



Paternal origin of Tungusic-speaking populations: Insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86

Bing-Li Liu¹ | Peng-Cheng Ma² | Chi-Zao Wang³ | Shi Yan⁴ |
Hong-Bing Yao⁵ | Yong-Lan Li⁶ | Yong-Mei Xie⁶ | Song-Lin Meng⁷ |
Jin Sun⁸ | Yan-Huan Cai⁸ | Sarengaowa Sarengaowa⁸ | Hui Li^{3,4,9} |
Hui-Zhen Cheng⁸ | Lan-Hai Wei^{8,9}

¹Institute of Chinese and Culture Education Studies, Huaqiao University, Xiamen, China

²School of Life Sciences, Jilin University, Changchun, China

³MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

⁴Human Phenome Institute, Fudan University, Shanghai, China

⁵Key Laboratory of Evidence Science of Gansu Province, Gansu University of Political Science and Law, Lanzhou, China

⁶Laboratory for Human Biology and Human Genetics, Institute of Ethnology and Anthropology, School of Ethnology and Anthropology, Inner Mongolia Normal University, Hohhot, China

⁷School of History and Ethnic Culture, Hulunbuir University, Hulunbuir, China

⁸Department of Anthropology and Ethnology, Institute of Anthropology, Xiamen University, Xiamen, China

⁹B&R International Joint Laboratory for Eurasian Anthropology, Fudan University, Shanghai, China

Correspondence

Hui-Zhen Cheng and Lan-Hai Wei,
Department of Anthropology and
Ethnology, Xiamen University, #422
South Siming Rd., Siming District,
Xiamen 361005, China.
Email: hellen@xmu.edu.cn (H.-Z. C.) and
ryan.lh.wei@gmail.com (L.-H. W.)

Funding information

Fundamental Research Funds for the
Central Universities, Grant/Award
Numbers: 19SKGC-QG04, 20720191047;
National Key Research and Development
Project, Grant/Award Number:
2016YFC0900300; National Natural
Science Foundation of China, Grant/
Award Numbers: 31671297, 31900406,
91731303; Scientific and Technology
Committee of Shanghai Municipality,
Grant/Award Number: 18490750300

Abstract

Objectives: Haplogroup C2a-M48 is the predominant paternal lineage of Tungusic-speaking populations, one of the largest population groups in Siberia. Up until now, the origins and dispersal of Tungusic-speaking populations have remained unclear. In this study, the demographic history of Tungusic-speaking populations was explored using the phylogenetic analysis of haplogroup C2a-M86, the major subbranch of C2a-M48.

Materials and methods: In total, 18 newly generated Y chromosome sequences from C2a-M48 males and 20 previously available Y-chromosome sequences from this haplogroup were analyzed. A highly revised phylogenetic tree of haplogroup C2a-M86 with age estimates was reconstructed. Frequencies of this lineage in the literature were collected and a comprehensive analysis of this lineage in 13 022 individuals from 245 populations in Eurasia was performed.

Results: The distribution map of C2a-M48 indicated the most probable area of origin and diffusion route of this paternal lineage in North Eurasia. Most C2a-M86 samples from Tungusic-speaking populations belonged to the sublineage C2a-F5484, which emerged about 3300 years ago. We identified six unique

Bing-Li Liu, Peng-Cheng Ma, and Chi-Zao Wang contributed equally to this study.

sublineages corresponding to the Manchu, Evenks, Evens, Oroqen, and Daur populations; these sublineages diverged gradually over the past 1900 years. Notably, we observed a clear north-south dichotomous structure for sublineages derived from C2a-F5484, consistent with the internal north-south divergence of Tungusic languages and ethnic groups.

Conclusions: We identified the important founding paternal haplogroup, C2a-F5484, for Tungusic-speaking populations as well as numerous unique subgroups of this haplogroup. We propose that the timeframe for the divergence of C2a-F5484 corresponds with the early differentiation of ancestral Tungusic-speaking populations.

1 | INTRODUCTION

The origin and cultural traditions of Tungusic-speaking populations in Siberia have long fascinated linguists, archeologists, anthropologists, and geneticists (Shirokogoroff, 1966). The interest in these populations lies in their special physical characteristics (a.k.a the Baykal type of human phenotypes; Alexeev, 1979; Debets, 1974; Levin, 1963), adaptation to cold climates (Cardona et al., 2014), Shamanism (J. Kwak, 2010; Meng, 2011; Winkelman, 2012), and social structure (Shirokogoroff, 1924, 1966). From the perspective of anthropology, the origin of Tungusic-speaking populations is still under debate. Some scholars have proposed that Tungusic-speaking populations expanded from the Eastern Cis-Baikal region of Siberia (Derevyanko, 1976; Helimski, 1985; Menges, 1968; Okladnikov, 1955; Vasilevič, 1969). Archeologist A. P. Okladnikov suggested that the remains associated with the Bronze Age Glazkovo culture (~4000-3300 before present [BP]) along the Lake Baikal are close to those of modern Tungusic-speaking populations (Tugolukov, 1980). Ethnologist Vasilevič proposed that an area south of Lake Baikal may be the Urheimat of all Tungusic-speaking populations and the south and north parts of modern Tungusic-speaking populations may have separated some 3500 years BP (Vasilevič, 1969). Other studies have argued that Evenk and Even populations in Siberia may have arisen during northward migrations from the Amur River basin (Feng, 2002; Gao, 2012; Janhunen, 1996; Pugach et al., 2016; Tugolukov, 1980; Wang & Robbeets, 2020). Besides, the mid-lower Amur River region is suggested to have been a refugial source for human re peopling regions across Siberia after the Last Glacial Maximum (Derenko, Malyarchuk, Grzybowski, et al., 2007; Derenko et al., 2010; Volodko et al., 2008; Wang & Robbeets, 2020). The Uril culture (~3300-2400 BP) in the Amur River region was considered to be the direct ancestor of the Poltse Culture (~2400-1800 BP), the Talakan Culture

(~2200-1600 BP), and the Mohe Culture (~1800-1300 BP) (Feng, 2002). The Mohe are the direct ancestors of the Jurchen and Manchu (Derevianko, 2015; Diakova, 2016; Elliott, 2001; Gorelova, 2002; Twitchett & Franke, 1994). Owing to limited written records and crucial archeological evidence, the origin and detailed evolutionary process of Tungusic-speaking population is yet to be resolved.

Previous genetic studies have provided clues to the origin of Tungusic-speaking populations (Duggan et al., 2013; Karafet, Osipova, Savina, Hallmark, & Hammer, 2018; Malyarchuk et al., 2010; Pugach et al., 2016). Y-chromosome haplogroup C2a1a2a-M86 (formerly C3c1-M86; see www.isogg.org, 15.34, accessed February 3, 2020), a major sublineage of C2a1a2-M48, is widespread in North Asia (Malyarchuk et al., 2010; Pakendorf et al., 2006), from the Kamchatka peninsula (Karmin et al., 2015) in the east to the Republic of Kalmykia in the west (Balnova et al., 2019; Malyarchuk et al., 2013; Nasidze et al., 2005). The frequency of this haplogroup is particularly high in the paternal gene pool of many Tungusic-speaking populations, such as Evenks, Evens, Negidal, Udegey, Ulchi, Ewenki, Oroqen, and Hezhen (Derenko, Malyarchuk, Denisova, et al., 2006; Duggan et al., 2013; Hammer et al., 2006; Karafet et al., 2001; T. M. Karafet et al., 2018; Khar'kov et al., 2008; Lell et al., 2002; Pakendorf et al., 2006; Puzyrev et al., 2003; Xue et al., 2006; Zerjal, Wells, Yuldasheva, Ruzibakiev, & Tyler-Smith, 2002; Zhong et al., 2010). Additionally, haplogroup C2a1a2a-M86 is also an important paternal lineage in some Mongolic- and Turkic-speaking populations (Chen et al., 2011; Di Cristofaro et al., 2013; Haber et al., 2012; Katoh et al., 2005; Wells et al., 2001). C2a1a2a-M86 is fairly recent based on divergence time estimates (Karmin et al., 2015). However, only three publications have mentioned the downstream structure of the phylogeny of C2a1a2-M48 (Balnova et al., 2019; Karafet et al., 2002; Karmin et al., 2015). Additional sequences of haplogroup C2a1a2-M48 are needed to reveal the downstream structure

of the phylogeny of this haplogroup and discover what its phylogenetic structure indicates about the origin and formation of modern Tungusic-speaking populations.

