method provided in the *qpAdm* function.

Table 1. Distal *qpAdm* modeling of key ancient Siberians. The results of *qpAdm* modeling composed of distal source, ANE, ANA, and Native American are summarized. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

Target	Ref1	Ref2	Ref3	p-value	Coeff1	Coeff2	Coeff3
UKY	AG3	DevilsCave_N	-	3.31E-03	0.265±0.018	0.735±0.018	-
	AG3	AR_EN	-	1.40E-03	0.257±0.024	0.743±0.024	-
	AG3	DevilsCave_N	USR1	9.27E-01	0.184±0.025	0.573±0.041	0.242±0.055
	AG3	AR_EN	USR1	8.32E-01	0.163±0.031	0.538±0.049	0.299±0.064
Kolyma_M	AG3	DevilsCave_N	-	2.19E-05	0.279±0.017	0.721±0.017	-
	AG3	AR_EN	-	3.53E-03	0.268±0.023	0.732±0.023	-
	AG3	DevilsCave_N	USR1	6.02E-01	0.188±0.024	0.541±0.038	0.271±0.053
	AG3	AR_EN	USR1	9.99E-01	0.171±0.03	0.55±0.045	0.28±0.062
irk030	AG3	DevilsCave_N	-	3.27E-05	0.295±0.017	0.705±0.017	-
	AG3	AR_EN	-	2.28E-04	0.304±0.022	0.696±0.022	-
	AG3	DevilsCave_N	USR1	1.33E-01	0.214±0.024	0.535±0.04	0.251±0.054
	AG3	AR_EN	USR1	8.77E-01	0.198±0.03	0.474±0.048	0.329±0.064
Dzhylinda-1	AG3	DevilsCave_N	-	1.52E-05	0.233±0.019	0.767±0.019	-
	AG3	AR_EN	-	1.21E-04	0.227±0.026	0.773±0.026	-
	AG3	DevilsCave_N	USR1	7.10E-01	0.136±0.025	0.556±0.041	0.307±0.056
	AG3	AR_EN	USR1	8.44E-01	0.113±0.032	0.513±0.053	0.374±0.07
Yakutia_MN	AG3	DevilsCave_N	-	7.04E-01	0.238±0.021	0.762±0.021	-
	AG3	AR_EN	-	1.66E-01	0.21±0.035	0.79±0.035	-
	AG3	DevilsCave_N	USR1	7.17E-01	0.216±0.031	0.722±0.046	0.061±0.063
	AG3	AR_EN	USR1	2.72E-01	0.161±0.046	0.676±0.077	0.162±0.1
WestBaikal_LNBA	AG3	DevilsCave_N	-	1.54E-03	0.242±0.01	0.758±0.01	-
	AG3	AR_EN	-	9.81E-04	0.238±0.014	0.762±0.014	-
	AG3	DevilsCave_N	USR1	1.01E-02	0.219±0.013	0.706±0.022	0.076±0.03
	AG3	AR_EN	USR1	1.18E-01	0.19±0.018	0.656±0.029	0.154±0.039
Yakutia_LN	AG3	DevilsCave_N	-	1.15E-03	0.129±0.011	0.871±0.011	-
	AG3	AR_EN	-	2.22E-03	0.129±0.017	0.871±0.017	-
	AG3	DevilsCave_N	USR1	5.35E-03	0.101±0.017	0.81±0.028	0.089±0.038
	AG3	AR_EN	USR1	7.44E-02	0.078±0.023	0.756±0.037	0.166±0.05
WestBaikal_EN	AG3	DevilsCave_N	-	1.56E-01	0.111±0.01	0.889±0.01	-
	AG3	AR_EN	-	1.96E-01	0.113±0.015	0.887±0.015	-
	AG3	DevilsCave_N	USR1	9.82E-02	0.111±0.014	0.887±0.025	0.002±0.033
	AG3	AR_EN	USR1	4.04E-01	0.085±0.02	0.829±0.033	0.086±0.044

Table 2. The genetic cladality tests between APS, APS−like Siberians and ANE−like Siberians. Genetic cladality between all kinds of pairs between APS, APS−like and ANE−like Siberian individuals are tested by *qpWave*. Non−significant p−values (p > 0.05) are highlighted in grey and support cladality between target individuals against base outgroup.

target 1	target 2	target 3	target 4	p-value
UKY	irk030	-	-	5.3E-02
irk030	Kolyma_M	-	-	3.31E-01
irk030	Dzhylinda-1	-	-	2.81E-03
UKY	Kolyma_M	-	-	3.18E-01
UKY	Dzhylinda-1	-	-	7.52E-01
Dzhylinda-1	Kolyma_M	-	-	2.3E-01
UKY	Dzhylinda-1	irk030	-	1.97E-02
UKY	Kolyma_M	irk030	-	1.02E-01
Dzhylinda-1	Kolyma_M	irk030	-	2.37E-02
UKY	Kolyma_M	Dzhylinda-1	-	3.86E-01
UKY	Kolyma_M	Dzhylinda-1	irk030	3.27E-02
irk030	WestBaikal_LNBA	-	-	5.62E-06
irk030	Dzhylinda-1	-	-	2.81E-03
irk030	Yakutia_MN	-	-	1.86E-03
Dzhylinda-1	WestBaikal_LNBA	-	-	7.79E-06
Dzhylinda-1	Yakutia_MN	-	-	1.03E-01
Yakutia_MN	WestBaikal_LNBA	-	-	1.40E-01
Saqqaq	irk030	-	-	4.13E-05
Saqqaq	Dzhylinda-1	-	-	3.96E-02
Saqqaq	WestBaikal_LNBA	-	-	4.08E-02
Saqqaq	Yakutia_MN	-	-	1.64E-01

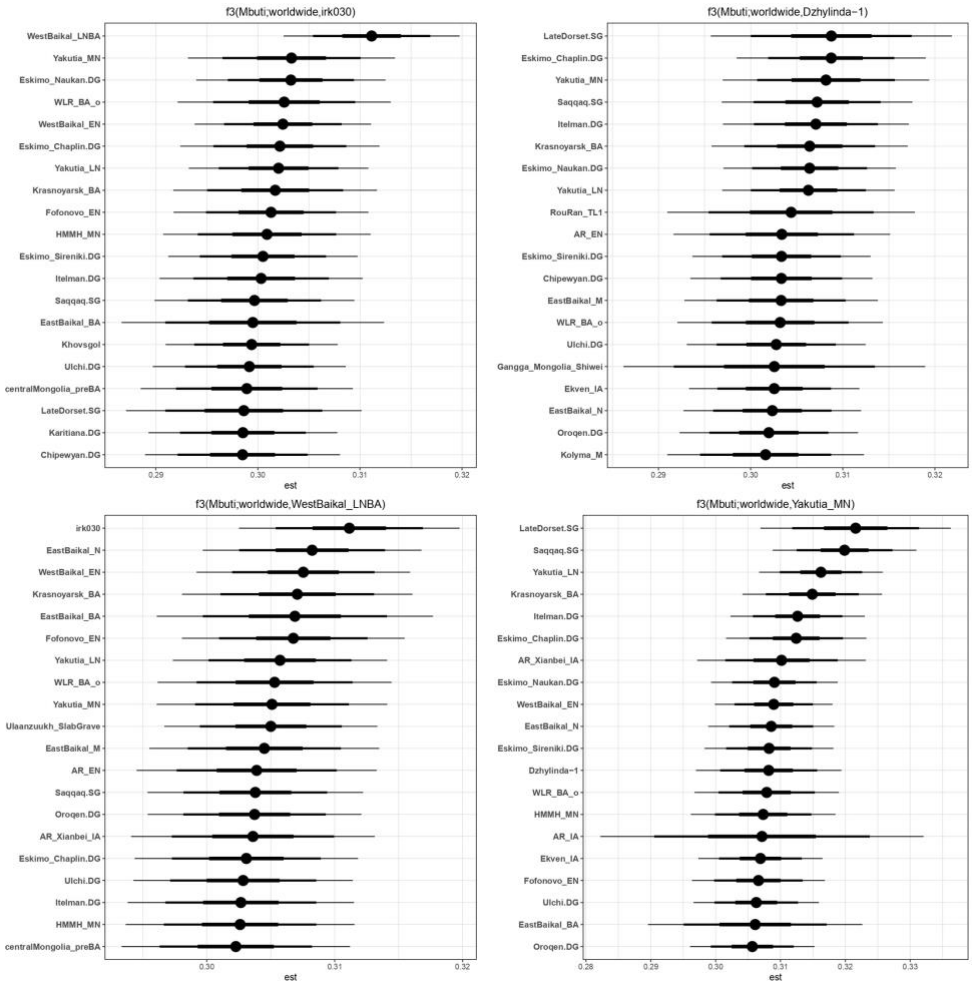

Figure 6. Genetic affinity measured by outgroup−F_3 statistics for APS−like and ANE−like Siberians. *Outgroup*−F_3 statistics are calculated for each APS−like and ANE−like Siberians to explore genetically close populations. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

Table 3. The genetic cladality tests between ANA and ANA−like Siberians. Genetic cladality between all kinds of pairs between ANA (AR_EN, DevilsCave_N, Yumin), EastBaikal_N and ANA−like Siberian (WestBaikal_EN, Yakutia_LN) are tested by *qpWave*. Non−significant p−values (p > 0.05) are highlighted in grey and support symmetricity between target individuals against base outgroup.

target1	target2	p-value
EastBaikal_N	AR_EN	7.65E-01
EastBaikal_N	DevilsCave_N	1.64E-07
EastBaikal_N	Yumin	3.83E-03
AR_EN	DevilsCave_N	1.47E-02
AR_EN	Yumin	3.13E-02
DevilsCave_N	Yumin	2.58E-06
EastBaikal_N	Yakutia_LN	2.18E-14
EastBaikal_N	WestBaikal_EN	7.01E-12
Yakutia_LN	WestBaikal_EN	5.42E-02

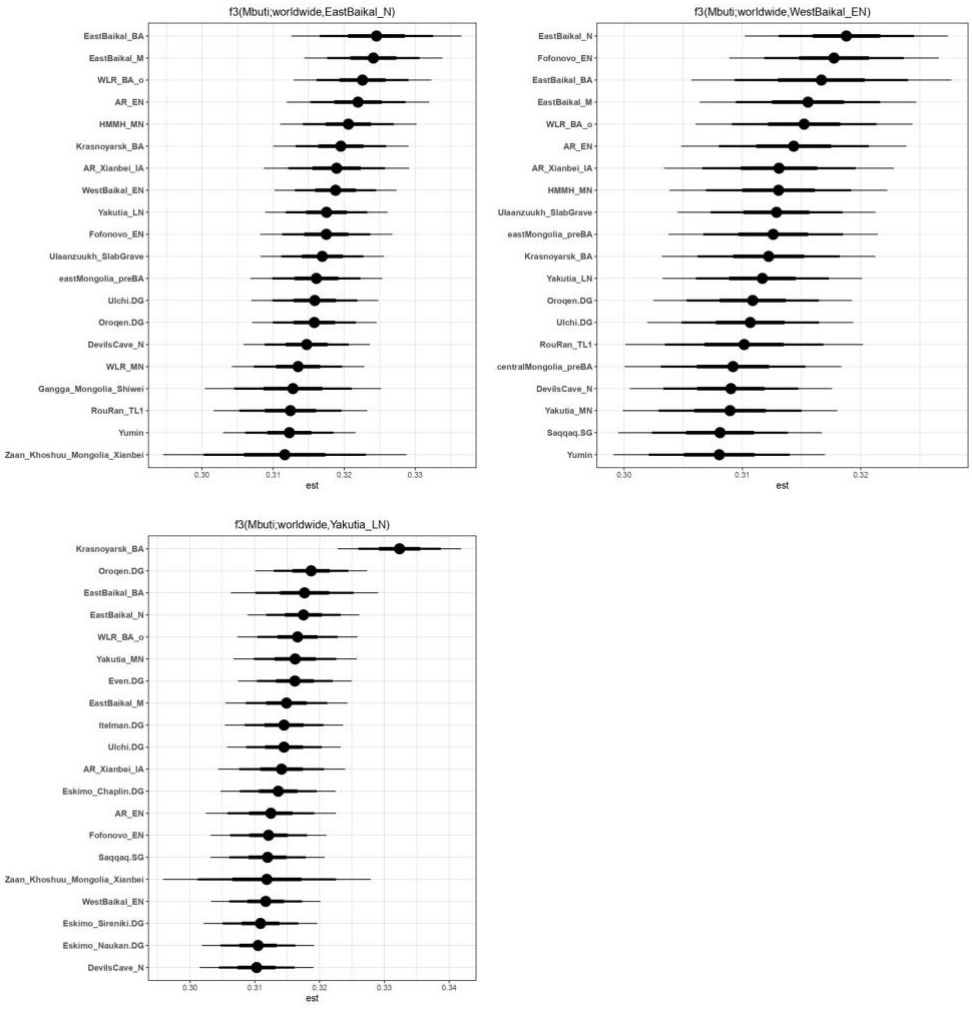

Figure 7. Genetic affinity measured by *outgroup*−F_3 statistics for EastBaikal_N and ANA−like Siberians. *Outgroup*−F_3 statistics are calculated for each EastBaikal_N and ANA−like Siberians to explore genetically close populations, and 20 closest populations are shown. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

3.1.4 Complex and multiple admixture events occurred during Holocene

Population dynamics in the Lake Baikal and Yakutia region were complicated. ANA-related ancestry increased, and the major Y-haplogroup was changed from Q to N over time in the Yakutia region, but a reversed pattern was observed in Lake Baikal. Thus, the proximal admixture model for each Middle Holocene Siberian population was tested to propose more high-resolution dynamics.

Since Yakutia_MN and WestBaikal_LNBA were undistinguishable by base outgroup in *qpWave* results (Table 2), proximal sources which can dissect Yakutia_MN and WestBaikal_LNBA were necessary. Dzhylinda-1 and irk030 were tested as proximal sources by qpWave analysis. As expected, each *qpWave* p-values decreased to 2.71E-02 and 2.14E-04 when Dzhylinda-1 and irk030 were added to the outgroup, respectively. Next, direction of gene flow was tested by calculating F_4 statistics, and asymmetric gene flows were confirmed from irk030 to WestBaikal_LNBA (F_4 (Mbuti, irk030; WestBaikal_LNBA, Yakutia_MN): Z=-3.83), and from Dzhylinda-1 to Yakutia_MN (F_4 (Mbuti, Dzhylinda-1; WestBaikal_LNBA, Yakutia_MN); Z=3.25). Thus, Yakutia_MN and WestBaikal_LNBA were found to have different proximal APS ancestry, which Dzhylinda-1 and irk030 represent, respectively.

As confirmed by *qpWave* (Table 2), an extra genetic stream was unnecessary for (Yakutia_MN, Dzhylinda-1), which means Yakutia_MN could be explained by only genetic drift from

Dzhylinda−1. This symmetricity was further tested against worldwide populations by F_4 statistics of the form F_4 (Mbuti, worldwide; Dzhylinda−1, Yakutia_MN), and results were symmetric to all populations except two, Yakutia_LN and WestBaikal_EN (Figure 8). This asymmetricity implies asymmetric gene flow between ANA−like Siberian and Yakutia_MN. Thus, *qpWave* was implemented again for (Dzhylinda−1, Yakutia_MN) by adding ANA−like Siberians to outgroups (Table 4). Yakutia_LN and WestBaikal_EN distinguished between Yakutia_MN and Dzhylinda−1, and even Yakutia_MN = Dzhylinda−1 + Yakutia_LN/WestBaikal_EN admixture models were found to be feasible by *qpAdm* analysis (Table 4). Then, WestBaikal_EN and Yakutia_LN were compared by *qpAdm* rotating approach, and Yakutia_LN was better than WestBaikal_EN (Table 4).

