Reconstructing the Genetic Relationship between Ancient and Present-Day Siberian Populations

Haechan Gill [1], Juhyeon Lee [1], and Choongwon Jeong [1,*]

[1]School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea

*Corresponding author: E-mail: cwjeong@snu.ac.kr.

Abstract

Human populations across a vast area in northern Eurasia, from Fennoscandia to Chukotka, share a distinct genetic component often referred to as the Siberian ancestry. Most enriched in present-day Samoyedic-speaking populations such as Nganasans, its origins and history still remain elusive despite the growing list of ancient and present-day genomes from Siberia. Here, we reanalyze published ancient and present-day Siberian genomes focusing on the Baikal and Yakutia, resolving key questions regarding their genetic history. First, we show a long-term presence of a unique genetic profile in southern Siberia, up to 6,000 yr ago, which distinctly shares a deep ancestral connection with Native Americans. Second, we provide plausible historical models tracing genetic changes in West Baikal and Yakutia in fine resolution. Third, the Middle Neolithic individual from Yakutia, belonging to the Belkachi culture, serves as the best source so far available for the spread of the Siberian ancestry into Fennoscandia and Greenland. These findings shed light on the genetic legacy of the Siberian ancestry and provide insights into the complex interplay between different populations in northern Eurasia throughout history.

Key words: Siberian ancestry, Syalakh–Belkachi cultures, Yakutia, Baikal, ancient genome.

Significance

The investigation into the origin and spread of Siberian ancestry, a unique genetic component found in human populations spanning from Fennoscandia to Chukotka, has been understudied. To address this gap, we delve into the population dynamics of Middle Holocene Siberian hunter–gatherers and establish the connections between present-day and Middle Holocene Siberian populations. Significantly, we find that the Middle Neolithic Yakutian Belkachi culture played a pivotal role in the spread of the Siberian ancestry.


Introduction

Migration and admixture are key demographic events that have influenced the genetic structure of modern human populations (Patterson et al. 2012). The genetic diversity of inhabitants in Inner Eurasia, a vast geographic region encompassing Siberia and the Eurasian Steppe, has been shaped by a complex history of mixture between diverse source populations of both eastern and western Eurasian origins (Jeong et al. 2019). As a result of this complex history, present-day Inner Eurasian populations are stratified into three distinct admixture clines mirroring geography.

The northernmost one among these clines, composed of populations from the boreal forest and tundra regions who mostly speak the Uralic and Yeniseian languages, share a distinct type of Eastern Eurasian ancestry, frequently referred to as the Siberian ancestry in recent archeogenetics literature (Tambets et al. 2018). Among the present-day populations, it is most enriched in Nganasans and other Samoyedic-speaking ones such as Nenets, Enets, and Selkups.

The Samoyedic-speaking populations inhabit the northernmost region of Siberia (Nganasans, Enets, and Nenets)
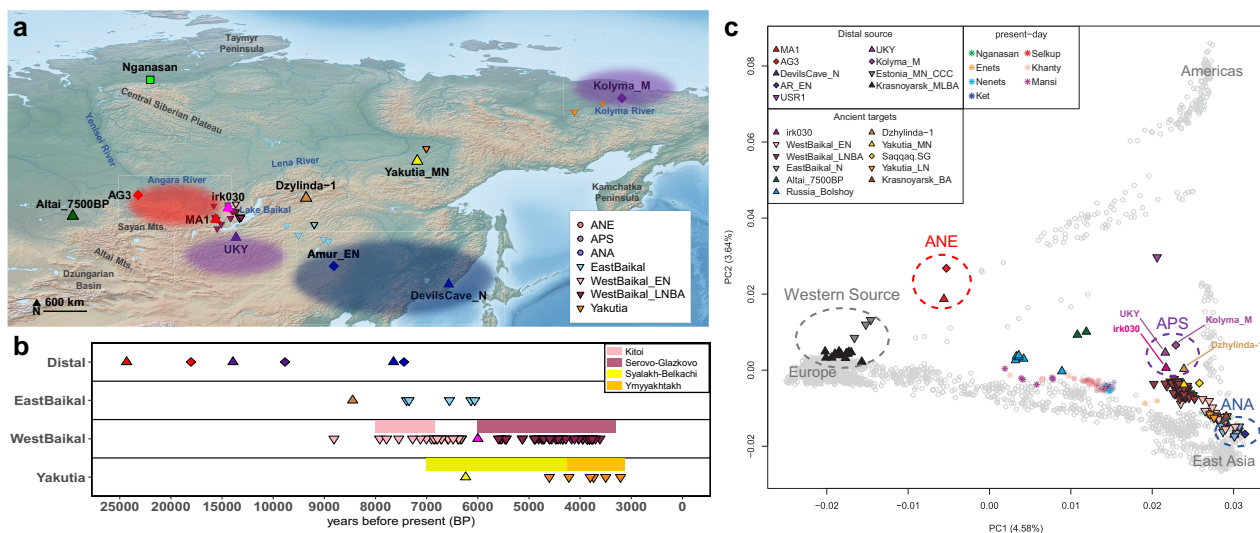
FIG. 1.—a) Geographic sites of analyzed samples in this study. Three distal source populations (AG3, MA1 for ANE; UKY, Kolyma_M for APS; and Amur_EN, DevilsCave_N for ANA), which are used for explaining the formation of Middle Holocene Siberians, are highlighted by color-shaded ellipses on the map. Among Middle Holocene Siberians, key individuals, irk030, Dzylinda-1, Altai_7500BP, and Yakutia_MN, are marked separately, and the rest of the Middle Holocene and present-day populations are marked based on geographic locations and archeological period. The base map data are downloaded from Natural Earth (https://www.naturalearthdata.com/downloads/). b) The radiocarbon dating of three distal sources and Middle Holocene Siberians. The dating results of key ancient samples are presented in before the present (BP). c) PCA performed with present-day Eurasian and American individuals. Each present-day individual is placed on PC 1 and 2 coordinates by colored asterisk symbols (key Siberian populations) or gray circles (others). Key ancient individuals are projected on precalculated PCs and marked with color-filled polygon symbols.

as well as the Yenisei River basin to the south (Selkups). Although ancient genomes have been only scarcely reported in these regions, the Siberian ancestry was present in a larger area in the past, including early Metal Age individuals from Bolshoy Oleni Ostrov in the Kola Peninsula, Iron Age individuals from the eastern Baltic Sea, and Iron Age individuals from the Volga–Oka interfluve (Lamnidis et al. 2018; Saag et al. 2019; Peltola et al. 2023). Together with the genetic analysis of present-day populations, these studies suggest that the populations with the Siberian ancestry once occupied a large area in Siberia and formed a substratum for the genetic profile of present-day populations in the regions. Therefore, it is crucial to trace the origins and spreads of the Siberian ancestry for understanding the formation of present-day human populations and languages in northern Eurasia. Despite the recent accumulation of ancient genome data in Siberia, centered on southern Siberia, it remains obscure how these populations are related to present-day inhabitants of Siberia, such as Nganasans, calling for a careful reinvestigation of previously published data.

Among the previously published ancient genomes, the Middle Holocene (ca. 8,000 to 3,000 yr ago; Sandweiss et al. 1999) southern Siberians represent a pivotal ancient lineage connected to the present-day Siberian ancestry. Of particular interest are those from the archeological sites at

Lake Baikal and Yakutia because they harbored Y haplogroups N and Q, which are prevalent in present-day Siberians (Karafet et al. 2018). These archeological sites also belong to multiple interrelated but distinct archeological cultures of hunter–gatherers: the Lake Baikal region was inhabited by the Kitoi culture during the Early Neolithic period (ca. 8,000 to 6,800 yr ago), followed by the Serovo–Glazkovo culture during the Late Neolithic period and Early Bronze Age (ca. 6,000 to 3,400 yr ago), while the Yakutia region was inhabited by the Syalakh–Belkachi culture during the Early–Middle Neolithic periods (ca. 7,000 to 4,300 yr ago), succeeded by the Late Neolithic Ymyyakhtakh culture (ca. 4,300 to 3,300 yr ago; Weber and Bettinger 2010; Coutouly 2016). Overall, they show genome-wide genetic profiles closely related to present-day Siberians but still different enough to reject a simple ancestor–descendant relationship. Therefore, the Middle Holocene Siberians in Yakutia and Lake Baikal provide an excellent starting point from which we can build a historical model for the origins of the Siberian ancestry and populations harboring it.

