Received: 20 May 2018 | Revised: 26 August 2018 | Accepted: 23 September 2018

DOI: 10.1002/ajhb.23194

WILEY American Journal of Human Biology

ORIGINAL RESEARCH ARTICLE

Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations

Tatiana M. Karafet[1] | Ludmila P. Osipova[2,3] | Olga V. Savina[1] | Brian Hallmark[4] | Michael F. Hammer[1,5]

[1]ARL Division of Biotechnology, University of Arizona, Tucson, Arizona

[2]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

[3]Novosibirsk State University, Novosibirsk, Russia

[4]Interdisciplinary Program in Statistics, University of Arizona, Tucson, Arizona

[5]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona

Correspondence
Michael F. Hammer, ARL Division of Biotechnology, University of Arizona, Tucson, AZ 85721.
Email: mfh@email.arizona.edu

Funding information
State Research Project, Grant/Award Number: 0324-2018-0016; National Science Foundation, Grant/Award Number: PLR-1203874

Abstract

Objectives: We examined autosomal genome-wide SNPs and Y-chromosome data from 15 Siberian and 12 reference populations to study the affinities of Siberian populations, and to address hypotheses about the origin of the Samoyed peoples.

Methods: Samples were genotyped for 567 096 autosomal SNPs and 147 Y-chromosome polymorphic sites. For several analyses, we used 281 093 SNPs from the intersection of our data with publicly available ancient Siberian samples. To examine genetic relatedness among populations, we applied PCA, F_{ST}, TreeMix, and ADMIXTURE analyses. To explore the potential effect of demography and evolutionary processes, the distribution of ROH and IBD sharing within population were studied.

Results: Analyses of autosomal and Y-chromosome data reveal high differentiation of the Siberian groups. The Siberian populations have a large proportion of their genome in ROH and IBD segments. Several populations (ie, Nganasans, Evenks, Yukagirs, and Koryaks) do not appear to have experienced admixture with other Siberian populations (ie, producing only positive f3), while for the other tested populations the composition of mixing sources always included Nganasans or Evenks. The Nganasans from the Taymyr Peninsula demonstrate the greatest level of shared shorter ROH and IBD with nearly all other Siberian populations.

Conclusions: Autosomal SNP and Y-chromosome data demonstrate that Samoyedic populations differ significantly in their genetic composition. Genetic relationship is observed only between Forest and Tundra Nentsi. Selkups are affiliated with the Kets from the Yenisey River, while the Nganasans are separated from their linguistic neighbors, showing closer affinities with the Evenks and Yukagirs.

1 | INTRODUCTION

The peoples of northwest Siberia speak languages belonging to the Finno-Ugric and Samoyedic (or Samoyed) branches of the Uralic linguistic family. Samoyed people lead nomadic or semi-nomadic lives in the Taymyr Peninsula and in the tundra or forest zone between the Ural Mountains and the Yenisey River. The Samoyedic languages fall into two main branches. The northern Samoyedic branch includes the Nentsi, Entsi, and Nganasan languages. The Nentsi are divided into two groups, which exhibit differing lifestyles. The nomadic reindeer breeders of the tundra are called Tundra Nentsi, while the semi-nomadic Forest Nentsi live in the taiga region north of the Middle Ob. The main occupations of the Forest Nentsi include fishing, hunting and reindeer breeding. Despite their low mutual intelligibility, their languages are considered to be dialects of a single language. Nganasans occupy the central tundra zone of the Taymyr Peninsula. They are the most northerly people in Russia. The Entsi people with less than 200 people (census 2010)

are nearly extinct. The only southern Samoyedic language with a considerable number of speakers is Selkup, which is spoken by people who used to live in the Sayan region in southern Siberia. The ancestors of the Selkups are believed to have migrated to the taiga belt in the middle of the Ob' River during the early part of the first millennium AD (Prokof'yeva, 1964). After the arrival of the Russians in Siberia, many Selkups fled to the north and occupied the Taz River and basins in the 17th century to avoid taxes and baptism. This group is known as the northern Selkups. The southern Selkups in Tomsk Oblast were greatly assimilated with the migrants. All other southern Samoyedic-speaking populations from the Sayan region were mixed with Turkic-speaking tribes. They abandoned their own original languages in favor of Turkic, a process that was completed in the first half of the 19th century (Helimsky, 1990).

There are several different opinions on the original homeland of Samoyeds. The assumption that all extant Samoyed-speaking peoples were related to Sayan tribes was put forward back in the 18th century by M. A. Castren (Prokof'yeva, 1964). Castren argued that the ancient Samoyeds and Finno-Ugric peoples lived on the slope of the Altai and Sayan Mountains and, from there, they moved to the north and to the west. Péter Hajdú claimed that the ancient Uralic home was not in southern Siberia but on the western side of the Ural Mountains, around the bend of the Volga, (Hajdú, 1963). According to P. Hajdú the ancestors of the Samoyeds left the Uralic community in the fourth millennium BC and moved to the forest regions of western Siberia. Based on the linguistic historical reconstruction, E. Helimsky suggested a wide area between Middle Ob and Yenisey as the proto-Samoyed original home matching the Kulay archeological culture of the 500 BC-500 AD (Helimsky, 1990). J. Janhunen placed the homeland of Proto-Samoyed in the area of the Minusinsk basin on the Upper Yenisey, which is bounded on the east by the Sayan Mountains and on the west by Kuznetsk Alatau (Junhunen, 2009).

The majority of previous Siberian population genetic studies have been mostly limited to classical markers, Y-chromosome and mtDNA surveys, and usually did not include all existing Samoyed populations (Derbeneva, Starikovskaya, Volodko, Wallace, & Sukernik, 2002; Derenko et al., 2014; Derenko et al., 2007; Derenko et al., 2010; Duggan et al., 2013; Karafet et al., 2002; Karafet, Osipova, Kazakovtseva, Posukh, & Hammer, 2001; Karafet et al., 1999; Lell et al., 1997; Sukernik, Abanina, Karafet, Osipova, & Galaktionov, 1979; Sukernik, Karafet, & Osipova, 1977; Volodko et al., 2008). Recent studies have analyzed Siberian genome-wide data, focusing mostly on the history of particular population(s) (Fedorova et al., 2013; Flegontov et al., 2016; Yunusbayev et al., 2015), the evidence of admixture in the history of Siberian aboriginal populations (Pugach et al., 2016; Wong et al., 2016), or cold adaptation (Cardona et al., 2014). The only genetic study that has specifically focused on the history of the Tundra Nentsi, Forest Nentsi, and Selkups from the perspective of male lineage demonstrated that their genetic and linguistic affiliations were not entirely concordant (Karafet, Osipova, Posukh, Wiebe, & Hammer, 1999).

Ancient DNA data have recently shed light on the history of Siberian populations and have corroborated different hypotheses. The oldest anatomically modern human genome, a 45 000-year-old Ust'-Ishim modern human male from Siberia, likely represents an early modern human radiation into Europe and Central Asia (Fu et al., 2014; Kuzmin, Kosintsev, Razhev, & Hodgins, 2009). His genome diverged from the ancestors of present-day west Eurasian and east Eurasian populations before—or simultaneously with—their divergence from each other. Recently it was discovered that Siberian and East Asian populations share 38% of their ancestry with the Ust'-Ishim individual (Wong et al., 2016). A 24 000-year-old Siberian from south-central Siberia (Mal'ta) was found to be basal to modern-day western Eurasians and genetically closely related to modern-day Native Americans, with little if any genetic affinity with East Asians (Raghavan et al., 2014). Analyses of ancient DNA from the Bronze and Iron Ages in south Siberia and Mongolia demonstrated overwhelmingly predominant European settlement, suggesting an eastward migration of people (probably Kurgan culture) across the Russian and Kazakh steppes (Allentoft et al., 2015; Gubina et al., 2016; Haak et al., 2015; Hollard et al., 2014; Keyser et al., 2009; Lazaridis et al., 2014; Mooder, Schurr, Bamforth, Bazahiski, & Savel'ev, 2006).

In the present study, we examined autosomal genome-wide SNPs, and Y-chromosome data from 15 Siberian, Khalka Mongolian and 11 reference populations from Europe, Near East, Central Asia and South Asia to infer the affinities of Siberian populations and to address hypotheses on the origin of the Samoyed peoples. The sub-Saharan Dinka from southern Sudan, who were genotyped using the same Affymetrix platform, were used as an out-group for some analyses. The pattern of genetic variation of Samoyed-speaking groups was compared and contrasted with contemporary populations as well as with published ancient DNA data from Siberia.