For modeling of WestBaikal_LNBA, F_4 statistics of the form F_4 (Mbuti, worldwide; irk030, WestBaikal_LNBA) were calculated to explore possible admixture source of WestBaikal_LNBA against irk030. irk030 had deep ANE ancestry represented by AG3 and MA1 whereas WestBaikal_LNBA had more ANA ancestry represented by AR_EN and EastBaikal_N (Figure 8). Therefore, WestBaikal_LNBA could be modeled adequately by two−way admixture: irk030 + EastBaikal_N/WestBaikal_EN, and EastBaikal_N was demonstrated as the best ANA proxy by *qpAdm* rotating approach (Table 5). To exclude the possibility of gene flow from WestBaikal_LNBA to irk030, a two−way admixture: irk030 = WestBaikal_LNBA + ANE was tested by *qpAdm* and found to be infeasible (Table 5). As a result, the gene would have flowed from irk030 to WestBaikal_LNBA because

irk030 not only predates time estimates of all WestBaikal_LNBA individuals but is not modeled as WestBaikal_LNBA + ANE.

Next, admixture models of ANA-like Siberians (i.e., Yakutia_LN and WestBaikal_EN) were tested. A simple one-way model of EastBaikal_N was infeasible for both Yakutia_LN and WestBaikal_EN (Table 3), and they were found to need more ANE sources by F_4 (Mbuti, worldwide; EastBaikal_N, Yakutia_LN/WestBaikal_EN) (Figure 9). Although Yakutia_LN and WestBaikal_EN were cladal against the base outgroup in *the qpWave* result (Table 3), they could be distinguished by adding Yakutia_MN and Dzhylinda-1 to the outgroup (Table 6). In addition, F_4 statistics found that Dzhylinda-1 and Yakutia_MN were more close to Yakutia_LN against WestBaikal_EN (F_4 (Mbuti, Dzhylinda-1; WestBaikal_EN, Yakutia_LN): Z=4.92, F_4 (Mbuti, Yakutia_MN; WestBaikal_EN, Yakutia_LN): Z=5.09). Since this implies that WestBaikal_EN and Yakutia_LN are not sister clades against Yakutia_MN and Dzhylinda-1, plausible tree topologies were tested by F_4 statistics (Table 7). Both Yakutia_LN and WestBaikal_EN were closer to Yakutia_MN than Dzhylinda-1, and Dzhylinda-1 was symmetric to Yakutia_LN and Yakutia_MN which suggests the tree topology, (Dzhylinda-1, (Yakutia_MN, Yakutia_LN)). On the other hand, Dzhylinda-1 was significantly closer to Yakutia_MN than WestBaikal_EN. This pattern is possible only if Yakutia_MN is an admixture of Dzhylinda-1 and WestBaikal_EN.

As expected, Yakutia_LN could be modeled as Dzhylinda-1/Yakutia_MN/Saqqaq + EastBaikal_N, and Yakutia_MN was found to be the best source by *qpAdm* rotating approach (Table 8). In

addition, Dzhylinda−1 was not a feasible source for WestBaikal_EN, and irk030 was found to be the best proximal source (Table 9). Although Yakutia_LN was found to be the best proximal source for Yakutia_MN by rotating approach (Table 4), this could be explained even when Yakutia_LN was a close descendant of Yakutia_MN. Considering the F_4 results and cultural contexts, Yakutia_LN is more likely to be a descendant of Yakutia_MN.

Lastly, all proximal models of the Middle Holocene Siberians were tested comprehensively by graph−based method, *qpgraph*. Basal graph including Mbuti, MA1, WHG, EastBaikal_N, and USR1 was constructed referring to the previous study (Yu et al., 2020), and ancient Siberian populations were added to the basal graph one by one based on each proximal model. The final admixture graph was statistically plausible (i.e., worst $|Z| < 3$) and needed seven admixture events (Figure 10).

Table 4. Proximal admixture modeling of Yakutia_MN. The results of *qpAdm* and *qpWave* modeling for Yakutia_MN are summarized. To find the best proximal source, candidate sources are systematically added to base outgroup and change of significance is observed. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

			Base outgroup			Base outgroup + WestBaikal_EN			Base outgroup + Yakutia_LN		
Target	**Ref1**	**Ref2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**
Yakutia_MN	Dzhylinda-1	-	1.03E-01	-	-	7.09E-04	-	-	2.75E-04	-	-
	Dzhylinda-1	WestBaikal_EN	3.30E-01	0.708±0.121	0.292±0.121	-	-	-	1.56E-03	0.672±0.132	0.328±0.132
	Dzhylinda-1	Yakutia_LN	2.53E-01	0.688±0.146	0.312±0.146	1.04E-01	0.535±0.103	0.465±0.103	-	-	-

Table 5. Proximal admixture model of irk030 and WestBaikal_LNBA. The results of *qpAdm* and *qpWave* modeling for WestBaikal_LNBA and irk030 are summarized. To find the best proximal source, candidate sources are systematically added to base outgroup and change of significance is observed. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

			Base outgroup			Base outgroup + WestBaikal_EN			Base outgroup + EastBaikal_N		
Target	**Ref1**	**Ref2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**
WestBaikal_LNBA	irk030	EastBaikal_N	9.26E-02	0.751±0.040	0.249±0.040	1.14E-01	0.735±0.034	0.265±0.034	-	-	-
	irk030	WestBaikal_EN	1.43E-01	0.667±0.047	0.333±0.047	-	-	-	1.63E-02	0.606±0.034	0.394±0.034
	irk030	Yakutia_LN	3.48E-02	0.650±0.059	0.350±0.059	1.04E-02	0.612±0.053	0.388±0.053	5.30E-03	0.577±0.040	0.423±0.04
irk030	AG3	WestBaikal_LNBA	1.39E-02	0.920±0.020	0.080±0.020	2.54E-02	0.920±0.017	0.080±0.017	7.32E-03	0.903±0.018	0.097±0.018
	MA1	WestBaikal_LNBA	1.53E-03	0.921±0.021	0.079±0.021	2.51E-03	0.911±0.017	0.089±0.017	2.57E-05	0.884±0.018	0.116±0.018

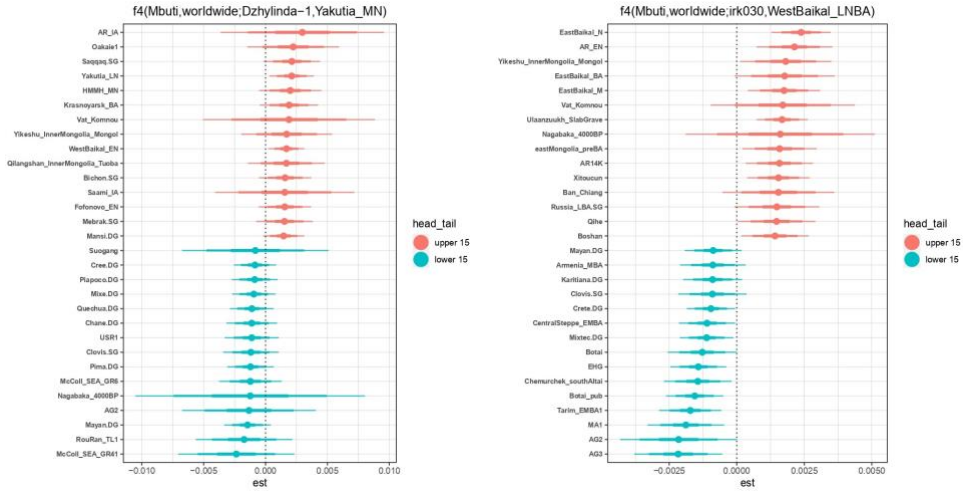

Figure 8. Genetic symmetricity test between ANE−like Siberians and each of proximal source. F_4 statistics of the form F_4 (Mbuti, worldwide; Dzhylinda−1, Yakutia_MN) and F_4 (Mbuti, worldwide; irk030, WestBaikal_LNBA) are calculated for exploring extra admixture sources. 15 most positive and 15 most negative Z score results are shown in red and blue, respectively. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

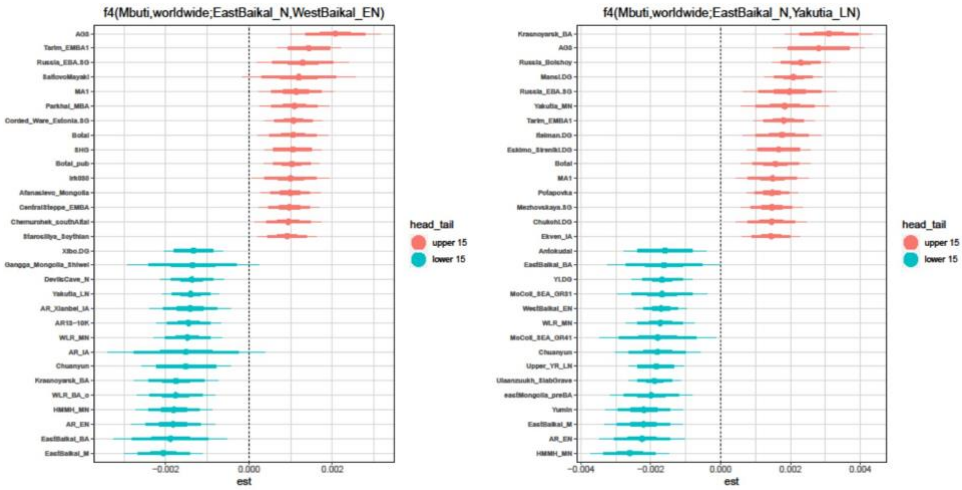

Figure 9. Genetic symmetricity test between ANA−like Siberians and proximal ANA source. F_4 statistics of the form F_4 (Mbuti, worldwide; EastBaikal_N, WestBaikal_EN) and F_4 (Mbuti, worldwide; EastBaikal_N, Yakutia_LN) are calculated for exploring extra admixture sources. 15 most positive and 15 most negative Z score results are shown in red and blue, respectively. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

Table 6. The genetic cladality tests between ANA−like Siberians. Genetic cladality between WestBaikal_EN and Yakutia_LN are tested by adding Dzhylinda−1 or Yakutia_MN to outgroup sets. Non−significant p−values (p > 0.05) are highlighted in grey and support symmetricity between target individuals against outgroup.

Target 1	Target 2	outgroup	p-value
WestBiakal_EN	Yakutia_LN	base outgroup	5.42E-02
WestBiakal_EN	Yakutia_LN	base outgroup + Dzhylinda-1	5.50E-05
WestBiakal_EN	Yakutia_LN	base outgroup + Yakutia_MN	2.77E-04

Table 7. The genetic cladality tests between Middle Holocene Siberians. F_4 statistics are calculated for detecting direction of gene flow and determining plausible tree topology for Dzhylina−1, Yakutia_MN, Yakutia_LN and WestBaikal_EN. Significant results (|Z| > 3) are highlighted in grey.

pop1	pop2	pop3	pop4	Z
Mbuti	WestBaikal_EN	Dzhylinda-1	Yakutia_MN	3.47
	Dzhylinda-1	WestBaikal_EN	Yakutia_MN	3.13
	Yakutia_MN	Dzhylinda-1	WestBaikal_EN	-0.227
	Yakutia_LN	Dzhylinda-1	Yakutia_MN	3.53
	Dzhylinda-1	Yakutia_LN	Yakutia_MN	0.447
	Yakutia_MN	Dzhylinda-1	Yakutia_LN	2.92

Table 8. Proximal admixture modeling of Yakutia_LN. The results of qpAdm modeling for Yakutia_LN are summarized. To find the best proximal source, candidate sources are systematically added to base outgroup and change of significance is observed. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

			Base outgroup			Base outgroup + Yakutia_MN			Base outgroup + Dzhylinda-1		
Target	**Ref1**	**Ref2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**
Yakutia_LN	Dzhylinda-1	EastBaikal_N	3.25E-01	0.344±0.043	0.656±0.043	3.58E-03	0.357±0.060	0.643±0.060	-	-	-
	Yakutia_MN	EastBaikal_N	8.90E-01	0.454±0.057	0.546±0.057	-	-	-	9.55E-01	0.443±0.061	0.557±0.061

Table 9. Proximal admixture modeling of WestBaikal_EN. The results of qpAdm modeling for WestBaikal_EN are summarized. To find the best proximal source, candidate sources are systematically added to base outgroup and change of significance is observed. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

			Base outgroup			Base outgroup + irk030			Base outgroup + UKY		
Target	**Ref1**	**Ref2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**	**p-value**	**Coeff1**	**Coeff2**
WestBaikal_EN	irk030	EastBaikal_N	1.26E-01	0.232±0.030	0.768±0.030	-	-	-	1.39E-01	0.236±0.029	0.764±0.029
	Dzhylinda-1	EastBaikal_N	7.61E-03	0.278±0.040	0.722±0.040	2.74E-05	0.260±0.045	0.740±0.045	8.46E-03	0.279±0.041	0.721±0.041
	UKY	EastBaikal_N	2.89E-01	0.290±0.034	0.710±0.034	4.50E-05	0.259±0.039	0.741±0.039	-	-	-
	Kolyma_M	EastBaikal_N	1.02E-01	0.249±0.034	0.751±0.034	1.65E-05	0.223±0.039	0.777±0.039	6.17E-02	0.260±0.035	0.740±0.035

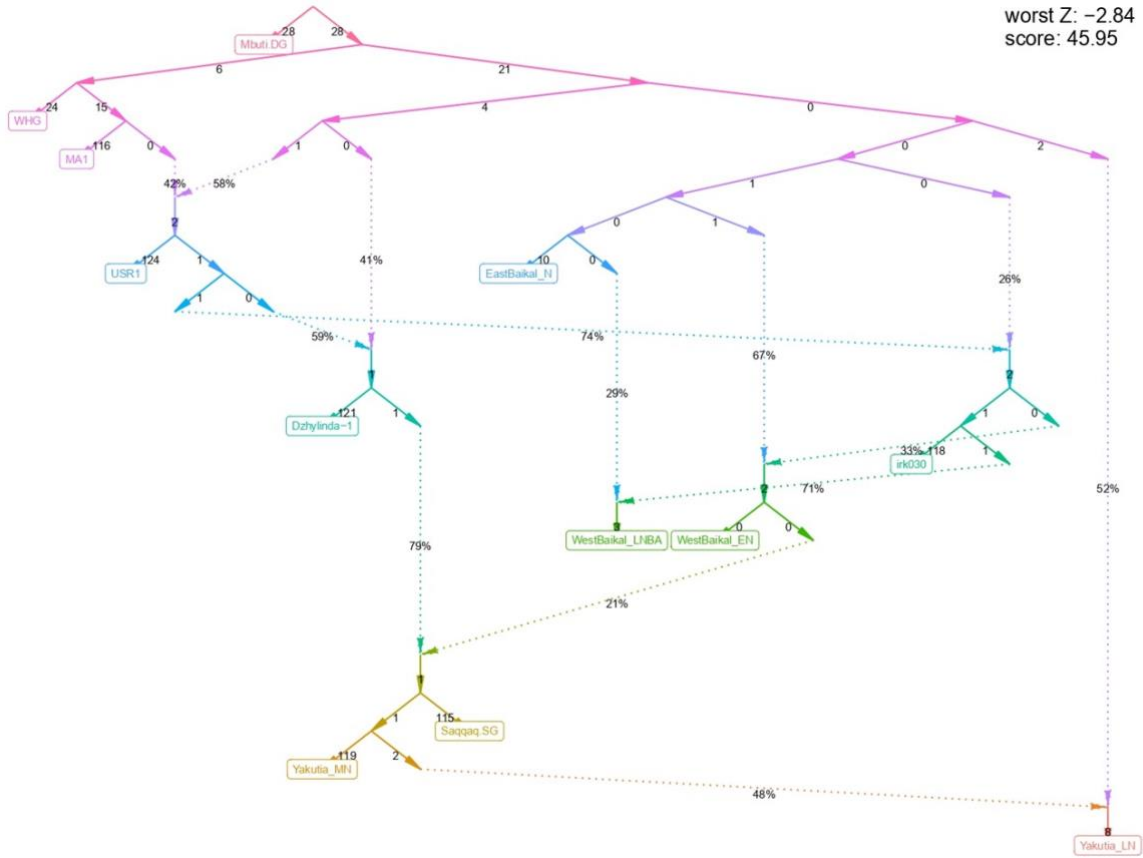

Figure 10. Manually fitted admixture graph explaining population dynamics of ancient Siberian populations. Admixture graph is based on proximal admixture modeling results and manually fitted by *qpgraph* function.