In this study, we analyzed the Y-chromosome sequences of 38 C2a1a2-M48 males from Tungusic-speaking populations and closely related populations. Our objectives were (a) to reconstruct a phylogeny and obtain age estimates for C2a1a2a-M86 and (b) to explore the origin, diversification, and expansion of this paternal lineage and its sublineages. In particular, we focused on the phylogenetic branching pattern of C2a1a2-M48 and its correlation with the divergence of Tungusic-speaking populations. Overall, we generated a refined Y-chromosome phylogenetic tree that allowed us to refine understanding of the origin and diffusion of paternal C2a1a2-M48 lineages and the role of this haplotype in the formation of modern Tungusic-speaking populations.

2 | MATERIALS AND METHODS

Blood samples were collected from unrelated healthy males from populations across China after obtaining informed consent. The samples used in this study include Mongolian, Mongolian_Buryats, Mongolian_Olots, Daur, and Evenks from HulunBuir, Inner Mongolia, China. Samples of Daur from Ili, Xinjiang, China were also included. Manchu samples were collected in Inner Mongolia, Liaoning, and Jilin of China. The ethics committee for biological research at the School of Life Sciences in Fudan University approved the study.

A series of Y-single-nucleotide polymorphism (Y-SNP) markers were genotyped to identify M48+ and M86+ individuals, including M130, P54, M105, M48, M208, M407, P33, M93, P39, P92, P53.1, M217, M38, M210, M356, P55, M347, and M86. Eighteen M48 samples were sent for DNA sequencing (Table S1). Full genome sequencing and targeted sequencing of Y-chromosomes were performed using the Illumina HiSeq 2000 (San Diego, California). Our previously described procedure was used for steps prior to next-generation sequencing (ie, DNA shearing, adaptor addition, gel electrophoresis, library preparation, bait design, and liquid-phase capture technology) (Yan et al., 2014).

Mapping of reads and SNP calling were conducted with the human reference hg38 using GATK packages (<https://github.com/broadinstitute/gatk/releases>) (Poplin et al., 2018) and standard procedures for next-generation sequencing (Li et al., 2009; Li & Durbin, 2010). The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Wang et al., 2017) at BIG Data Center, Beijing Institute of Genomics (BIG) (BIG Data Center Members, 2017), Chinese Academy of Sciences, under accession numbers PRJCA001362,

PRJCA001618, and PRJCA000430 and are publicly accessible at <http://bigd.big.ac.cn/gsa>.

To obtain a comprehensive overview of the paternal lineage, available Y-chromosome data from the literature were also included. In total, 38 sequences of haplogroup C2a-M48 were analyzed as well as 7 other samples of the closely related haplogroup (C2a1a1a3-M504 and C2a1a1a1-F1756) from the Human Genome Diversity Project (Bergström et al., 2020) as an outgroup. The sequences of haplogroup C2a-M48 included 18 new sequences from the current study, 6 from the HGDP Project (Bergström et al., 2020), 12 from Estonia Biocenter (Karmin et al., 2015), 1 from the Simons Genome Diversity Project (Mallick et al., 2016), and 1 from Balinova et al. (2019). The reference genome hg38 was used for all analyses. SNPs and haplogroup names were assigned according to www.isogg.org (15.34, accessed February 3, 2020) (Y Chromosome Consortium, 2002).

Standard procedures (bwa + samtools) were used to analyze the next-generation sequencing results (Li et al., 2009; Li & Durbin, 2009). To obtain a confident Y-SNP data set for reconstruction of a phylogenetic tree and age estimation, a series of strict filters were applied to the original variants file, including: (a) restriction to the 7.034M BP region widely used by Poznik et al. (2016), Karmin et al. (2015), and Adamov, Guryanov, Karzhavin, Tagankin, and Urasin (2015); (b) restriction to single-nucleotide variants (Y-SNP); (c) removal of all positions with a call rate <80%; (d) removal of positions with a heterozygous call rate >5%; (e) restriction to positions with a base coverage of ≥ 3 , a base quality of >20, and a distance between SNPs of >10 BP; and (f) removal of all recurrent or triadic variants, even if their positions are considered to have high confidence on the current phylogenetic tree.

Bayesian evolutionary analyses and age estimation were performed using BEAST (v. 2.0.0) (Bouckaert et al., 2014). A Bayesian skyline coalescent tree and a strict clock were used. The bModelTest package, which allows the BEAST program to infer the most optimistic substitution model for input sequences (Bouckaert & Drummond, 2017), was employed. An age of 35 383 years was assigned to haplogroup C2-M217 (95% confidence interval = 33 305-37 537 years) for calibration (Karmin et al., 2015). The calculation was performed with 20 million iterations and sampling every 10 000 steps. Results were visualized in Tracer v.1.6 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) and FigTree v1.4.2 (Rambaut, 2018), with a burn-in of 30% and more than 200 effective sample sizes.

To more comprehensively characterize lineage C2a-M48, Y-chromosome haplogroup frequencies were collected in 245 East Eurasian populations from 44 previous publications (Abilev et al., 2012; Biro, Zalan, Volgyi, & Pamjav, 2009; Cai et al., 2011; Derenko, Malyarchuk,

Wozniak, et al., 2007; Di Cristofaro et al., 2013; Dulik, Osipova, & Schurr, 2011; Dulik et al., 2012; Fechner et al., 2008; Haber et al., 2012; Hallast et al., 2015; Hammer et al., 2006; Kang et al., 2011; Karmin et al., 2015; Katoh et al., 2005; Khar'kov, Khamina, Medvedeva, Shtygasheva, & Stepanov, 2011; Khar'kov et al., 2009; Kharkova et al., 2014; Kim et al., 2011; K. D. Kwak et al., 2006; Lell et al., 2002; Lippold et al., 2014; Malyarchuk et al., 2010; Nasidze et al., 2005; Pakendorf et al., 2006; Pakendorf, Novgorodov, Osakovskij, & Stoneking, 2007; Pimenoff et al., 2008; Qamar et al., 2002; Roewer et al., 2007; Sabitov & Daulet, 2012; Sengupta et al., 2006; Tarlykov et al., 2013; Turuspekov, Sabitov, Daulet, Sadykov, & Khalidullin, 2011; Wozniak et al., 2006; Xue et al., 2006; Zerjal et al., 2003; Zheng et al., 2009; Zhong et al., 2010; Zhou et al., 2008; Zhu et al., 2005) (see Table S2). The frequencies of C2a-M48 in populations were plotted on a geographic map using Surfer 7.0 (Golden Software, Inc., Golden, Colorado).

3 | RESULTS

Among 13,022 samples from 245 populations throughout eastern Eurasia, we identified 703 samples belonging to

haplogroup C2a-M48. The distribution of C2a-M48 in East Eurasian populations is shown in Figure 1. We observed high frequencies of C2a-M48 in most Tungusic-speaking populations and some Mongolic- and Turkic-speaking populations (see Table S2).

We reconstructed a fine phylogenetic tree of C2a-M48, with a focus on haplogroup C2a-M86 (Figures 2 and 3). The genotypes of variants of Y-chromosome sequences can be found in Table S3. The results of age estimations can be found in Table S4. We detected 32 subclades, 470 nonprivate Y-SNP variants, and 368 private variants (Table S3). Haplogroup C2a-M48 diverged from C2a-L1373 about 14.6 kya. Two sublineages (C2a1a2b-M86 and C2a1a2b-B90) diverged from haplogroup C2a-M48 about 11.6 kya. C2a-M86 underwent a continuous and significant expansion beginning 3300 years ago (Figure 3 and Table S4).