The genetic makeup of the Middle Holocene Siberians resulted from the admixture of three ancestries (Fig. 1): Ancient North Eurasian (ANE), Ancient Paleo-Siberian (APS; an ancestry closely related to the Native American ancestries), and Ancient North Asian (ANA). ANE ancestry is represented by the Upper Paleolithic individuals from the Mal'ta (MA1) and Afontova Gora sites (AG2 and AG3;

Raghavan et al. 2014; Fu et al. 2016). During the Last Glacial Maximum (LGM), ANE ancestry intermixed with populations of Eastern Eurasian origin and formed the ancestral population of Native Americans. This ancestral population left its genetic legacy in later APS populations, e.g. 14,000-yr-old terminal Pleistocene individual from Ust-Kyakta-3 in southern Siberia (UKY) and 9,800-yr-old Mesolithic individual from the Duvanny Yar site at the Kolyma River in northern Siberia (Kolyma_M; Sikora et al. 2019; Yu et al. 2020). At the beginning of the Middle Holocene, individuals of ANA ancestry already appeared in both East and West Baikal (Kılınç et al. 2021), presumably expanded from the neighboring regions in northeastern China where ANA ancestry was present at least 14,000 yr ago (Siska et al. 2017; Ning et al. 2020; Mao et al. 2021). Although the genetic profile and the geographic distribution of these three ancestries in Siberia have been actively investigated using ancient genomes, it remains unexplored how, when, and where the genetic profiles of present-day Siberians were formed out of these three ancestral populations. In this study, we provide a proximal historical model for the genetic relationship between a comprehensive set of published Siberian genome data. Our findings demonstrate that the APS population was present in the Baikal and Yakutia during the Middle Holocene, and in each region, the local APS population formed a genetic substratum for the later populations. Yakutian populations from three different time points are genetically distinct due to multiple streams of ANA-related gene flow between them. Finally, we show that the Middle Neolithic Yakutia (Yakutia_MN) population serves as a better-fitting source than the Lake Baikal one for the Siberian ancestry found in the genetic makeup of northern Siberians, northeastern Europeans, Paleo-Eskimos, and ancient Athabaskans.

Results

The Genetic Profile of Ancient Siberian Individuals

We curated published genomes and genome-wide data of ancient Siberian populations focusing on Lake Baikal and Yakutia. Most ancient individuals date to the Middle Holocene, ranging 8,800 to 3,100 BP (Fig. 1). To overview the genetic profile of these ancient individuals, we performed principal component analysis (PCA; Patterson et al. 2006) and projected them onto the principal components (PCs) calculated from 2,270 present-day Eurasian and American individuals (Fig. 1). PC1 separates individuals from west to east, and PC2 separates from Eurasians to Native Americans. While most Middle Holocene Siberian individuals fall on a cline in the PC space between the ANE and ANA populations, two individuals deviate from this cline: Dzhylinda-1, the earliest individual among east Baikal individuals (6,564 to 6,429 cal. BCE), and irk030,

the earliest Late Neolithic West Baikal individual (4,150 to 3,950 cal. BCE; Kılınç et al. 2021). They are shifted upward along PC2, suggesting an extra affinity with Native Americans.

For the group-based analyses, we removed PCA outliers and first-degree relatives and allocated the remaining ancient Siberian individuals into five analysis supergroups according to their geographic location, archeological period, and PCA pattern: Early Neolithic East Baikal (EastBaikal_N; $n = 5$), Early Neolithic West Baikal (WestBaikal_EN; $n = 21$), Late Neolithic to Early Bronze Age West Baikal (WestBaikal_LNBA; $n = 45$), Middle Neolithic Yakutia (Yakutia_MN; $n = 1$), and Late Neolithic Yakutia (Yakutia_LN; $n = 5$; de Barros Damgaard, Martiniano, et al. 2018; Sikora et al. 2019; Yu et al. 2020; Kılınç et al. 2021; supplementary table S1, Supplementary Material online).

The Ancestral Native American Genetic Legacy Diverged in Holocene Siberia

We first model the genetic profile of the two ancient Siberian individuals, Dzhylinda-1 and irk030, who potentially show a link to the ancestral Native American gene pool (Fig. 1). Of note, two earlier-period APS individuals, 14,000-yr-old UKY and 9,800-yr-old Kolyma_M, show a similar shift in the PC space toward Native Americans to a greater degree. We formally tested if an ancestry component related to Native Americans is required to explain the genetic profile of these Siberian individuals using qpAdm (Lazaridis et al. 2016). While the two-way admixture models of neither ANE + ANA nor Native American + ANA fit them, the three-way model of ANE + ANA + Native American adequately fits them with similar ancestry proportions with the earlier APS individuals (Fig. 2; supplementary table S4, Supplementary Material online). Indeed, using qpWave (Reich et al. 2012), we show that (UKY, Kolyma_M, irk030) and (UKY, Kolyma_M, Dzhylinda-1) can be modeled as a clade, respectively (supplementary table S4, Supplementary Material online). However, Dzhylinda-1 and irk030 do not form a clade in qpWave analysis, suggesting a difference in the admixture proportions between these two Holocene individuals. Our analysis extends the presence of the APS population at least to the Late Neolithic (irk030) in Siberia.

By utilizing outgroup-f_3 statistics (Patterson et al. 2012) in the form of f_3(Mbuti; irk030/Dzhylinda-1, X), we searched for the genetic link of irk030 and Dzhylinda-1 with later populations (supplementary fig. S2 and table S8, Supplementary Material online). Dzhylinda-1 and irk030 showed a high genetic affinity with Yakutia_MN and WestBaikal_LNBA, respectively. f_4 statistics (Patterson et al. 2012) in the form of f_4(Mbuti, irk030/Dzhylinda-1; WestBaikal_LNBA, Yakutia_MN) confirm that Dzhylinda-1 and irk030 are
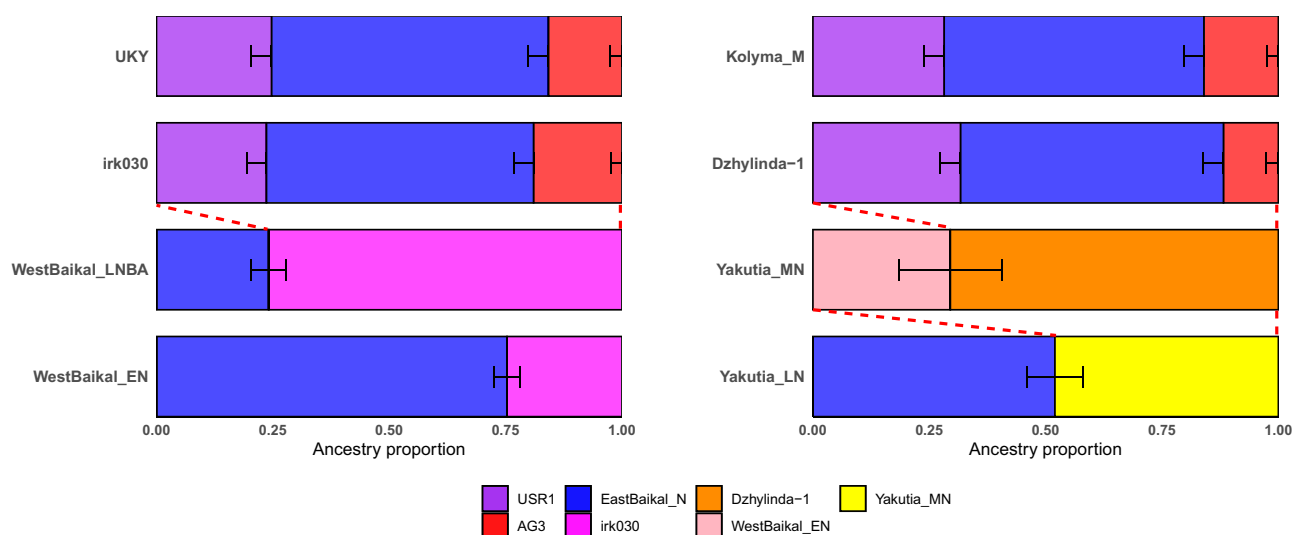
Fig. 2.—The genetic profiles of ancient Siberians estimated by *qpAdm*. Target populations are modeled as a mixture of two or three populations. The ancestry proportion of each source is represented by the box size on the *x* axis. Horizontal bars represent ±1 SE estimated by *qpAdm* using 5 cM block jack-knifing. Detailed results are presented in supplementary tables S4 and S5, Supplementary Material online.

closer to Yakutia_MN and WestBaikal_LNBA, respectively (supplementary fig. S3 and table S7, Supplementary Material online). Notably, irk030 and Dzhylinda-1 are the only populations that precede WestBaikal_LNBA and Yakutia_MN in their regions, respectively, and distinguish them. These findings suggest that the difference between the Yakutia and West Baikal populations trace back to the distinct Middle Holocene APS populations represented by Dzhylinda-1 and irk030, respectively. We note that Dzhylinda-1 is from the northern part of the East Baikal region, close to but not in Yakutia, but utilize him for modeling Yakutian populations based on this genetic affinity.