3.2.1. The origin of present-day Uralic speaking people

Based on the population dynamics of Siberia during the Middle Holocene, genetic profiles of the present-day Uralic-speaking populations, Nganasan and Selkup, were studied. According to the PCA result, Yakutia genomes were getting closer to Nganasan over time, and the Bronze Age genome from Krasnoyarsk Krai (Krasnoyarsk_BA) was also close to Nganasan (Figure 4). As expected, Nganasan showed a high genetic affinity with Yakutia_LN and Krasnoyarsk_BA in *outgroup*-F_3 results (Figure 11). To confirm the necessity of an additional admixture source, *qpWave* was implemented, and an extra gene flow was necessary for (Nganasan, Yakutia_LN) but unnecessary for (Nganasan, Krasnoyarsk_BA) (Table 10). Then, *qpWave* found Yakutia_LN and Krasnoyarsk_BA were cladal even if Nganasan was added to the outgroup, implying tree topology, (Nganasan, (Yakutia_LN, Krasnoyarsk_BA)) (Table 10). In addition, both Nganasan and Krasnoyarsk_BA were symmetric to Yakutia_MN and EastBaikal_N and adequately modeled as Yakutia_MN + EastBaikal_N (Table10, Table 11). Thus, the genetic makeup of Nganasan occurred concurrently with that of Yakutia_LN, and then Nganasan has isolated from other populations.

Another Uralic-speaking people, Selkup, also showed a high *outgroup*-F_3 value with ancient Yakutia populations (Figure 4, Figure 11). In order to model Selkup, additional admixture sources were searched by F_4 statistics of the form F_4 (Mbuti, worldwide;

Nganasan, Selkup), and ANE populations represented by AG3 and Eastern European hunter-gatherers (EHG) were necessary for Selkup (Figure 12). However, both Nganasan + ANE and Yakutia_LN + ANE were infeasible for Selkup (Table 11). Thus, Yakutia_LN was replaced with Yakutia_MN as a representative of Yakutia_LN + ANE, and then F_4 (Mbuti, worldwide; Yakutia_MN, Selkup) found an extra gene flow from Sintashta_MLBA, the Bronze Age Steppe populations (Figure 12). Then, two-way admixture models, Selkup = Yakutia_MN + Bronze Age Steppe populations, were tested, and only Yakutia_MN + Krasnoyarsk_MLBA was feasible (Table 11).

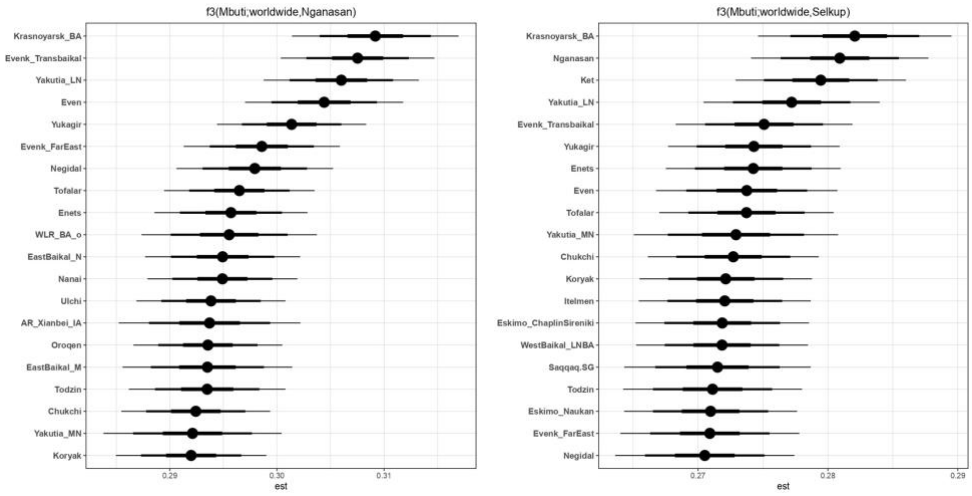

Figure 11. Genetic affinity measured by *outgroup*−F_3 statistics for present−day Uralic−speaking populations. *Outgroup*− F_3 statistics are calculated for Nganasan and Selkup population to explore genetically close populations, and 20 closest populations are shown. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

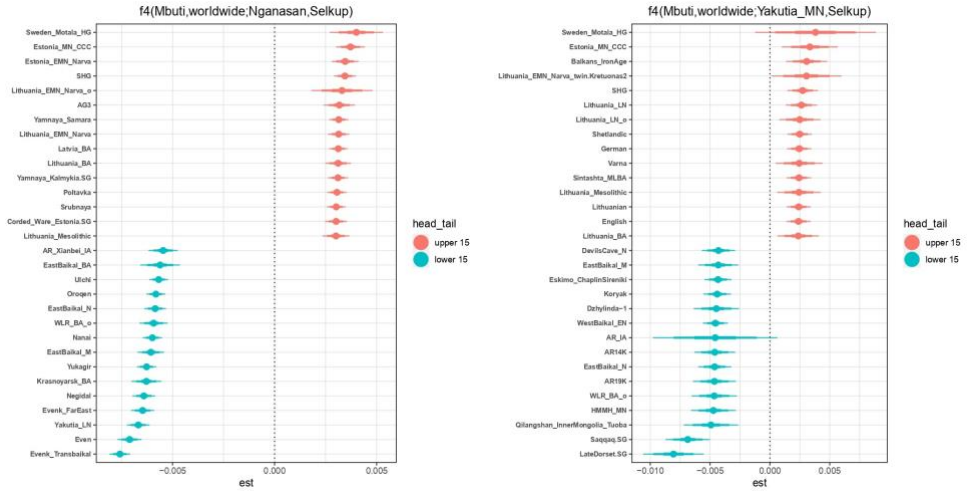

Figure 12. Genetic symmetricity test between Selkup and Yakutia lineage populations. F_4 statistics of the form F_4 (Mbuti, worldwide; Nganasan, Selkup) and F_4 (Mbuti, worldwide; Yakutia_MN, Selkup) are calculated for exploring extra admixture sources of Selkup. 15 most positive and 15 most negative Z score results are shown in red and blue, respectively. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

Table 10. Cladality tests of present-day Nganasan. The results of *qpWave* cladality tests between Nganasan and close ancient relatives, Krasnoyarsk_BA and Yakutia_LN are summarized. To find plausible topology, targets and other proximal sources are systematically added to outgroup and change of significance is observed. Models with p-value > 0.05 are highlighted.

Target 1	Target 2	outgroup	p-value
Nganasan	Yakutia_LN	base outgroup	1.38E-02
	Yakutia_LN	base outgroup + Krasnoyarsk_BA	2.93E-03
	Krasnoyarsk_BA	base outgroup	1.63E-01
	Krasnoyarsk_BA	base outgroup + Yakutia_LN	2.56E-05
	Krasnoyarsk_BA	base outgroup + Yakutia_MN	8.85E-02
	Krasnoyarsk_BA	base outgroup + EastBaikal_N	5.90E-02
Yakutia_LN	Krasnoyarsk_BA	base outgroup	9.83E-01
	Krasnoyarsk_BA	base outgroup + Nganasan	1.97E-01

Table 11. Proximal admixture modeling of present-day Uralic speaking populations. Proximal admixture models for Nganasan, Krasnoyarsk_BA and Selkup are tested by *qpAdm*. Models with p-value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

Target	Ref1	Ref2	P-value	Coeff1	Coeff2
Nganasan	Yakutia_MN	EastBaikal_N	7.80E-01	0.585±0.062	0.415±0.062
Krasnoyarsk_BA	Yakutia_MN	EastBaikal_N	7.79E-01	0.481±0.100	0.519±0.100
Selkup	Yakutia_LN	AG3	4.71E-68	0.666±0.015	0.334±0.015
	Yakutia_LN	Estonia_MN_CCC	6.99E-09	0.658±0.009	0.342±0.009
	Yakutia_LN	Sintashta_MLBA	2.72E-34	0.682±0.007	0.318±0.007
	Yakutia_LN	Krasnoyarsk_MLBA	4.26E-29	0.672±0.007	0.328±0.007
	Nganasan	AG3	8.50E-79	0.673±0.013	0.327±0.013
	Nganasan	Estonia_MN_CCC	8.29E-16	0.672±0.008	0.328±0.008
	Nganasan	Sintashta_MLBA	2.70E-51	0.700±0.007	0.300±0.007
	Nganasan	Krasnoyarsk_MLBA	1.35E-43	0.688±0.007	0.312±0.007
	Dzhylinda-1	Sintashta_MLBA	3.55E-02	0.730±0.012	0.270±0.012
	Dzhylinda-1	Krasnoyarsk_MLBA	2.82E-02	0.723±0.013	0.277±0.013
	Yakutia_MN	Sintashta_MLBA	4.16E-02	0.736±0.013	0.264±0.013
	Yakutia_MN	Krasnoyarsk_MLBA	5.10E-02	0.727±0.014	0.273±0.014
	Saqqaq	Sintashta_MLBA	7.87E-02	0.711±0.011	0.289±0.011
	Saqqaq	Krasnoyarsk_MLBA	1.62E-01	0.703±0.012	0.297±0.012

3.2.2 Siberian ancestry in north-eastern Europe originated from Neolithic Yakutia culture

The early metal Age individuals from Bolshoy Oleni Ostrov in west Russia (Russia_Bolshoy) were the oldest samples with Siberian ancestry in north-east Europe. *Outgroup*-F_3 statistics were calculated first to explore the proximal Siberian sources of Russia_Bolshoy. Russia_Bolshoy showed high genetic affinities with Bronze Age genomes from Tarim Basin mummies (Tarim_EMBA1) (Zhang et al., 2021), ancient Yakutia genomes and the Eastern European hunter-gatherers (EHG) (Figure 13). Although Russia_Bolshoy showed the highest genetic affinity with Tarim_EMBA1, they showed pretty asymmetric F_4 statistics against worldwide populations (Figure 13). Since Middle Holocene Yakutia genomes were asymmetrically close to Russia_Bolshoy than Tarim_EMBA1, Russia_Bolshoy = Yakutia lineage + Tarim_EMBA1 was tested by *qpAdm*. However, this model was insufficient to explain Russia_Bolshoy and was found to need Eastern European hunter-gatherers sources by F_4 (Mbuti, worldwide; Yakutia_MN, Russia_Bolshoy) (Table 12, Figure 13). Then, three-way admixture: Yakutia_LN + Estonia_MN_CCC + Tarim_EMBA1 was feasible for Russia_Bolshoy, and the more parsimonious admixture model Yakutia_MN + Estonia_MN_CCC was also feasible (Table 12). Although the two-way admixture model is inconsistent with the Y-haplogroup pattern, the number of Yakutia_MN samples is too small (n=1) that it is difficult to determine which of the two sources is

better.

Finally, the genetic makeup of the Russia_Bolshoy population and migration of Yakutia ancestry to the Paleo-Eskimo population were comprehensively tested by *qpgraph*. As a result, Yakutia ancestry was found to be transferred to north-eastern Europe after it diverged from the Paleo-Eskimo population (Figure 14).

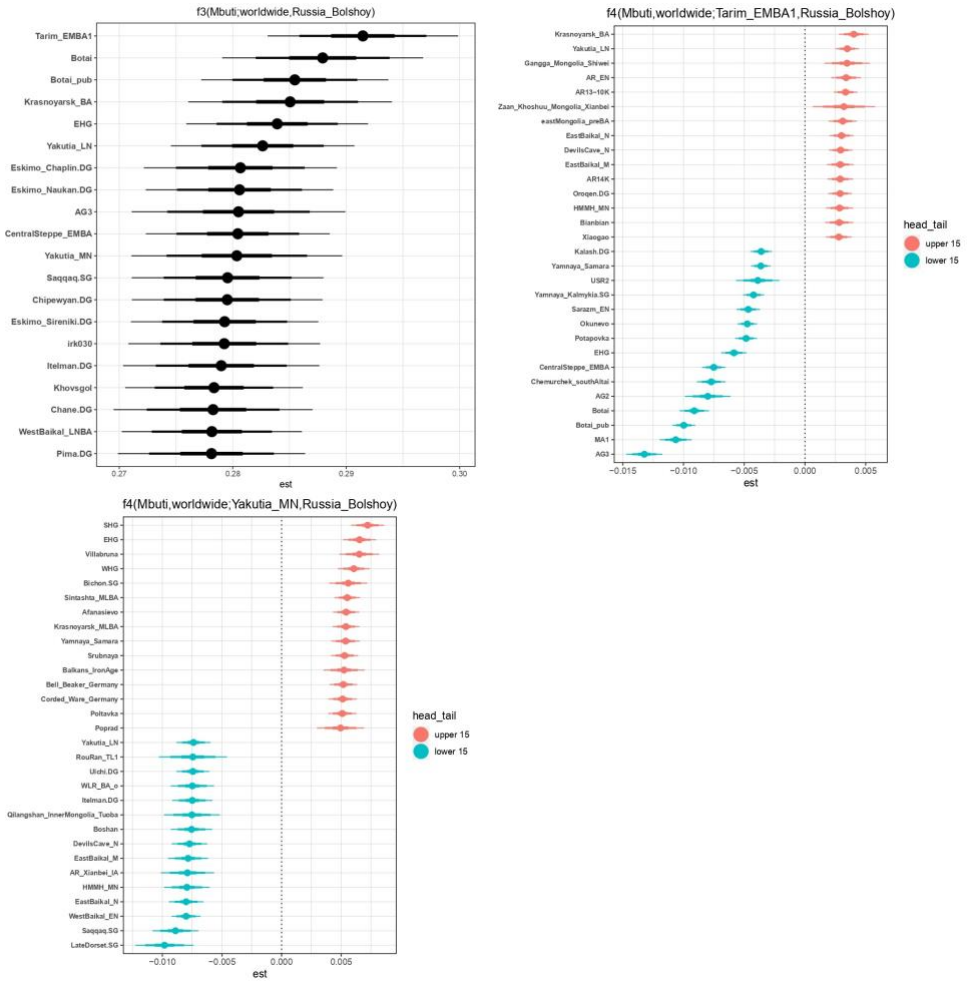

Figure 13. F−statistics results for exploring admixture model of Russia Bolshoy. *Outgroup*−F_3 statistics are calculated for Russia Bolshoy population to explore genetically close populations, and 20 closest populations are shown (Upper left). F_4 statistics of the form F_4 (Mbuti, worldwide; Tarim_EMBA1, Russia_Bolshoy) (Upper right) and F_4 (Mbuti, worldwide; Yakutia_MN, Russia_Bolshoy) (Bottom left) are calculated for exploring extra admixture sources of Russia_Bolshoy. 15 most positive and 15 most negative Z score results are shown in red and blue, respectively. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