Haplogroup C2a-B90 was first discovered in Koryak populations (Karmin et al., 2015). In this study, we detected two samples assigned to lineage C2a-B90 in the Evenks population (Figure 3 and Table S1). Therefore, we proposed that this haplogroup (C2a-M48x86, including C2a-B90) may be a relic of an ancient population who were once scattered widely in Eastern Siberia, the

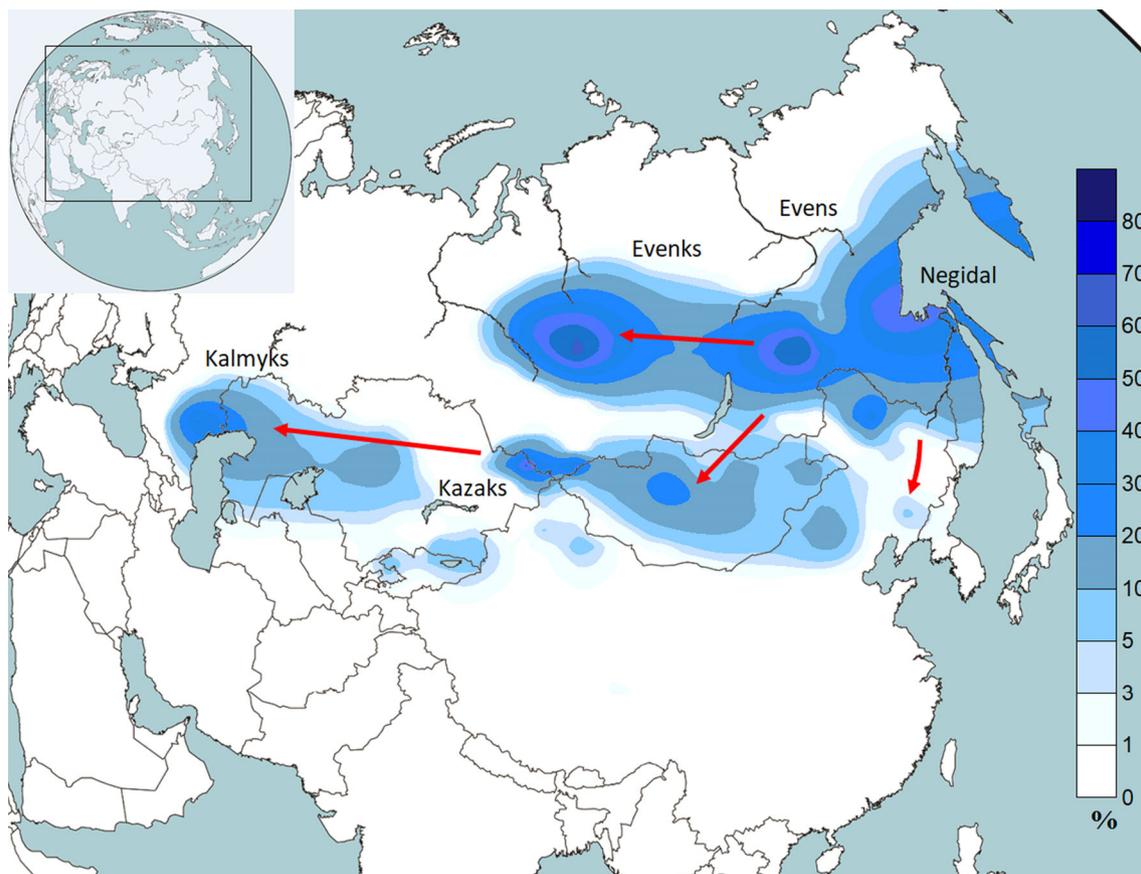
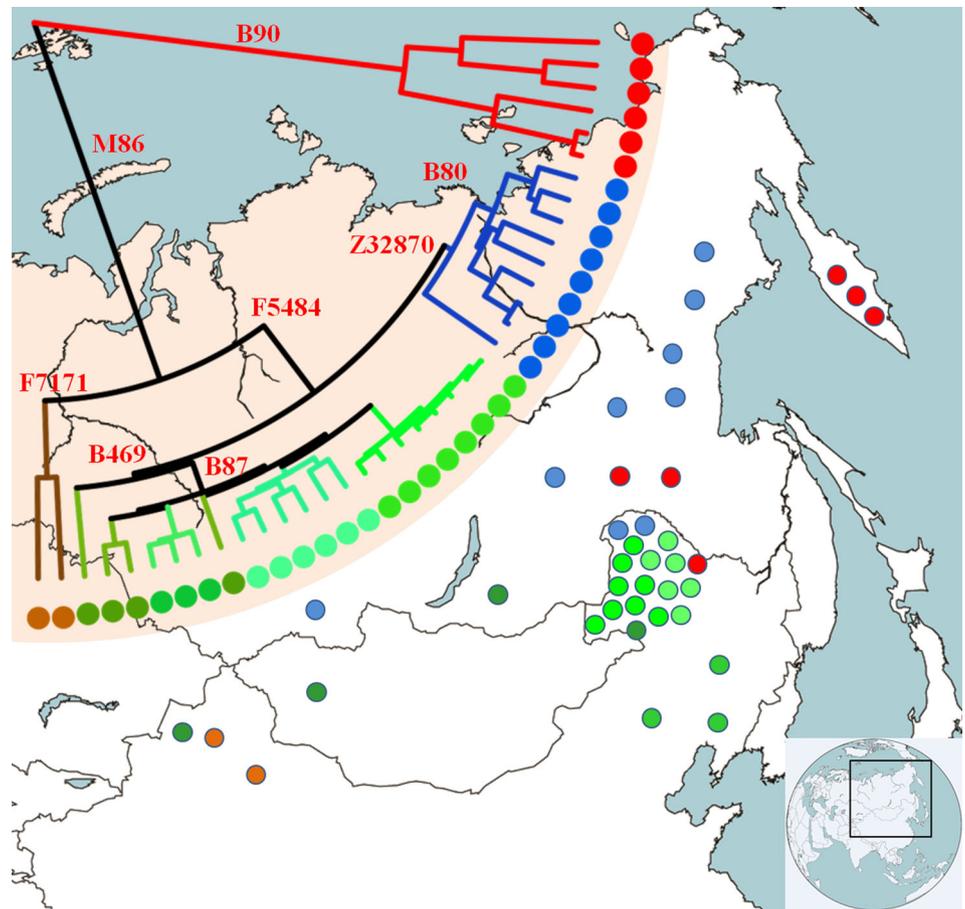


FIGURE 1 Distribution of the Y-chromosome lineage C2*-Star Cluster across Eurasia

FIGURE 2 Phylogeographic analysis of sublineages of C2a-M86. The circles with different colors correspond to the different sublineages



Amur River region, and the regions around the sea of Okhotsk (Lell et al., 2002). Thus, it is highly possible that C-M48xM86 in Iengra Evenks, Yakut-speaking Evenks, and Yukaghir populations also belong to this lineage (Pakendorf et al., 2006, 2007).

At approximately 3300 years ago, haplogroup C2a-M86 split into two subbranches, C2a-F7171 and C2a-F5484. The frequency of C2a-M48 is high in present-day Mongolic- and Turkic-speaking populations (Chen et al., 2011; Di Cristofaro et al., 2013; Haber et al., 2012; Katoh et al., 2005; Wells et al., 2001). In this study, the C2a-F7171 subbranch included two samples from Uyghur and Xibo populations in Xinjiang, China. This may be the dominant lineage for C2a-M86 samples from Mongolic- and Turkic-speaking populations. In contrast, all C2a-M86 samples from Tungusic-speaking populations belong to the C2a-F5484 lineage (Figures 2 and 3). Based on the distinct dichotomous structure of C2a-M86, we refer to C2a-F5484 as the eastern branch of C2a-M86 and C2a-F7171 as the western branch. The appearance of C2a-F7171 in some Mongolic- and Turkic-speaking populations could be either a relic of a common ancestor of these two populations with

Tungusic-speaking populations or the result of a recent admixture at an unknown historical period. More data and analyses are needed to explore the history of C2a-F7171 in Mongolic- and Turkic-speaking populations.

A key finding of our study is the identification of sublineages representing different subgroups of Tungusic-speaking populations, as revealed by the downstream structure of haplogroup C2a-F5484 (Figures 2 and 3). C2a-F5484 diverged to form sublineages C2a-Z32870 and C2a-B469 about 1900 years ago (Table S4). Within the C2a-Z32870 lineage, subbranch B80 was unique to the Even, Oroqen, and Evenki populations. This haplogroup underwent continuous expansion beginning 1300 years ago. The other sublineage, C2a-B469, underwent an even greater expansion dating from 1,200 years ago. We detected several major sublineages of C2a-B469. For example, C2a-F31686, C2a-F10073, and Y25271 lineages were nearly exclusively found in Manchu, Daur, and Oroqen populations respectively. F15201, found in a Xibe individual (HGDP01247) and a Daur individual (DaurE14), may also be an important paternal lineage of ancient Southern Tungusic-speaking populations.