2 | MATERIALS AND METHODS

2.1 | Populations and samples

All samples from four Samoyed-speaking populations: the Nganasans, Tundra Nentsi, Forest Nentsi, and northern Selkups, were collected in traditional villages or tundra camps. We also included samples from an additional 24 populations from Siberia, Central Asia, East Asia, South Asia, Europe, the Near East, sub-Saharan Africa (ie, the Dinka of southern

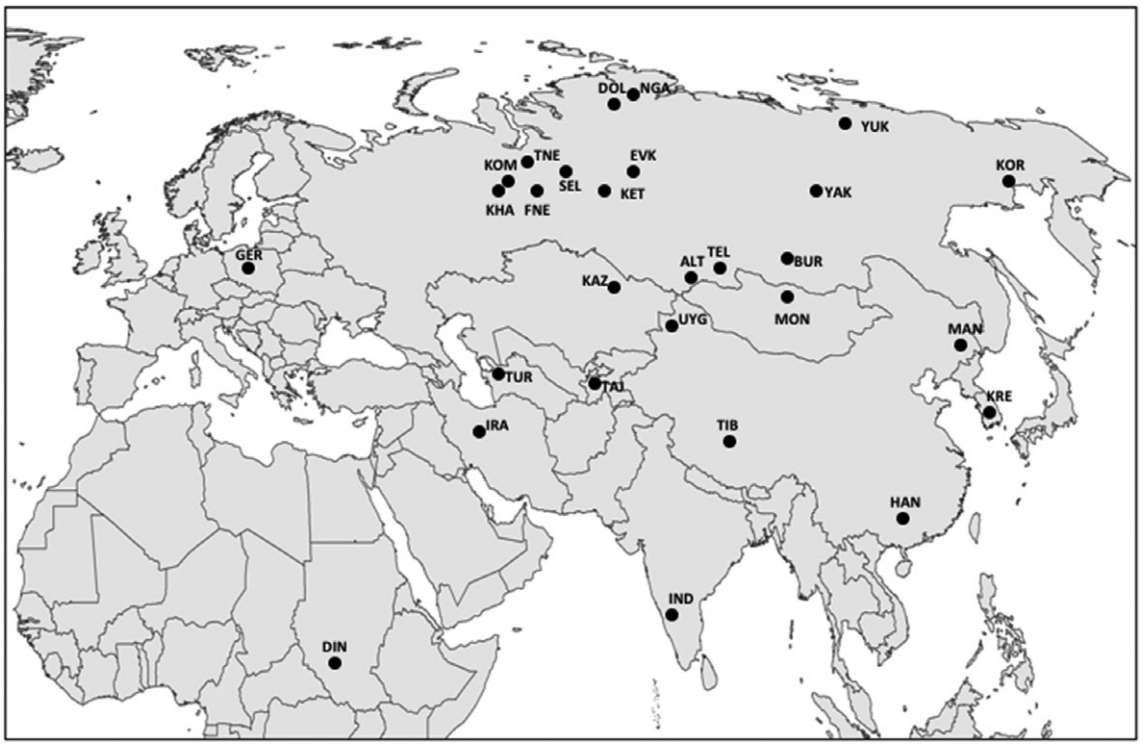

FIGURE 1 Approximate geographic location of sampling sites. For populations collected in different sites the center of geographic distribution is shown. See Supporting Information Table S1 for population codes, sample sizes, and linguistic affiliation

Sudan) (Figure 1**,** Table 1). Additional information on sampling locations, sample and population sizes, and language classification is available in Supporting Information Table S1. All samples analyzed here were included in our previous studies (Hammer et al., 2001; Karafet et al., 2002; Karafet, Osipova, & Hammer, 2008).

2.2 | Data sets, genotyping, and statistical analyses

A total of 330 samples from Siberia, Mongolia, East Asia (EAS), and southern Sudan were genotyped for 567 096 single nucleotide polymorphisms (SNPs) on the Affymetrix (Axiom) platform using standard protocols. Data were submitted to the NCBI/GEO (Accession # GSE73996). After removing close relatives, the total number of samples in our Axiom data set was 311. For several analyses we added samples from Central Asia (CAS), as well as from Iran, Germany and India as representatives of the Near East (NEA), Europe (EUR), and South Asia (SAS), respectively; all of which were previously genotyped on the same platform and published elsewhere (Karafet et al., 2015) (Figure 1).

To assess the extent of inbreeding in Siberia we used an individual-based estimate of the distribution of ROH using PLINK (Purcell et al., 2007) with the parameters identical to Karafet et al., 2015 (Karafet et al., 2015). To estimate effective population sizes and divergence time between populations, we evaluated the decay of linkage disequilibrium (LD) with recombination distance for each chromosome using the genotypic-based r^2 statistic estimated in PLINK (Purcell et al., 2007). Analyses of LD and IBD (identically by descent) were performed on full data set of 548 935 SNPs (details described in Karafet et al., 2015). Divergence time between populations was estimated as $T = 2N_eF_{ST}$, where N_e is effective population size as the harmonic means between the two populations (McEvoy, Powell, Goddard, & Visscher, 2011). To detect shared IBD pairwise segments for all pairs of samples, we ran GERMLINE 1.5 on the phased unpruned data with default parameters. We divided the genome into nonoverlapping 1 Mb blocks, removed blocks with <100 SNPs, and kept only the shared IBD segments with a length exceeded 3 cM. The mean length of IBD sharing among populations and genetic distances based on shared IBD were computed similar to Behar et al. (2013), and Karafet et al. (2016).

For several analyses we used 281 093 SNPs from the intersection of our data with publicly available ancient Siberian samples (Allentoft et al., 2015; Fu et al., 2014; Raghavan et al., 2014). The merged autosomal data set was applied to principal components analysis (PCA) (Patterson, Price, & Reich, 2006), ADMIXTURE analysis (Alexander, Novembre, & Lange, 2009), and the TreeMix (version 1.12) program (Pickrell & Pritchard, 2012). The TreeMix analyses were performed to generate trees and to estimate the f3 statistic to find an affinity between populations. A window size of 500 SNPs was used for accommodating LD. We used D statistics (Patterson et al., 2012) implemented in a software ADMIXTOOL to estimate the relatedness of ancient ancestry with modern Siberian groups.

TABLE 1 Populations, their geographic locations, and the number of individuals analyzed

		Autosomal SNPs	Y-chromosome SNPs
Population	Code	N	N
Siberia			
Altai-Kizhi	ALT	18	98
Buryat	BUR	16	80
Dolgan	DOL	10	57
Evenk	EVK	13	91
Forest Nentsi	FNE	26	82
Ket	KET	17	44
Khant	KHA	12	165
Komi	KOM	13	78
Koryak	KOR	8	11
Mongol	MON	10	75
Nganasan	NGA	18	34
Selkup	SEL	29	129
Teleut	TEL	10	40
Tundra Nentsi	TNE	11	47
Yakut	YAK	13	62
Yukagir	YUK	7	10
Total	231	1103	
Central Asia			
Kazakh	KAZ	10	29
Tajik	TAJ	11	15
Turkmen	TUR	21	44
Uygur	UYG	12	66
Total		54	154
East Asia			
Han Chinese	HAN	13	40
Korean	KRE	18	63
Manchurian	MAN	14	50
Tibetan	TIB	18	71
Total		63	224
Near East			
Iranian	IRA	27	136
Europe			
German	GER	22	37
South Asia			
Indian	IND	27	51
Africa			
Dinka	YRI	17	
Totals		441	1705

2.3 | Y chromosome analyses

A total of 1705 samples belonging to the abovementioned 27 populations were analyzed for 147 polymorphic sites on the nonrecombining portion of the human Y chromosome by TaqMan, allele-specific PCR or RFLP (Supporting Information Table S2)**.** We use the mutation-based naming system that keeps the major haplogroup information followed by the name of the terminal mutation that defines a given haplogroup (Karafet et al., 2008). Genotyping data resulted in 97 Y haplogroups (Supporting Information Table S3). We also analyzed 12 short tandem repeats (STRs): *DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426*, and *DYS439* as described by Redd et al. (2002). Y chromosome STR data are provided in Supporting Information Table S4. Genetic distances between populations, F_{ST} and Mantel tests were computed using Arlequin v. 3.5.1.12 (Excoffier & Lischer, 2010).

3 | RESULTS

3.1 | Genetic variation: PCA, F_{ST}, and ADMIXTURE analyses

To examine genetic relatedness among the samples, we applied PCA on the merged autosomal data set using a "drop one in" procedure (Veeramah et al., 2011), incorporating samples from modern populations along with ancient samples from Siberia (Figure 2). This procedure helps to avoid the potential effect of high relatedness among individuals in isolated populations from Siberia, as well as uneven sample sizes. Our PCA analysis to some extent recaps previously reported patterns of genetic structure across Eurasia (Pugach et al., 2016; Wong et al., 2016). PC1 corresponds to west-east differentiation, while the second PC generally differentiates northern and southern populations. Overall, the positions of populations roughly correspond to their geographical location. PCA demonstrates high differentiation of the Siberian groups, which might be explained by founder effect, endogamy, and genetic drift due to small population sizes and limited gene flow. Consistent with their origin, Mongolic-speaking Buryats demonstrate genetic similarity with Mongols, and Turkic-speaking Altai-Kizhi and Teleuts are drawn close to CAS groups. The Tungusic-speaking Evenks collected in central and eastern Siberia cluster together and overlap with Yukagirs. Dolgans are widely scattered in the plot, justifying their recent origin from one Evenk clan, Yakuts, and Russian peasants in the 18th century (Popov, 1964). Uralic-speaking populations comprise a very wide cluster with Komi drawn to Europe, and Khants showing a closer affinity with Selkups, Tundra and Forest Nentsi. Yenisey-speaking Kets are intermingled with Selkups. Interestingly, Samoyedic-speaking Nganasans from the Taymyr Peninsula form a separate tight cluster closer to Evenks, Yukagirs, and Koryaks. Similar to previous results (Fu et al., 2014), the ancient Ust'-Ishim sample discovered on the banks of the River Irtysh does not reveal genetic affinities with Siberian groups, rather it clusters with South Asian Indians. In the first two principal components, the Mal'ta sample has an intermediate position between modern western Eurasians, CAS groups and western Siberians. Samples from the Eneolithic (ENE) and Bronze Age (BA) Afanasievo, Okunevo, and Andronovo cultures in the Altai-Sayan region form a rather tight cluster with closer