Table 12. Proximal admixture models for Bolshoy individuals are tested by *qpAdm*. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

Target	Ref1	Ref2	Ref3	p-value	Coeff1	Coeff2	Coeff3
Bolshoy	Yakutia_LN	Tarim_EMBA1	-	1.42E-174	0.370 ±0.020	0.630 ±0.020	-
	Yakutia_LN	Tarim_EMBA1	EHG	6.17E-03	0.413 ±0.013	-0.035 ±0.027	0.623 ±0.023
	Yakutia_LN	Tarim_EMBA1	Estonia_MN_CCC	1.13E-01	0.389 ±0.014	0.158 ±0.023	0.454 ±0.018
	Yakutia_LN	AG3	-	1.43E-79	0.391 ±0.019	0.609 ±0.019	-
	Yakutia_LN	AG3	EHG	2.58E-01	0.409 ±0.012	-0.0373 ±0.031	0.628 ±0.030
	Yakutia_LN	AG3	Estonia_MN_CCC	4.43E-01	0.405 ±0.015	0.152 ±0.029	0.443 ±0.025
	Yakutia_LN	EHG	-	6.14E-03	0.404 ±0.011	0.596 ±0.011	-
	Yakutia_LN	Estonia_MN_CCC	-	1.80E-10	0.445 ±0.012	0.555 ±0.012	-
	Yakutia_MN	EHG	-	1.42E-03	0.452 ±0.016	0.548 ±0.016	-
	Yakutia_MN	Estonia_MN_CCC	-	3.83E-01	0.513 ±0.019	0.487 ±0.019	-
	Saqqaq	EHG	-	6.53E-06	0.437 ±0.014	0.563 ±0.014	-
	Saqqaq	Estonia_MN_CCC	-	9.06E-04	0.476 ±0.014	0.524 ±0.014	-

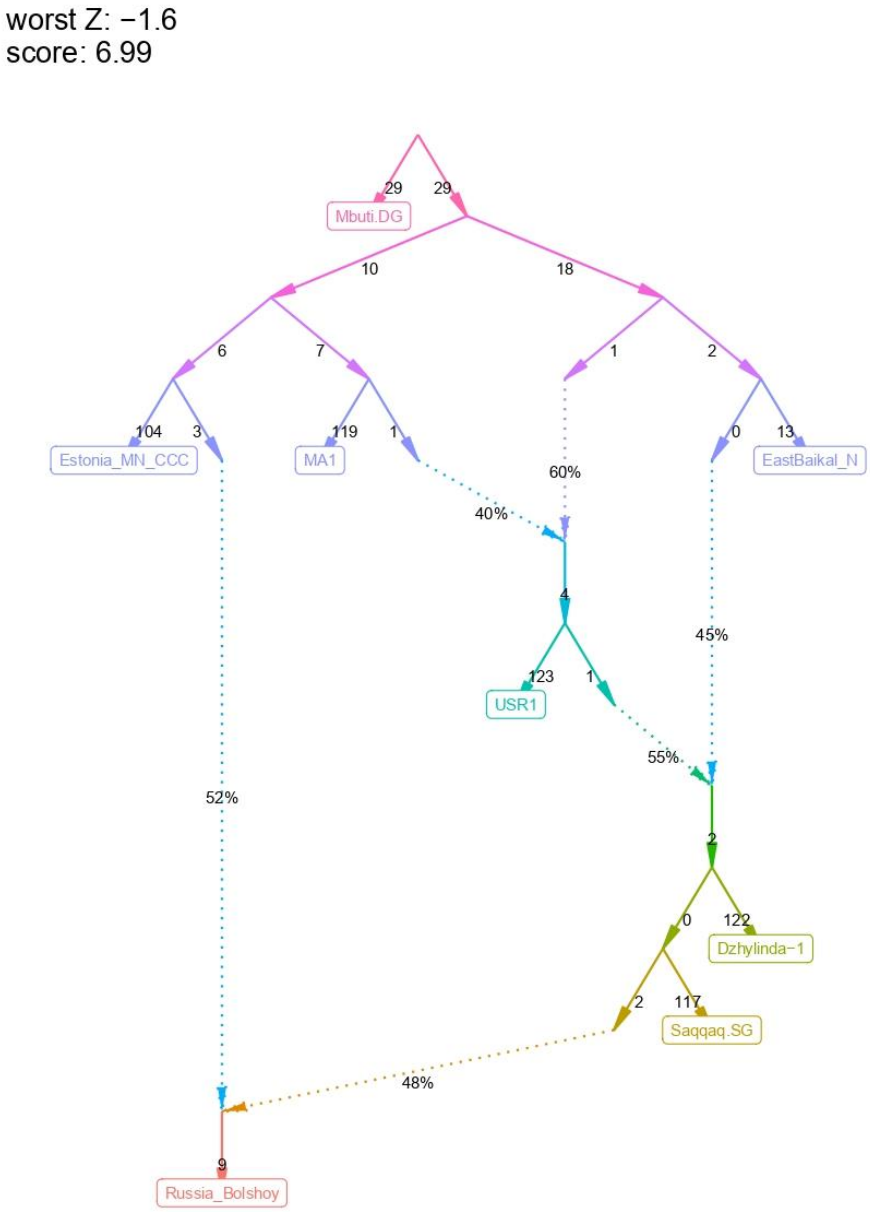

Figure 14. Manually fitted admixture graph explaining impact of ancient Yakutia population on north−eastern European. Admixture graph is manually fitted by *qpgraph* based on proximal admixture modeling.

3.2.3. Common Siberian ancestors of Dene-Yenisei speaking people originated from Yakutia

The origin of the Yenisei language was explored by reconstructing the admixture history of Ket. According to PCA and *outgroup*-F_3 results, Ket showed close relation to Selkup people (Figure 3, Figure 4, Figure 15). As expected, a two-way admixture: Yakutia_MN + Krasnoyarsk_MLBA was also feasible (Table 13).

To propose common Siberian ancestors of Yenisei-speaking people and Na-Dene-speaking people, ancient Athabaskans were modeled as Yakutia_MN + First American represented by Late Pleistocene genome from Clovis burial site (Rasmussen et al., 2014), and Eskimo Aleuts were modeled as Yakutia_MN + Athabaskan. As expected, Yakutia_MN could replace the Saqqaq ancestry in both Na-Dene-speaking ancient Athabaskan and Aleut populations (Table 13). Consequently, Ket could be explained as a two-way admixture: Saqqaq + Krasnoyarsk_MLBA (Table 13). As a result, Yenisei-speaking Ket people and Na-Dene-speaking Athabaskan share their Siberian ancestry, which Yakutia_MN represents.

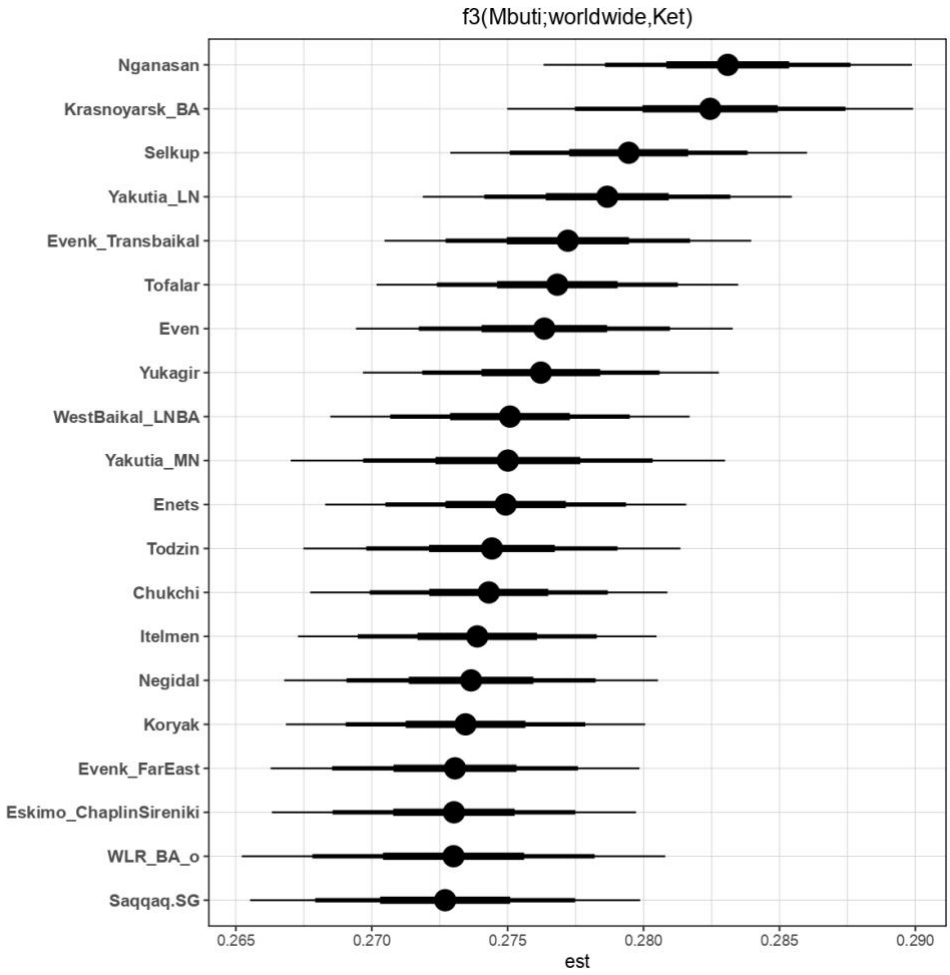

Figure 15. Genetic affinity measured by *outgroup*−F_3 statistics for present−day Yenisei−speaking populations. *Outgroup*− F_3 statistics are calculated for Ket population to explore genetically close populations, and 20 closest populations are shown. Horizontal bars represent the point estimate ±3 (thin), ±2 (intermediate), ±1 (thick) standard error, respectively. Standard errors are calculated by 5cM block jackknifing.

Table 13. Proximal admixture models for Ket, ancient Athabaskan and Eskimo Aleut populations are tested by *qpAdm*. Models with p−value > 0.05 are highlighted in grey and estimates of coefficients are represented with ± standard error.

Target	Ref1	Ref2	p-value	Coeff1	Coeff2
Ket	Yakutia_MN	Krasnoyarsk_MLBA	1.03E-01	0.771±0.014	0.229±0.014
	Saqqaq	Krasnoyarsk_MLBA	8.16E-02	0.745±0.012	0.255±0.012
Athabaskan	Yakutia_MN	Clovis	2.68E-01	0.376±0.029	0.624±0.029
	Saqqaq	Clovis	4.54E-01	0.376±0.029	0.624±0.029
Eskimo_Aleut	Yakutia_MN	Athabaskan	7.89E-01	0.147±0.040	0.853±0.040
	Saqqaq	Athabaskan	8.45E-01	0.179±0.035	0.821±0.035

4. Discussion

This study suggests a spatiotemporal demographic model of Middle Holocene Siberian populations. Before Holocene, APS ancestry occupied the Siberia region widely from Lake Baikal to the Kolyma river. Even though the previous study (K ı l ı nç et al., 2021) did not report about remained APS ancestry during Middle Holocene, two individuals, irk030 and Dzhylinda-1, supported that APS ancestry remained at least until the Neolithic period. Moreover, irk030 and Dzhylinda-1 not only support the existence of APS ancestry but also dissect the Middle Holocene Siberian populations into Yakutia and Lake Baikal lineage. Thus, the Yakutia and Lake Baikal seem to be split at least before Middle Holocene began.

After the Yakutia and Lake Baikal lineage diverged, ANA ancestors migrated from the Amur river basin, affecting both Yakutia and Lake Baikal lineage independently. As a result, the early Neolithic Kitoi culture appeared in the west Baikal region. Then, Kitoi people would have interacted with Middle Neolithic Belkachi culture in the Yakutia region and exchanged genetic materials. This result has been supported by archaeological evidence (Kuzmin & Bellwood, 2015). However, it has never been supported by genomic evidence. Thus, this result is the first genomic evidence of interaction between Kitoi culture and Belkachi culture, and the extent of the influence of Kitoi culture in north Siberian needs to be further studied.

It is also essential to confirm whether genetic materials would have flowed from Yakutia to the west Baikal region. Indeed,

WestBaikal_LNBA individuals, mostly belonging to the Serevo-Glazkovo culture, showed a similar genetic profile to the Belkachi culture individuals on the PCA. However, no significant gene flow from Yakutia to west Baikal was found in this study, and it seems that the formation of Serevo-Glazkovo culture was unrelated to Belkachi culture or Ymyiakhtakh culture. In other words, the west Baikal and Yakutia populations were split again after the Early Bronze Age. This result will help to understand the descendants of west Baikal populations in the Central Steppe or Altai Mountain (Jeong et al., 2018; Zhang et al., 2021).

Based on the high-resolution dynamics of Siberian populations, the impact of the Yakutia lineage was explored on global populations. Although it was impossible to determine which population was more appropriate for Bolshoy, at least it is clear that Siberian ancestry migrated from the Yakutia region to north-eastern Europe during the Neolithic period. Moreover, Yakutia lineage was also found in present-day Nganasan and Selkup, suggesting that the common origin of the Uralic-speaking people was from Yakutia.

Another exciting result of this study is also related to language families. The origin of the Yenisei-speaking people, a long-debated topic in linguistics, is suggested by studying Ket people. Yenisei-speaking people are found to be related to Na-Dene-speaking people through Belkachi culture and Paleo Eskimos. This result suggests that two language families originated from common Siberian ancestors, and this scenario is more parsimonious than the back migration of Paleo-Eskimo ancestry from America to Siberia. However, since Yakutia lineage also influenced Uralic-speaking people, it is still contentious what makes Yenisei and the Uralic

languages differ. This problem is interesting and vital for reconstructing the precise formation of north Eurasian and understanding the evolution of language. Therefore, to solve this problem, Siberian genomes must be paid attention to and studied further in the future.

References

Abondolo, D. (2015). *The Uralic languages*. Routledge.

Bouwman, A., & Rühli, F. (2016). Archaeogenetics in evolutionary medicine. *Journal of Molecular Medicine*, *94*(9), 971–977.

Cocks, L. R. M., & Torsvik, T. H. (2007). Siberia, the wandering northern terrane, and its changing geography through the Palaeozoic. *Earth-Science Reviews*, *82*(1–2), 29–74.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J. V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., & Baimukhanov, N. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*, *360*(6396), eaar7711.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, *28*(8), 2239–2252.

Flegontov, P., Altınışık, N. E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D. A., Broomandkhoshbacht, N., Candilio, F., & Culleton, B. J. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*, *570*(7760), 236–240.

Flegontov, P., Altınışık, N. E., Changmai, P., Vajda, E. J., Krause, J., & Schiffels, S. (2016). Na-Dene populations descend from the Paleo-Eskimo migration into America. *bioRxiv*, 074476.

Flegontov, P., Changmai, P., Zidkova, A., Logacheva, M. D., Altınışık, N. E., Flegontova, O., Gelfand, M. S., Gerasimov, E. S., Khrameeva, E. E., & Konovalova, O. P. (2016). Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Scientific reports*, *6*(1), 1–12.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., Johnson, P. L., Aximu-Petri, A., Prüfer, K., & De Filippo, C. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, *514*(7523), 445–449.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D.,

Furtwängler, A., Haak, W., Meyer, M., & Mittnik, A. (2016). The genetic history of ice age Europe. *Nature*, *534*(7606), 200–205.

