

1-19-2023

Paternal Origin of Mongolic-Speaking Populations: A Review of Studies from Recent Decades (1999–2019) and their Implications for Multidisciplinary Research in the Future

Sharengaowa

Social and Human Evolution Laboratory, School of Ethnology and Anthropology, Inner Mongolia Normal University, Hohhot, China

Peng-Cheng Ma

School of Life Sciences, Jilin University, Changchun, China

Wen-Jiao Yang

Department of Anthropology and Ethnology, School of Sociology and Anthropology, Xiamen University, Xiamen, China

Altangoo Ochirbat

School of Mathematic and Natural Sciences, Mongolian National University of Education, Ulaanbaatar, Mongolia

Maxat Zhabagin

National Center for Biotechnology, Nur-Sultan, Kazakhstan

Follow this and additional works at: https://digitalcommons.wayne.edu/humbiol_preprints

See next page for additional authors

Recommended Citation

Sharengaowa; Ma, Peng-Cheng; Yang, Wen-Jiao; Ochirbat, Altangoo; Zhabagin, Maxat; Sun, Na; Xie, Yong-Mei; Li, Yong-Lan; and Wei, Lan-Hai, "Paternal Origin of Mongolic-Speaking Populations: A Review of Studies from Recent Decades (1999–2019) and their Implications for Multidisciplinary Research in the Future" (2023). *Human Biology Open Access Pre-Prints*. 201.

https://digitalcommons.wayne.edu/humbiol_preprints/201

This Article is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState.

Authors

Sharengaowa, Peng-Cheng Ma, Wen-Jiao Yang, Altangoo Ochirbat, Maxat Zhabagin, Na Sun, Yong-Mei Xie, Yong-Lan Li, and Lan-Hai Wei

Paternal Origin of Mongolic-Speaking Populations: A Review of Studies from Recent Decades (1999–2019) and their Implications for Multidisciplinary Research in the Future

Sharengaowa,^{1,#} Peng-Cheng Ma,^{2,#} Wen-Jiao Yang,³
Altangoo Ochirbat,⁴ Maxat Zhabagin,⁵ Na Sun,⁶
Yong-Mei Xie,¹ Yong-Lan Li,^{7,1,*} and Lan-Hai
Wei^{1,7,8,*}

¹ Social and Human Evolution Laboratory, School of Ethnology and Anthropology, Inner Mongolia Normal University, Hohhot, China.

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

³ Department of Anthropology and Ethnology, School of Sociology and Anthropology, Xiamen University, Xiamen, China.

⁴ School of Mathematic and Natural Sciences, Mongolian National University of Education, Ulaanbaatar, Mongolia.

⁵ National Center for Biotechnology, Nur-Sultan, Kazakhstan.

⁶ College of Foreign Languages, Huaqiao University, Quanzhou, China.

⁷ College of Life Science and Technology, Inner Mongolia Normal University, Hohhot, China.

⁸ B&R International Joint Laboratory for Eurasian Anthropology, School of Life Science, Fudan University, Shanghai, China.

[#]These authors contributed equally to the study.

^{*}Correspondence to:

Lan-Hai Wei, School of Ethnology and Anthropology, Inner Mongolia Normal University, Hohhot 010022, China. E-mail: Ryan.lh.wei@gmail.com.