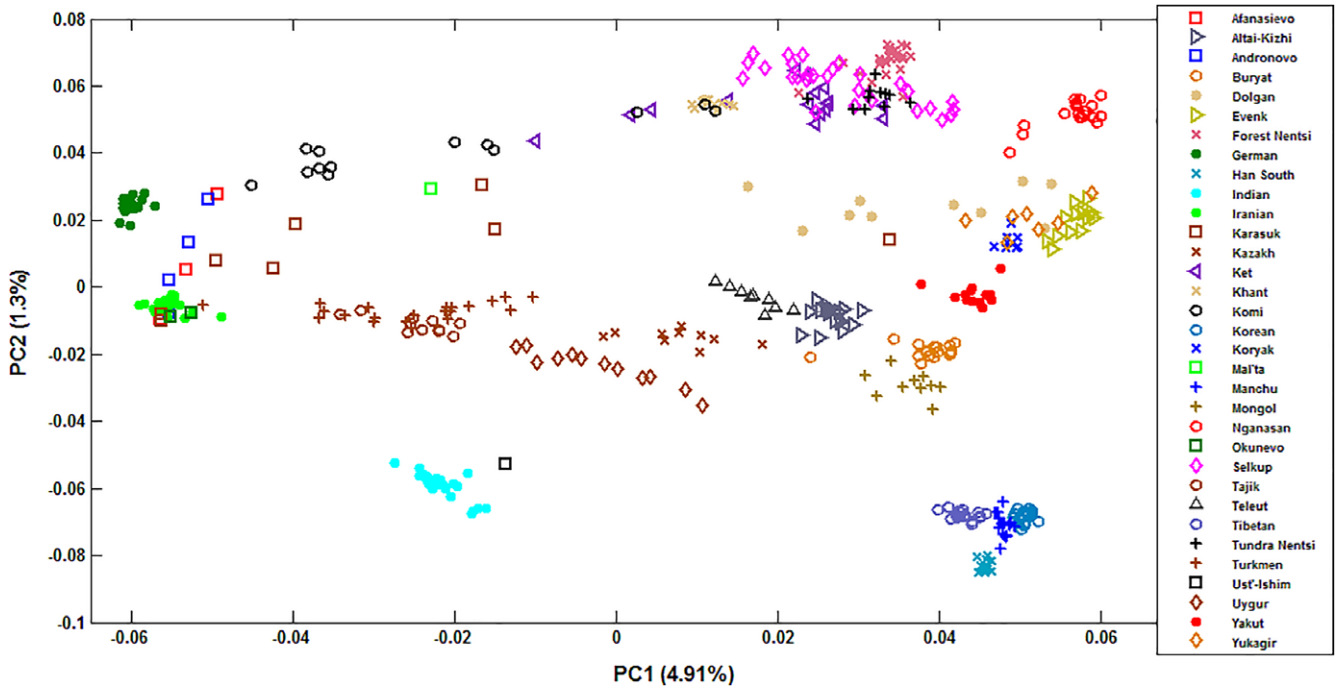

FIGURE 2 Principal component analysis (PCA) analysis using the "drop one in" technique for 27 present-day (N = 424) and 6 ancient populations (N = 20). PCA was performed on 281 093 SNPs from the intersection of our data with publicly available ancient Siberian samples

genetic relationships with present-day Iranians and Germans. Conversely, ancient samples from Karasuk BA culture are broadly dispersed among Europeans, Central Asians, and Siberians.

We performed structure-like analysis with the algorithm ADMIXTURE. This analysis infers the genetic ancestry of each individual, where the individuals are assumed to have originated from K ancestral clusters. The plots for K = 3-7 are shown in Figure 3**,** with the cross-validation error being the lowest for K = 6. At K = 3 clusters broadly correspond to EUR, EAS, and ancient BA culture regions. Among Siberians, the Komi are primarily Europeans, while Nganasans, Evenks, Yukagirs, and Koryaks are nearly 100% East Asians. At K = 4 finer scale subcontinental structure can be distinguished with the emergence of a "Siberian" component. This component is highly pronounced in the Nganasans. Outside Siberia, this component is present in Germany and in CAS at low frequency. Within ancient cultures, this component has the highest frequency in three BA Karasuk samples. It is also found in Mal'ta, ENE Afanasievo and BA Andronovo, but not in Ust'-Ishim and BA Okunevo. At K = 5, the "Siberian" component is roughly subdivided into two components with different geographic distributions. The "Nganasan" component is frequent in nearly all Siberian populations, except the Komi, Kets and Selkups. The newly derived "Selkup-Ket" component is found at high frequencies in western Siberian populations. It is observed in BA Karasuk and in Mal'ta. At K = 6, the western Siberian "Nentsi-Khant" ancestry component was developed in Forest and Tundra Nentsi, Khants. This component is also present at low levels in EUR, CAS, Tibet, and southern Siberia. With K = 7, a single ancestry component dominated in South-Asian Indians. In agreement with PCA, the Ust'-Ishim sample is genetically closely affiliated to Indians. Differentiation among Siberian populations measured by F_{ST} is very high. For example, the mean F_{ST} (0.036) for Siberia was almost identical to the F_{ST} (0.039) for 11 worldwide populations (Wilcoxon rank sum test$P$ value = .582). Samoyedic-speaking populations show a substantial level of variation with an F_{ST} (0.029) that does not differ significantly from the whole of Siberia (P value = .762).

3.2 | Population gene flow: TreeMix, f3, and D statistics

To reconstruct the patterns of population splits and mixtures in the history of the ancestors of present-day Siberian populations, ENE, BA ancient Siberian groups, and populations from other geographic regions, we built a tree using TreeMix using the Dinka as an out-group (Pickrell & Pritchard, 2012). The maximum-likelihood population tree inferred without admixture events places ENE and BA ancient Siberians as a separate branch that is basal to Eurasians (Figure 4). The population tree splits the majority of Siberian populations in two branches: one consisting of Samoyed-speaking Tundra Nentsi, Forest Nentsi, and Selkups with Kets; and a second branch grouping Nganasans, Evenks, Yukagirs, Dolgans, Koryaks, and Yakuts with Mongolians and EAS populations. These two branches are leading to the Altai-Kizhi and Teleuts from the Altai-Sayan region. Allowing three admixture events (Supporting Information Figure S1) finds evidence of admixture between Europeans

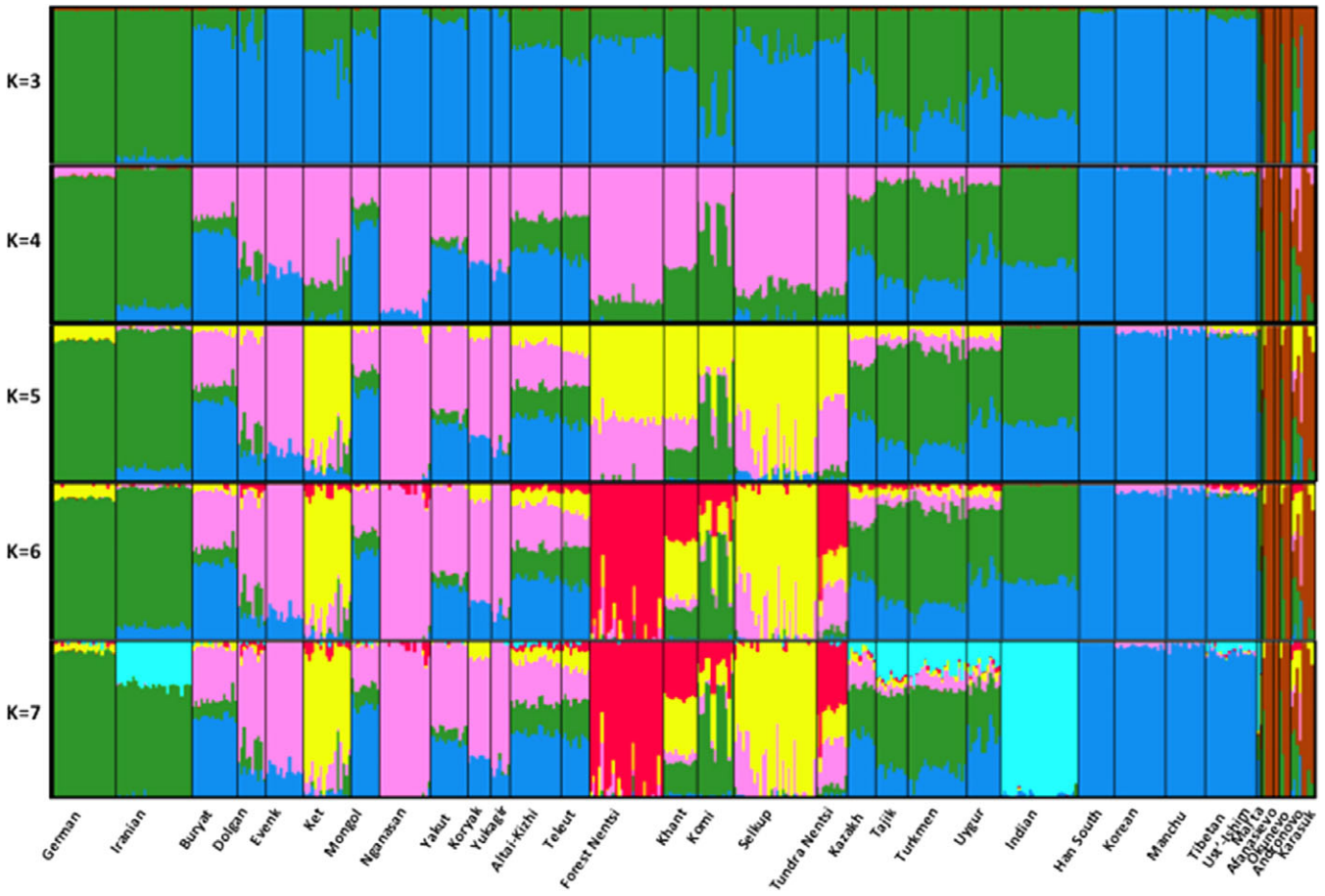

FIGURE 3 ADMIXTURE plot. Clustering of 444 individuals from 27 present-day and 6 ancient populations (281 093 SNPs) assuming K3 to K7 clusters. Individuals are shown as vertical bars colored in ratio to their estimated ancestry within each cluster