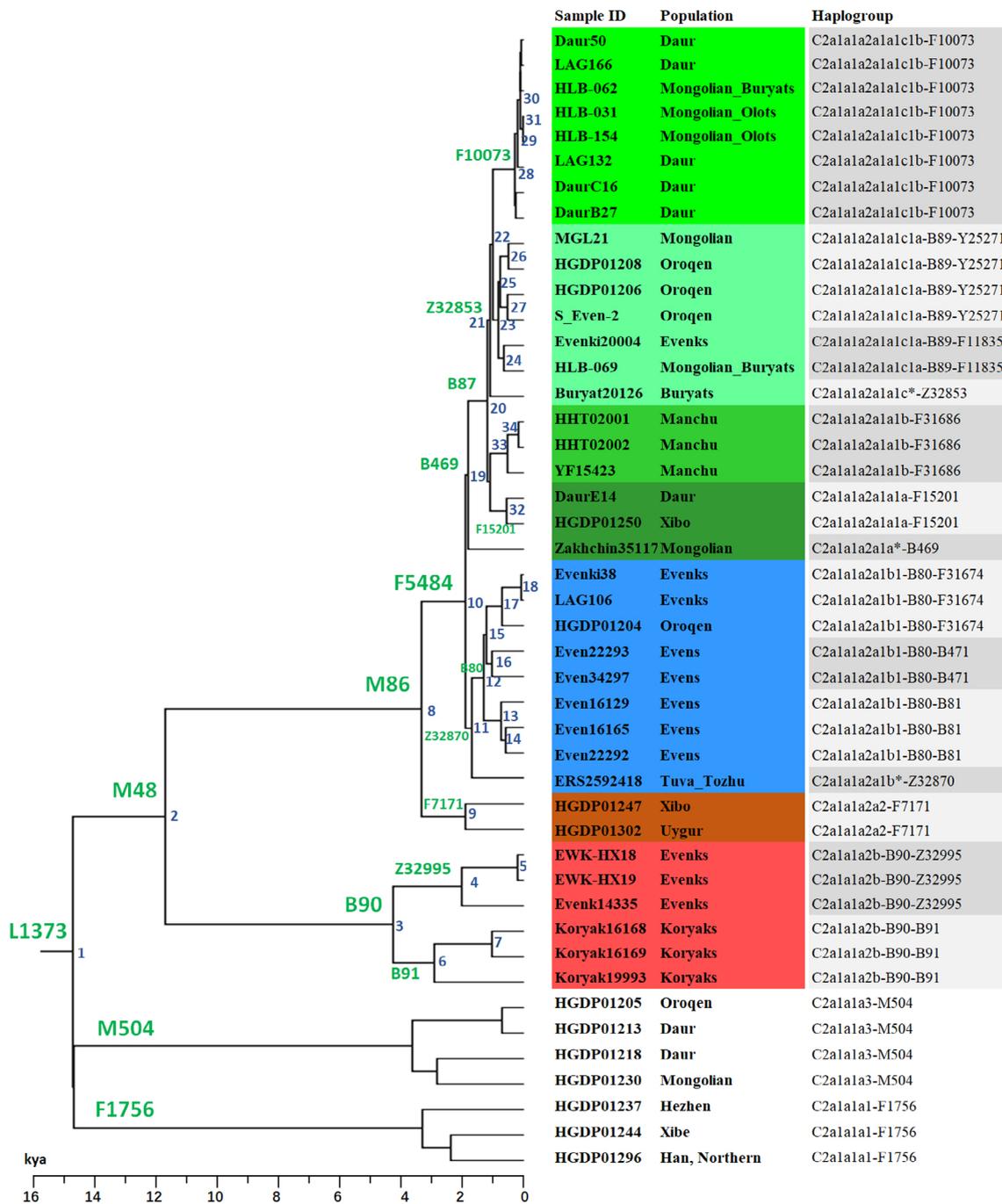


FIGURE 3 Detailed phylogeny of C2a-M86. The divergence time estimates for each bifurcation point with numbers are listed in Table S3

4 | DISCUSSION

4.1 | Dispersal of C2-M48 in Eastern Eurasia and its implications for origin of Tungusic-speaking populations

Based on the achievements of ethnology and historical analysis in the past and the genetic evidence in this

study, we propose four general migration routes of haplogroup C2a-M86 in Eastern Eurasia (red arrows in Figure 1). We also propose that the middle and upper reaches of the Amur River is likely the area of origin of haplogroup C2a-M86. The northward diffusion of ancestral Tungusic-speaking populations may have led to the appearance of Evenk and Even populations in Siberia. Some ancestral Tungusic-speaking populations may have

migrated to the Mongolia Plateau and contributed to the high frequencies of C2a-M86 in present-day Mongolic- and Turkic-speaking populations. On the other hand, the majority of ancestral Tungusic-speaking populations may have scattered gradually in the lower reaches of the Amur River and Northeast China before eventually evolving into the modern Tungusic-speaking populations in these region. For the first time, we provide a clear distribution pattern for the paternal lineage C2a-M48 in Eurasia and suggest that this lineage can be used to date the formation of modern Tungusic-speaking populations.

From the perspective of linguistics, some linguists propose that the similarity between Tungusic languages and Mongolic languages is a heritage of a common ancestor while other scholars have argued that the similarity is the result of recent contact between these two language groups in a recent historical period (Clauson, 1956, 1972; Doerfer, 1963; Georg, 2003; Ligeti, 1975; Poppe, 1965, 1972; Ramstedt, 1952, 1957; Robeets, 2011; Starostin, Dybo, & Mudrak, 2003; Vovin, 2005). We found many subbranches of C2-M48 among samples from Mongolians in this study (Figure 3). However, previous studies suggested that haplogroups C2a1a1a3-M504, C2a1a1a1-F1756, and C2b1a1a1a-M407 resemble the founding paternal lineages of Mongolic-speaking populations (Huang et al., 2018; Wei et al., 2017, 2018) while varied subbranches of Q-M242, R-M207, and N-231 are more frequent in the paternal gene pool of Turkic-speaking populations (Derenko, Malyarchuk, Denisova, et al., 2006; Malyarchuk et al., 2011; Rootsi et al., 2007). Therefore, we suggest that the lineage C2a-F7171 in Mongolic- and Turkic-speaking populations, rather than being a relic of a genetic component in a common ancestor, is the result of an admixture that occurred in the recent historical period. In summary, we suggest that the sharing of C2-M48 by Tungusic-, Mongolic-, and Turkic-speaking populations cannot be used as evidence to support a common origin of these three population groups.

4.2 | C2a-F5484 is one of the most important founding paternal lineages of Tungusic-speaking populations

In this study, we characterized the precise pattern of differentiation of M48 and detected C2a-F5484, the eastern sublineage of M48, in most M48 samples from Tungusic-speaking populations. The frequencies of different Y-SNP haplogroups in Tungusic-speaking populations reported in the literature can be found in Table S5. Many other paternal haplogroups have been detected in Tungusic-speaking populations. The N-P43, N-M46, and P-M45 haplogroup frequencies are high in paternal gene pools

of some Tungusic-speaking populations like the Evens, Evenks, Nanais, and Negidal populations in Siberia (Lell et al., 2002; Pakendorf et al., 2006, 2007). This may be explained by admixture with other indigenous populations in Northern Asia during the formation of Tungusic-speaking populations in the past 4000 years (Karafet et al., 2001, 2018). Nevertheless, it is possible that subbranches of N-M231 were also an important paternal lineage of the common ancestor of Tungusic-speaking populations. Previous studies found high frequencies and special subbranches of N-M231 (Karmin et al., 2015) in some Tungusic-speaking populations (Table S5) (Derenko, Malyarchuk, Denisova, et al., 2006; Derenko, Malyarchuk, Wozniak, et al., 2006; Duggan et al., 2013; Hammer et al., 2006; Karafet et al., 2001; Lell et al., 2002; Malyarchuk et al., 2010; Pakendorf et al., 2006; Rootsi et al., 2007). The low frequencies of N-M231 in some modern Tungusic-speaking populations may be the result of long-distance migration and a founder effect during the formation process of these populations.