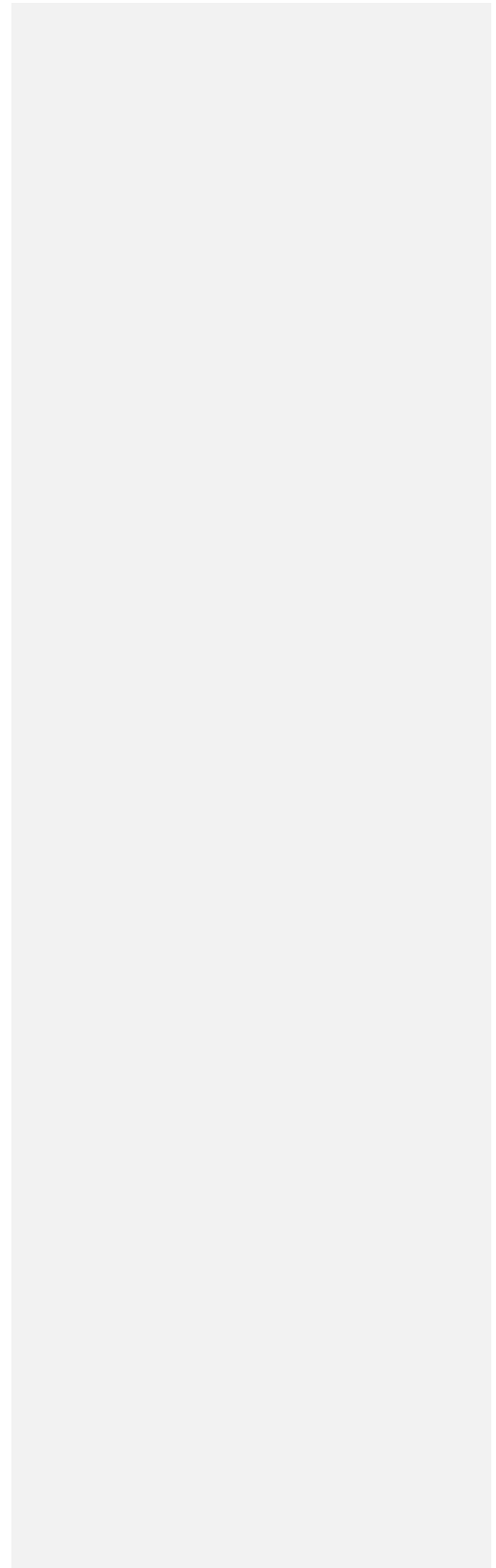
Yong-Lan Li, College of Life Science and Technology, Inner Mongolia Normal University, Hohhot 021002, China. E-mail: liyonglan2005@126.com.

KEY WORDS: MONGOLIC-SPEAKING POPULATION, PATERNAL GENE POOL, COMMON COMPONENTS,

Preprint version. Visit <http://digitalcommons.wayne.edu/humbiol/> after publication to acquire the final version.

EVOLUTION HISTORY, MULTIDISCIPLINARY RESEARCH

Preprint version. Visit <http://digitalcommons.wayne.edu/humbiol/> after publication to acquire the final version.



Abstract

The activities of Mongolic-speaking populations, a large group of people in eastern Eurasia, have important impact on the history of East Asia and other parts of Eurasia. Most previous genetic research of East Asian populations, including ancient DNA studies, have involved samples from Mongolic-speaking populations or their ancient relatives. Here, we summarized frequency data of paternal Y-chromosome haplogroups from all available literature about Mongolic-speaking populations from 1999 to 2019. Fourteen paternal components were identified and six of them were proposed as major and common components in ancestor groups of Mongolic-speaking populations. We thoroughly discussed the possible origin, migration patterns, and the roles of these six components in the evolution history of Mongolic-speaking populations. Meanwhile, we discussed the implications of the present achievements of human genetics for multidisciplinary research in ethnology, history, archaeology and linguistics in the future.

Introduction

Modern Mongolic-speaking population are widely distributed in the middle and east of Eurasian steppe. The ancient Donghu, Xianbei and Shiwei tribes are considered to be the “proto-Mongols” (Lin 2007). They probably originated in the eastern part of the Mongolian Plateau. They deeply involved in the historical process of the Mongolian Plateau and Northern China between 3000 and 1000 years ago (Lin 2007). About 1000 A.D., the Mongols rose from the northeast of the Mongolian Plateau to the middle and upper reaches of Heilongjiang River (a.k.a. Amur River) (Zhang 1998). They rose rapidly and established a powerful Mongol Empire since the beginning of 13 centuries (Howorth 1880). The Mongolian tribes that spread to different regions have gradually formed a modern Mongolian population through evolution and mixing with the local populations (Di Cosmo, Frank, and Golden 2009). At the same time, many ancient Mongol tribes integrated into other groups in Eurasia (Amitai 2007). The analyses about the genetic history of Mongolian populations and their ancestors is helpful to explore the historical process of human populations in the east of Eurasia.