The Late Neolithic Local APS Population in West Baikal

The genetic profile of the West Baikal populations underwent a transition from ANA-rich WestBaikal_EN population to APS-rich WestBaikal_LNBA, coinciding with the cultural shift from the Kitoi to the Serovo–Glazkovo culture (Weber et al. 2002). Previous studies have suggested that the resurgence of the APS ancestry in Serovo–Glazkovo was due to gene flow from an APS source like UKY, Kolyma_M, and Altai hunter–gatherers (Altai_HG; Sikora et al. 2019; Yu et al. 2020; Wang et al. 2023). Utilizing irk030, the earliest Serovo–Glazkovo-related individual with the APS genetic profile distinct from later Serovo–Glazkovo ones, we examined the so far unexplored hypothesis that irk030 represents the source of the APS ancestry found in the West Baikal populations.

Using f_4 symmetry test, we found that irk030 has higher affinity with the ANE ancestry represented by AG3 and

Tarim_EMBA1, while other West Baikal populations, both the Early Neolithic Kitoi and the later Serovo–Glazkovo ones, have more ANA ancestry represented by East Baikal populations (supplementary fig. S3, Supplementary Material online; Table 1; supplementary table S7, Supplementary Material online). We further explored the genetic makeup of the West Baikal populations by investigating the admixture model between ANA and irk030 and found EastBaikal_N + irk030 adequately fit both West Baikal populations (Fig. 2; supplementary table S5, Supplementary Material online). WestBaikal_EN showed a relatively lower contribution from irk030 ($P = 0.088$; 25% from irk030), while WestBaikal_LNBA showed a higher contribution from irk030 ($P = 0.289$; 76% from irk030; Fig. 2 and Table 1; supplementary table S5, Supplementary Material online). Importantly, previously reported admixture models for the West Baikal populations, using Altai_HG as the APS source instead of irk030, turn infeasible when irk030 is added to the outgroup population set following the *qpAdm* rotating approach (Harney et al. 2021; $P < 0.004$; supplementary table S5, Supplementary Material online). Additionally, we tested if irk030 provides a better APS proxy for the APS ancestry in Altai_HG, who were modeled as a mixture of ANE and APS (Wang et al. 2023). We show that Altai_HG is adequately modeled as irk030 + ANE while APS + ANE models with other APS sources break when irk030 is added to the outgroup population set (supplementary table S5, Supplementary Material online). Based on these results, we propose that the connection between West Baikal and the Altai was mediated by a gene flow of the APS ancestry from West Baikal to the Altai, flipping the direction of gene flow previously suggested (Wang et al. 2023).

Table 1

Key genetic symmetricity test and admixture modeling results

Key f_4 statistic results between target populations

pop1	pop2	pop3	pop4	f_4	Z-score	Table
Mbuti	irk030	WestBaikal_LNBA	Yakutia_MN	−0.00187	−3.833	S7
Mbuti	Dzhylidna-1	WestBaikal_LNBA	Yakutia_MN	0.00190	3.256	S7
Mbuti	EastBaikal_N	irk030	WestBaikal_EN	0.00487	13.032	S7
Mbuti	EastBaikal_N	irk030	WestBaikal_LNBA	0.00239	6.557	S7
Mbuti	WestBaikal_EN	Dzhylinda-1	Yakutia_MN	0.00161	3.356	S7
Mbuti	EastBaikal_N	Yakutia_MN	Yakutia_LN	0.00238	5.525	S7

Key *qpAdm* and *qpWave* results of target populations

Target	ref1	ref2	ref3	coeff1	coeff2	coeff3	P	Table
irk030	UKY	…	…	…	…	…	0.085	S4
WestBaikal_EN	irk030	EastBaikal_N	…	0.247 ± 0.027	0.753 ± 0.027	…	0.088	S5
WestBaikal_LNBA	irk030	EastBaikal_N	…	0.759 ± 0.037	0.241 ± 0.037	…	0.289	S5
Dzhylinda-1	UKY	…	…	…	…	…	0.911	S4
Yakutia_MN	Dzhylinda-1	WestBaikal_EN	…	0.705 ± 0.111	0.295 ± 0.111	…	0.235	S5
Yakutia_LN	Yakutia_MN	EastBaikal_N	…	0.480 ± 0.060	0.520 ± 0.060	…	0.942	S5
Krasnoyarsk_BA	Yakutia_MN	EastBaikal_N	…	0.543 ± 0.092	0.457 ± 0.092	…	0.675	S6
Nganasan	Yakutia_MN	EastBaikal_N	…	0.590 ± 0.057	0.410 ± 0.057	…	0.920	S6
Enets	Yakutia_MN	Krasnoyarsk_MLBA	…	0.884 ± 0.019	0.116 ± 0.019	…	0.681	S6
Nenets	Yakutia_MN	Krasnoyarsk_MLBA	…	0.748 ± 0.021	0.252 ± 0.021	…	0.084	S6
Selkup	Yakutia_MN	Krasnoyarsk_MLBA	irk030	0.364 ± 0.074	0.268 ± 0.010	0.367 ± 0.071	0.886	S6
Ket	Yakutia_MN	Krasnoyarsk_MLBA	irk030	0.346 ± 0.094	0.218 ± 0.011	0.436 ± 0.090	0.290	S6
Mansi	Yakutia_MN	Krasnoyarsk_MLBA	…	0.577 ± 0.012	0.423 ± 0.012	…	0.714	S6
Khanty	Yakutia_MN	Krasnoyarsk_MLBA	…	0.600 ± 0.013	0.400 ± 0.013	…	0.146	S6
Russia_Bolshoy	Yakutia_MN	Estonia_MN_CCC	…	0.506 ± 0.018	0.494 ± 0.018	…	0.309	S6
Saqqaq	Yakutia_MN	…	…	…	…	…	0.150	S6
Athabaskan_1100BP	Yakutia_MN	Anzick	…	0.379 ± 0.028	0.621 ± 0.028	…	0.233	S6
	Saqqaq	Anzick	…	0.370 ± 0.028	0.630 ± 0.028	…	0.517	S6
Ekven_IA	Yakutia_MN	Anzick	…	0.606 ± 0.021	0.394 ± 0.021	…	0.308	S6
	Saqqaq	Anzick	…	0.607 ± 0.020	0.393 ± 0.020	…	0.932	S6

f_4 (pop1, pop2, pop3, and pop4) for key genetic tests. Z-scores were calculated by dividing f_4 by the s.e.m. estimated by 5 cM block jackknifing. Key qpAdm and qpWave results: coeff1, coeff2, and coeff3 represent the ancestry proportion (±1 s.e.m.) contributed by each reference population. The "*P*" column indicates the *P* value, and the "Table" column indicates the supplementary table that includes the result.

## Repeated Introduction of ANA Ancestry into Yakutia Populations

In Yakutia, we first show that Dzhylinda-1, Yakutia_MN, and Yakutia_LN do not form a simple time series of local population continuity without gene flow from other sources. While a close relationship between Dzhylinda-1 and Yakutia_MN was suggested in the previous study, their relationship was not formally modeled (Kılınç et al. 2021). We show that WestBaikal_EN is closer to Yakutia_MN than its preceding Dzhylinda-1 as shown in f_4(Mbuti, WestBaikal_EN; Dzylinda-1, Yakutia_MN) = 3.4 standard error measures (s.e.m.; supplementary fig. S3 and table S7, Supplementary Material online). Likewise, many ANA-related populations are closer to Yakutia_LN than its preceding Yakutia_MN, e.g. f_4(Mbuti, EastBaikal_N; Yakutia_MN, Yakutia_LN) = 5.5 s.e.m. (supplementary fig. S3 and table S7, Supplementary Material online). However, ancient West Baikal populations are symmetrically related to Yakutia_MN and Yakutia_LN: f_4(Mbuti, WestBaikal_EN/WestBaikal_LNBA; Yakutia_MN, Yakutia_LN) = 2.6 and 1.4 s.e.m., respectively (supplementary table S7, Supplementary Material online). Formally modeling this relationship with *qpAdm*, we show that Yakutia_MN and Yakutia_LN are adequately modeled as Dzhylinda-1 + WestBaikal_EN (P = 0.235; 30% contribution from WestBaikal_EN) and Yakutia_MN + EastBaikal_N (P = 0.942; 52% contribution from EastBaikal_N), respectively (Fig. 2 and Table 1; supplementary table S5, Supplementary Material online). In summary, our findings illustrate three distinct genetic strata in Yakutia, represented by a time series from Dzhylinda-1 to Yakutia_MN to Yakutia_LN, involving multiple episodes of ANA ancestry introduction.