High frequencies of subbranches of O-M175 were also found in Southern Tungusic-speaking populations, especially those in China (Hammer et al., 2006; Xue et al., 2006). Since populations from East Asia are frequently assigned to subbranches of O-M175 in phylogenetic analyses (Zhong et al., 2010). O-M175 subbranches in Tungusic-speaking populations are the result of a recent admixture occurring during the southward diffusion process of Tungusic-speaking populations and their long-term residence in northeast China (Karafet et al., 2001).

In summary, we have made several observations about the lineage C2a-M86 (more exactly C2a-F5484) that assist in determining the origin and differentiation of all Tungusic-speaking populations. Firstly, lineage C2a-M86 occurred with high or moderate frequency in nearly all Tungusic-speaking populations (Table S5). Secondly, this lineage underwent expansion in the recent 4000 years and the expansion of this branch roughly laid the foundation for Tungusic-speaking populations. Thirdly, lineage C2a-M86 was only a minor percentage of the paternal gene pool of some Tungusic-speaking populations, such as the Manchu and Xibe. On the other hand, historic studies have suggested that modern Manchu and Xibe populations result from an admixture between ancient Tungusic-speaking tribes and local populations in northeast China. All lineages beside C2a-M86 in Tungusic-speaking populations, such as N-P43, N-M46, P-M45, and O-M175, can be considered as either relics of a common ancestor or the result of an admixture with other populations after the expansion of ancient Tungusic-speaking populations. Therefore, we propose that the Y-chromosome haplogroup C2a-F5484 is one of the most important founding paternal lineages of all Tungusic-speaking populations.

4.3 | A clear north-south dichotomous structure of Tungusic-speaking populations

The two major sublineages of C2a-F5484 exhibited distinct geographic distributions (Figures 2 and 3, Table S1). The northern sublineage C2a-Z32870 was almost exclusively found in Evens and Evenki populations in Siberia and B80, the major subbranch of Z32870, was not detected in non-Tungusic-speaking populations. The southern sublineage, C2a-B469, showed a more complex pattern of contact with other populations. A large number of C2a-B469 samples were also found in non-Tungusic-speaking populations (Figure 3). Nevertheless, we identified unique subbranches corresponding to subgroups of Tungusic-speaking populations, such as F31686 for the Manchu and Y25271 for the Oroqen. It is worth noting that the Evens and Evenks are found in Siberia while other Tungusic-speaking populations generally live in the Amur River region and Northeastern China. The north-south divide in the distribution of the C2a-F5484 sublineage (Z32870 and B469, Figure 2) corresponds well with two major geographic distribution regions of Tungusic-speaking populations. Advanced genetic evidence in this study do not support the previous opinion that the northern part of Tungusic-speaking populations is originally a subset of the southern part, or vice versa (see Section 1). Genetic evidence tends to support the conjecture that an adjacent region of the modern distribution range of the northern and southern parts of Tungusic-speaking populations was most likely the location of the common ancestor.

Ethnologists (Shirokogoroff, 1924, 1966), historians (Gao, 2012), and linguists (Avrorin, 1959; Baskakov, 1981; Benzing, 1955; Cincius, 1977; Doerfer, 1978; Ikegami, 2002; Poppe, 1965; Ramstedt, 1957; Skorik, 1968; Sunik, 1982; Vovin, 2009) have previously agreed that the geographical north-south separation largely shaped the ethnic groups and languages of Tungusic-speaking populations. We identified the C2a-F5484 haplogroup as one of the most important founding paternal lineages of Tungusic-speaking populations, suggesting that the north-south dichotomous structure in the phylogeny of C2a-F5484 is closely related to the north-south differentiation of Tungusic-speaking populations from the perspective of ethnology, history, and linguistics.

In general, based on our results and observations from other disciplines, we propose that the most recent common ancestor of modern Tungusic-speaking populations lived in the northern region of the upstream area of the Amur River region, also known as the region south of the Stanovoy Range in Russia, between 4 and 3 kya. The initial expansion of Tungusic-speaking populations involved migration in multiple directions. The B80 subbranch may

be the dominant paternal lineage in the subgroup that migrated toward the Kamchatka peninsula, eventually giving rise to modern Evens. Secondly, the Evenks and Evenki are likely the direct descendants of the proposed ancestral group, as evidenced by the high frequencies of haplogroup C2a-M48 and large number of subbranches of C2a-M48 in these two populations. Additionally, the B87 subbranch may be the dominant paternal lineage in the subgroup that migrated in a southeastern direction. The substantial dispersal in the middle and downstream area of the Amur region led to modern Tungusic-speaking populations in this region. Tungusic-speaking populations living in the southern parts of this region were recorded as Mohe, Jurchen, and Manchu in the historical records materials of East Asia.

The long-term migratory habits of reindeer may have been an important factor in the diffusion of Tungusic-speaking populations across Siberia. The names of many Tungusic-speaking populations shared the meaning “reindeer breeder,” including the Oroqen, Orok, and Orochi (Levin & Potapov, 1964; Shirokogoroff, 1966). Reindeer is the only large domesticated animal in Northern Eurasia (Helskog & Indrelid, 2011). Reindeer feed mainly on plants in permafrost regions, including grass, leaves, fresh mushrooms, lichens, and mosses. Reindeers and their breeders need to migrate repeatedly across a vast geographic region to get sufficient foods (Bjørklund, 2013; Forbes & Kumpula, 2009). Therefore, ethnologists have suggested that the lifestyle of reindeer breeders is one of the main reasons for the vast distribution of Tungusic-speaking populations (Levin & Potapov, 1964; Shirokogoroff, 1966).

In this study, we have provided a clear distribution map for paternal haplogroup C2a-M86 and reconstructed a highly revised phylogenetic tree for this lineage with age estimates. Our results revealed that C2a-M86 samples from Tungusic-speaking populations shared a common ancestor carrying haplogroup C2a-F5484 about 3300 years ago. We identified four sublineages that are highly specific to the Manchu, Evenks, Evens, and Oroqen populations. We propose that the emergence and expansion of sublineages of C2a-F5484 correspond to the expansion of subgroups of Tungusic-speaking populations and their languages across North Eurasia. These results will guide the exploration of the population history of North Eurasia in the future.

ACKNOWLEDGMENTS

We thank all donors for providing DNA samples and/or DNA sequences. This work was supported by the National Natural Science Foundation of China (31900406 to L.-H. W., 31671297 and 91731303 to H. L.), the Fundamental Research Funds for the Central Universities (20720191047 for L.-H. W. and 19SKGC-QG04 for B.-L. L.), and the Scientific and Technology Committee of Shanghai Municipality

(18490750300 to L.-H. W. and H. L.). H. L. was also supported by National Key Research and Development Project (2016YFC0900300).

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest. The authors alone are responsible for the content and writing of the article.

AUTHOR CONTRIBUTIONS

Bing-li Liu: Formal analysis; funding acquisition; investigation; writing-original draft; writing-review and editing. **Peng-cheng Ma:** Data curation; formal analysis; investigation; methodology. **Chi-Zao Wang:** Data curation; formal analysis; resources. **Shi Yan:** Data curation; resources. **Hong-Bing Yao:** Data curation; resources. **yonglan Li:** Data curation; resources. **Yong-Mei Xie:** Data curation; resources. **Song-Lin Meng:** Data curation; resources. **Jin Sun:** Formal analysis; visualization. **Yan-Huan Cai:** Formal analysis; investigation. **Sarengaowa Sarengaowa:** Formal analysis. **Hui Li:** Data curation; resources. **Hui-zhen Cheng:** Conceptualization; data curation; formal analysis; investigation; visualization; writing-original draft. **Lan-Hai Wei:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; resources; supervision; validation; visualization; writing-original draft; writing-review and editing.