After more than two decades of research, Y chromosome was considered to be one of the most powerful tools for studying the evolution history of human populations (Underhill 2003; Jobling and Tyler-Smith 2003). The non-recombination region of Y chromosome (NRY) follows strict paternal inheritance. Stable mutations in the Y-NRY region can be used to construct a phylogenetic tree with strict upstream and downstream relationships (Karafet et al. 2008). Since patrilineal society is the main social form of most modern human populations, one population or one male family clan tend to have unique paternal Y-chromosome lineage/haplogroup (Wu et al. 2020; Balanovsky et al. 2015; Zhabagin et al. 2020; Zhabagin et al. 2021). In 1997, Underhill et al. reported that they had identified 19 new

single nucleotide polymorphism on human Y chromosomes (Y-SNP) by using high performance liquid chromatography (DHPLC) (Underhill et al. 1997). Since then, the number of Y-SNP has increased rapidly and become an important tool for studying the history of human populations (Karmin et al. 2015; Karafet et al. 2008; Jobling and Tyler-Smith 2003; Y Chromosome Consortium 2002). In total, more than 800,000 Y-SNP have been discovered by the end of 2021 (<http://ybrowse.org/gbrowse2/gff>). In 1999, Su et al. reported the first dataset of paternal gene pool of East Asia populations and samples of Mongolians and Buryats were included in this study (Su et al. 1999). Since then, scholars all over the world continued to study populations in different regions. Up till now, the paternal genetic structure of most of human populations have been investigated by scholars.

Previous studies have shown high diversity in paternal gene pool of populations from northern Asia. The main source of the Japanese and Korean contains the indigenous people of the Japanese archipelago in the Paleolithic period, the agricultural people of the Neolithic Period, the Bronze culture people that spread from northeast China around A.D., and the immigrants from mainland China in recent two thousand years (Wang et al. 2020; Hammer et al. 2006; Tanaka et al. 2004). The ancestor groups may have brought varied sub-branches of D-M64, C-M8, C2-F1067, O1b-M176, and O2-M122 in modern Korean and Japanese. The Y-chromosome haplogroup C2a-M48 is the only predominant paternal lineage in most Tungusic-speaking populations (Liu et al. 2020; Lell et al. 2002; Malyarchuk et al. 2010; Pakendorf et al. 2006), nevertheless, intense admixed components can also be found in paternal gene pool of southern Tungusic-speaking populations (Xue et al. 2006; Zhong et al. 2010). For Turkic-speaking populations, different dominant lineages were identified in different ethnic groups, including C2-M217 (Zhabagin et al. 2021), N-M231 (Khar'kov et al. 2008), R1b-M269 (Malyarchuk et al. 2011), R1a1a-M17 (Underhill et al. 2015) and Q-M242 (Dulik et al. 2012). The multiple sub-lineages of N-M231 are the predominant paternal haplogroups of the varied Uralic-speaking populations (Ilumäe et al. 2016; Karmin et al. 2015; Rootsi et al. 2007). The downstream clades of C2-M217, N-M231 and Q-M242 are the predominant paternal lineages in most of Paleo-Siberian populations (Lell et al. 2002; Flegontov et al. 2016; Karafet et al. 2002; Karmin et al. 2015).

Previous studies have also provided data of paternal genetic structure for almost all Mongolic-speaking populations. Early studies found that C2-M217 reached high frequencies in most Mongolic-speaking populations (Su et al. 1999; Karafet et al. 2001; Katoh et al. 2005; Zerjal et al. 2003). Further studies with better resolution of phylogeny of Y-chromosome haplogroup provided more insights about related demographic history. Firstly,