Lastly, we comprehensively tested all proximal models of the Middle Holocene Siberian populations by graph-based analysis, *qpGraph* (Fig. 3). We construct a basal graph including Mbuti, MA1, Western European hunter–gatherers

worst Z = -2.77
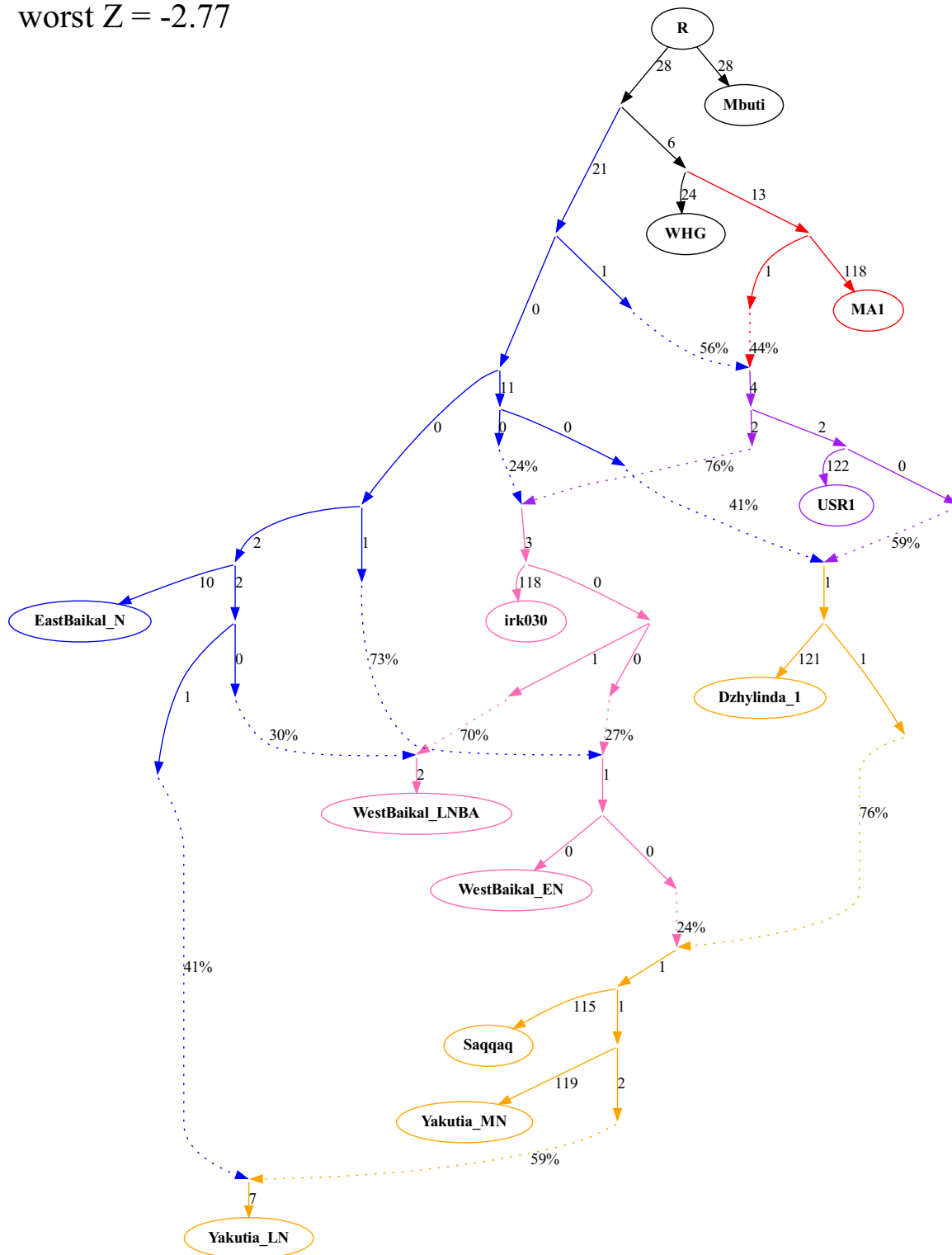
Fig. 3.—Manually fitted admixture graph explaining population dynamics of ancient Siberian populations. The admixture graph is based on proximal admixture modeling results and manually fitted by *qpGraph* function. ANE, ANA, Native American, West Baikal, and Yakutia lineages are shown in red, blue, purple, light pink, and yellow, respectively. The worst *Z*-score between fitted and observed f_4 statistics is −2.77.

(WHG), EastBaikal_N, and USR1, referring to the previous study (Yu et al. 2020). We added irk030 and Dzhylinda-1 as an independent mixture of Native American and ANA branches. Then, we modeled West Baikal and Yakutia populations as successors of irk030 and Dzhylinda-1, respectively. The proximal admixture models are well explained in the statistically feasible final admixture graph (worst Z-score = −2.77). We caution that we did not perform a comprehensive search over possible graph topologies and that the presented graph includes several zero-length branches. We speculate that this is likely due to lack of statistical resolution due to limited number of ancient genomes per group but may also reflect the demographic history of rapid branching events, such as a rapid expansion of the ANA-related population into Siberia.

The Siberian Ancestry Originated from Neolithic Yakutia Populations

Our detailed examination of the Middle Holocene Siberian populations has provided significant insights into the origin of the Siberian ancestry. We identified two primary lineages within the Middle Holocene Siberians: the Lake Baikal lineage and the Yakutia lineage. Interestingly, we observed an increasing affinity between ancient Yakutia populations and present-day Nganasan over time, while this trend was not observed in the West Baikal populations (supplementary fig. S1, Supplementary Material online). Further analysis using the outgroup-f_3 statistic revealed a strong genetic affinity between Nganasan and Yakutia, as well as a closely related individual in southern Siberia, Krasnoyarsk_BA (supplementary fig. S4 and table S9, Supplementary Material online). Similar to Yakutia_LN and Krasnoyarsk_BA, Nganasan could be modeled as Yakutia_MN + EastBaikal_N (P = 0.675; 54% contribution from Yakutia_MN), but its Yakutia_MN ancestry proportion was slightly higher than that of Yakutia_LN (48%; Fig. 4 and Table 1; supplementary table S6, Supplementary Material online). This model breaks when Yakutia_LN was added to the outgroup population set, supporting a strong affinity between Nganasan and Yakutia_LN (P = 2.38 × 10^{-11}; supplementary table S6, Supplementary Material online). Therefore, we suggest that Nganasan descended from a metapopulation to which Yakutia_LN and Krasnoyarsk_BA belonged but its direct ancestor had less contribution from the EastBaikal_N-related gene flow than Yakutia_LN.

To understand better the genetic legacy of ancient Yakutian populations in present-day Siberians, we explored the admixture models of various present-day Siberian populations, including Enets, Nenets, Selkups, Kets, Mansi, and Khanty. Previous studies modeled these populations as either two-way admixture of Nganasan + Srubnaya (Jeong et al. 2019) or three-way admixture of Kolyma_M + DevilsCave_N + Afanasievo (Sikora et al. 2019). However, the former used present-day Nganasan as a source and the latter provided models with three or more sources only. In contrast, we find that a simple two-way admixture model of Yakutia_MN + Krasnoyarsk_MLBA fits all of the above six populations (P > 0.08; Fig. 4; supplementary table S6, Supplementary Material online). Competing models with earlier Yakutian source (Dzhylinda-1 + Krasnoyarsk_MLBA; P < 0.034) or with a later one (Yakutia_LN + Krasnoyarsk_MLBA; P < 0.006) uniformly fail (supplementary table S6, Supplementary Material online).

For Selkups and Kets who live most closely to the Baikal region among the above, the Yakutia_MN + Krasnoyarsk_MLBA model turned infeasible when WestBaikal_LNBA was included in outgroups (P < 0.006; supplementary table S6, Supplementary Material online). To quantify the WestBaikal_LNBA-related ancestry component in Selkups and Kets, we applied a three-way admixture model of Yakutia_MN + irk030 + Krasnoyarsk_MLBA, considering that WestBaikal_LNBA and Yakutia_MN are too similar to be used together as sources while keeping the resolution of the model sufficiently high. The indirect model with irk030 fits both Selkups and Kets adequately even when WestBaikal_LNBA is included as an extra outgroup with 37% to 44% contribution from irk030 (P > 0.289; Fig. 4; supplementary table S6, Supplementary Material online). The irk030-related contribution was not detected in the other four populations (supplementary table S6, Supplementary Material online).

While Yakutia_LN does not provide fitting models for the above-listed present-day Siberian populations, all present-day Siberian populations have higher outgroup-f_3 values with Yakutia_LN/Nganasan than with the sources in the model, Yakutia_MN and Krasnoyarsk_MLBA (supplementary fig. S4 and tables S8 and S9, Supplementary Material online). Based on our results, we hypothesize that the true source of the Siberian ancestry in these populations likely stemmed from an unsampled population more closely related to Yakutia_LN/Nganasan than the sampled Yakutia_MN is, but not sharing the ANA gene flow with Yakutia_LN/ Nganasan.