ORCID

Jin Sun  <https://orcid.org/0000-0003-0348-6715>

Hui-Zhen Cheng  <https://orcid.org/0000-0002-3965-9895>

REFERENCES

- Abilev, S., Malyarchuk, B., Derenko, M., Wozniak, M., Grzybowski, T., & Zakharov, I. (2012). The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan. *Human Biology*, 84(1), 79–89. <https://doi.org/10.3378/027.084.0106>
- Adamov, D., Guryanov, V., Karzhavin, S., Tagankin, V., & Urasin, V. (2015). Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. *The Russian Journal of Genetic Genealogy*, 7(1), 68–89.
- Alexeev, V. P. (1979). The differential geography of races. In W. A. Stini (Ed.), *Physiological and morphological adaptation and evolution*. Berlin: De Gruyter.
- Avrorin, V. A. (1959–1961). *The grammar of the nanai language: phonetics and grammar* (Vol. 1–2). Moscow; Leningrad: Nauka.
- Balinova, N., Post, H., Kushniarevich, A., Flores, R., Karmin, M., Sahakyan, H., ... Rootsi, S. (2019). Y-chromosomal analysis of clan structure of Kalmyks, the only European Mongol people, and their relationship to Oirat-Mongols of Inner Asia. *European Journal of Human Genetics*, 27(9), 1466–1474. <https://doi.org/10.1038/s41431-019-0399-0>
- Baskakov, N. A. (1981). *Altai language family and its learning*. Moscow: Nauka.
- Benzing, J. (1955). Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik. In *Abhandlungen der geistes- und sozialwissenschaftlichen Klasse*. Wiesbaden: Akademie der Wissenschaften und der Literatur in Mainz in Kommission bei Franz Steiner Verlag.
- Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., ... Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367(6484), eaay5012. <https://doi.org/10.1126/science.aay5012>
- BIG Data Center Members. (2017). The BIG data center: From deposition to integration to translation. *Nucleic Acids Research*, 45(D1), D18–D24. <https://doi.org/10.1093/nar/gkw1060>
- Biro, A. Z., Zalan, A., Volgyi, A., & Pamjav, H. (2009). A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *American Journal of Physical Anthropology*, 139(3), 305–310. <https://doi.org/10.1002/ajpa.20984>
- Björklund, I. (2013). Domestication, reindeer husbandry and the development of Sámi Pastoralism. *Acta Borealia*, 30(2), 174–189. <https://doi.org/10.1080/08003831.2013.847676>
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42. <https://doi.org/10.1186/s12862-017-0890-6>
- Cai, X., Qin, Z., Wen, B., Xu, S., Wang, Y., Lu, Y., ... Huang, X. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One*, 6(8), e24282. <https://doi.org/10.1371/journal.pone.0024282>
- 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>
- Chen, Z., Zhang, Y., Fan, A., Zhang, Y., Wu, Y., Zhao, Q., ... Rao, S. (2011). Y-chromosome haplogroup analysis indicates that Chinese Tuvans share distinctive affinity with Siberian Tuvans. *American Journal of Physical Anthropology*, 144(3), 492–497. <https://doi.org/10.1002/ajpa.21453>
- Cincius, V. I. (1977). *Comparative dictionary of Tungus-Manchu languages: Materials for an etymological dictionary (in Russian)* (Vol. 1). Leningrad: Nauka.
- Clauson, G. (1956). The case against the Altaic theory. *Central Asiatic Journal*, II(3), 181–187.
- Clauson, G. (1972). *An etymological dictionary of pre-thirteenth-century Turkish*. Oxford: Oxford University Press.
- Debets, G. (1974). Essay on the graphical presentation of the genealogical classification of human races. In Y. Bromley (Ed.), *Soviet ethnology and anthropology today*. Berlin: De Gruyter Mouton.
- Derenko, M., Malyarchuk, B., Denisova, G. A., Wozniak, M., Dambueva, I., Dorzhu, C., ... Zakharov, I. (2006). Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Human Genetics*, 118(5), 591–604. <https://doi.org/10.1007/s00439-005-0076-y>
- Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., ... Zakharov, I. (2007).

- Phylogeographic analysis of mitochondrial DNA in Northern Asian populations. *The 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>
- Derenko, M. V., Malyarchuk, B. A., Wozniak, M., Dambueva, I. K., Dorzhu, C. M., Luzina, F. A., ... Zakharov, I. A. (2006). The diversity of Y-chromosome lineages in indigenous population of South Siberia. *Doklady Biological Sciences*, 411, 466–470.
- Derenko, M. V., Malyarchuk, B. A., Wozniak, M., Denisova, G. A., Dambueva, I. K., Dorzhu, C. M., ... Zakharov, I. A. (2007). Distribution of the male lineages of Genghis Khan's descendants in northern Eurasian populations. *Russian Journal of Genetics*, 43(3), 334–337. <https://doi.org/10.1134/S1022795407030179>
- Derevianko, E. I. (2015). Mohe—Tunguses. *Problems in Archaeology, Ethnography, and Anthropology of Siberia and Adjacent Areas*, 21, 224–227.
- Derevyanko, A. (1976). *Priamur'e in antiquity. One thousand B.C.* Novosibirsk: Nauka.
- Di Cristofaro, J., Pennarun, E., Mazieres, S., Myres, N. M., Lin, A. A., Temori, S. A., ... Chiaroni, J. (2013). Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. *PLoS One*, 8(10), e76748. <https://doi.org/10.1371/journal.pone.0076748>
- Diakova, O. V. (2016). *Koryo-Jurchen's borderlands: Contacts and conflicts.* Moscow, Russia: Russian Acad Sciences, Inst Far Eastern Studies.
- Doerfer, G. (1963). *Turkish and Mongol elements in the New Persian. I-IV.* Wiesbaden: Franz Steiner Verlag.
- Doerfer, G. (1978). Classification problems of Tungus. In G. Doerfer & M. Weiers (Eds.), *Tungusica, Band 1: Beiträge zur Nordasiatischen Kulturgeschichte* (pp. 1–26). Wiesbaden: Otto Harrassowitz.
- 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>
- Dulik, M. C., Osipova, L. P., & Schurr, T. G. (2011). Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One*, 6(3), e17548. <https://doi.org/10.1371/journal.pone.0017548>
- Dulik, M. C., Zhadanov, S. I., Osipova, L. P., Askapuli, A., Gau, L., Gokcumen, O., ... Schurr, T. G. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and indigenous Altaians. *American Journal of Human Genetics*, 90(2), 229–246. <https://doi.org/10.1016/j.ajhg.2011.12.014>
- Elliott, M. C. (2001). *The Manchu way: The eight banners and ethnic identity in late Imperial China.* Redwood: Stanford University Press.
- Fechner, A., Quinque, D., Rychkov, S., Morozowa, I., Naumova, O., Schneider, Y., ... Nasidze, I. (2008). Boundaries and clines in the west Eurasian Y-chromosome landscape: Insights from the European part of Russia. *American Journal of Physical Anthropology*, 137(1), 41–47. <https://doi.org/10.1002/ajpa.20838>
- Feng, E.-X. (2002). *Archaeology in Russian Far East and Eastern Siberia (in Chinese).* Changchun, China: Jinlin University Press.
- Forbes, B., & Kumpula, T. (2009). The ecological role and geography of reindeer (*Rangifer tarandus*) in Northern Eurasia. *Geography Compass*, 3, 1356–1380. <https://doi.org/10.1111/j.1749-8198.2009.00250.x>
- Gao, K. J. (2012). *The rising of Tungusic-speaking populations.* Beijing, China: China Publishing House.
- Georg, S. (2003). Japanese, the Altaic theory, and the limits of language classification. In A. Vovin & T. Osada (Eds.), *Perspectives on the origins of the Japanese language* (pp. 429–449). Kyoto, Japan: International Research Center for Japanese Studies.
- Gorelova, L. M. (2002). *Handbook of oriental studies. Section 8 Uralic & Central Asian Studies, Manchu grammar.* Leiden, The Netherlands: Brill.
- Haber, M., Platt, D. E., Ashrafian Bonab, M., Youhanna, S. C., Soria-Hernanz, D. F., Martinez-Cruz, B., ... Genographic, C. (2012). Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events. *PLoS One*, 7(3), e34288. <https://doi.org/10.1371/journal.pone.0034288>
- Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J. H., Arroyo-Pardo, E., ... Jobling, M. A. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Molecular Biology and Evolution*, 32(3), 661–673. <https://doi.org/10.1093/molbev/msu327>
- Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., & Horai, S. (2006). Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *Journal of Human Genetics*, 51, 47–58. <https://doi.org/10.1007/s10038-005-0322-0>
- Helimski, E. (1985). Samodiisko-tungusskie leksicheskie sviazi i ikh etno-istoricheskie implikatsii [Samoyed-Tungusic lexical relations and their ethno-historical implications]. In E. I. Ubriatova (Ed.), *Uralo-Altaiistika.* Novosibirsk: Nauka.
- Helskog, K., & Indrelid, S. (2011). Humans and reindeer. *Quaternary International*, 238(1), 1–3. <https://doi.org/10.1016/j.quaint.2011.03.018>
- Huang, Y. Z., Wei, L. H., Yan, S., Wen, S. Q., Wang, C. C., Yang, Y. J., ... Li, H. (2018). Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407. *Molecular Genetics and Genomics*, 293(3), 657–663. <https://doi.org/10.1007/s00438-017-1403-4>
- Ikegami, J. (2002). *Materials for the study of the Tungus-Manchu languages: A collection of Uilta, Evenki and Manchu texts and Evenki and early recorded Uilta and Olcha vocabularies [in Japanese].* Sapporo, Japan: Hokkaido University Press.
- Janhunen, J. (1996). *Manchuria. An ethnic history.* Helsinki, Finland: The Finno-Ugrian Society.
- Kang, L., Lu, Y., Wang, C., Hu, K., Chen, F., Liu, K., ... Li, H. (2011). Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas. *Annals of Human Genetics*, 76, 92–99. <https://doi.org/10.1111/j.1469-1809.2011.00690.x>
- Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R. S., ... Hammer, M. F. (2001). Paternal population history of East Asia: Sources, patterns, and microevolutionary processes.