Georg, S. (2008). Yeniseic languages and the Siberian linguistic area. In *Evidence and Counter-Evidence: Essays in Honour of Frederik Kortlandt, Volume 2* (pp. 151–168). Brill.

Harney, E., Patterson, N., Reich, D., & Wakeley, J. (2021). Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics*, *217*(4), iyaa045.

Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., & Khussainova, E. (2019). The genetic history of admixture across inner Eurasia. *Nature ecology & evolution*, *3*(6), 966–976.

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A. S., Taylor, W. T. T., Hagan, R. W., Bromage, S., Tsolmon, S., Trachsel, C., & Grossmann, J. (2018). Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proceedings of the National Academy of Sciences*, *115*(48), E11248–E11255.

Kennett, D. J., Plog, S., George, R. J., Culleton, B. J., Watson, A. S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., & Kistler, L. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nature communications*, *8*(1), 1–9.

Khabtagaeva, B. (2019). *Language Contact in Siberia: Turkic, Mongolic, and Tungusic Loanwords in Yeniseian*. Brill.

Kılınç, G. M., Kashuba, N., Koptekin, D., Bergfeldt, N., Dönertaş, H. M., Rodríguez-Varela, R., Shergin, D., Ivanov, G., Kichigin, D., & Pestereva, K. (2021). Human population dynamics and Yersinia pestis in ancient northeast Asia. *Science advances*, *7*(2), eabc4587.

Kim, T., Woo, E. J., & Pak, S. (2018). Principles of Archaeogenetics and the Current Trends of Ancient Genome Studies. *Korean Journal of Physical Anthropology*, *31*(4), 105–119.

Krupnik, I. (2000). Reindeer pastoralism in modern Siberia: research and survival during the time of crash. *Polar Research*, *19*(1), 49–56.

Kuzmin, Y. V., & Bellwood, P. (2015). Northern and North-eastern Asia: archaeology. *The global prehistory of human migration*, 191–196.

Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., & Weihmann,

A. (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature communications*, *9*(1), 1–12.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., & Sirak, K. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature*, *536*(7617), 419–424.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., & Tandon, A. (2016). The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206.

Massilani, D., Skov, L., Hajdinjak, M., Gunchinsuren, B., Tseveendorj, D., Yi, S., Lee, J., Nagel, S., Nickel, B., & Devièse, T. (2020). Denisovan ancestry and population history of early East Asians. *Science*, *370*(6516), 579–583.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., & Cheronet, O. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., & Novak, M. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503.

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J. A., Albrechtsen, A., Malaspinas, A.-S., & Sikora, M. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*, *553*(7687), 203–207.

Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., & Lipson, M. (2019). The formation of human populations in South and Central Asia. *Science*, *365*(6457), eaat7487.

Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., & Hudson, M. J. (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nature communications*, *11*(1), 1–9.

Pääbo, S., Higuchi, R. G., & Wilson, A. C. (1989). Ancient DNA and the polymerase chain reaction: the emerging field of molecular archaeology (Minireview). *The Journal of biological chemistry*, *264*(17), 9709–9712.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, *2*(12), e190.

Peter, B. M. (2016). Admixture, population structure, and F-statistics. *Genetics*, *202*(4), 1485–1501.

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1–1.

Pinotti, T., Bergström, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., Lacerda, D. R., Solli, A., Norstedt, J., & Reed, K. (2019). Y chromosome sequences reveal a short Beringian Standstill, rapid expansion, and early population structure of Native American founders. *Current Biology*, *29*(1), 149–157. e143.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford Jr, T. W., Orlando, L., & Metspalu, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91.

Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., Rasmussen, S., Moltke, I., Albrechtsen, A., & Doyle, S. M. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*, *506*(7487), 225–229.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., & Mesa, N. (2012). Reconstructing native American population history. *Nature*, *488*(7411), 370–374.

Saag, L., Varul, L., Scheib, C. L., Stenderup, J., Allentoft, M. E., Saag, L., Pagani, L., Reidla, M., Tambets, K., & Metspalu, E. (2017). Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Current Biology*, *27*(14), 2185–2193. e2186.

Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., & Renaud, G. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature*, *570*(7760), 182–188.

Siska, V., Jones, E. R., Jeon, S., Bhak, Y., Kim, H.-M., Cho, Y. S., Kim, H., Lee, K., Veselovskaya, E., & Balueva, T. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago.

Science advances, *3*(2), e1601877.

Slobodin, S. B. (2019). Neolithic of the Northeast Asia and the Arctic small tool tradition of the North America.

Stepanov, A., Kuzmin, Y. V., Hodgins, G., & Jull, A. (2012). Kyordyughen site, Ymyiakhtakh culture, Yakutia: An interpretation of burial rite. *Archaeology, Ethnology and Anthropology of Eurasia*, *40*(4), 51–61.

Vajda, E. (2019). Yeniseian and Dene hydronyms.

Vajda, E. J. (2013). *Yeniseian peoples and languages: a history of Yeniseian studies with an annotated bibliography and a source guide*. Routledge.

Vovin, A. (2000). Did the Xiong-nu speak a Yeniseian language? *Central Asiatic Journal*, *44*(1), 87–104.

Weber, A. W., & Bettinger, R. (2010). Middle Holocene hunter-gatherers of Cis-Baikal, Siberia: an overview for the new century. *Journal of Anthropological Archaeology*, *29*(4), 491–506.

Weber, A. W., Link, D. W., & Katzenberg, M. A. (2002). Hunter-gatherer culture change and continuity in the Middle Holocene of the Cis-Baikal, Siberia. *Journal of Anthropological Archaeology*, *21*(2), 230–299.

Yu, H., Spyrou, M. A., Karapetian, M., Shnaider, S., Radzevičiūtė, R., Nägele, K., Neumann, G. U., Penske, S., Zech, J., & Lucas, M. (2020). Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. *Cell*, *181*(6), 1232–1245. e1220.

Zhang, F., Ning, C., Scott, A., Fu, Q., Bjørn, R., Li, W., Wei, D., Wang, W., Fan, L., & Abuduresule, I. (2021). The genomic origins of the Bronze Age Tarim Basin mummies. *Nature*, *599*(7884), 256–261.

국문 초록

시베리아는 우랄산맥에서부터 태평양까지 이어지는 넓은 북아시아를 차지하는 지역으로 툰드라와 타이가 삼림지대로 구성되어 있다. 시베리아는 비록 적은 밀도의 사람들이 살았던 지역이지만, 그 역사는 굉장히 복잡하고 흥미롭다. 그 이유는 여러 차례에 걸친 이주와 혼합이 시베리아에서 있었기 때문이며 고유전체 정보가 이 사건들을 밝히는데 기여해왔다. 실제로 시베리아에서는 동유라시아인과 서유라시아인 사이의 혼합이 빈번히 일어났으며, 아메리카 원주민의 조상 집단이 형성되어 아메리카 대륙으로 이주하기도 하였다. 그러나, 최근 몇 년간 보고된 시베리아의 고유전체들이 아메리카 원주민과의 관계에 집중되어 분석된 나머지 고대 시베리아인 집단들 사이의 포괄적인 관계에 대한 이해는 현재 부족한 상황이다. 따라서 본 연구에서는 이를 해결하기 위해 출판된 고유전체를 종합해 재분석하여, 고대 시베리아 집단의 자세한 유전적 역사를 재구성하고 다른 집단에 끼친 영향을 연구하였다. 그 결과, 시베리아 집단은 중기 홀로세가 시작되기 전부터 야쿠티아와 바이칼 호, 두 개 지역의 고유 집단으로 분기되어 있었다. 그리고 신석기 시대에 두 집단 사이의 유전적 흐름이 있었고 이는 각 지역에서의 유전적 프로필 및 문화적 변화와 일치하였다. 다음으로 야쿠티아 지역 집단이 다른 지역 집단에 끼친 영향을 연구하였고, 이들은 굉장히 활발하게 이주하여 응가나산, 셀쿱 사람과 더불어 북동유럽인과 같은 현대 우랄어족(語族)의 유전자 풀 형성에 큰 영향을 준 것으로 밝혀졌다. 또한, 야쿠티아 집단은 예니세이어를 사용하는 시베리아의 케트 사람과 나데네어를 사용하는 북아메리카의 애서베스카 사람들에게도 영향을 준 것으로 밝혀져 예니세이어족과 나데네어족이 시베리아계 조상을 공유함을 시사하였다. 이 결과는 언어적 유사성과 유전적 유사성 사이의 상관관계를 연구할 수 있는 중요한 연구사례가 될

것이며 북유라시아 집단의 형성과 아메리카 대륙으로의 이주를 더 구체적으로 재구성하는데 중요한 정보를 제공할 것이다.

주요어 : 시베리아, 중기 홀로세, 고고유전체학, 인류 집단유전학, 우랄어, 데네-예니세이어, 야쿠티아
학번 : 2021-23420

Appendix 1. A list of Middle Holocene Siberian genomes used in this study. This list includes 92 Middle Holocene Siberian genomes which is extracted from AADR dataset and used in this study. 1st degree relatives and duplicates were excluded in further analysis and PCA outliers were also excluded in further analysis.

Genetic ID	Publication	Date	Group ID	Sex	Y hg	remark
UKY001	YuCell2020	13929	UKY	M	Q1b1	
Kolyma_River_noUDG.SG	SikoraNature2019	9775	Kolyma_M	M	Q1a1	
irk051_noUDG.SG	KilincSciAdv2021	7922	Baikal_EN	F	n/a	
irk00x_noUDG.SG	KilincSciAdv2021	8419	Dzhylinda-1	M	C2a1a	
cta016_noUDG.SG	KilincSciAdv2021	8264	EastBaikal_M	F	n/a	
irk007_noUDG.SG	KilincSciAdv2021	8801	WestBaikal_EN	M	N1a2	
DA357_noUDG.SG	DamgaardScience2018	7785	WestBaikal_EN	M	C2a1a	
DA341_noUDG.SG	DamgaardScience2018	7577	WestBaikal_EN	F	n/a	
DA359_noUDG.SG	DamgaardScience2018	7545	WestBaikal_EN	M	N1a2	
DA340_noUDG.SG	DamgaardScience2018	7023	WestBaikal_EN	M	n/a	
DA245_noUDG.SG	DamgaardScience2018	7944	WestBaikal_EN	M	N1a2	
DA249_noUDG.SG	DamgaardScience2018	7840	WestBaikal_EN	F	n/a	
DA246_noUDG.SG	DamgaardScience2018	7723	WestBaikal_EN	F	n/a	
DA247_noUDG.SG	DamgaardScience2018	7689	WestBaikal_EN	M	N	
brn008_noUDG.SG	KilincSciAdv2021	7379	EastBaikal_N	M	N1a1a1~	
DA248_noUDG.SG	DamgaardScience2018	7646	WestBaikal_EN	M	N1a2	
brn001_noUDG.SG	KilincSciAdv2021	7334	EastBaikal_N	F	n/a	
DA250_noUDG.SG	DamgaardScience2018	7377	WestBaikal_EN	M	N1a2	
DA252_noUDG.SG	DamgaardScience2018	7317	WestBaikal_EN	F	n/a	
DA251_noUDG.SG	DamgaardScience2018	7299	WestBaikal_EN	M	N1a2	
DA253_noUDG.SG	DamgaardScience2018	7240	WestBaikal_EN	F	n/a	
DA362_noUDG.SG	DamgaardScience2018	7230	WestBaikal_EN	M	N1a2	
ANG001	YuCell2020	7274	WestBaikal_EN_o	M	C2a	PCA outlier
STB002	YuCell2020	6587	WestBaikal_EN	F	n/a	
IUO001	YuCell2020	6905	WestBaikal_EN	M	C2a1a1	PCA outlier
KAG001	YuCell2020	6856	WestBaikal_EN	M	C2a1a1	PCA outlier

Appendix 1. Continued

Genetic ID	Publication	Date	Group ID	Sex	Y hg	remark
NEO229_noUDG.SG	SikoraNature2019	6577	WestBaikal_EN	F	n/a	I7336 3rd degree relatives or further
brn003_noUDG.SG	KilincSciAdv2021	6550	EastBaikal_N	M	N	
N5a.SG	KilincSciAdv2021	6232	Yakutia_MN	M	n/a	
brn012_noUDG.SG	KilincSciAdv2021	6115	EastBaikal_N	F	n/a	brn002.SG 2nd or 3rd deg
brn002_noUDG.SG	KilincSciAdv2021	6061	EastBaikal_N	M	N1a2	brn012.SG 2nd or 3rd deg
irk030_noUDG.SG	KilincSciAdv2021	5969	irk030	M	Q1b1a3	
irk036_noUDG.SG	KilincSciAdv2021	4720	WestBaikal_LNBA	M	Q1b1b~	
irk057_noUDG.SG	KilincSciAdv2021	4380	WestBaikal_LNBA	M	Q1b	
irk025_noUDG.SG	KilincSciAdv2021	4350	WestBaikal_LNBA	M	Q1b1b~	irk022 3rd to 5th deg
irk050_noUDG.SG	KilincSciAdv2021	4350	WestBaikal_LNBA	F	n/a	
irk022_noUDG.SG	KilincSciAdv2021	4251	WestBaikal_LNBA	M	Q1b1a3	irk025 3rd to 5th deg
irk040.SG	KilincSciAdv2021	5559	WestBaikal_LNBA	M	Q1b1b~	
irk034_noUDG.SG	KilincSciAdv2021	5525	WestBaikal_LNBA	M	n/a	
irk033_noUDG.SG	KilincSciAdv2021	4858	WestBaikal_LNBA	M	Q1b1b~	
DA358_noUDG.SG	DamgaardScience2018	4713	WestBaikal_LNBA	F	n/a	
DA360_noUDG.SG	DamgaardScience2018	4704	WestBaikal_LNBA	F	n/a	
DA354_noUDG.SG	DamgaardScience2018	4591	WestBaikal_LNBA	M	Q1b1b~	
irk075_noUDG.SG	KilincSciAdv2021	5397	WestBaikal_LNBA	M	Q1b1b~	
irk071_noUDG.SG	KilincSciAdv2021	5132	WestBaikal_LNBA	M	Q1b1b~	
irk068_noUDG.SG	KilincSciAdv2021	4468	WestBaikal_LNBA	M	Q1b1b~	
irk061_noUDG.SG	KilincSciAdv2021	4341	WestBaikal_LNBA	M	Q1b1b~	
irk008_noUDG.SG	KilincSciAdv2021	3787	WestBaikal_LNBA	F	n/a	
BZK002	YuCell2020	4709	WestBaikal_LNBA_o	M	Q1b1a3	
irk017_noUDG.SG	KilincSciAdv2021	5132	WestBaikal_LNBA	F	n/a	
mak026_noUDG.SG	KilincSciAdv2021	4759	WestBaikal_LNBA	F	n/a	
DA337_noUDG.SG	DamgaardScience2018	4328	WestBaikal_LNBA	M	Q1b1b~	
DA335_noUDG.SG	DamgaardScience2018	4200	WestBaikal_LNBA	M	Q1b1b~	DA334 2nd or 3rd deg
DA336_noUDG.SG	DamgaardScience2018	4200	WestBaikal_LNBA	M	Q1b1b~	DA338 siblings (brothers)
GLZ001	YuCell2020	4556	WestBaikal_LNBA_o	M	C2a1a	
N4a1_noUDG.SG	KilincSciAdv2021	4505	Yakutia_LN	M	N1a1a1a	
DA334_noUDG.SG	DamgaardScience2018	4129	WestBaikal_LNBA	M	Q1b1b~	DA335 2nd or 3rd deg
DA339_noUDG.SG	DamgaardScience2018	4040	WestBaikal_LNBA	M	Q1b1a3	
GLZ002	YuCell2020	4430	WestBaikal_LNBA_o	M	C2a1	