Dispersal of the Middle Neolithic Yakutian Ancestry beyond Siberia

The Siberian ancestry extends beyond the Ural Mountains into northeastern European populations, indicating a historical migration from Siberia to the west. To pinpoint the genetic strata from which Siberian ancestry expanded and initially appeared in northeastern Europe, we examined early Metal Age individuals in Bolshoy Oleni Ostrov (Russia_Bolshoy). Russia_Bolshoy displayed high outgroup-f_3 values with Tarim_EMBA1, Yakutia populations,
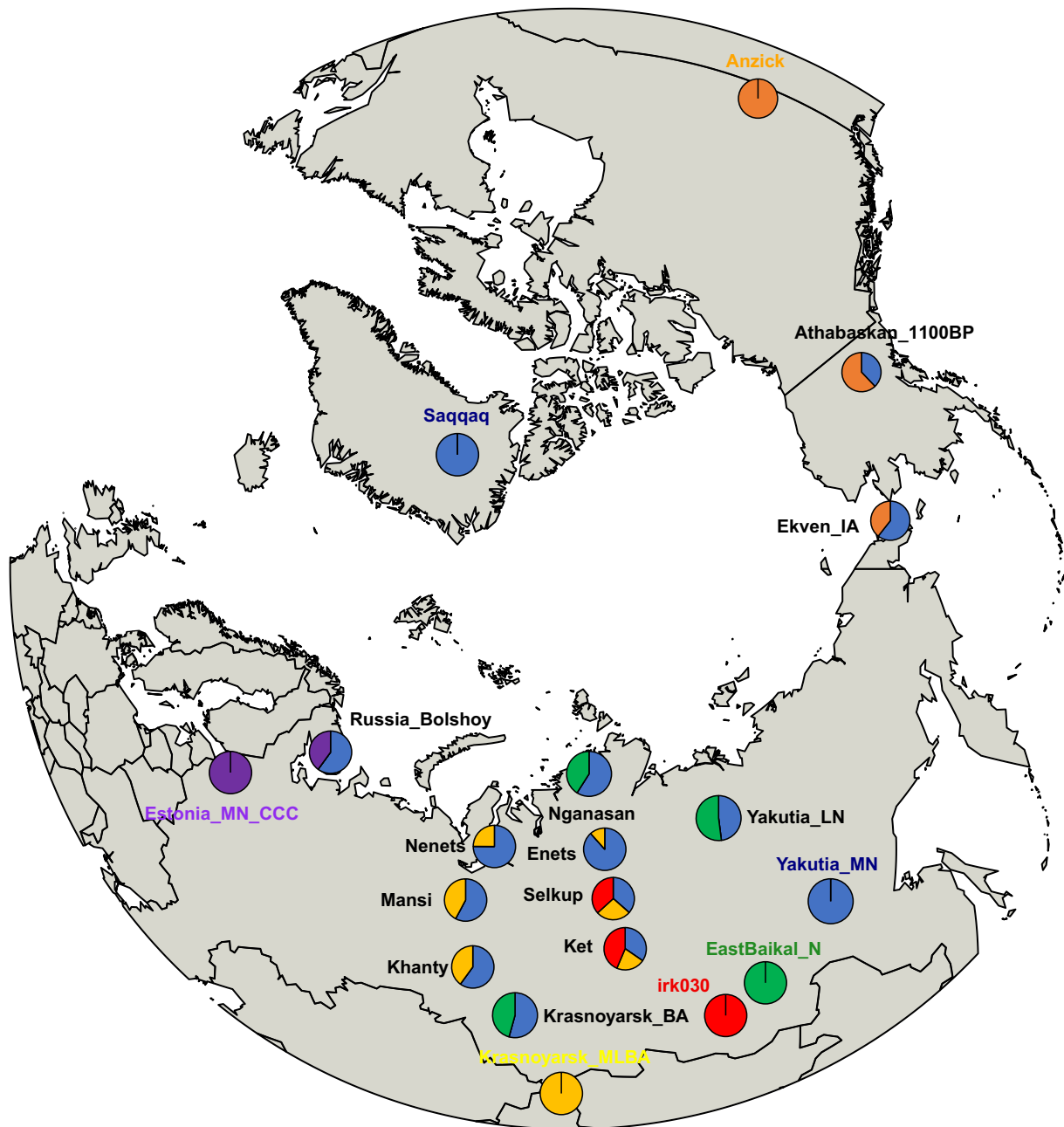
Fig. 4.—The genetic profiles of Siberian ancestry–related populations on a worldwide map. Yakutia_MN is used to represent Siberian ancestry, and each target is modeled as a mixture of Yakutia_MN and its local proximal source. All admixture models are feasible, and ancestry proportion is shown in pie charts (supplementary table S6, Supplementary Material online).

and Eastern European hunter–gatherers (EHG; supplementary fig. S4 and table S8, Supplementary Material online). A parsimonious two-way admixture model, Yakutia_MN + Middle Neolithic Combed Ceramic Culture (Estonia_MN_CCC), adequately explains Russia_Bolshoy (P = 0.309; 51% contribution from Yakutia_MN; Fig. 4 and Table 1; supplementary table S6, Supplementary Material online). It is worth noting that Russia_Bolshoy exhibits a different major Y haplogroup

pattern from Yakutia_MN (i.e. two Russia_Bolshoy males have Y haplogroup N, while Yakutia_MN has Y haplogroup Q; supplementary table S1, Supplementary Material online). We keep a possibility open that this is simply due to limited sampling of Yakutia_MN and suggest that further sampling of ancient genomes from Yakutia_MN is essential to fully understand the origins of Y haplogroup N in Russia_Bolshoy.

Regarding the origins of Paleo-Eskimos in Greenland, it has been suggested that they shared a cultural similarity with the Belkachi culture to which Yakutia_MN belongs (Powers and Jordan 1990; Coutouly 2016; Flegontov et al. 2019; Kılınç et al. 2021). However, the genetic connection between Belkachi and Paleo-Eskimo has not been thoroughly explored, especially in comparison to the earlier APS populations (Kılınç et al. 2021). We confirm that Yakutia_MN, who belonged to the Belkachi culture, is cladal with a Paleo-Eskimo individual Saqqaq even when earlier APS populations are included in the outgroups (P > 0.086; Table 1; supplementary table S6, Supplementary Material online). In addition, we confirm that Yakutia_MN can replace Saqqaq as a source in previously reported models for Athabaskans and Neo-Eskimos with comparable ancestry proportions: Yakutia_MN + Anzick fits ancient Athabaskans (Flegontov et al. 2019; P = 0.233; 38% contribution from Yakutia_MN) and Neo-Eskimos from Ekven site (Sikora et al. 2019; Wang et al. 2023; P = 0.308; 61% contribution from Yakutia_MN; Fig. 4; supplementary table S6, Supplementary Material online).

Discussion

In this study, we reconstruct detailed population dynamics of ancient Siberian populations and propose the Neolithic Yakutia populations as the origin of the Siberian ancestry. While ancient individuals with the APS ancestry are sporadically found across Siberia from Baikal to Arctic Siberia, it remains unclear how the APS populations related to each other and to later populations in the region (Sikora et al. 2019; Yu et al. 2020). By showing that two ancient individuals, irk030 and Dzhylinda-1, belong to the APS metapopulation, we confirm a long-term presence of the APS population in southern Siberia between 14,000 and 6,000 yr ago. Moreover, these two individuals represent two sublineages within the APS metapopulation, forming the genetic substratum for the later Siberian populations in the West Baikal and Yakutia, respectively, thus suggesting that the divergence between the two regions dates back at least ca. 8,500 yr ago. However, it is worth noting that this divergence does not mean a complete separation between the two regions, as shown by a gene flow from WestBaikal_EN to Yakutia_MN, indicating an expansion of Kitoi culture–related populations to the Yakutia region. Such a genetic connection aligns with archeological studies (McKenzie 2009; Kuzmin 2015) describing the spread of net-impressed pottery style from West Baikal to Yakutia.