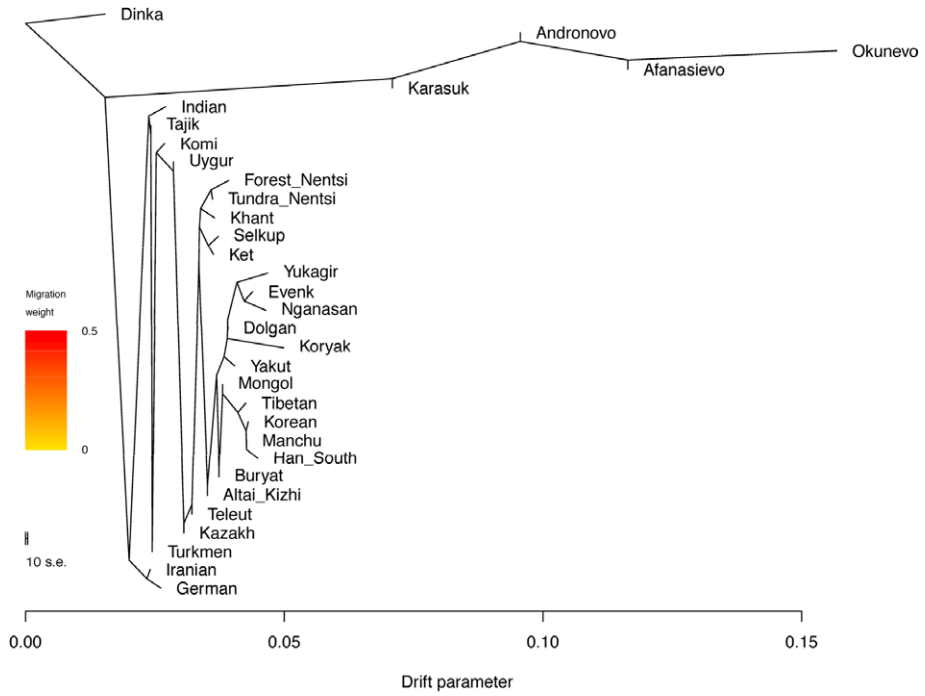

FIGURE 4 Autosomal TreeMix admixture graph (281 093 SNPs) was built for 27 present-day groups from Siberia, Asia, Europe, near east and for four Siberian bronze age ancient populations. Dinka populations were used as an out-group. Tree with admixture events are shown in supporting information Figure S1

and the BA Andronovo culture, between BA Karasuk and the whole cluster of Siberian and East Asian groups, and between the Nganasan and the Samoyed-speaking Tundra and Forest Nentsi, Selkups, and Ket branch.

We also calculated the population divergence time *TF* based on *Ne* and F_{ST} information (Supporting Information Table S5). In agreement with the phylogenetic tree the smallest *TF* estimates within Siberian populations are found between Kets and Selkups (~1.7 KYA) and between Forest and Tundra Nensti (~2 KYA), while Nganasans represent the most diverged Siberian population with the time of separation that varied from the smallest of ~3.0 KYA with Evenks to ~14.3 KYA with the Komi.

We calculated the f3-statistic for all possible triple combinations of 27 modern populations and 6 ancient Siberian cultures to infer evidence of admixture with the *Test* population. A significantly negative value of the f3 statistic implies that the population is admixed. The most negative and the most significant f3-statistics are presented in Supporting Information Table S6. When ancient cultures were used as the *Test group,* Ust'-Ishim, Mal'ta and Okunevo do not produce any negative f3 values suggesting the absence of admixture and/or substantial postadmixture drift. On the other hand, Afanasievo, Andronovo, and Karasuk reveal a signal of admixture. For Afanasievo, Andronovo the most negative f3 statistics (~−0.021-0.027) are observed when we model them as a mixture of Okunevo and Germans, Iranians, or Komi. We should note that the Komi people do not represent autochthonous Siberian populations. Their homeland lies in the north-east of European Russia. Genetically they are close to eastern Slavic populations (Khrunin, Verbenko, Nikitina, & Limborska, 2007). In contrast, the BA Karasuk culture shows a very different pattern of admixture demonstrating substantial shared genetic history with Siberian populations; in particular with the Nganasans, Evenks, Koryaks, and Yukagirs (f3 ≈ −0.036, $Z \approx -192$). This result is consistent with the allele frequency-based D-statistic tests that showed all Siberian populations to be closer to the Karasuk than to the Andronovo culture (Supporting Information Figure S2). While investigating the genetic history of modern Siberian populations, we found that several populations (ie, Nganasans, Evenks, Yukagirs, and Koryaks) do not exhibit a history of admixture events, producing only positive f3 values (Supporting Information Table S6). For the rest of Siberian populations the composition of mixing sources is strikingly similar: one source always includes Nganasans or Evenks, while the other is dominated by Germans or Iranians. Among Siberian populations the most negative statistics were found for the Altai-Kizhi and Teleuts from the Altai-Sayan region, due to a multilayered history of admixture, and Dolgans because of their complex origin.

3.3 | Identity-by-descent analysis and the level of inbreeding (ROH estimation) in Siberian populations

The analysis of shared IBD segments complements analyses of population structure and gene flow. Our analysis was restricted to the segments longer than 3 cM to minimize background LD. The mean IBD sharing between Siberian populations and 11 groups from CAS, EAS, SAS, EUR, and NEA is reported in Supporting Information Tables S7A-E. Overall, the total genome-wide sharing for an average pair of samples from the same population varies greatly from 7.54 cM in Buryats to 163.4 cM in Yukagirs. The majority of shared IBD segments were found within the same population with some exceptions. The Dolgans share more segments with the Nganasans than within themselves (54.13 vs 41.72, Mann-Whitney test, $P = .000000000001562546$). The result is not surprising as the demographic data showed that the Nganasans were subjected to intense assimilation by the Dolgans in the second half of the 20th century (Goltsova, Osipova, Zhadanov, & Villems, 2005). Tundra Nentsi share more IBD with Forest Nentsi than within themselves (83.96 vs 50.3, $P = .000055$) possibly due to the common origin and long-term gene flow. The Ket and Selkup populations allocate significantly more IBD blocks between populations than with individuals from their own population (121.2 cM vs 85.9 cM for Kets, $P = .000008$, and 121.2 cM vs 114.9 cM for Selkups, $P = .043$). Buryats share more IBD segments with Yakuts (9.54 cM vs 8.54 cM, $P = .000117$).

The extensive IBD sharing within populations may be the result of rather small Siberian population sizes. We calculated the Pearson correlation coefficient between IBD sharing and effective population sizes (N_e) estimated with LD data (McEvoy et al., 2011). As expected, the correlation coefficient was found to be negative and highly significant ($r = -0.825$, $P < .0001$). A negative, but not statistically significant correlation ($r = -0.4248$, $P = .1$ for two-tailed probability, $P = .051$ for one-tailed probability) is observed between IBD sharing and census sizes. The distribution of length of shared IBD segments might be a consequence of an isolation-by-distance process. We performed a Mantel test comparing pairwise geographic distances and genetic distances based on the mean IBD segments among populations normalized by sample sizes. A higher and statistically significant negative correlation was observed between geographic distances and the amount of shared IBD ($r = -0.349$, $P = .0006$). To infer the presence of a north-south or east-west gradient of shared IBD segments, we performed correlation analyses for each group between latitude or longitude and the average IBD sharing within all Siberian populations and Mongolia. Interestingly, we found a statistically significant north-south gradient ($r = 0.681$, $P = .002$), and no evidence of an east-west gradient ($r = -0.142$, $P = .3$).

Overall, other than the Altais, Teleuts, and Koryaks, populations within Siberia are characterized by high levels of pairwise IBD sharing (mean total IBD >5 cM). Elevated

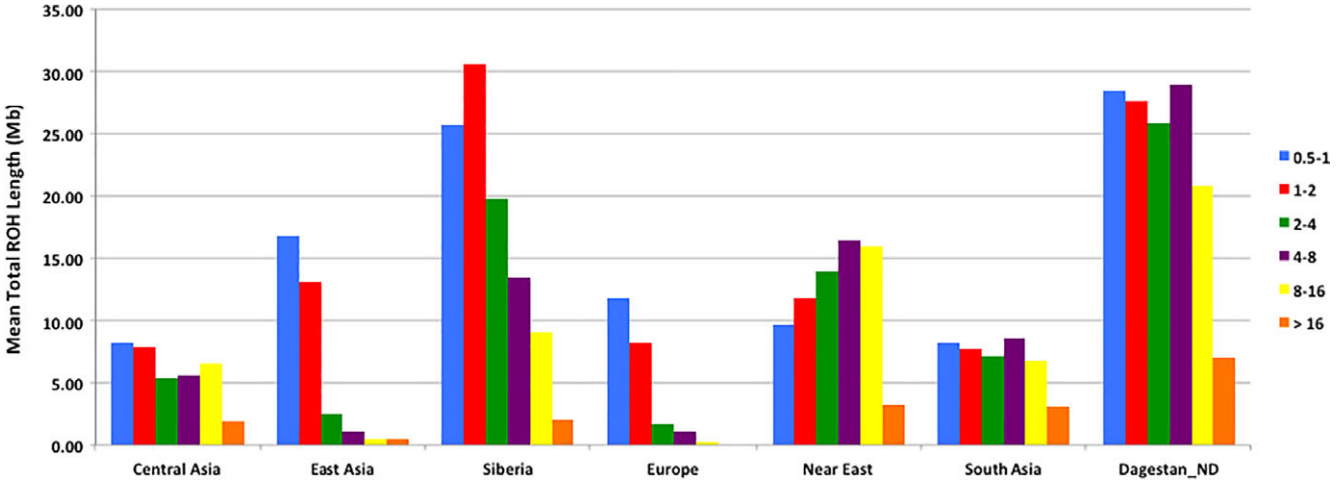

FIGURE 5 Distribution of runs of homozygosity in different geographic regions

IBD sharing in Komi population is due to high genetic relatedness to the Khants through interpopulation marriages. When the Khants are excluded from the analysis, the total average shared IBD is decreased from 5.32 to 3.1 cM. Within Samoyedic-speaking populations, an unusual pattern of sharing IBD is observed. Only two Samoyedic-speaking populations, the Tundra and Forest Nentsi, shared long segments of IBD (84 cM). The Selkups reveal extremely high IBD sharing with the Kets (121.2 cM). Interestingly, the small Nganasan population demonstrates the greatest level of IBD sharing not only with Samoyedic-speaking populations and neighboring Dolgans, but also with nearly all other Siberian populations (average IBD - 17.84 cM, Supporting Information Table S7). The highest sharing is especially evident for smaller IBD segments (<6 cM). It is worth noting that individuals from all Siberian populations, except the Komi, have a very low proportion of their genomes shared with other geographic regions. The greater sharing is found only between Kazakh and Uygur from CAS.