Appendix 1. Continued

Genetic ID	Publication	Date	Group ID	Sex	Y hg	remark
GLZ003	YuCell2020	4478	WestBaikal_LNBA	M	Q1b1	
STB001	YuCell2020	4604	WestBaikal_LNBA	M	Q1b1	
ZPL002	YuCell2020	4168	WestBaikal_LNBA	M	Q1b1	
ZPL001	YuCell2020	3951	WestBaikal_LNBA	F	n/a (female)	
KPT003	YuCell2020	3945	WestBaikal_LNBA	M	Q1b1	
KPT002_petrous	YuCell2020	3931	WestBaikal_LNBA	M	Q1b1b~	
KPT004	YuCell2020	3773	WestBaikal_LNBA	M	Q1b1	
KPT001	YuCell2020	3759	WestBaikal_LNBA	F	n/a	
KAG002	YuCell2020	3748	WestBaikal_LNBA	M	Q1b1	
KPT006	YuCell2020	3614	WestBaikal_LNBA	F	n/a	
DA338_noUDG.SG	DamgaardScience2018	4200	WestBaikal_LNBA_rel	M	Q1b1b~	DA338 siblings (brothers)
N4b2_noUDG.SG	KilincSciAdv2021	4200	Yakutia_LN	M	n/a	
kra001_noUDG.SG	KilincSciAdv2021	4170	Krasnoyarsk_BA	M	N1a1a1a1a	
NEO232_noUDG.SG	SikoraNature2019	4812	WestBaikal_LNBA	M	Q1b1	I7780 3rd deg or further
NEO231_noUDG.SG	SikoraNature2019	4665	WestBaikal_LNBA	F	n/a	
NEO298_noUDG.SG	SikoraNature2019	4630	WestBaikal_LNBA	F	n/a	
irk076_noUDG.SG	KilincSciAdv2021	4104	WestBaikal_LNBA_rel	M	Q1b	DA336 1st deg
NEO230_noUDG.SG	SikoraNature2019	4875	WestBaikal_LNBA	F	n/a	
KPT005	YuCell2020	4034	WestBaikal_LNBA_o	M	Q1b1	
DA343_noUDG.SG	DamgaardNature2018	4819	WestBaikal_LNBA	U	n/a	
DA353_noUDG.SG	DamgaardNature2018	4285	WestBaikal_LNBA	U	n/a	
KPT002_tooth	YuCell2020	3931	WestBaikal_LNBA_dup	M	Q1b1b~	KPT002_petrous dup
DA356_noUDG.SG	DamgaardNature2018	4277	WestBaikal_LNBA	U	n/a	
yak022_noUDG.SG	KilincSciAdv2021	3799	Yakutia_LN	F	n/a	yak023 2nd or 3rd deg
DA361_noUDG.SG	DamgaardNature2018	4130	WestBaikal_LNBA	U	n/a	
DA342_noUDG.SG	DamgaardScience2018	5617	WestBaikal_LNBA	F	n/a	
DA344_noUDG.SG	DamgaardScience2018	5595	WestBaikal_LNBA	M	n/a	
DA355_noUDG.SG	DamgaardScience2018	5469	WestBaikal_LNBA	M	Q1b1b~	
yak023_noUDG.SG	KilincSciAdv2021	3735	Yakutia_LN_rel	F	n/a	yak024.SG 1st deg
DA345_noUDG.SG	DamgaardScience2018	5454	WestBaikal_LNBA	M	N	
yak024_noUDG.SG	KilincSciAdv2021	3500	Yakutia_LN	F	n/a	yak023.SG 1st deg; yak021 1st deg
yak021_noUDG.SG	KilincSciAdv2021	3225	Yakutia_LN_rel	F	n/a	yak024 1st deg
irk078_noUDG.SG	KilincSciAdv2021	3107	EastBaikal_BA	F	n/a	

Appendix 2. A list of worldwide populations used in this study. This table summarizes for which analysis each group was used. "N" column represents the total number of individuals.

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Afanasievo	23	O	O	O		O			Narasimhan et al. (2019) Science	I6711, I6713, I6715, I2069, I2071, I3388, I3950, I3954, I1829, I3952, I5269, I5270, I5271, I5272, I5273, I5277, I5278, I5279, I3387, I10565, I10564, I11112, I11752
Afanasievo_Mongolia	2	O	O	O		O			Jeong et al. (2020) Cell	SHT001, SHT002
AG2	1	O	O	O		O			Raghavan et al. (2014) Nature	AfontovaGora2_noUDG.SG
AG3	1	O	O	O	O	O	O		Fu et al. (2016) Nature	AfontovaGora3_noUDG_d
Alalakh_MLBA	25	O	O	O		O			Skourtanioti et al. (2020) Cell	ALA001_d, ALA002, ALA004, ALA008, ALA011, ALA013, ALA014_d, ALA015, ALA016, ALA017, ALA018, ALA020, ALA023, ALA024, ALA025, ALA026, ALA028, ALA029, ALA030, ALA034, ALA035, ALA037, ALA039, ALA084, ALA095
Alan	5	O	O	O		O			Damgaard et al. (2018) Nature	DA146_noUDG.SG,DA160_noUDG.SG,DA162_noUDG.SG,DA164_noUDG.SG,DA243_noUDG.SG
Altai_MLBA	7	O	O	O		O			Jeong et al. (2020) Cell	BIL001, ULI001,ULI002,ULI,003, BER002,SBG001, ULZ001
Anatolia_BA	3	O	O	O		O			Lazaridis et al. (2017) Nature	I2683, I2495, I2499
Anatolia_ChL	1	O	O	O		O			Lazaridis et al. (2016) Nature	I1584
Anatolia_N	24	O	O	O		O			Mathieson et al. (2015) Nature	I0726,I1099,I1100,I1102,I1103,I0707,I0708,I0709,I0736,I0744,I0745,I0746,I1096,I1097,I1098,I1101,I0723,I0724,I0727, I1579,I1580,I1581,I1583,I1585
AR_EN	2	O	O	O	O	O	O		Ning et al. (2020) Nat. Commun.	WQM4_noUDG, ZLNR-2_noUDG
AR_IA	1	O	O	O		O			Ning et al. (2020) Nat. Commun.	ZLNR-1_noUDG
AR_Xianbei_IA	2	O	O	O		O			Ning et al. (2020) Nat. Commun.	MGS-M6_noUDG,MGS-M7R_noUDG
Armenia_ChL	5	O	O	O		O			Lazaridis et al. (2016) Nature	I1634, I1632, I1631, I1409, I1407

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Armenia_EBA	3	O	O	O		O			Lazaridis et al. (2016) Nature	I1635, I1633, I1658
Armenia_MBA	1	O	O	O		O			Lazaridis et al. (2016) Nature	I1656
Arslantepe_LC	17	O	O	O		O			Skourtanioti et al. (2020) Cell	ART004,ART005,ART012,ART014,ART015,ART017,ART018,ART019,ART020_d,ART022,ART023,ART026,ART027_d,ART032,ART038,ART039,ART042
Balkans_BronzeAge	6	O	O	O		O			Mathieson et al. (2018) Nature	I2163, Bul10, Bul6,Bul8, I2175,I2520
Balkans_Chalcolithic	12	O	O	O		O			Mathieson et al. (2018) Nature	I0781,I0785,I2423,I2424,I2425,I2427,I2430,I2431,I2509,I2519,I2526,I2181,I2426
Balkans_IronAge	1	O	O	O		O			Mathieson et al. (2018) Nature	I5769
Ban_Chiang	3	O	O	O		O			Lipson et al. (2018) Science	I8970, I4458,I8974
Bell_Beaker_Germany	41	O	O	O		O			Mathieson et al. (2015) Nature	I0108,I0111,I0112,I0113,I0806,I1546,I1549,I3588,I3589,I3590,I3594,I3600,I3601,I4134,I4143,I4144,I5014,I5023,I5523,I5524,I5525,I5527,I5529,I5531,I5658,I5659,I5833,I5834,I5836,I6482,I6590,I6591,I6624,I0060,I3602,I3604,I4249,I4250,I5655,I5661,I5663
Bianbian	1	O	O	O		O			Yang et al. (2020) Science	Bianbian_noUDG
Boncuklu	5	O	O	O		O			Feldman et al. (2019) Nat. Commun.	ZHAG_BON004.A0101_Luk10,ZHAJ_BON034.A0101_Luk9,ZHJ_BON024.A0101_Luk84,ZKO_BON001.A0101_Luk7,ZMOJ_BON014.A0101_Luk21
Boncuklu.SG	4	O	O	O		O			Kılınç et al. (2016) Curr. Biol.	Bon001_noUDG.SG,Bon002_noUDG.SG,Bon004_noUDG.SG
Boshan	1	O	O	O		O			Yang et al. (2020) Science	BS
Botai_pub	3	O	O	O		O			Damgaard et al. (2018) Science	BOT14_noUDG.SG,BOT15_noUDG.SG,BOT2016_noUDG.SG
Bustan_BA	10	O	O	O		O			Narasimhan et al. (2019) Science	I11025,I11026,I11027,I11519,I4156,I4157,I4159,I4899,I5604,I5605
Bustan_EN	1	O	O	O		O			Narasimhan et al. (2019) Science	I11028
BZK002	1	O	O	O		O			Yu et al. (2020) Science	BZK002
CamlibelTarlasi_LC	9	O	O	O		O			Skourtanioti et al. (2020) Cell	CBT001,CBT003,CBT004,CBT005,CBT010,CBT011,CBT014,CBT015,CBT016
CentralSteppe_EMBA	2	O	O	O		O			Damgaard et al. (2018) Science	EBA1_noUDG.SG,EBA2_noUDG.SG

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Chandman_IA	9	O	O	O		O			Jeong et al. (2020) Cell	CHN001,CHN003,CHN004,CHN006,CHN007,CHN008,CHN012,CHN015,CHN016
Chemurchek_northAlai	2	O	O	O		O			Jeong et al. (2020) Cell	KUR001, KUM001
Chemurchek_southAltai	2	O	O	O		O			Jeong et al. (2020) Cell	YAG001, IAG001
CHG	2	O	O	O		O			Jones et al. (2015) Nat. Commun.	KK1_noUDG.SG, SATP_noUDG.SG
Chokhopani.SG	1	O	O	O		O			Jeong et al. (2016) Proc. Natl. Acad. Sci. USA	C1_noUDG.SG
Chuanyun	1	O	O	O		O			Yang et al. (2020) Science	L5694_noUDG
WestBaikal_EN	22	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv. Damgaard et al. (2018) Science Yu et al. (2020) Science Sikora et al. (2019) Nature	irk051_noUDG.SG, irk007_noUDG.SG, DA340_noUDG.SG, DA341_noUDG.SG, DA357_noUDG.SG, DA359_noUDG.SG, DA245_noUDG.SG, DA246_noUDG.SG, DA247_noUDG.SG, DA248_noUDG.SG, DA249_noUDG.SG, DA250_noUDG.SG, DA251_noUDG.SG, DA252_noUDG.SG, DA253_noUDG.SG, DA362_noUDG.SG, ANG001, IUO001, KAG001, NEO229_noUDG.SG, STB002
WestBaikal_LNBA	45	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv. Damgaard et al. (2018) Science Yu et al. (2020) Science Sikora et al. (2019) Nature	irk022_noUDG.SG,irk025_noUDG.SG,irk036_noUDG.SG,irk050_noUDG.SG,irk057_noUDG.SG, irk033_noUDG.SG,irk034_noUDG.SG,irk040.SG, DA358_noUDG.SG,DA360_noUDG.SG, DA354_noUDG.SG, irk075_noUDG.SG,irk071_noUDG.SG, irk068_noUDG.SG,irk061_noUDG.SG,irk008_noUDG.SG, irk017_noUDG.SG,mak026_noUDG.SG, DA337_noUDG.SG,DA335_noUDG.SG,DA336_noUDG.SG,DA334_noUDG.SG,DA339_noUDG.SG, GLZ003, KAG002,KPT001,KPT002_petrous,KPT003,KPT004,KPT006,STB001,ZPL001,ZPL002,KPT002_tooth,KPT005, NEO232_noUDG.SG,NEO231_noUDG.SG, NEO298_noUDG.SG, NEO230_noUDG.SG, DA343_noUDG.SG,DA353_noUDG.SG,DA356_noUDG.SG,DA361_noUDG.SG, DA342_noUDG.SG,DA344_noUDG.SG,DA345_noUDG.SG,DA355_noUDG.SG
Clovis.SG	1	O	O	O		O	O		Rasmussen et al. (2014) Nature	Anzick_realigned_noUDG.SG M

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Corded_Ware_Germany	10	O	O	O		O			Mathieson et al. (2015) Nature	I0049,I0103,I0104,I0106,I1532,I1536,I1538,I1539,I1542,I1544
DevilsCave_N	4	O	O	O	O	O	O		Sikora et al. (2019) Nature	NEO236_noUDG.SG,NEO238_noUDG.SG,NEO239_noUDG.SG,NEO240_noUDG.SG
Dzharkutan1_BA	6	O	O	O		O			Narasimhan et al. (2019) Science	I4161,I4163,I4312,I4313,I4315,I7411
Dzhylinda-1	1	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv.	irk00x_noUDG.SG
eastMongolia_preBA	1	O	O	O		O			Jeong et al. (2020) Cell	SOU001
Ebla_EMBA	11	O	O	O		O			Skourtanioti et al. (2020) Cell	ETM001,ETM004,ETM005,ETM006,ETM010,ETM012,ETM014,ETM016,ETM018,ETM023,ETM026
EHG	4	O	O	O		O	O		Damgaard et al. (2018) Science Mathieson et al. (2015) Nature	Sidelkino_noUDG.SG, I0211,I0061, I0124
ElMiron	1	O	O	O		O			Fu et al. (2016) Nature	ElMiron_d
Estonia_MN_CCC	3	O	O	O	O	O	O	O	Saag et al. (2017) CurrentBiology Mittnik et al. (2018)NatureCommunications	MA974_noUDG.SG, Tamula3, Tamula1
Fofonovo_EN	4	O	O	O		O			Jeong et al. (2020) Cell	FNO001,FNO003,FNO006,FNO007
Geoksyur_EN	21	O	O	O		O			Narasimhan et al. (2019) Science	I12478,I12479,I12480,I12481,I12482,I12483,I12485,I12486,I12487,I8503,I8504,I8505,I8510,I8524,I8526,I8527,I8528,I8529,I8530,I8532,I12484
Globular_Amphora	9	O	O	O		O			Mathieson et al. (2018) Nature	I2405, I2433, I2434, I2440, I2441, I2403, ILK001, ILK002, ILK003
Gonur1_BA	19	O	O	O		O			Narasimhan et al. (2019) Science	I10410,I10411,I11042,I1781,I1782,I1784,I1787,I1788,I1790,I1793,I2085,I2087,I2116,I2121,I2125,I2128,I3374,I6119,I6217
Hajji_Firuz_C	5	O	O	O		O			Narasimhan et al. (2019) Science	I2323,I2328,I4241,I4349_noUDG,I4351_noUDG
HMMH_MN	1	O	O	O		O			Ning et al. (2020) Nat. Commun.	HMF32_noUDG
Hun_TianShan_1700BP	16	O	O	O		O			Damgaard et al. (2018) Nature	DA100_noUDG.SG,DA101_noUDG.SG,DA104_noUDG.SG,DA385_noUDG.SG,DA52_noUDG.SG,DA54_noUDG.SG,DA65_noUDG.SG,DA66_noUDG.SG,DA68_noUDG.SG,DA69_noUDG.SG,DA72_noUDG.SG,DA73_noUDG.SG,DA74_noUDG.SG,DA80_noUDG.SG,DA82_noUDG.SG,DA96_noUDG.SG,DA71_noUDG.SG,DA105_noUDG.SG