- American Journal of Human Genetics*, 69(3), 615–628. <https://doi.org/10.1086/323299>
- 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.
- Karafet, T. M., Osipova, L. P., Savina, O. V., Hallmark, B., & Hammer, M. F. (2018). Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations. *American Journal of Human Biology*, 30(6), e23194. <https://doi.org/10.1002/ajhb.23194>
- Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Jarve, M., Talas, U. G., ... Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, 25(4), 459–466. <https://doi.org/10.1101/gr.186684.114>
- Katoh, T., Munkhbat, B., Tounai, K., Mano, S., Ando, H., Oyungere, G., ... Inoko, H. (2005). Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene*, 346, 63–70. <https://doi.org/10.1016/j.gene.2004.10.023>
- Khar'kov, V. N., Medvedeva, O. F., Luzina, F. A., Kolbasko, A. V., Gafarov, N. I., Puzyrev, V. P., & Stepanov, V. A. (2009). Comparative characteristics of the gene pool of Teleuts inferred from Y-chromosomal marker data. *Genetika*, 45(8), 994–1003. <https://doi.org/10.1134/s1022795409080158>
- Khar'kov, V. N., Khamina, K. V., Medvedeva, O. F., Shtygasheva, O. V., & Stepanov, V. A. (2011). Genetic diversity of the Khakass gene pool: Subethnic differentiation and the structure of Y-chromosome haplogroups. *Molekuliarnaia Biologiia*, 45(3), 404–416. <https://doi.org/10.1134/s0026893311020117>
- Khar'kov, V. N., Stepanov, V. A., Medvedev, O. F., Spiridonova, M. G., Maksimova, N. R., Nogovitsyna, A. N., & Puzyrev, V. P. (2008). The origin of Yakuts: Analysis of Y-chromosome haplotypes. *Molekuliarnaia Biologiia (Mosk)*, 42(2), 226–237.
- Kharkova, V. N., Khamina, K. V., Medvedeva, O. F., Simonova, K. V., Ereminab, E. R., & Stepanova, V. A. (2014). Gene Pool of Buryats: Clinal variability and territorial subdivision based on data of Y-chromosome markers. *Russian Journal of Genetics*, 50(2), 203–213.
- Kim, S. H., Kim, K. C., Shin, D. J., Jin, H. J., Kwak, K. D., Han, M. S., ... Kim, W. (2011). High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: A genetic perspective on the peopling of Korea. *Investigative Genetics*, 2(1), 10. <https://doi.org/10.1186/2041-2223-2-10>
- Kwak, J. (2010). A study on the beliefs and myths of the Evenks in Siberia. [시베리아에벤크족의 신앙과 신화에 대한 연구]. *Journal of North-East Asian Cultures*, 1(22), 345–363.
- Kwak, K. D., Suren, G., Tundewrentsen, S., Jin, H. J., Hong, S. B., Han, M. S., & Kim, W. (2006). Y-chromosome STR haplotype profiling in the Mongolian population. *Legal Medicine (Tokyo, Japan)*, 8(1), 58–61. <https://doi.org/10.1016/j.legalmed.2005.08.005>
- Lell, J. T., Sukernik, R. I., Starikovskaya, Y. B., Su, B., Jin, L., Schurr, T. G., ... Wallace, D. C. (2002). The dual origin and Siberian affinities of Native American Y chromosomes. *American Journal of Human Genetics*, 70(1), 192–206. <https://doi.org/10.1086/338457>
- Levin, M. G. (1963). *Ethnic origins of the peoples of Northeastern Asia* [Anthropology of the North: Translation from Russian Source, 1. Translated By Henry M. Michael]. Toronto, ON: University of Toronto Press.
- Levin, M. G., & Potapov, L. P. (1964). *The peoples of Siberia*. Chicago, IL: University of Chicago Press.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Ligeti, L. (1975). La théorie altaïque et la lexico-statistique. In L. Ligeti (Ed.), *Researches in Altaic languages* (pp. 99–115). Budapest, Hungary: Akademiai Kiado.
- Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., ... Stoneking, M. (2014). Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genetics*, 5, 13. <https://doi.org/10.1186/2041-2223-5-13>
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., ... Reich, D. (2016). The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, 538(7624), 201–206. <https://doi.org/10.1038/nature18964>
- Malyarchuk, B., Derenko, M., Denisova, G., Khoyt, S., Wozniak, M., Grzybowski, T., & Zakharov, I. (2013). Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels. *Journal of Human Genetics*, 58(12), 804–811. <https://doi.org/10.1038/jhg.2013.108>
- Malyarchuk, B., Derenko, M., Denisova, G., Maksimov, A., Wozniak, M., Grzybowski, T., ... Zakharov, I. (2011). Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *Journal of Human Genetics*, 56(8), 583–588. <https://doi.org/10.1038/jhg.2011.64>
- Malyarchuk, B., Derenko, M., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., & Zakharov, I. (2010). Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Annals of Human Genetics*, 74(6), 539–546. <https://doi.org/10.1111/j.1469-1809.2010.00601.x>
- Meng, H. Y. (2011). *Characteristics of Shammanism of the Tungusic-speaking people* (Vol. 1). Pa Leiden: E J Brill.
- Menges, K. H. (1968). The Tungusic language. In W. Fuchs, et al. (Eds.), *Tungusology* (pp. 21–256). Leiden: Brill.
- Nasidze, I., Quinque, D., Dupanloup, I., Cordaux, R., Kokshunova, L., & Stoneking, M. (2005). Genetic evidence for the Mongolian ancestry of Kalmyks. *American Journal of Physical Anthropology*, 128(4), 846–854. <https://doi.org/10.1002/ajpa.20159>
- Okladnikov, A. P. (1955). *The Neolithic and Bronze Age in the Baikal region (in Russian)*. Moscow & Leningrad: AN SSSR.
- Pakendorf, B., Novgorodov, I. N., Osakovskij, V. L., Danilova, A. P., Protod'jakonov, A. P., & Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: Genetic variation