To infer the level of inbreeding and to explore the potential effect of demographic processes on Siberian population structure, we examined the length distribution of ROH in PLINK. In the expanded sample of 424 individuals the population mean of the cumulative ROH length per individual varied from the smallest of 17.5 Mb (SE = 2.9 Mb) in Uygurs to 201.1 Mb (SE = 11.6 Mb) in the Koryaks from northeastern Siberia (Supporting Information Table S8). The Siberian populations have a large proportion of their genome in ROH, relative to the other populations examined here (mean ROH length = 100.7 Mb, SE =9.43). To evaluate the pattern of ROH distribution, we compared Siberian populations with Nakh-Daghestanian-speaking (ND) isolates from Caucasus, well-known for their exceptionally high coefficient of inbreeding (mean ROH length = 128.1 Mb, SE = 6.54), who were genotyped using the same Affymetrix (Axiom) platform (Karafet et al., 2015). The distribution of ROHs according to size (in Mb) clearly shows a high proportion of individuals with extended regions of autozygosity (ROH length 0.5-2 Mb) in both ND-speaking and Siberian populations compared with the individuals in other geographic regions (Figure 5). However, they differ in their frequency of distinctive predefined ROH length categories. While shorter ROH (<4 Mb) are common in both regions, the longer ROH are increasingly prevalent only in ND-speaking samples and those from the Near East, but not in Siberia. Within Siberia, populations also differ in the length of ROH in predefined classes (Supporting Information Figure S3). Nganasans, Koryaks, Yukagirs, and Evenks reveal very large numbers of shorter ROH, but few long ROH, while Forest Nentsi stand out in having more long runs (>4 Mb) than any other populations.

3.4 | Distribution of Y-chromosome haplogroups

Y-chromosome haplogroups are presented in Supporting Information Table S3. Although Siberia exhibits 42 haplogroups, the vast majority of Siberian Y-chromosomes belong only to 4 of the 18 major clades (N = 46.2%; C = 20.9%; Q = 14.4%; and R = 15.2%). The Y-chromosome haplogroup N is widely spread across Siberia and Eastern Europe (Ilumae et al., 2016; Karafet et al., 2002; Wong et al., 2016) and reaches its maximum frequency among Siberian populations such as Nganasans (94.1%) and Yakuts (91.9%). Within Siberia, two sister subclades N-P43 and N-L708 show different geographic distributions. N-P43 and derived haplogroups N-P63 and N- P362 (phylogenetically identical to N-B478* and N-B170, respectively) (Ilumae et al., 2016) are extremely rare in other major geographic regions. Likely originating in western Siberia, they are limited almost entirely to northwest Siberia, the Volga-Uralic regions, and the Taymyr Peninsula (ie, do not extend to eastern Siberia). Conversely, clade N-L708 is frequent in all Siberian populations except the Kets and Selkups, reaching its highest frequency in the Yakuts (91.9%). Haplogroups derived from N-L708 are broadly spread not only in Siberia, but also in other regions of northern Eurasia

(Ilumae et al., 2016). Haplogroup C is dominant in Buryats, Evenks, and Yukagirs, reaching frequencies of 61.3%, 69.2% and 80%, respectively. This haplogroup is distributed at low or moderate frequency in CAS and EAS. The major Native American haplogroup Q was observed in 10 Siberian populations. However, the majority of haplogroup Q chromosomes occurred only in two Siberian populations, the Kets and the Selkups, with frequencies of 97.7% and 66.7%, respectively. The occurrence of haplogroup R in Siberia is mostly limited to the Altai region and western Siberia, with the highest frequency found in the Teleuts (70%) and Altai-Kizhi (46.8%). This haplogroup is widely distributed in EUR, CAS, and SAS.

It is interesting to note that Samoyedic-speaking populations are not characterized by a distinct set of founder Y haplogroups. The frequency of N-P63 is high in Nganasans (88.8%) and moderate in Forest Nentsi (37.8%). Haplogroups N-P362 and N-L1034 are frequent in Tundra Nentsi (51.1% and 23.4%, respectively). Haplogroup Q-L54 with a frequency of 66.6% is dominant in Selkups. To explore regional relationships, multidimensional scaling based on Y-chromosome SNPs was performed on 27 populations (Supporting Information Figure S4). Similar to autosomal SNPs, Y chromosome markers reveal close relationships between Forest and Tundra Nentsi, and between Sekups and Kets. Contrary to analysis of autosomal SNPs, Nganasans represent an outlier not showing genetic similarity to Dolgans and Evenks.

We assessed associations between autosomal and Y-chromosomal population structure of Siberia by correlating matrices of genetic distances. For Y-chromosomal data, we used distances based on Y-SNP and Y-STR frequencies. No significant correlations were found between Y-chromosome and autosomal SNP structure for both data sets ($r = 0.028$, $P = .557$; $r = 0.073$, $P = .302$). Similar to autosomal data, F_{ST} statistics based on Y chromosome data indicated a significant degree of population differentiation within Siberia (0.367 for SNPs, 0.296 for STRs). When all 27 populations were included, the F_{ST} value rose to 0.409 for SNP haplogroups, but lowered for STR haplotypes (0.236). F_{ST} increased to 0.544 when only Samoyedic populations were used for analysis.

4 | DISCUSSION

In this study, we used genome-wide and Y-chromosome data to describe the population structure of indigenous Siberian groups and to clarify questions about their origins with a particular interest in Samoyedic-speaking populations.

4.1 | Genetic structure of modern Samoyeds at the Siberian landscape

The PCA, ADMIXTURE, and F_{ST} results based on autosomal SNPs (Figures 2 and 3) and Y-chromosome data demonstrate that Samoyedic populations differ significantly in their genetic composition. Genetic affinities are observed between Forest and Tundra Nentsi; Selkups are related to Kets from Yenisey River, while Nganasans are separated from their linguistic neighbors, showing closer relationships with the Evenks and Yukagirs (Figure 4).

The Siberian pattern of high-level differentiation among populations may be a general feature of indigenous groups that have small effective population sizes and that have been isolated for long periods of time. The long-tract ROH and IBD analyses seem to be particularly relevant to a history of isolation and the reduced population sizes of Siberian populations. Recent parental relatedness within four to six generations affects ROH over 5 Mb in length (O'Dushlaine et al., 2010). Thus, it is not surprising that Forest Nentsi have the longest runs of homozygosity among Siberian populations (Supporting Information Figure S3). Eight percent of all Forest Nentsi marriages were those between close relatives such as uncle-niece, first cousins and double first cousins, with a genealogy-based coefficient of inbreeding of 0.012 (Abanina & Sukernik, 1980). Nonetheless, the majority of Siberian populations—in particular the Nganasans, Koryaks, Yukagirs, and Evenks—have large numbers of shorter ROH, but few long ROH. The increased numbers of shorter ROH might be the result of common extended haplotypes in small, isolated communities with reduced N_e in the past and with little recent inbreeding as they are frequently inherited from both parents (Kirin et al., 2010). The sharp reduction in numbers of Siberian populations in the 16th-17th centuries was documented as related to the incidence of epidemic diseases, especially smallpox, which had been unknown in Siberia before the advent of the Russians (Forsyth, 1992). In agreement with ROH analysis, the majority of Siberian populations also reveal an increased number of within-population shared IBD segments with high correlation coefficient between IBD sharing and *Ne*. Similar to previous studies (Pugach et al., 2016; Yunusbayev et al., 2015), Nganasans demonstrates unusually high IBD sharing with other Siberian populations. The fact that a majority of modern Siberian people share excessive IBD sharing (and possibly the same set of ancestors) seems to contradict long-term isolation and very high genetic differentiation observed within Siberia. It was suggested that large-scale population dispersals and admixture are likely to have played a substantial role in the prehistory and population structure in Siberia and might explain the unusually high IBD sharing (Pugach et al., 2016). However, the large number of short ROH and excess IBD sharing may also be a sign of positive selection. It was shown that IBD sharing might detect not only selection acting on a new allele, but also selection acting on standing variation (ie, selection on alleles that were already segregating in the population when the selective advantage was introduced) (Albrechtsen, Moltke, & Nielsen, 2010; Di Gaetano et al., 2014). When the selected allele has achieved high frequencies in the population, there is a marked increase in

IBD (Albrechtsen et al., 2010). The successful long-term settlement of Siberia undoubtedly demanded multiple adaptations to cope with the extended severe cold stress, dramatic variation in photoperiod, and limited food resources experienced by these populations. Several studies found evidence of natural selection in Siberia and identified candidate gene sets with Siberian-specific signals of selection (Cardona et al., 2014; Clemente et al., 2014; Fumagalli et al., 2015; Hallmark, Karafet, Hsieh, Osipova, & Hammer, 2017; Hsieh et al., 2017). We found that the local density of IBD blocks of all lengths is not constant across the genome (data not shown). Of particular interest may be two very commonly shared regions in Siberian populations: first one on chromosome 15q26.1 and a region on chromosome 1p36.13. A total of 123 out of 221 Siberian samples revealed IBD overlapping blocks from 3.5 to 43.4 cM, which encompass the gene PLIN1 on chromosome 15: the encoded protein is a modulator of adipocyte lipid metabolism. The region on chromosome 1p36.13 had IBD segments with the length 3.23-7.05 cM in 107 samples from all Siberian populations except Koryaks. This block covers the gene PLA2G2A, which participates in the regulation of phospholipid metabolism. Both genes are related to diet, particularly to fat metabolism, and demonstrate genomic and spatial patterns consistent with selection for cold climate and/or diet in Siberia (Hallmark et al., 2017). A statistically significant correlation between latitude and the average IBD sharing within all Siberian populations and Mongolia might be indirect support for the action of widespread, polygenic selection in Siberia.