Contrary to previous suggestions of the Serovo–Glazkovo culture resulting from admixture between the preceding local Kitoi culture and an incoming ANE-rich population (de Barros Damgaard, Martiniano, et al. 2018; Yu et al. 2020), our study presents findings that challenge this hypothesis. Specifically, our analysis of the earliest Serovo–Glazkovo individual, irk030, belongs to the APS ancestry and has no direct connection with the Kitoi culture–related populations. This raises an alternative scenario that Serovo–Glazkovo culture originated from a local APS substratum possibly tracing back to the pre-Kitoi period, not from the Early Neolithic Kitoi culture, although more ancient genomes are required to verify the APS genetic profile of the earliest Serovo–Glazkovo population. Previous studies also suggested a discontinuity between the Kitoi and Serovo–Glazkovo cultures based on a gap of archeological records between them or a significantly different mitochondrial haplogroup composition (Weber et al. 2002; Mooder et al. 2006). Instead, our research indicates that region from West Baikal to Altai Mountains could be the refugium of APS ancestry, providing a promising avenue for hypothesis testing. Accurate assessment can be achieved when early Serovo–Glazkovo individuals are unearthed in this region in the future.

It is noteworthy that the Siberian ancestry found across Eurasia and North America can be traced to a single gene pool best represented by Yakutia_MN. The genetic difference between the Middle Neolithic Belkachi culture (Yakutia_MN) and Late Neolithic Ymyyakhtakh culture (Yakutia_LN) allows us to reason that the spread of the Siberian ancestry happened prior to Yakutia_LN. This fits with the time range given by the admixture date for the earliest presence of the Siberian ancestry in northeastern Europe (Bolshoy Oleni Ostrov; ca. 4,000 BP) and by the initial appearance of the Paleo-Eskimo culture (ca. 4,500 BP). This genetic evidence also aligns well with the dissemination of ceramic and lithic technologies, as documented in previous studies (Coutouly 2016; Kozlov et al. 2020). Interestingly, evidence of the second wave of Siberian ancestry expansion, associated with Ymyyakhtakh culture, is discernible in Nganasan individuals but not in another Samoyedic-speaking population, Selkup. This suggests that the divergence within Samoyedic-speaking populations may go back to the Neolithic period.

Despite our effort to construct a detailed historical model for the relationship between ancient and present-day Siberian populations, there are certain aspects that require further investigation. First, our understanding on the distribution and impact of the APS ancestry in Siberia is based only on a handful of ancient genomes, thus leaving it unknown how it became superseded by later migrant populations across Siberia. Second, Yakutia_MN and Russia_Bolshoy belong to different Y haplogroups, N and Q, respectively, which may be attributed to the limited availability of Yakutia_MN genomes. Third, we hypothesize an unsampled population whose genetic profile is similar to Yakutia_MN but has a higher genetic affinity with later Yakutia_LN, leaving it to be tested in future studies. We call for future paleogenomic studies to produce Middle Neolithic ancient genomes across Siberia, especially from

Yakutia, to enhance our understanding on the details of the spread of the Siberian ancestry.

Materials and Methods

Genotype Data Preparation

We compiled previously published genome-wide genotype data of ancient individuals for the "1240K" panel, a set of 1,233,013 ancestry-informative single nucleotide polymorphisms (SNPs; Mathieson et al. 2015; Fu et al. 2016). First, we took publicly available random pseudo-haploid pulldown genotype data for the 1240K panel from the Allen Ancient DNA Resource v37.2 and other individual studies (supplementary table S2, Supplementary Material online; Rasmussen et al. 2010, 2014, 2015; Fu et al. 2014, 2016; Lazaridis et al. 2014, 2016, 2017; Raghavan et al. 2014, 2015; Allentoft et al. 2015; Jones et al. 2015; Mathieson et al. 2015; Jeong et al. 2016, 2018, 2020; Kılınç et al. 2016; Saag et al. 2017; Unterländer et al. 2017; Yang et al. 2017; Harney et al. 2018; Lipson et al. 2018; Mathieson et al. 2018; Mittnik et al. 2018; Moreno-Mayar et al. 2018; Narasimhan et al. 2019; Ning et al. 2020; Skourtanioti et al. 2020; Yang et al. 2020; Yu et al. 2020; Wang et al. 2021). Second, for ancient individuals whose genotype data are not available but aligned reads (BAMs) are, we obtained BAM files from the European Nucleotide Archive (https://www.ebi.ac.uk/ena), using the accession numbers provided in the original publications (de Barros Damgaard, Marchi, et al. 2018; de Barros Damgaard, Martiniano, et al. 2018; Krzewińska et al. 2018; Lamnidis et al. 2018; Flegontov et al. 2019; Sikora et al. 2019; Kılınç et al. 2021; Wang et al. 2023). Finally, we obtained FASTQ files of present-day Khanty and Nenets from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA; Wong et al. 2017). Supplementary tables S2 and S3, Supplementary Material online, provide a summary of the reference and raw data type (i.e. 1240K pulldown haploid genotype, BAM, or FASTQ) for each ancient genome data.

For present-day Khanty and Nenets, we aligned raw reads to the human reference genome with decoy sequence (hs37d5) using the bwa-aln and bwa-samse modules in Burrows-Wheeler Aligner (BWA; Li and Durbin 2009). Before genotyping with BAM file, we eliminated polymerase chain reaction duplicates using Picard MarkDuplicates module v2.20.0 for present-day data (https://broadinstitute.github.io/picard/) and dedup v0.12.8 for ancient data (Peltzer et al. 2016) and removed low-quality reads with Phred-scaled mapping quality score lower than 30 using SAMtools v1.9 (Li et al. 2009). We then examined the pattern of post-mortem chemical damages for ancient samples using mapDamage program v2.2.1 (Jónsson et al. 2013) to ensure that it matched the expected pattern from the reported library preparation method. To minimize the impact of chemical damages in genotyping, we trimmed 3 and 10 bps at both ends of reads for partial-UDG and non-UDG treated double-strand libraries, respectively, using bamUtils v1.0.15 (Jun et al. 2015).

Using these BAM files, we recalibrated the BAM files of Khanty and Nenets using BaseRecalibrator module and calculated genotype likelihoods for variants in the 1240K panel using UnifiedGenotyper module in GATK v3.8.1.0 (DePristo et al. 2011). Then, we calculated the genotype posterior probability with genotype likelihood and a prior of (0.9985, 0.0010, 0.0005) and took the highest probability genotype call when its probability exceeds 0.9. For ancient samples, we produced random pseudo-haploid genotype data for the 1240K panel by randomly choosing one high-quality base (Phred-scaled base quality score 30 or higher) using the pileupCaller v1.5.2 with "randomHaploid" option (https://github.com/stschiff/sequenceTools; v1.5.2 last accessed at April 19, 2023). For double-stranded library data, we used ends-masked BAM files for transition SNPs and nonmasked BAM files for transversions. Excluding bases that are enriched in post-mortem damages, end-masking substantially reduces incorporation of post-mortem damages into the genotype calls while still being unable to fully eliminate damages in double-stranded non-UDG libraries. For single-stranded library data, we used nonmasked BAM files and applied the "singleStrandMode" option: this minimizes the impact of chemical damages by disregarding forward strand reads for C/T SNPs and reverse strand reads for G/A SNPs. We intersected the 1240K genotype data of ancient individuals with two sets of present-day worldwide individuals: (i) 1240K genotype data of individuals from the Simons Genome Diversity Project (Mallick et al. 2016) and (ii) a broader set of individuals genotyped on the Affymetrix Axiome Genome-wide Human Origins 1 ("HumanOrigins"; 593,124 autosomal SNPs; Patterson et al. 2012; Lazaridis et al. 2016; Flegontov et al. 2019; Jeong et al. 2019). For allele frequency–based analyses, our primary data set of choice was the 1240K set. However, we utilized the HumanOrigins data set for PCA and allele frequency–based analyses in cases where present-day Nganasan, Enets, Selkup, Ket, and Mansi populations were included.

We extracted metainformation about key ancient genome data including latitude, longitude, sex, radiocarbon date, mean coverage, haplogroup, and relatedness from each publication. Although almost samples showed reliable results with downloaded data, we found some samples showed inappropriate mean coverage and Y haplogroup (i.e. Yakutia_MN shows relatively higher mean coverage than reported one). Thus, we manually calculated the mean coverage using qualimap v2.2.1 (Okonechnikov et al. 2016), genotyped on Y chromosome SNPs from the ISOGG using pileupCaller v1.4.0.5 option "majorityCall,"

and assigned the Y haplogroup using yHaplo program (Poznik, 2016; https://github.com/alexhbnr/yhaplo; version 2016.01.08, last accessed at April 28, 2022), respectively. In addition, we calculated pairwise genotype mismatch rate between individuals who allocated into same groups to check relatedness (Kennett et al. 2017). For each first-degree pair or duplicate, we removed one of the individuals with lower sequencing coverage for further analysis.