C2-F1756 is the predominant paternal lineage of ancient Donghu and Xianbei populations (Zhang et al. 2018; Li et al. 2018a; Wang et al. 2012), who are considered to be “proto-Mongols” (Lin 2007; Wei et al. 2017). Secondly, research of C2c1a1a1-M407 suggested that this haplogroup may originate in southern part of northeast China and it eventually became one of the predominant lineages of the western Mongolic-speaking populations (Balnova et al. 2019; Huang et al. 2018; Malyarchuk et al. 2016; Malyarchuk et al. 2010). Furthermore, scholars believed that the spread of lineage C2-star cluster in Eurasia was directly related to the expansion of Mongols tribes since the 13th century (Zerjal et al. 2003; Wei, Yan, et al. 2018; Abilev et al. 2012). Also, haplogroup C2-F7171 is an important lineage in western Mongolic-speaking populations (Balnova et al. 2019; Malyarchuk et al. 2013; Di Cristofaro et al. 2013; Liu et al. 2020), but the history of this clade remains unclear. Haplogroup N-M231 also reached high frequencies in some Mongolic-speaking populations (Hammer et al. 2006; Balnova et al. 2019). In addition, O-M175 was identified with high frequency among Mongolic-speaking populations (Balnova et al. 2019; Malyarchuk et al. 2013; Katoh et al. 2005; Xue et al. 2006). On the other side, the frequencies of haplogroups D, E, I, J, Q, R1a and R1b were relatively low among Mongolic-speaking populations (see the references cited in this paragraph).

The genetic studies revealed the evolution history of human beings and this kind of results provided evidence for research in other disciplines. However, most of previous studies focused on the genetic data they generated (Katoh et al. 2005) or the studied populations (Wang et al. 2019) with evidence from ethnology, history, archaeology and linguistics. Some research only discussed a particular sub-clade of genetic lineage (Huang et al. 2018). So far, there has not been a comprehensive summary of the research over the past 20 years, discussing the gains and losses of these studies, and analyzing the benefits and implications of these studies to the research of other disciplines. In addition, the current molecular anthropology research has entered the era of massive whole genome and ancient DNA, so it is necessary to conduct a comprehensive summary on the previous research and provide some prospects for the future research.

In this study, we summarized the paternal Y-chromosome genetic data of Mongolic-speaking populations in all publicly available literature to date. Our first objective is to describe the overall pattern of the paternal genetic structure of the all Mongolic-speaking populations. Secondly, the origin of all identified paternal components and their roles in the evolution history of Mongolic-speaking populations will be fully discussed. Thirdly, we will explore the possible origin and the mixing process of varied ancestral groups, as well as the

evolution path through which they evolved into modern Mongolic-speaking populations. Finally, we will discuss possible implications of the current genetic achievements for multidisciplinary research in ethnology, history, archaeology and linguistics in the future.

Materials and methods

Summarizing of data

The phylogenetic tree of Y-SNP changed rapidly and intensively in the past decades (Karmin et al. 2015; Karafet et al. 2008; Underhill et al. 2000; Y Chromosome Consortium 2002). The same Y-SNP marker may be listed with different haplogroup name in different publications. Also, researchers from different institutes may test different Y-SNP markers in different publications. Therefore, it is a challenge to summarize the frequency data from publications in the past decades. The following steps were used to summarize the frequency data from publication with a unified phylogenetic tree (Figure 1).

Firstly, we followed regulations proposed by the Y Chromosome Consortium (Karafet et al. 2008; Y Chromosome Consortium 2002) and adopted the latest phylogeny tree from ISOGG (www.isogg.org). We selected a set of markers that were widely used in most of previous publications. Then we constructed a unified phylogenetic tree with haplogroup name (Figure 1).

Frequency data of Y-SNP haplogroup were adopted from previous literatures (Su et al. 1999; Wells et al. 2001; Shi et al. 2008; Karafet et al. 2001; Katoh et al. 2005; Hong et al. 2006; Xue et al. 2006; Zerjal et al. 2002; Hammer et al. 2006; Derenko, Malyarchuk, Denisova, et al. 2006; Derenko, Malyarchuk, Wozniak, et al. 2006; Malyarchuk et al. 2010; Derenko et al. 2007; Zhong et al. 2010; Zhong et al. 2011).

The frequency data in the previous literature were adjusted according to the unified phylogenetic tree in this research (Figure 1).