4.2 | The origin of Samoyed populations

Most Russian ethnographers and anthropologists believe that all Samoyed-speaking peoples are descendants of indigenous Paleo-Asiatic populations that were assimilated by ancient Samoyeds who came from southern Siberia in the middle of the first millennium AD (Napolskikh, 1995; Prokof'ev, 1940; Vasil'ev, 1985). The dispersal of ancient Samoyedic-speaking reindeer herders and their languages took place in the lands occupied by local hunters in a vast territory of western Siberia and the Taymyr Peninsula. The interactions between aboriginal peoples and the Samoyedic newcomers might have differed in various areas, resulting in diverse degrees of genetic mixing between these groups and cultural assimilation of the Samoyedic language. Prokof'ev (1940) suggested that Nganasans, Nentsi, and Entsi were descendants of one aboriginal tribe, while Selkups shared distinct Paleo-Asiatic components with Ugric-speaking Khants and Mansi. However, questions like, who were those autochthonous tribes, how different were they, and when did they inhabit these geographic areas, remain open.

It was generally accepted that the initial C^{14}-dated Asian Upper Paleolithic industries occurred in the Altai Mountains of southwest Siberia at 43 300 ± 1600 years BP and spread to the east and north much later (Goebel, Derevianko, & Petrin, 1993; Kuzmin, Jull, Orlova, & Sulerzhitsky, 1998). The earliest evidence of human presence in the territory occupied by contemporary Samoyedic-speaking populations is sparse. It was believed that in the Upper Paleolithic (since ca. 30 000 BP) people occupied just the periphery of West Siberian Plain and Trans-Urals (Zenin, 2002). Only in the early Holocene did people inhabit the whole of western Siberia. The initial human settlement of Taymyr was previously dated not earlier than 6000 years BP (Khlobystin, 2005). Two recent archeological findings suggest that people lived across northern Siberia much earlier than the previously thought. The oldest anatomically modern human genome, found recently at 57°N in western Siberia, produced a direct radiocarbon date of 45 000 years BP (Fu et al., 2014; Kuzmin et al., 2009). A mammoth kill site on the Taymyr Peninsula, dated to 45 000 years BP, expanding the populated area to almost 72°N (Pitulko et al., 2016). This is a rare case of indisputable evidence for human involvement without artifact association.

These considerations, along with the emerging picture of culture similarities, environmental conditions and lifeways among Paleolithic Siberians, suggest that "a continental culture of sedentary Arctic Paleolithic hunters" existed in the north of Siberia and Eastern Europe by the end of the glacial period (Bogoraz, 1929; Okladnikov, 1964). The advancement of mammoth hunting probably allowed people to survive and spread widely across northernmost Arctic Siberia (Pitulko et al., 2016). Local groups unquestionably could not survive in isolation, as they needed to access crucial resources clustered irregularly over landscapes. Benefitting from information preserved over generations and contacts with the neighboring groups, it is likely that they resorted to foraging-areas exploration (Rolland, 2014). Simchenko (1968) developed this hypothesis further and suggested that by the end of the glacial period all ancient Arctic populations from the Kola Peninsula up to the Bering Strait represent an ethnically homogeneous ancient aboriginal population. He posited that modern northern Samoyeds (Nganasans, Entsi, and Nentsi), Yukagirs, Evenks, Evens, Chukchi, and Koryaks originated from the same or closely related ancestral tribe(s). The language of those ancient populations is believed to be related to the Uralic linguistic branch.

Our results are compatible with this hypothesis. PCA and a maximum likelihood tree (with and without migration events) cluster northern populations (Nganasans, Yukagirs, Evenks, and Koryaks) together despite the large geographic distances separating them (Figures 2 and 4). ADMIXTURE analysis demonstrates that the "Siberian" component that is highly pronounced in Nganasans, is also frequent in northern Siberian populations (Figure 3**, K = 4**). Positive f3 values do not support a history of admixture between these arctic and subarctic populations (Supporting Information Table S6). At the same time, this analysis implicated Nganasans and/or Evenks as a mixing source population for the

other Siberian groups. These results are consistent with our finding that the Nganasans and Evenks exhibit the highest level of smaller-IBD-segment sharing with nearly all other Siberian populations (Supporting Information Table S7). Indirect support for the existence of prehistoric Arctic populations also derives from Y chromosome and mtDNA data. The oldest modern human genome, Ust'-Ishim, belongs to the NO clade (Wong et al., 2016), the predecessor of the most frequent north Eurasian haplogroup N. Similarly, mtDNA data revealed an affinity between samples from ancient Mesolithic Yuzhnyy Oleni Island, Karelia, from the periphery of Eurasian Arctic (uncal. 7000-7500 years BP) and present-day northwestern Siberian populations (Sarkissian et al., 2013).

Linguistic historical reconstructions point to two areas as the proto-Samoyed original home: a wide forest region between the Middle Ob and Yenisey Rivers that matches the Kulay archeological culture of 500 BC-500 AD (Helimsky, 1990) and the Tagar culture in the Minusinsk basin on the Upper Yenisey (Junhunen, 2009). Interestingly, it remains debated whether there were migrations among settled groups of Kulay and Tagar cultures (Bobrov, 2011). The Minusinsk basin region has an exceptionally well-documented direct sequence of archeological cultures, extending from the Eneolithic Afanasievo culture (3500-2500 BC) through the Bronze and Iron Age Okunevo (2500-2000 BC), to the Andronovo (2000-1500 BC), Karasuk (1500-800 BC), Tagar (800-100 BC), and Tashtyk (100 BC-400 AD) cultures, and continuing up to the historical Yenisei Kirghiz (from 400 AD), and Mongols (from 1300 AD). Autosomal genome data from 18 ancient samples from the Afanasievo, Okunevo, Andronovo, and Karasuk archeological sites have recently become available (Allentoft et al., 2015). Yeniseian-speaking Kets were suggested to be connected to Karasuk culture (Chlenova, 1972; van Driem, 2007).

The most controversial among archeologists is the origin of the Karasuk culture. It is commonly accepted that Karasuk culture represents the continuation and transformation of the Okunevo and Andronovo cultures, along with invasion of new migrants. Indeed, mtDNA and autosomal SNP data indicate a significant shift in the genetic structure of Karasuk in comparison with Afanasievo, Okunevo, and Andronovo cultures (Allentoft et al., 2015; Keyser et al., 2009). However, there is no consensus about the origin of migrants in the Karasuk culture. Many suggestions of the southern source population have been put forward, including Central Asia, the steppes of Kazakhstan, Mongolia, and China (Chlenova, 1972; Jettmar, 1950). A northern forest origin of Karasuk has been suggested, based on similarities with artistic traditions of the mature Bronze Age (Martynov, 1991). In our study, we found no support for migrations from East Asia to the Karasuk. Quite the opposite, ADMIXTURE, PCA, f3, and D statistical analyses provided several lines of evidence that show closer affinity of the Karasuk culture to modern Siberian populations (Figures 2 and 3**,** Supporting Information Figure S2, Table S6). However, any results based on the comparison of ancient cultures with modern populations have to be taken with caution since samples from the actual ancestral populations currently are not available.

The Tagar culture is commonly believed to have developed out of the traditions of the previous Bronze Age Karasuk culture without significant immigration (Bokovenko, 2006). Autosomal data on ancient Tagar samples are not yet available. We cannot determine whether Karasuk and Tagar populations were genetically close. However, the proposed descendant populations of Kets and Selkups show very close connections. They clustered together in the PCA plot and tree, they shared a significant amount of long and short IBD blocks, and their divergence is estimated at ~1.7 KYA. This is likely an underestimate because the divergence model does not take into consideration migrations between the Kets and Selkups. However, this divergence time does coincide with the suggested time when ancestors of Selkups migrated to the north during the early part of the first millennium AD (Prokof'yeva, 1964).

In sum, our data are consistent with the following hypotheses about the origin of modern Samoyedic populations: (1) The Nganasans are very likely direct descendants of the ancient sedentary Arctic Paleolithic hunters of northern Asia and Neolithic hunters of wild deer who were culturally assimilated by various Samoyedic peoples. The Nganasan language and material culture suggest prehistoric contacts with proto Tungus and Yukagir populations (Anikin & Helimsky, 2007; Simchenko, 1968). (2) Our data are compatible with the hypothesis that Tundra and Forest Nentsi are descended from the same aboriginal tribe that gave rise to the Nganasans who were assimilated by ancient Samoyeds (Prokof'ev, 1940). PCA, ADMIXTURE analyses, and treemix tree reveal close genetic ties between the Tundra and Forest Nentsi, while ROH and IBD analyses indicate that these two populations share more short IBD segments (<6 cM) with other populations than with themselves—most likely as a result of a common origin and long-term gene flow. In structure analysis, the "Nganasan" component is present at a higher frequency in the Tundra Nentsi (51%), than in the Forest Nentsi (46%) (Figure 3). Both populations also share the common Y-chromosome P63 mutation with Nganasans (Tundra Nentsi - 63%, Forest Nentsi - 49%). Moreover, evidence of higher Samoyedic than autochthonous proportion in the Forest Nentsi comes from information about their clan affiliation. The patrilineal clan system was universal for almost all Siberian populations. Based on the names of clans, toponymics and legends, ethnographers traced clan origins either to ancient Samoyeds or to aboriginal tribes in the Forest Nentsi population (Vasil'ev, 1977). Among 1552 Forest Nentsi studied in 1978 (Abanina &

Sukernik, 1980) 58% traced back to Samoyedic clans, while 42% belonged to the clans of aboriginal origin.