PCA

We conducted PCA with present-day individuals genotyped on the autosomal part of the HumanOrigins array ($n =$ 593,124) using *smartpca* v18140 from EIGENSOFT v8.0.0 (Patterson et al. 2006). We used two population sets, the first including present-day Eurasian and American (2,270) and the second including Eurasian only (2,077). We projected ancient individuals not included in the PC calculation using the "lsqrproject: YES" option. Samples used in PCA are listed in supplementary table S2, Supplementary Material online.

f Statistics

We calculated the *f* statistics by *qp3pop* and f_4 functions from the R library ADMIXTOOLS2 v2.0.0. (https://github.com/uqrmaie1/admixtools, publication pending). We calculated outgroup-f_3 using the Central African population Mbuti as an outgroup to measure shared genetic drift between target populations. Likewise, Mbuti was used as an outgroup to calculate f_4 statistics in the form of f_4(Mbuti, X; target1, target2) for testing symmetricity between targets or searching additional admixture sources. Populations used in *f* statistics are listed in supplementary table S2, Supplementary Material online, and the results of *f* statistics are summarized in supplementary tables S7 to S9, Supplementary Material online.

qpWave and *qpAdm* Analysis

We used *qpWave* and *qpAdm* functions from the R library ADMIXTOOLS2 v2.0.0. for admixture modeling analysis. We used the following populations as a base outgroup set for both *qpWave* and *qpAdm* analyses: present-day Central African hunter–gatherers Mbuti (1240K: $n = 5$; HumanOrigins: $n = 10$), Taiwanese Aborigines Ami (1240K: $n = 2$; HumanOrigins: $n = 10$), Native Americans Mixe (1240K: $n = 5$; HumanOrigins: $n = 10$), indigenous Andamanese islander Onge (1240K: $n = 2$; HumanOrigins: $n = 11$), early Neolithic Iranians from the Ganj Dareh site Iran_N ($n = 8$; Lazaridis et al. 2016; Narasimhan et al. 2019), Epipaleolithic European Villabruna ($n = 1$; Fu et al. 2016), early Neolithic farmers from western Anatolia Anatolia_N ($n = 23$; Mathieson et al. 2015), early Neolithic northern East Asian Yumin from Inner Mongolia ($n = 1$; Yang et al. 2020), 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 source to outgroup, was used to find the best proximal source.

Graph-Based Analysis

In order to test component-wise admixture models, graph-based analysis was implemented using the *qpGraph* function from the R library ADMIXTOOLS2 v2.0.0. Before graph fitting, f_2 statistics between all pairs of targets were calculated by the *extract_f2* function in ADMIXTOOLS2 with the "max_miss=0" option, the same as the "allsnps: NO" option from the previous version. The number of SNPs remaining by applying this option 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; USR1 for Native Americans; and 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. The estimated branch length and admixture proportions were converted to dot file by in-house code, and we plotted admixture graph using Graphviz 6.0.1.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Author Contributions

C.J. conceived and supervised the study. H.G. and J.L. curated and analyzed data. C.J. and H.G. wrote the manuscript with the input from J.L.

Funding

Conflict of Interest

The authors declare no competing interests.

Data Availability

Allen Ancient DNA Resource (AADR) v37.2 data sets were derived from sources in the public domain: https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data. The public domain and accession numbers of other present-day and

ancient genome data are listed in supplementary table S3, Supplementary Material online. The genotype data of ancient individuals for the 1240K panel, excluding those for whom we took publicly available 1240K panel genotype calls, have been deposited in the Edmond Data Repository of the Max Planck Society (https://edmond.mpg.de/dataset.xhtml?persistentId=doi:10.17617/3.QZBM1X). All analyses performed in this study are based on publicly available programs. Program names, versions, and nondefault options are described in the Materials and Methods section. All scripts used for the analyses presented in this study are publicly available via Github repository (https://github.com/CWJeongLab/Siberia).

Literature Cited

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. Population genomics of bronze age Eurasia. Nature 2015:522(7555):167–172. https://doi.org/10.1038/nature14507.

Coutouly YAG. Migrations and interactions in prehistoric Beringia: the evolution of Yakutian lithic technology. Antiquity 2016:90(349):9–31. https://doi.org/10.15184/aqy.2015.176.

de Barros Damgaard, P., Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, et al. 137 ancient human genomes from across the Eurasian steppes. Nature 2018:557(7705):369–374. https://doi.org/10.1038/s41586-018-0094-2.

de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, Barjamovic G, Rasmussen S, Zacho C, Baimukhanov N, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science 2018:360(6396):eaar7711. https://doi.org/10.1126/science.aar7711.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011:43(5):491–498. https://doi.org/10.1038/ng.806.

Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Culleton BJ, et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. Nature 2019:570(7760):236–240. https://doi.org/10.1038/s41586-019-1251-y.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prüfer K, De Filippo C, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 2014:514(7523):445–449. https://doi.org/10.1038/nature13810.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, et al. The genetic history of ice age Europe. Nature 2016:534(7606):200–205. https://doi.org/10.1038/nature17993.

Harney É, May H, Shalem D, Rohland N, Mallick S, Lazaridis I, Sarig R, Stewardson K, Nordenfelt S, Patterson N, et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. Nat Commun. 2018:9(1):3336. https://doi.org/10.1038/s41467-018-05649-9.

Harney E, Patterson N, Reich D, Wakeley J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. Genetics 2021:217(4):iyaa045. https://doi.org/10.1093/genetics/iyaa045.

Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, Immel A, Wang C-C, Ixan O, Khussainova E, et al. The genetic history of admixture across inner Eurasia. Nat Ecol Evol. 2019:3(6):966–976. https://doi.org/10.1038/s41559-019-0878-2.

Jeong C, Ozga AT, Witonsky DB, Malmström H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, et al. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. Proc Natl Acad Sci U S A. 2016:113(27):7485–7490. https://doi.org/10.1073/pnas.1520844113.

Jeong C, Wang K, Wilkin S, Taylor WTT, Miller BK, Bemmann JH, Stahl R, Chiovelli C, Knolle F, Ulziibayar S, et al. A dynamic 6,000-year genetic history of Eurasia's eastern Steppe. Cell 2020:183(4):890–904.e29. https://doi.org/10.1016/j.cell.2020.10.015.

Jeong C, Wilkin S, Amgalantugs T, Bouwman AS, Taylor WTT, Hagan RW, Bromage S, Tsolmon S, Trachsel C, Grossmann J, et al. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proc Natl Acad Sci U S A. 2018:115(48):E11248–E11255. https://doi.org/10.1073/pnas.1813608115.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun. 2015:6(1):8912. https://doi.org/10.1038/ncomms9912.

Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013:29(13):1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. 2015:25(6):918–925. https://doi.org/10.1101/gr.176552.114.

Karafet TM, Osipova LP, Savina OV, Hallmark B, Hammer MF. Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations. Am J Hum Biol. 2018:30(6):e23194. https://doi.org/10.1002/ajhb.23194.

Kennett DJ, Plog S, George RJ, Culleton BJ, Watson AS, Skoglund P, Rohland N, Mallick S, Stewardson K, Kistler L, et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat Commun. 2017:8(1):14115. https://doi.org/10.1038/ncomms14115.

Kılınç GM, Kashuba N, Koptekin D, Bergfeldt N, Dönertaş HM, Rodríguez-Varela R, Shergin D, Ivanov G, Kichigin D, Pestereva K, et al. Human population dynamics and Yersinia pestis in ancient northeast Asia. Sci Adv. 2021:7(2):eabc4587. https://doi.org/10.1126/sciadv.abc4587.

Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, Baird D, Dönertaş HM, Ghalichi A, Yaka R, et al. The demographic development of the first farmers in Anatolia. Curr Biol. 2016:26(19):2659–2666. https://doi.org/10.1016/j.cub.2016.07.057.

Kozlov AI, Vershubskaya GG, Borinskaya SA. The divergence of genetic complexes in anthropologically related populations with different types of management of natural resources. Mosc Univ Anthropol Bull. 2020:4:99–110. https://doi.org/10.32521/2074-8132.2020.4.099–110. In Russian: Дивергенция генетических комплексов у антропологически родственных популяций при разных типах хозяйствования.

Krzewińska M, Kılınç GM, Juras A, Koptekin D, Chyleński M, Nikitin AG, Shcherbakov N, Shuteleva I, Leonova T, Kraeva L, et al. Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. Sci Adv. 2018:4(10):eaat4457. https://doi.org/10.1126/sciadv.aat4457.

Kuzmin YV. Northern and north-eastern Asia: archaeology. In: Bellwood P, editors. The global prehistory of human migration. Chichester: Wiley-Blackwell; 2015. p. 191–196.

Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V, Khartanovich V, Balanovsky O, Ongyerth M, Weihmann A, et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat Commun. 2018:9(1):1–12. https://doi.org/10.1038/s41467-018-07483-5.

Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, Furtwängler A, Peltzer A, Posth C, Vasilakis A, et al. Genetic origins of the Minoans and Mycenaeans. Nature 2017:548(7666): 214–218. https://doi.org/10.1038/nature23310.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. Genomic insights into the origin of farming in the ancient Near East. Nature 2016:536(7617):419–424. https://doi.org/10.1038/nature19310.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 2014:513(7518):409–413. https://doi.org/10.1038/nature13673.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietrusewsky M, Pryce TO, Willis A, Matsumura H, Buckley H, et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science 2018:361(6397):92–95. https://doi.org/10.1126/science.aat3188.

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 2016:538(7624):201–206. https://doi.org/10.1038/nature18964.

Mao X, Zhang H, Qiao S, Liu Y, Chang F, Xie P, Zhang M, Wang T, Li M, Cao P, et al. The deep population history of northern East Asia from the Late Pleistocene to the Holocene. Cell 2021:184(12):3256–3266.e13. https://doi.org/10.1016/j.cell.2021.04.040.

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Candilio F, Cheronet O, et al. The genomic history of southeastern Europe. Nature 2018:555(7695):197–203. https://doi.org/10.1038/nature25778.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 2015:528(7583):499–503. https://doi.org/10.1038/nature16152.

McKenzie HG. Review of early hunter-gatherer pottery in eastern Siberia. In: Peter Jordan MZ, editors. Ceramics before farming: the dispersal of pottery among prehistoric Eurasian hunter-gatherers. Walnut Creek: Left Coast Press; 2009. p. 167–208.

Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, Allmäe R, Khartanovich V, Moiseyev V, Tõrv M, et al. The genetic prehistory of the Baltic Sea region. Nat Commun. 2018:9(1):442. https://doi.org/10.1038/s41467-018-02825-9.

Mooder KP, Schurr TG, Bamforth FJ, Bazaliiski VI, Savel'ev NA. Population affinities of Neolithic Siberians: a snapshot from prehistoric Lake Baikal. Am J Phys Anthropol. 2006:129(3): 349–361. https://doi.org/10.1002/ajpa.20247.

Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, Kamm JA, Albrechtsen A, Malaspinas A-S, Sikora M, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature 2018:553(7687): 203–207. https://doi.org/10.1038/nature25173.

Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, et al. The formation of human populations in South and Central Asia. Science 2019:365(6457):eaat7487. https://doi.org/10.1126/science.aat7487.

Ning C, Li T, Wang K, Zhang F, Li T, Wu X, Gao S, Zhang Q, Zhang H, Hudson MJ, et al. Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat Commun. 2020:11(1):1–9. https://doi.org/10.1038/s41467-020-16557-2.

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2016:32(2):292–294. https://doi.org/10.1093/bioinformatics/btv566.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics 2012:192(3):1065–1093. https://doi.org/10.1534/genetics.112.145037.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006:2(12):e190. https://doi.org/10.1371/journal.pgen.0020190.

Peltola S, Majander K, Makarov N, Dobrovolskaya M, Nordqvist K, Salmela E, Onkamo P. Genetic admixture and language shift in the medieval Volga-Oka interfluve. Curr Biol. 2023:33(1):174–182.e10. https://doi.org/10.1016/j.cub.2022.11.036.

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. EAGER: efficient ancient genome reconstruction. Genome Biol. 2016:17(1):1–14. https://doi.org/10.1186/s13059-016-0918-z.

Powers WR, Jordan RH. Human biogeography and climate change in Siberia and Arctic North America in the fourth and fifth millennia BP. Philos Trans A Math Phys Sci. 1990:330(1615): 665–670.

Poznik GD. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv 2016. https://doi.org/10.1101/088716, preprint: not peer reviewed.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Orlando L, Metspalu E, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 2014:505(7481):87–91. https://doi.org/10.1038/nature12736.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 2015:349(6250):aab3884. https://doi.org/10.1126/science.aab3884.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature 2014:506(7487):225–229. https://doi.org/10.1038/nature13025.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 2010:463(7282):757–762. https://doi.org/10.1038/nature08835.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CP, Ponce de León MS, Allentoft ME, Moltke I, et al. The ancestry and affiliations of Kennewick Man. Nature 2015:523(7561):455–458. https://doi.org/10.1038/nature14625.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. Reconstructing native American population history. Nature 2012:488(7411):370–374. https://doi.org/10.1038/nature11258.

Saag L, Laneman M, Varul L, Malve M, Valk H, Razzak MA, Shirobokov IG, Khartanovich VI, Mikhaylova ER, Kushniarevich A, et al. The arrival of Siberian ancestry connecting the Eastern Baltic to Uralic speakers further East. Curr Biol. 2019:29(10):1701–1711.e16. https://doi.org/10.1016/j.cub.2019.04.026.

Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag L, Pagani L, Reidla M, Tambets K, Metspalu E, et al. Extensive farming in Estonia started through a sex-biased migration from the Steppe. Curr Biol. 2017:27(14):2185–2193.e6. https://doi.org/10.1016/j.cub.2017.06.022.

Sandweiss DH, Maasch KA, Anderson DG. Transitions in the mid-Holocene. Science 1999:283(5401):499–500. https://doi.org/10.1126/science.283.5401.499.

Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, Margaryan A, de Barros Damgaard P, de la Fuente C, Renaud G, et al. The population history of northeastern Siberia since the Pleistocene. Nature 2019:570(7760):182–188. https://doi.org/10.1038/s41586-019-1279-z.

Siska V, Jones ER, Jeon S, Bhak Y, Kim H-M, Cho YS, Kim H, Lee K, Veselovskaya E, Balueva T, et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. Sci Adv. 2017:3(2):e1601877. https://doi.org/10.1126/sciadv.1601877.

Skourtanioti E, Erdal YS, Frangipane M, Restelli FB, Yener KA, Pinnock F, Matthiae P, Özbal R, Schoop U-D, Guliyev F, et al. Genomic history of Neolithic to Bronze Age Anatolia, northern Levant, and southern Caucasus. Cell 2020:181(5):1158–1175.e28. https://doi.org/10.1016/j.cell.2020.04.044.

Tambets K, Yunusbayev B, Hudjashov G, Ilumäe A-M, Rootsi S, Honkola T, Vesakoski O, Atkinson Q, Skoglund P, Kushniarevich A, et al. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. 2018:19(1):1–20. https://doi.org/10.1186/s13059-018-1522-1.

Unterländer M, Palstra F, Lazaridis I, Pilipenko A, Hofmanová Z, Groß M, Sell C, Blöcher J, Kirsanow K, Rohland N, et al. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nat Commun. 2017:8(1):14615. https://doi.org/10.1038/ncomms14615.

Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, et al. Genomic insights into the formation of human populations in East Asia. Nature 2021:591(7850):413–419. https://doi.org/10.1038/s41586-021-03336-2.

Wang K, Yu H, Radzevičiūtė R, Kiryushin YF, Tishkin AA, Frolov YV, Stepanova NF, Kiryushin KY, Kungurov AL, Shnaider SV, et al. Middle Holocene Siberian genomes reveal highly connected gene pools throughout North Asia. Curr Biol. 2023:33(3):423–-433.e5. https://doi.org/10.1016/j.cub.2022.11.062.

Weber AW, Bettinger R. Middle Holocene hunter-gatherers of Cis-Baikal, Siberia: an overview for the new century. J Anthropol Archaeol. 2010:29(4):491–506. https://doi.org/10.1016/j.jaa.2010.08.002.

Weber AW, Link DW, Katzenberg MA. Hunter-gatherer culture change and continuity in the Middle Holocene of the Cis-Baikal, Siberia. J Anthropol Archaeol. 2002:21(2):230–299. https://doi.org/10.1006/jaar.2001.0395.

Wong EH, Khrunin A, Nichols L, Pushkarev D, Khokhrin D, Verbenko D, Evgrafov O, Knowles J, Novembre J, Limborska S, et al. Reconstructing genetic history of Siberian and Northeastern European populations. Genome Res. 2017:27(1):1–14. https://doi.org/10.1101/gr.202945.115.

Yang MA, Fan X, Sun B, Chen C, Lang J, Ko Y-C, Tsang C-h, Chiu H, Wang T, Bao Q, et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. Science 2020:369(6501):282–288. https://doi.org/10.1126/science.aba0909.

Yang MA, Gao X, Theunert C, Tong H, Aximu-Petri A, Nickel B, Slatkin M, Meyer M, Pääbo S, Kelso J, et al. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Curr Biol. 2017:27(20):3202–3208.e9. https://doi.org/10.1016/j.cub.2017.09.030.

Yu H, Spyrou MA, Karapetian M, Shnaider S, Radzevičiūtė R, Nägele K, Neumann GU, Penske S, Zech J, Lucas M, et al. Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. Cell 2020:181(6):1232–1245.e20. https://doi.org/10.1016/j.cell.2020.04.037.

Associate editor: Adam Eyre-Walker