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
irk030	1	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv.	irk030_noUDG.SG
Iberia_ChL	82	O	O	O		O			Gunther et al. (2015) PNAS, Mathieson et al. (2015) Nature, Lipson et al. (2017) Nature, Olalde et al. (2018) Nature, Olalde et al. (2019) Science, Valdiosera et al. (2018) PNAS	I0257,I0258,I0260,I0261,I0262,I0263,I0455,I0456,I0457,I0459,I0460,I0823,I0825,I0826,I1271,I1272,I1276,I1277,I1280,I1281,I1282,I1284,I1300,I1303,I1314,I1553,I1838,I1842,I1843,I1845,I1846,I1975,I1976,I1978,I1981,I2467,I2473,I3269,I3270,I3271,I3272,I3277,I4245,I4247,I4565,I5838,I6475,I6543,I6584,I6587,I6604,I6609,I6612,I6617,I6622,I6628,I6629,I6630,I7587,I8048,I8131,I8141,I8148,I8149,I8150,I8153,I8154,I8156,I8158,I8197,I8198,I8199,I8364,I8365,I8566,I8569,I0581_d,I3276,I6542,I6596,I6608,I8140_d
Iberia_EN	11	O	O	O		O			Mathieson et al. (2015) Nature, VillalbaMouco et al. (2019) CurrentBiology, Lipson et al. (2017) Nature, Valdiosera et al. (2018) PNAS, Fregel et al. (2018) PNAS	CHA001_merged,CHA002_merged,CHA003_merged,CHA004.A0102,FUC003_merged,I0409,I0410,I0412,I0413,I1972,I2199
Iberia_MN	46	O	O	O		O			Mathieson et al. (2015) Nature, VillalbaMouco et al. (2019) CurrentBiology, Lipson et al. (2017) Nature, Olalde et al. (2019) Science	ELT002_merged,ELT006_merged,I0405,I0406,I0407,I0408,I10277,I10278,I10280,I10283,I10284,I10285,I10286,I11248,I11249,I11300,I11301,I11303,I11305,I11306,I3214,I7547,I7549,I7550,I7594,I7598,I7600,I7601,I7602,I7603,I7604,I7605,I7606,I7642,I7643,I7644,I7645,I7646,I7647,I8134,I8567,I8568,I10287,I11304,I3273,I7679
Iran_ChL	5	O	O	O		O			Lazaridis et al. (2016) Nature	I1661, I1670, I1662, I1674, I1665
Iran_N	8	O	O	O		O	O		Lazaridis et al. (2016) Nature; Naramsimhan et al. (2019) Science	I7527,I1290,I1944,I1945,I1947,I1949,I1951,I1954,I1946,I1952
Israel_C	19	O	O	O		O			Harney et al. (2018) Nat. Commun.	I0644,I1152,I1154,I1160,I1164,I1165,I1168,I1169,I1170,I1171,I1172,I1178,I1179,I1182,I1184,I1187,I1155,I1177,I1181
Jordan_EBA	3	O	O	O		O			Lazaridis et al. (2016) Nature	I1705, I1706, I1730
Kangju_1800BP	6	O	O	O		O			Damgaard et al. (2018) Nature	DA121_noUDG.SG,DA229_noUDG.SG,DA123_noUDG.SG,DA206_noUDG.SG,DA125_noUDG.SG,DA226_noUDG.SG
Karasuk.SG	6	O	O	O		O			Allentoft et al. (2015) Nature	RISE494_noUDG.SG,RISE493_noUDG.SG,RISE495_noUDG.SG,RISE496_noUDG.SG,RISE499_noUDG.SG,RISE502_noUDG.SG
Karluk_TianShan	2	O	O	O		O			Damgaard et al. (2018) Nature	DA222_noUDG.SG, DA230_noUDG.SG

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Kennewick_WA.SG	1	O	O	O		O			Rasmussen et al. (2015) Nature	kennewick_noUDG.SG
Khovsgol	20	O	O	O		O			Jeong et al. (2018) Proc. Natl. Acad. Sci. USA	ARS001.A0101,ARS002.B0101,ARS003.B0101,ARS004.A0101,ARS005.A0101,ARS006.A0101,ARS007.B0101,ARS008.A0101,ARS009.A0101,ARS011.A0101,ARS012.A0101,ARS013.A0101,ARS014.B0101,ARS015.A0101,ARS016.A0101,ARS017.A0101,ARS018.A0101,ARS024.A0101,ARS025.A0101,ARS026.A0101
Kimak_Kazakhstan_1350BP	1	O	O	O		O			Damgaard et al. (2018) Nature	DA87_noUDG.SG
Kipchak	2	O	O	O		O			Damgaard et al. (2018) Nature	DA179_noUDG.SG, DA23_noUDG.SG
Krasnoyarsk_BA	1	O	O	O	O	O	O		Kılınç et al. (2021) Sci. Adv.	kra001_noUDG.SG
Krasnoyarsk_MLBA	16	O	O	O	O	O	O		Narasimhan et al. (2019) Science	I11214,I1821,I1828,I1851,I1852,I1853,I1856,I3389,I3390,I3391,I3392,I3394,I3395,I3396,I6716,I6718,I3393,I6717
Kolyma_M	1	O	O	O	O	O	O		Sikora et al. (2019) Nature	Kolyma_River_noUDG.SG
LateDorset.SG	1	O	O	O		O			Raghavan et al. (2015) Science	LateDorset-XIV-H_126_noUDG.SG
LBK_EN	69	O	O	O		O			Lazaridis et al. (2014) Nature, Mathieson et al. (2015) Nature, Lipson et al. (2017) Nature, Rivollat et al. (2020) ScienceAdvance	HBS002,HBS004,HBS005,HBS006,HBS007,HBS008,HBS009,I0018,I0022,I0025,I0026,I0054,I0056,I0100,I0795,I0797,I2014,I2022,SCH001,SCH004,SCH007,SCH009,SCH010,SCH011,SCH014,SCH015,SCH016,SCH018,SMH004,XN164,XN165,XN166,XN167,XN168,XN169,XN170,XN171,XN172,XN173,XN174,XN175,XN178,XN180,XN182,XN183,XN188,XN191,XN205,XN206,XN207,XN211,XN215,XN224,XN225,I0046,I0048,I0057,I0659,I0821,I1550,I2008,I2017,I2026,I2029,I2030,I2032,I2036,I2037,I2038
Levant_N	8	O	O	O		O			Lazaridis et al. (2016) Nature	I1707,I1710,I1415,I1701,I1414,I1700, I1699,I1679
Liangdao1	1	O	O	O		O			Yang et al. (2020) Science	LD1
Liangdao2	1	O	O	O		O			Yang et al. (2020) Science	LD2
Loebanr_IA	32	O	O	O		O			Narasimhan et al. (2019) Science	I10000,I10001,I10974,I12134,I12137,I12455,I12456,I12457,I12458,I12459,I12979,I12980,I12981,I12982,I12984,I12987,I12988,I13220,I13221,I13222,I13223,I13224,I13226,I13227,I13228,I5400,I6292,I6553,I6556,I8998,I8999,I12983
MA1	1	O	O	O		O	O	O	Raghavan et al. (2014) Nature	MA1_noUDG.SG

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Man_Bac	7	O	O	O		O			Lipson et al. (2018) Science	I0626_noUDG2,I0627,I1135,I1859,I2731,I2947,I1137
Mebrak.SG	2	O	O	O		O			Jeong et al. (2016) Proc. Natl. Acad. Sci. USA	M344_noUDG.SG,M63_noUDG.SG
Miaozigou_MN	3	O	O	O		O			Ning et al. (2020) Nat. Commun.	MZGM10-1_noUDG,MZGM16_noUDG,MZGM25-2_noUDG
Minoan_Lasithi	5	O	O	O		O			Lazaridis et al. (2017) Nature	I0070,I0071,I0073,I0074,I9005
Minoan_Odigitria	5	O	O	O		O			Lazaridis et al. (2017) Nature	I9127,I9128,I9129,I9130,I9131
Mongolia_EIA_SlabGrave_1	13	O	O	O		O			Wang et al. (2021) Nature	I12960,I12969,I12971,I13178,I13963,I6349,I6352,I6353,I6357,I6359,I6365,I6369,I7032
Mongolia_LBA_Ulaanzukh_2	3	O	O	O		O			Wang et al. (2021) Nature	I12972,I13964,I14037
Mycenaean	4	O	O	O		O			Lazaridis et al. (2017) Nature	I9010, I9006, I9041
Natufian	4	O	O	O		O			Lazaridis et al. (2016) Nature	I1072_d,I0861,I1069,I1690
Nomad_Kazakhstan_1700BP	2	O	O	O		O			Damgaard et al. (2018) Nature	DA177_noUDG.SG,DA95_noUDG.SG
Nomad_Kazakhstan_IA	1	O	O	O		O			Damgaard et al. (2018) Nature	DA92_noUDG.SG, DA129_noUDG.SG
Nui_Nap	2	O	O	O		O			Lipson et al. (2018) Science	I2497, I2948
Okunevo	14	O	O	O		O			Damgaard et al. (2018) Science	RISE515_noUDG.SG,RISE667_noUDG.SG,RISE670_noUDG.SG,RISE671_noUDG.SG,RISE672_noUDG.SG,RISE674_noUDG.SG,RISE675_noUDG.SG,RISE677_noUDG.SG,RISE680_noUDG.SG,RISE681_noUDG.SG,RISE683_noUDG.SG,RISE684_noUDG.SG,RISE685_noUDG.SG,RISE719_noUDG.SG
Parkhai_EBA	1	O	O	O		O			Narasimhan et al. (2019) Science	I6671
Parkhai_EN	4	O	O	O		O			Narasimhan et al. (2019) Science	I4259, I4634, I4635, I6669
Parkhai_LBA	1	O	O	O		O			Narasimhan et al. (2019) Science	I6668
Parkhai_MBA	1	O	O	O		O			Narasimhan et al. (2019) Science	I6674
Pinarbasi_HG	1	O	O	O		O			Feldman et al. (2019) Nat. Commun.	ZBC_IPB001.B-C0101_Luk2-Pinarbasi

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Qihe	1	O	O	O		O			Yang et al. (2020) Science	Qihe2_noUDG_d
Russia_Bolshoy	6	O	O	O	O	O	O	O	Lamnidis et al. (2018) Nat. Commun.	BOO001.A0101, BOO002.A0101, BOO003.A0101, BOO004.A0101, BOO005.A0101, BOO006.A0101
Saka_Kazakhstan_2500 BP	5	O	O	O		O			Damgaard et al. (2018) Nature	DA11_noUDG.SG,DA13_noUDG.SG,DA14_noUDG.SG,DA15_noUDG.SG,DA16_noUDG.SG,MJ-51_noUDG.SG
Saka_TianShan_2150BP	10	O	O	O		O			Damgaard et al. (2018) Nature	DA47_noUDG.SG,DA48_noUDG.SG,DA49_noUDG.SG,DA50_noUDG.SG,DA51_noUDG.SG,DA55_noUDG.SG,DA57_noUDG.SG,DA58_noUDG.SG, DA130_noUDG.SG
Samdzong.SG	4	O	O	O		O			Jeong et al. (2016) Proc. Natl. Acad. Sci. USA	S10_noUDG.SG,S35_noUDG.SG,S40_noUDG.SG,S41_noUDG.SG
Sappali_Tepe_BA	11	O	O	O		O			Narasimhan et al. (2019) Science	I4285,I4286,I4288,I4289,I7414,I7416,I7419,I7420,I7421,I7492,I7542
Saqqaq.SG	1	O	O	O	O	O	O	O	Rasmussen et al. (2010) Nature	Inuk_noUDG.SG
Sarazm_EN	2	O	O	O		O			Narasimhan et al. (2019) Science	I4910, I4290
Sarmatian	20	O	O	O		O			Krzewińska et al. (2018) Sci. Adv Jarve et al. (2019) CurrentBiology Damgaard et al. (2018) Nature Mathieson et al. (2015) Nature	chy001_noUDG.SG,chy002_noUDG.SG,tem001_noUDG.SG,tem002_noUDG.SG,tem003_noUDG.SG, DA134_noUDG.SG,DA136_noUDG.SG,DA139_noUDG.SG,DA141_noUDG.SG,DA143_noUDG.SG,DA144_noUDG.SG,DA145_noUDG.SG,MJ-38_noUDG.SG, DA26_noUDG.SG,DA30_noUDG.SG,DA202_noUDG.SG, DA20_noUDG.SG,DA27_noUDG.SG, I0574,I0575
Scythian_Hungary_2400BP	5	O	O	O		O			Damgaard et al. (2018) Nature	DA191_noUDG.SG,DA194_noUDG.SG,DA197_noUDG.SG, DA195_noUDG.SG,DA198_noUDG.SG
Shahr_I_Sokhta_BA1	8	O	O	O		O			Narasimhan et al. (2019) Science	I11462,I11474,I11476,I11478,I11479,I11483,I8724,I8725
SHG	7	O	O	O		O			Mathieson et al. (2015) Nature, Mittnik et al. (2018) NatureCommunications	I0011,I0012,I0013,I0014,I0015,I0017,MotalaAA
Shimao_LN	3	O	O	O		O			Ning et al. (2020) Nat. Commun.	SM-SGDLM27_noUDG,SM-SGDLM6_noUDG,SM-SGDLM7X_noUDG
Sintashta_MLBA	36	O	O	O		O	O		Narasimhan et al. (2019) Science	I0937,I0938,I0939,I0942,I0943,I0982,I0984,I0986,I0989,I1003,I1006,I1008,I1011,I1013,I1018,I1019,I1022,I1024,I1027,I1029,I1053,I1055,I1060,I1061,I1062,I1063,I1064,I1065,I1082,I1084,I1086,I1088,I1089,I1090,I7480,I1012
Srubnaya	11	O	O	O		O			Mathieson et al. (2015) Nature	I0232,I0234,I0235,I0358,I0359,I0361,I0422,I0423,I0424,I0430,I0431