(3) The composition of Selkup ancestry was suggested to be different from that of the Nganasans or of the Nentsi. Prokof'ev (1940) proposed that the Selkups shared the same Paleo-Asiatic components with Ugric-speaking Khants and Mansi. Pelikh (1972) argued that in addition to Paleo-Asiatic people in western Siberia and Samoyeds from southern Siberia some other ancestral components could be identified in the Selkups. Vasil'ev (1983) hypothesized that the Selkups did not assimilate the aboriginal people of western Siberia but rather represent the direct descendants of southern Samoyeds. We found some support for Vasil'ev's hypothesis. The Selkups differ significantly from the Nganasans, Forest and Tundra Nentsi, sharing common genetic features not with the Khants as proposed by Prokof'ev (1940), but with the Kets who speak the only surviving form of the Yeneseian linguistic family (Figures 2–4). Y-chromosome data also demonstrate the close genetic affinity of the Selkups and the Kets (Supporting Information Figure S4). IBD analysis revealed that the Selkups exhibit the highest IBD sharing with the Kets in all segment classes (Supporting Information Table S7). This may suggest not only recent gene flow between populations, but also their common ancestry or ancient admixture.

ACKNOWLEDGMENTS

This study was supported by National Science Foundation grant to T.M.K. and M.F.H. (PLR-1203874), and by the State Research Project (No 0324-2018-0016) to L.P.O.

AUTHOR CONTRIBUTIONS

TK and LO collected samples. TK and MH designed the study. TK, OS, and BH analyzed the data. TK drafted the manuscript. MH and LP edited the manuscript and provided critical comments on the manuscript.

ORCID

Michael F. Hammer https://orcid.org/0000-0003-0172-429X

LITERATURE CITED

Abanina, T. A., & Sukernik, R. I. (1980). Population structure of the Forest Nentsi. Results of genealogical research. *Genetika*, *16*(1), 156–164.

Albrechtsen, A., Moltke, I., & Nielsen, R. (2010). Natural selection and the distribution of identity-by-descent in the human genome. *Genetics*, *186*(1), 295–U486. https://doi.org/10.1534/genetics.110.113977

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., & Willerslev, E. (2015). Population genomics of bronze age Eurasia. *Nature*, *522*(7555), 167–172. https://doi.org/10.1038/nature14507

Anikin, A. E., & Helimsky, E. A. (2007). *Samoyedic-Tungus-Manchu lexical relations*. Moscow: Nauka.

Behar, D. M., Metspalu, M., Baran, Y., Kopelman, N. M., Yunusbayev, B., Gladstein, A., & Rosenberg, N. A. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Human Biology*, *85*(6), 859–900.

Bobrov, V. V. (2011). *Tagar Culture in North Forest-Steppe*. Novosibirsk: Paper presented at the Terra Scythica.

Bogoraz, W. G. (1929). Elements of the culture of circumpolar zone. *American Anthropologist*, *31*(4), 579–601.

Bokovenko, N. (2006). The emergence of the Tagar culture. *Antiquity*, *80*(310), 860–879.

Cardona, A., Pagani, L., Antao, T., Lawson, D. J., Eichstaedt, C. A., Yngvadottir, B., … Kivisild, T. (2014). Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS One*, *9*(5), e98076. https://doi.org/10.1371/journal.pone.0098076

Chlenova, N. L. (1972). *The Chronology of the monuments in Karasuk Epoch*. Moscow: Nauka.

Clemente, F. J., Cardona, A., Inchley, C. E., Peter, B. M., Jacobs, G., Pagani, L., & Kivisild, T. (2014). A selective sweep on a deleterious mutation in CPT1A in Arctic populations. *American Journal of Human Genetics*, *95*(5), 584–589. https://doi.org/10.1016/j.ajhg.2014.09.016

Derbeneva, O. A., Starikovskaya, E. B., Volodko, N. V., Wallace, D. C., & Sukernik, R. I. (2002). Mitochondrial DNA variation in the Kets and Nganasans and its implications for the initial peopling of northern Eurasia. *Russian Journal of Genetics*, *38*(11), 1316–1321. https://doi.org/10.1023/A:1021111530654

Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Litvinov, A., Grzybowski, T., … Zakharov, I. (2014). Western Eurasian ancestry in modern Siberians based on mitogenomic data. *BMC Evolutionary Biology*, *14*, 1–11. https://doi.org/10.1186/s12862-014-0217-9

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., & Zakharov, I. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *American Journal of Human Genetics*, *81*(5), 1025–1041. https://doi.org/10.1086/522933

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Rogalla, U., Perkova, M., … Zakharov, I. (2010). Origin and post-glacial dispersal of mitochondrial DNA Haplogroups C and D in northern Asia. *PLoS One*, *5*(12), e15214. https://doi.org/10.1371/journal.pone.0015214

Di Gaetano, C., Fiorito, G., Ortu, M. F., Rosa, F., Guarrera, S., Pardini, B., … Matullo, G. (2014). Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection. *PLoS One*, *9*(3), e91237. https://doi.org/10.1371/journal.pone.0091237

Duggan, A. T., Whitten, M., Wiebe, V., Crawford, M., Butthof, A., Spitsyn, V., … Pakendorf, B. (2013). Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One*, *8*(12), e83570. https://doi.org/10.1371/journal.pone.0083570

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fedorova, S. A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., … Villems, R. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of Northeast Eurasia. *BMC Evolutionary Biology*, *13*, 1–18. https://doi.org/10.1186/1471-2148-13-127

Flegontov, P., Changmai, P., Zidkova, A., Logacheva, M. D., Altinisik, N. E., Flegontova, O., … Tatarinova, T. V. (2016). Genomic study of the Ket: A paleo-Eskimo-related ethnic group with significant ancient north Eurasian ancestry. *Scientific Reports*, *6*, 1–12. https://doi.org/10.1038/srep20768

Forsyth, J. (1992). *History of the People of Siberia*. Cambridge: Cambridge University Press.

Fu, Q. M., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., … Paabo, S. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, *514*(7523), 445–449. https://doi.org/10.1038/nature13810

Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jorgensen, M. E., & Nielsen, R. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, *349*(6254), 1343–1347. https://doi.org/10.1126/science.aab2319

Goebel, T., Derevianko, A. P., & Petrin, V. T. (1993). Dating the middle-to-upper-Paleolithic transition at Kara-Bom. *Current Anthropology*, *34*(4), 452–458. https://doi.org/10.1086/204192

Goltsova, T. V., Osipova, L. P., Zhadanov, S. I., & Villems, R. (2005). The effect of marriage migration on the genetic structure of the Taimyr Nganasan population: Genealogical analysis inferred from MtDNA markers. *Russian Journal of Genetics*, *41*(7), 779–788. https://doi.org/10.1007/s11177-005-0160-3

Gubina, M. A., Kulikov, I. V., Babenko, V. N., Chikisheva, T. A., Romashchenko, A. G., Voevoda, M. I., & Molodin, V. I. (2016). The dynamics of the composition of mtDNA haplotypes of the ancient population of the Altai Mountains from the early bronze age (3rd millennium BC) to the iron age (2nd-1st centuries BC). *Russian Journal of Genetics*, *52*(1), 93–106. https://doi.org/10.1134/S1022795416010063

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., … Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211. https://doi.org/10.1038/nature14317

Hajdú, P. (1963). *The Samoyed Peoples and Languages*. The Hague, The Netherlands: Indiana University.

Hallmark, B., Karafet, T. M., Hsieh, P., Osipova, L. P., & Hammer, M. F. (2017). *Evidence of Local Adaptation at Three Loci in Indigenous Siberians* Molecular Biology and Evolution.

Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., & Zegura, S. L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Molecular Biology and Evolution*, *18*(7), 1189–1203.

Helimsky, E. A. (1990). The external connections and early contacts of the Uralic languages. *Problems of Uralistics*, *1*, 19–43.

Hollard, C., Keyser, C., Giscard, P. H., Tsagaan, T., Bayarkhuu, N., Bemmann, J., & Ludes, B. (2014). Strong genetic admixture in the Altai at the middle bronze age revealed by uniparental and ancestry informative markers. *Forensic Science International. Genetics*, *12*, 199–207. https://doi.org/10.1016/j.fsigen.2014.05.012

Hsieh, P., Hallmark, B., Watkins, J., Karafet, T., Osipova, L., Gutenkunst, R., & Hammer, M. (2017). *Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations.* (MBE-16-1199). Molecular Biology and Evolution.

Ilumae, A. M., Reidla, M., Chukhryaeva, M., Jarve, M., Post, H., Karmin, M., & Rootsi, S. (2016). Human Y chromosome Haplogroup N: A non-trivial time-resolved Phylogeography that cuts across language families. *American Journal of Human Genetics*, *99*(1), 163–173. https://doi.org/10.1016/j.ajhg.2016.05.025

Jettmar, K. (1950). The Karasuk culture and its south-eastern affinities. *Bulletin of the Museum of Far Eastern Antiquities*, *22*(1), 83–126.

Junhunen, J. (2009). Proto-Uralic—What, where, and when? In J. Ylikoski (Ed.), *The Quasquicentennial of the Finno-Ugrian Society* (pp. 79–94). Helsinki: Suomalais-Ugrilainen Seura.