- and the origins of Yakuts. *Human Genetics*, 120(3), 334–353. <https://doi.org/10.1007/s00439-006-0213-2>
- Pakendorf, B., Novgorodov, I. N., Osakovskij, V. L., & Stoneking, M. (2007). Mating patterns amongst Siberian reindeer herders: Inferences from mtDNA and Y-chromosomal analyses. *American Journal of Physical Anthropology*, 133(3), 1013–1027. <https://doi.org/10.1002/ajpa.20590>
- Pimenoff, V. N., Comas, D., Palo, J. U., Vershubsky, G., Kozlov, A., & Sajantila, A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics*, 16(10), 1254–1264. <https://doi.org/10.1038/ejhg.2008.101>
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., ... Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178. <https://doi.org/10.1101/201178>
- Poppe, N. (1965). *Introduction to Altaic linguistics*. Wiesbaden: Harrassowitz.
- Poppe, N. (1972). A new symposium on the Altaic theory. *Central Asiatic Journal*, 16, 37–58.
- Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A., ... Tyler-Smith, C. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*, 48(6), 593–599. <https://doi.org/10.1038/ng.3559>
- Pugach, I., Matveev, R., Spitsyn, V., Makarov, S., Novgorodov, I., Osakovskiy, 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>
- Puzryev, V. P., Stepanov, V. A., Golubenko, M. V., Puzryev, K. V., Maksimova, N. R., Khar'kov, V. N., ... Nogovitsyna, A. N. (2003). MtDNA and Y-chromosome lineages in the Yakut population. *Genetika*, 39(7), 975–981.
- Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., ... Mehdi, S. Q. (2002). Y-chromosomal DNA variation in Pakistan. *American Journal of Human Genetics*, 70(5), 1107–1124. <https://doi.org/10.1086/339929>
- Rambaut, A. (2018). FigTree download page. Retrieved from <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ramstedt, G. J. (1952). *Introduction to Altaic linguistics, Volume 1: Phonology* (Vol. 104.1). Helsinki: Suomalais-Ugrilainen Seura.
- Ramstedt, G. J. (1957). *Introduction to Altaic linguistics, Volume 2: Morphology* (Vol. 104.2). Helsinki: Suomalais-Ugrilainen Seura.
- Robeets, M. (2011). Six objections to the Transeurasian hypothesis. *Inalco Journée Altaïque*.
- Roewer, L., Kruger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L., ... Nasidze, I. (2007). Y-chromosomal STR haplotypes in Kalmyk population samples. *Forensic Science International*, 173(2–3), 204–209. <https://doi.org/10.1016/j.forsciint.2006.11.013>
- Rootsi, S., Zhivotovskiy, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R., ... Ilumae, A. M. (2007). A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics*, 15, 204–211. <https://doi.org/10.1038/sj.ejhg.5201748>
- Sabitov, Z., & Daulet, B. (2012). Haplogroups and haplotypes of 50 Karakalpak. *The Russian Journal of Genetic Genealogy*, 4(1), 83–90.
- Sengupta, S., Zhivotovskiy, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E., ... Underhill, P. A. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *American Journal of Human Genetics*, 78(2), 202–221. <https://doi.org/10.1086/499411>
- Shirokogoroff, S. M. (1924). *Social organization of the Manchus: A study of the Manchu clan organization*. Shanghai, China: Royal Asiatic Society.
- Shirokogoroff, S. M. (1966). *Social organization of the Northern Tungus: With introductory chapters concerning geographical distribution and history of these groups*. Oosterhout, The Netherlands: Anthropological Publications.
- Skorik, P. Y. (1968). *National language of the Soviet Union Vol.5: Mongolian, Manchu Tungusic and ancient Asian*. Leningrad: Nuaka.
- Starostin, S., Dybo, A., & Mudrak, O. (2003). *Etymological dictionary of the Altaic language*. Leiden, The Netherlands: E.J. Brill.
- Sunik, O. (1982). *Nouns of Manchu-Tungusic languages*. Leningrad: Nuaka.
- Tarlykov, P. V., Zholdybayeva, E. V., Akilzhanova, A. R., Nurkina, Z. M., Sabitov, Z. M., Rakhypbekov, T. K., & Ramanculov, E. M. (2013). Mitochondrial and Y-chromosomal profile of the Kazakh population from East Kazakhstan. *Croatian Medical Journal*, 54(1), 17–24. <https://doi.org/10.3325/cmj.2013.54.17>
- Tugolukov, V. A. (1980). Ethnic roots of the Tungus people. In I. S. Gurvich (Ed.), *Ethnogenesis of the peoples of the North* (pp. 152–176). Moscow: Nauka.
- Turuspekov, Y., Sabitov, Z., Daulet, B., Sadykov, M., & Khalidullin, O. (2011). The Kazakhstan DNA project hits first hundred Y-profiles for ethnic Kazakhs. *The Russian Journal of Genetic Genealogy*, 2(1), 69–84.
- Twitchett, D., & Franke, H. (1994). *The Cambridge history of China: Volume 6, Alien regimes and border states, 907-1368*. Cambridge: Cambridge University Press.
- Vasilevič, G. M. (1969). *Evenks. Historico-ethnographic sketches (XVIII-beg of the XXC)*. Leningrad: Nauka; Leningradskoe otdelenie.
- 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 Pleistocene peopling of the Americas. *American Journal of Human Genetics*, 82(5), 1084–1100. <https://doi.org/10.1016/j.ajhg.2008.03.019>
- Vovin, A. (2005). The end of the Altaic controversy. *Central Asiatic Journal*, 49(1), 71–132.
- Vovin, A. (2009). Tungusic languages. In K. Brown & S. Ogilvie (Eds.), *Concise encyclopedia of languages of the world* (pp. 1103–1105). New York: Elsevier.
- Wang, C.-C., & Robbeets, M. (2020). The homeland of Proto-Tungusic inferred from contemporary words and ancient

- genomes. *Evolutionary Human Sciences*, 2, e8. <https://doi.org/10.1017/ehs.2020.8>
- Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., ... Zhao, W. (2017). GSA: Genome sequence archive. *Genomics, Proteomics & Bioinformatics*, 15(1), 14–18. <https://doi.org/10.1016/j.gpb.2017.01.001>
- Wei, L. H., Huang, Y. Z., Yan, S., Wen, S. Q., Wang, L. X., Du, P. X., ... Li, H. (2017). Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. *Journal of Human Genetics*, 62, 915–918. <https://doi.org/10.1038/jhg.2017.60>
- Wei, L. H., Yan, S., Lu, Y., Wen, S. Q., Huang, Y. Z., Wang, L. X., ... Li, H. (2018). Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan. *European Journal of Human Genetics*, 26(2), 230–237. <https://doi.org/10.1038/s41431-017-0012-3>
- Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., ... Bodmer, W. F. (2001). The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), 10244–10249. <https://doi.org/10.1073/pnas.171305098>
- Winkelman, M. (2012). Shamanism in cross-cultural perspective. *International Journal of Transpersonal Studies*, 31(2), 47–62. <https://doi.org/10.24972/ijts.2012.31.2.47>
- Wozniak, M., Derenko, M., Malyarchuk, B., Dambueva, I., Grzybowski, T., & Miscicka-Sliwka, D. (2006). Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. *Forensic Science International*, 164(2-3), 271–275. <https://doi.org/10.1016/j.forsciint.2005.11.023>
- Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J., ... Tyler-Smith, C. (2006). Male demography in East Asia: A north–south contrast in human population expansion times. *Genetics*, 172(4), 2431–2439. <https://doi.org/10.1534/genetics.105.054270>
- Y Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research*, 12(2), 339–348. <https://doi.org/10.1101/gr.217602>
- Yan, S., Wang, C. C., Zheng, H. X., Wang, W., Qin, Z. D., Wei, L. H., ... Jin, L. (2014). Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One*, 9(8), e105691. <https://doi.org/10.1371/journal.pone.0105691>
- Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R., & Tyler-Smith, C. (2002). A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *American Journal of Human Genetics*, 71(3), 466–482. <https://doi.org/10.1086/342096>
- Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., ... Tyler-Smith, C. (2003). The genetic legacy of the Mongols. *American Journal of Human Genetics*, 72(3), 717–721. <https://doi.org/10.1086/367774>
- Zheng, L., Sun, H., Wang, J., Li, S., Bai, J., Jin, Y., ... Fu, S. (2009). Y chromosomal STR polymorphism in Northern Chinese populations. *Biological Research*, 42(4), 497–504. <https://doi.org/10.4067/S0716-97602009000400011>
- Zhong, H., Shi, H., Qi, X. B., Xiao, C. J., Jin, L., Ma, R. Z., & Su, B. (2010). Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *Journal of Human Genetics*, 55(7), 428–435. <https://doi.org/10.1038/jhg.2010.40>
- Zhou, R., Yang, D., Zhang, H., Yu, W., An, L., Wang, X., ... Xie, X. (2008). Origin and evolution of two Yugur sub-clans in Northwest China: A case study in paternal genetic landscape. *Annals of Human Biology*, 35(2), 198–211. <https://doi.org/10.1080/03014460801922927>
- Zhu, B., Li, X., Wang, Z., Wu, H., He, Y., Zhao, J., & Liu, Y. (2005). Y-STRs haplotypes of Chinese Mongol ethnic group using Y-PLEX 12. *Forensic Science International*, 153(2-3), 260–263. <https://doi.org/10.1016/j.forsciint.2004.11.002>

SUPPORTING INFORMATION

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

How to cite this article: Liu B-L, Ma P-C, Wang C-Z, et al. Paternal origin of Tungusic-speaking populations: Insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86. *Am J Hum Biol*. 2020:e23462. <https://doi.org/10.1002/ajhb.23462>