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Suogang	2	O	O	O		O			Yang et al. (2020) Science	SuogangB1_noUDG_d,SuogangB3_noUDG_d
Tagar	8	O	O	O		O			Damgaard et al. (2018) Nature	DA2_noUDG.SG,DA3_noUDG.SG,DA4_noUDG.SG,DA5_noUDG.SG,DA6_noUDG.SG,DA7_noUDG.SG,DA8_noUDG.SG,DA9_noUDG.SG
Tanshishan	7	O	O	O		O			Yang et al. (2020) Science	L5700_noUDG,L5701_noUDG_d,L5704_noUDG_d,L5705_noUDG,L5706_noUDG_d,L7415_noUDG,L7417_noUDG_d
TellKurdu_EC	5	O	O	O		O			Skourtanioti et al. (2020) Cell	KRD001, KRD003, KRD004, KRD005, KRD006
Tepe_Anau_EN	3	O	O	O		O			Narasimhan et al. (2019) Science	I4085, I4086, I4087
Tepe_Hissar_C	12	O	O	O		O			Narasimhan et al. (2019) Science	I2337,I2512,I2513,I2514,I2918,I2922,I2923,I2924,I2925,I2927,I2928,I2921
Tepecik.SG	3	O	O	O		O			Kılınç et al. (2016) Curr. Biol.	Tep006_noUDG.SG,Tep001_noUDG.SG,Tep002_noUDG.SG,Tep003_noUDG.SG,Tep004_noUDG.SG
Tianyuan	1	O	O	O		O			Yang et al. (2017) Curr. Biol.	Tianyuan
EastBaikal_BA	1	O	O	O	O	O			Kılınç et al. (2021) Sci. Adv.	irk078_noUDG.SG
EastBaikal_M	1	O	O	O	O	O			Kılınç et al. (2021) Sci. Adv.	cta016_noUDG.SG
EastBaikal_N	5	O	O	O	O	O	O	O	Kılınç et al. (2016) Sci. Adv.	brn003_noUDG.SG, brn001_noUDG.SG, brn002_noUDG.SG,brn012_noUDG.SG, brn008_noUDG.SG
UKY	1	O	O	O	O	O	O	O	Yu et al. (2020) Science	UKY001
Ulaanzuukh_SlabGrave	15	O	O	O		O			Jeong et al. (2020) Cell	BUL001, BOR001, DAR001,DAR002, BUL002, ULN001,ULN002,ULN003,ULN004,ULN005,ULN006,ULN007,ULN009,ULN010,ULN011,ULN015
Upper_YR_IA	3	O	O	O		O			Ning et al. (2020) Nat. Commun.	DCZM17IV_noUDG,DCZ-M21II_noUDG,DCZ-M22IV_noUDG,DCZ-M6_noUDG
Upper_YR_LN	4	O	O	O		O			Ning et al. (2020) Nat. Commun.	JCKM1-1_noUDG,LJM2_noUDG,LJM3_noUDG,LJM4_noUDG,LJM5_noUDG,LJM14_noUDG,LJM25_noUDG
USR1	1	O	O	O	O	O	O	O	Moreno-Mayar et al. (2018) Nature	USR1.SG
Ust_Ishim	1	O	O	O		O			Fu et al. (2014) Nature	Ust_Ishim.DG
Varna	2	O	O	O		O			Mathieson et al. (2018) Nature	ANI160, ANI159-ANI181
Villabruna	1	O	O	O		O	O		Fu et al. (2016) Nature	Villabruna_noUDG

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
West_Siberia_N	2	O	O	O		O	O		Narasimhan et al. (2019) Science	I1960, I5766
WHG	8	O	O	O		O		O	Lazaridis et al. (2014) Nature Mathieson et al. (2018) Nature Fu et al. (2016) Nature	Loschbour.DG, Iboussieres25-1_noUDG, Iboussieres31-2_noUDG, Falkenstein_noUDG, Rochedane, BerryAuBac_noUDG, I1875, I2158
WLR_BA	2	O	O	O		O			Ning et al. (2020) Nat. Commun.	91KLH11_noUDG, 91KLH18_noUDG
WLR_BA_o	1	O	O	O		O			Ning et al. (2020) Nat. Commun.	91KLM2_noUDG
WLR_LN	2	O	O	O		O			Ning et al. (2020) Nat. Commun.	EDM139_noUDG, EDM176_noUDG
WLR_MN	2	O	O	O		O			Ning et al. (2020) Nat. Commun.	BLSM27S_noUDG, BLSM41_noUDG
Wusun_TianShan_2200 BP	4	O	O	O		O			Damgaard et al. (2018) Nature	DA220_noUDG, DA223_noUDG
Xiaogao	1	O	O	O		O			Yang et al. (2020) Science	Xiaogao_noUDG
Xiaojingshan	2	O	O	O		O			Yang et al. (2020) Science	XJS1309_M4_noUDG,XJS1311_M16_noUDG
Xitoucun	7	O	O	O		O			Yang et al. (2020) Science	L5700_noUDG,L5701_noUDG_d,L5704_noUDG_d,L5705_noUDG,L5706_noUDG_d,L7415_noUDG,L7417_noUDG_d
Yakutia_LN	5	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv.	N4a1_noUDG.SG,N4b2_noUDG.SG, yak021_noUDG.SG,yak022_noUDG.SG,yak024_noUDG.SG
Yakutia_MN	1	O	O	O	O	O	O	O	Kılınç et al. (2021) Sci. Adv.	N5a.SG
Yakutia_UP	1	O	O	O	O	O			Kılınç et al. (2021) Sci. Adv.	yak025_noUDG.SG
Yamnaya_Samara	10	O	O	O		O			Mathieson et al. (2015) Nature	I0231,I0357,I0370,I0429,I0438,I0439,I0441,I0443,I0444,I7489
YR_LBIA	5	O	O	O		O			Ning et al. (2020) Nat. Commun.	HJTM115_noUDG,JXNTM2_noUDG,JXNTM23_noUDG,LGM41_noUDG,LGM79_noUDG
YR_LN	7	O	O	O		O			Ning et al. (2020) Nat. Commun.	HJTM107_noUDG,HJTM109_noUDG,PLTM310_noUDG,PLTM311_noUDG,PLTM313_noUDG,WD-WT1H16_noUDG,WD-WT5M2_noUDG,PLTM312_noUDG
YR_MN	7	O	O	O		O			Ning et al. (2020) Nat. Commun.	WGM20_noUDG,WGM35_noUDG,WGM43_noUDG,WGM70_noUDG,WGM76S_noUDG,WGM94_noUDG,XW-M1R18_noUDG
Yumin	1	O	O	O		O	O		Yang et al. (2020) Science	China_NEastAsia_Inland_EN

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Abazin	10			O	O	O				
Abkhasian	9			O	O	O				
Adygei	31			O	O	O				
Albanian	6			O	O	O				
Aleut	7			O	O	O				
Aleut_Tlingit	4			O	O	O				
Altaian	24			O	O	O				
Altaian_Chelkans	8			O	O	O				
Ami	10			O	O	O	O			
Armenian	10			O	O	O				
Armenian_Hemsheni	8			O	O	O				
Atayal	9			O	O	O				
Avar	10			O	O	O				
Azeri	17			O	O	O				
Balkar	10			O	O	O				
Bashkir	53			O	O	O				
Basque	29			O	O	O				
Belarusian	10			O	O	O				
Besermyan	6			O	O	O				
Borneo	9			O	O	O				
Bulgarian	10			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Buryat	37			O	O	O				
Cambodian	8			O	O	O				
Chechen	9			O	O	O				
Chukchi	20			O	O	O				
Chuvash	16			O	O	O				
Circassian	9			O	O	O				
Croatian	10			O	O	O				
Cypriot	8			O	O	O				
Czech	10			O	O	O				
Dai	10			O	O	O				
Darginian	8			O	O	O				
Daur	9			O	O	O				
Dolgan	4			O	O	O				
Dungan	13			O	O	O				
Enets	3			O	O	O				
English	10			O	O	O				
Eskimo_ChaplinSireniki	7			O	O	O				
Eskimo_Naukan	12			O	O	O				
Estonian	10			O	O	O				
Even	9			O	O	O				
Evenk_FarEast	5			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Evenk_Eastbaikal	8			O	O	O				
Ezid	8			O	O	O				
Finnish	8			O	O	O				
French	61			O	O	O				
Gagauz	9			O	O	O				
Georgian	23			O	O	O				
German	10			O	O	O				
Greek	20			O	O	O				
Han	43			O	O	O				
Hezhen	8			O	O	O				
Hungarian	20			O	O	O				
Icelandic	12			O	O	O				
Ingushian	10			O	O	O				
Irish	24			O	O	O				
Irish_Ulster	6			O	O	O				
Italian_North	20			O	O	O				
Italian_South	5			O	O	O				
Itelmen	6			O	O	O				
Japanese	29			O	O	O				
Jew_Ashkenazi	7			O	O	O				
Jew_Georgian	7			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Kabardinian	10			O	O	O				
Kaitag	8			O	O	O				
Kalmyk	10			O	O	O				
Karachai	12			O	O	O				
Karakalpak	14			O	O	O				
Karelian	15			O	O	O				
Kazakh	18			O	O	O				
Ket	22			O	O	O	O			
Khakass	16			O	O	O				
Khakass_Kachins	8			O	O	O				
Khamnegan	10			O	O	O				
Kinh	8			O	O	O				
Korean	6			O	O	O				
Koryak	9			O	O	O				
Kubachinian	6			O	O	O				
Kumyk	8			O	O	O				
Kurd	8			O	O	O				
Kyrgyz	9			O	O	O				
Lahu	8			O	O	O				
Lak	10			O	O	O				
Lezgin	9			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Lithuanian	10			O	O	O				
Maltese	8			O	O	O				
Mansi	8			O	O	O				
Miao	10			O	O	O				
Moldavian	10			O	O	O				
Mongol	34			O	O	O				
Mongola	6			O	O	O				
Mordovian	32			O	O	O				
Nanai	10			O	O	O				
Naxi	9			O	O	O				
Negidal	3			O	O	O				
Nganasan	33			O	O	O	O			
Nivh	10			O	O	O				
Nogai	25			O	O	O				
Norwegian	11			O	O	O				
Orcadian	13			O	O	O				
Oroqen	9			O	O	O				
Ossetian	16			O	O	O				
Polish	23			O	O	O				
Romanian	10			O	O	O				
Russian	71			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Russian_Archangelsk_ Krasnoborsky	6			O	O	O				
Russian_Archangelsk_ Leshukonsky	5			O	O	O				
Russian_Archangelsk_ Pinezhsky	5			O	O	O				
Saami	3			O	O	O				
Sardinian	27			O	O	O				
Scottish	33			O	O	O				
Selkup	25			O	O	O	O			
Semende	10			O	O	O				
She	10			O	O	O				
Sherpa	2			O	O	O				
Shetlandic	10			O	O	O				
Shor_Khakassia	6			O	O	O				
Shor_Mountain	7			O	O	O				
Sicilian	11			O	O	O				
Sorb	9			O	O	O				
Spanish	53			O	O	O				
Spanish_North	5			O	O	O				
Tabasaran	10			O	O	O				
Tajik	31			O	O	O				
Tatar_Kazan	13			O	O	O				
Tatar_Mishar	10			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Tatar_Siberian	19			O	O	O				
Tatar_Siberian_Zabolotniye	5			O	O	O				
Thai	10			O	O	O				
Tibetan	2			O	O	O				
Todzin	3			O	O	O				
Tofalar	13			O	O	O				
Tu	10			O	O	O				
Tubalar	27			O	O	O				
Tujia	10			O	O	O				
Turkish	50			O	O	O				
Turkish_Balikesir	5			O	O	O				
Turkmen	7			O	O	O				
Tuvinian	20			O	O	O				
Udmurt	10			O	O	O				
Ukrainian	21			O	O	O				
Ulchi	25			O	O	O				
Uygur	10			O	O	O				
Uzbek	27			O	O	O				
Veps	10			O	O	O				
Xibo	7			O	O	O				
Yakut	20			O	O	O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Yi	10			O	O	O				
Yukagir	19			O	O	O				
Abkhasian.DG	2		O			O				
Adygei.DG	2		O			O				
Albanian.DG	1		O			O				
Aleut.DG	2		O			O				
Altaian.DG	1		O			O				
Ami.DG	2		O			O	O			
Armenian.DG	2		O			O				
Atayal.DG	1		O			O				
Australian.DG	1		O			O				
Balochi.DG	2		O			O				
Basque.DG	2		O			O				
BedouinB.DG	2		O			O				
Bengali.DG	2		O			O				
Bergamo.DG	2		O			O				
Bougainville.DG	2		O			O				
Brahmin.DG	2		O			O				
Brahui.DG	2		O			O				
Bulgarian.DG	2		O			O				
Burmese.DG	2		O			O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Burusho.DG	2		O			O				
Cambodian.DG	2		O			O				
Chane.DG	1		O			O				
Chechen.DG	1		O			O				
Chipewyan.DG	2		O			O				
Chukchi.DG	1		O			O				
Cree.DG	2		O			O				
Crete.DG	2		O			O				
Czech.DG	1		O			O				
Dai.DG	5		O			O				
Daur.DG	1		O			O				
Druze.DG	2		O			O				
Dusun.DG	2		O			O				
English.DG	2		O			O				
Eskimo_Chaplin.DG	1		O			O				
Eskimo_Naukan.DG	2		O			O				
Eskimo_Sireniki.DG	2		O			O				
Estonian.DG	2		O			O				
Even.DG	3		O			O				
Finnish.DG	3		O			O				
French.DG	4		O			O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Georgian.DG	2		O			O				
Greek.DG	2		O			O				
Han.DG	4		O			O				
Hawaiian.DG	1		O			O				
Hazara.DG	2		O			O				
Hezhen.DG	2		O			O				
Hungarian.DG	2		O			O				
Icelandic.DG	2		O			O				
Igbo.DG	2		O			O				
Igorot.DG	2		O			O				
Iranian.DG	2		O			O				
Iraqi_Jew.DG	2		O			O				
Irula.DG	2		O			O				
Itelman.DG	1		O			O				
Japanese.DG	3		O			O				
Jordanian.DG	3		O			O				
Kalash.DG	2		O			O				
Kapu.DG	2		O			O				
Karitiana.DG	4		O			O				
Kashmiri_Pandit.DG	1		O			O				
Kharia.DG	1		O			O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Khonda_Dora.DG	1		O			O				
Kinh.DG	2		O			O				
Kongo.DG	1		O			O				
Korean.DG	2		O			O				
Korean_Ulsan	88		O			O				
Korean_Ulsan.SG	88		O			O				
Kurumba.DG	1		O			O				
Kusunda.DG	2		O			O				
Kyrgyz.DG	2		O			O				
Lahu.DG	2		O			O				
Lemande.DG	2		O			O				
Lezgin.DG	2		O			O				
Madiga.DG	2		O			O				
Makrani.DG	2		O			O				
Mala.DG	3		O			O				
Mansi.DG	2		O			O				
Maori.DG	1		O			O				
Mayan.DG	2		O			O				
Mbuti.DG	5		O			O	O	O		
Miao.DG	2		O			O				
Mixe.DG	3		O			O	O			

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Mixtec.DG	2		O			O				
Mongola.DG	2		O			O				
Nahua.DG	1		O			O				
Naxi.DG	3		O			O				
North_Ossetian.DG	2		O			O				
Onge.DG	2		O			O	O			
Orcadian.DG	2		O			O				
Oroqen.DG	2		O			O				
Palestinian.DG	3		O			O				
Papuan.DG	16		O			O				
Pathan.DG	2		O			O				
Piapoco.DG	2		O			O				
Pima.DG	2		O			O				
Polish.DG	1		O			O				
Punjabi.DG	4		O			O				
Quechua.DG	3		O			O				
Relli.DG	2		O			O				
Russian.DG	2		O			O				
Samaritan.DG	1		O			O				
Sardinian.DG	4		O			O				
She.DG	2		O			O				

Appendix 2. Continued

New.Group.Label	N	Ancient	1240K	HO	PCA	f statistics	qpWave /qpAdm	qpgraph	References	Ancient individual IDs
Sindhi.DG	2		O			O				
Spanish.DG	2		O			O				
Surui.DG	2		O			O				
Tajik.DG	2		O			O				
Thai.DG	2		O			O				
Tlingit.DG	2		O			O				
Tu.DG	2		O			O				
Tubalar.DG	2		O			O				
Tujia.DG	2		O			O				
Turkish.DG	2		O			O				
Tuscan.DG	2		O			O				
Ulchi.DG	2		O			O				
Uygur.DG	2		O			O				
Xibo.DG	2		O			O				
Yadava.DG	2		O			O				
Yakut.DG	2		O			O				
Yemenite_Jew.DG	2		O			O				
Yi.DG	2		O			O				
Zapotec.DG	2		O			O				