Karafet, T. M., Bulayeva, K. B., Bulayev, O. A., Gurgenova, F., Omarova, J., Yepiskoposyan, L., & Hammer, M. F. (2015). Extensive genome-wide autozygosity in the population isolates of Daghestan. *European Journal of Human Genetics*, *23*, 1405–1412. https://doi.org/10.1038/ejhg.2014.299

Karafet, T. M., Bulayeva, K. B., Nichols, J., Bulayev, O. A., Gurgenova, F., Omarova, J., & Hammer, M. F. (2016). Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan. *Journal of Human Genetics*, *61*(3), 181–191. https://doi.org/10.1038/jhg.2015.132

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*, *18*(5), 830–838. https://doi.org/10.1101/gr.7172008

Karafet, T. M., Osipova, L. P., Gubina, M. A., Posukh, O. L., Zegura, S. L., & Hammer, M. F. (2002). High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Human Biology*, *74*(6), 761–789. https://doi.org/10.1353/hub.2003.0006

Karafet, T. M., Osipova, L. P., & Hammer, M. F. (2008). The effect of history and life-style on genetic structure of north Asian populations. *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, *5*, 395–415.

Karafet, T. M., Osipova, L. P., Kazakovtseva, M. A., Posukh, O. L., & Hammer, M. H. (2001). A history of Siberian native people from the perspectives of male lineages. *American Journal of Physical Anthropology*, 88–88.

Karafet, T. M., Osipova, L. P., Posukh, O. L., Wiebe, V., & Hammer, M. F. (1999). Y chromosome microsatellite haplotypes and the history of Samoyed-speaking populations in Northwest Siberia. In B. D. Goldstein & C. Schlötterer (Eds.), *Microsatellites: Evolution and applications* (pp. 249–265). Oxford: University Press.

Keyser, C., Bouakaze, C., Crubezy, E., Nikolaev, V. G., Montagnon, D., Reis, T., & Ludes, B. (2009). Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Human Genetics*, *126*(3), 395–410. https://doi.org/10.1007/s00439-009-0683-0

Khlobystin, L. P. (2005). Taymyr: The archaeology of northernmost Eurasia. In W. W. Fitzhugh & V. V. Pitulko (Eds.), *Contributions to Circumpolar Anthropology* (Vol. 5). Washington, D.C.: Arctic Studies Center, National Museum of Natural History, Smithsonian Institution.

Khrunin, A., Verbenko, D., Nikitina, K., & Limborska, S. (2007). Regional differences in the genetic variability of Finno-Ugric speaking Komi populations. *American Journal of Human Biology*, *19*(6), 741–750. https://doi.org/10.1002/ajhb.20620

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One*, *5*(11), e13996. https://doi.org/10.1371/journal.pone.0013996

Kuzmin, Y. V., Jull, A. J. T., Orlova, L. A., & Sulerzhitsky, L. D. (1998). C-14 chronology of stone age cultures in the Russian Far East. *Radiocarbon*, *40*(2), 675–686.

Kuzmin, Y. V., Kosintsev, P. A., Razhev, D. I., & Hodgins, G. W. (2009). The oldest directly-dated human remains in Siberia: AMS 14C age of talus bone from the Baigara locality, west Siberian plain. *Journal of Human Evolution*, *57*(1), 91–95. https://doi.org/10.1016/j.jhevol.2009.04.003

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

Lell, J. T., Brown, M. D., Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B., Torroni, A., & Wallace, D. C. (1997). Y chromosome polymorphisms in native American and Siberian populations: Identification of native American Y chromosome haplotypes. *Human Genetics*, *100*(5–6), 536–543. https://doi.org/10.1007/s004390050548

Martynov, A. I. (1991). *The Ancient Art of Northern Asia*. Urbana (IL): University of Illinois Press.

McEvoy, B. P., Powell, J. E., Goddard, M. E., & Visscher, P. M. (2011). Human population dispersal "out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Research*, *21*(6), 821–829. https://doi.org/10.1101/gr.119636.110

Mooder, K. P., Schurr, T. G., Bamforth, F. J., Bazahiski, V. I., & Savel'ev, N. A. (2006). Population affinities of Neolithic Siberians: A snapshot from prehistoric Lake Baikal. *American Journal of Physical Anthropology*, *129*(3), 349–361. https://doi.org/10.1002/ajpa.20247

Napolskikh, V. V. (1995). *Uralic original home: History of studies (Vol. 1, p. 48)*. Izhevsk: Udmurt Institute for history, language and literature.

O'Dushlaine, C. T., Morris, D., Moskvina, V., Gill, M., Corvin, A., Wilson, J. F., & Consortium, I. S. (2010). Population structure and genome-wide patterns of variation in Ireland and Britain. *European Journal of Human Genetics*, *18*(11), 1248–1254. https://doi.org/10.1038/ejhg.2010.87

Okladnikov, A. P. (1964). Ancient population of Siberia and its culture. In M. G. Levin & L. P. Potapov (Eds.), *The Peoples of Siberia* (pp. 13–98). Chicago: University of Chicago Press.

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

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pelikh, G. I. (1972). *The Origin of Selkups*. Tomsk Tomsk University, Tomsk.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pitulko, V. V., Tikhonov, A. N., Pavlova, E. Y., Nikolskiy, P. A., Kuper, K. E., & Polozov, R. N. (2016). Paleoanthropology. Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. *Science*, *351*(6270), 260–263. https://doi.org/10.1126/science.aad0554

Popov, A. A. (1964). The Dolgans. In M. G. Levin & L. P. Potapov (Eds.), *The Peoples of Siberia* (pp. 655–669). Chicago: University of Chicago Press.

Prokof'ev, G. N. (1940). Ethnogony of the ethnic groups on the Ob-Yenisei basin. *Soviet Etnography*, *3*, 67–76.

Prokof'yeva, E. D. (1964). The Nentsy. In M. G. Levin & L. P. Potapov (Eds.), *The Peoples of Siberia* (pp. 547–570). Chicago: University of Chicago Press.

Pugach, I., Matveev, R., Spitsyn, V., Makarov, S., Novgorodov, I., Osakovsky, V., & Pakendorf, B. (2016). The complex admixture history and recent southern origins of Siberian populations. *Molecular Biology and Evolution*, *33*(7), 1777–1795. https://doi.org/10.1093/molbev/msw055

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., … Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of native Americans. *Nature*, *505*(7481), 87–91. https://doi.org/10.1038/nature12736

Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., & Hammer, M. F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Science International*, *130*(2–3), 97–111.

Rolland, N. (2014). The Pleistocene peopling of the north: Paleolithic milestones and thresholds horizons in northern Eurasia: Part I: Lower Paleolithic antecedents. *Archaeology, Ethnology and Anthropology of Eurasia*, *42*(1), 2–17.

Sarkissian, C. D., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., … Consortium, G. (2013). Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of north East Europe. *PLoS Genetics*, *9*(2), e1003296. https://doi.org/10.1371/journal.pgen.1003296

Simchenko, Y. B. (1968). [Some data on ancient ethnic substrate in the composition of North Eurasia peoples] *Problemi antropologii i istoricheskoy etnografii Azii.* Moscow.

Sukernik, R. I., Abanina, T. A., Karafet, T. M., Osipova, L. P., & Galaktionov, O. K. (1979). Population-structure of wooden Nentzi .1. Distribution of blood-groups in 6 sub-isolates. *Genetika*, *15*(2), 327–332.

Sukernik, R. I., Karafet, T. M., & Osipova, L. P. (1977). Genetic structure of Nganasans (Tavgiytsi), an isolated indigenous Group of Northern Siberia, from Taymyr peninsula .2. Analysis of genetic-variation. *Genetika*, *13*(10), 1855–1864.

van Driem, G. (2007). Endangered languages of South Asia. In M. Brenzinger (Ed.), *Language Diversity Endangered* (pp. 303–308). Berlin and New York: Mouton de Gruyter.

Vasil'ev, V. I. (1977). Problems in the ethnogenesis and ethnic history of the peoples of the north (based upon Samodiyan data). *Soviet Ethnography*, (4), 3–17.

Vasil'ev, V. I. (1983). Main problems of ethnic history of Samoyedic peoples (Nentsi, Entsi, Nganasans, Selkups). In *The problems of ethnic history of Samoyedic peoples* (pp. 3–7). Omsk: Nauka.

Vasil'ev, V. I. (1985). Main problems of ethnic history of North Samoyeds *Uralo-Altaistika*. In *Archeology, ethnography, language* (pp. 119–123). Novosibirsk: Nauka.

Veeramah, K. R., Tonjes, A., Kovacs, P., Gross, A., Wegmann, D., Geary, P., & Stumvoll, M. (2011). Genetic variation in the sorbs of eastern Germany in the context of broader European genetic diversity. *European Journal of Human Genetics*, *19*(9), 995–1001. https://doi.org/10.1038/ejhg.2011.65

Volodko, N. V., Starikovskaya, E. B., Mazunin, I. O., Eltsov, N. P., Naidenko, P. V., Wallace, D. C., & Sukernik, R. I. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *American Journal of Human Genetics*, *82*(5), 1084–1100. https://doi.org/10.1016/j.ajhg.2008.03.019

Wong, E. H., Khrunin, A., Nichols, L., Pushkarev, D., Khokhrin, D., Verbenko, D., & Valouev, A. (2016). Reconstructing genetic history of Siberian and northeastern European populations. *Genome Research*, *27*, 1–14.

Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., … Villems, R. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genetics*, *11*(4), e1005068. https://doi.org/10.1371/journal.pgen.1005068

Zenin, V. N. (2002). Major stages in the human occupation of the west Siberian plain during the Paleolithic. *Archaeology, Ethnology and Anthropology of Eurasia*, *3.4*(12), 22–44.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: 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;e23194. https://doi.org/10.1002/ajhb.23194