Several rules were used to re-align the frequency data.

- A) Symbol “-” was used to indicate that one marker of unified phylogenetic tree was not tested in the corresponding publication.
- B) The frequency of a haplogroup were imputed to be zero according to topology of unified phylogenetic tree and original data. For example, if both marker M231 and M46 were tested while the frequency of N-M231xM46 is zero, it means that the frequency of all other branches of N-M231 is zero, including N1a2b-P43 and N1*-LLY22g unified phylogenetic tree. If the total amount of C-M130, F-M89 and K-M9 reached to 100% percent of all samples of a population, then the frequency of DE-M1 of BT-SRY1532.1 will be determined as zero. This rule was applied
- C) if many sub-branches tested in a literature while we only list a macro-haplogroup in the unified phylogenetic tree in this research, the original frequency data will be accumulated into one number. For example, both O1-M119, O1b-M95, and O2-M122 were tested a literature. But we only used O-M175 in this study. Therefore, the three original frequency data will be stacked up to generate the frequency of O-M175.

The results of summary were listed in Table 1 and Table S1. ALL of the three rules above were applied throughout the re-alignment. The authors are familiar with the topology

Commented [JW1]: Trish: this is a numbered list; the styling removes the numbers, but they will be there when the article is typeset

Commented [JW2]: Trish: this is styled as an unnumbered list, because they used A,B,C for this grouping

of phylogenetic tree and had re-checked every number in Table 1 and Table S1.

General frequency of a haplogroup

Since different markers might be tested in different publications, it is impossible to calculate the accurate percentage of a paternal haplogroup in all publications. Therefore, we use a compromised method to calculate the overall proportion of a paternal lineage in all Mongolian populations.

For each haplogroup, the number of positive individuals of a marker was accumulated as numerator and the corresponding sampling size was accumulated as denominator. The quotient of these two numbers is the frequency of the analyzed haplogroup in all Mongolic-speaking populations. Due to the special calculation method, the sum of all frequencies is not equal to hundred percent. This kind of calculation aims to summarize all useful frequency data as much as possible. Overall, the frequencies data are listed in Table 1 and actual number of positive individuals are listed in Table S1. Also, the details of all references showed with series number in Table 1 are also listed in Table S1. Besides, we also collected the Y-SNP data of ancient DNA from the East Mongolian Plateau and adjacent region (Table S2) (Seguin-Orlando et al. 2014; Allentoft et al. 2015; Fu et al. 2016; Jeong et al. 2018; Damgaard et al. 2018; Lamnidis et al. 2018; de Barros Damgaard et al. 2018; Narasimhan et al. 2019; Sikora et al. 2019; Flegontov et al. 2019; Yang et al. 2020; Cui et al. 2020; Ning et al. 2020; Jeong et al. 2020; Yu et al. 2020; Ma et al. 2021; Wang et al. 2021; Kilinc et al. 2021; Gneccchi-Ruscione et al. 2021).

{~?~IM: insert 4233 Figure 1 here.}

Figure 1. The major paternal lineages of the people in Northern Asia and the categories in Table 2. The six common components of Mongolic-speaking populations discussed in the following sections were also indicated.

Classification of genetic components

The studied paternal genetic components are divided into two categories, or levels. The first level is the common components of ancestor group, which is the predominant component in the common ancestor group. If the proportion of a component in this population group is very low, but this component is almost unique to this population group, we also classify this component as common component of ancestor group. The second level is admixed components. Almost all modern population groups are admixture of several ancient populations. Having identified the common components, it is necessary to further study how

and when those admixed components became incorporated into modern populations. Subsequently, we will fully discuss all the paternal profiles that exist in the Mongolic-speaking populations and determine which level these paternal components belong to.

Results and Discussion

The general pattern of paternal gene pool of Mongolic-speaking populations

In Table 1, we collected paternal haplogroup data of all Mongolic-speaking populations from almost all related literature published between 1999 and 2019. Our summary is sufficient to interpret the overall picture of the paternal genetic structure of these populations. The sampling locations of summarized data are shown in Figure 1.

{~?~IM: insert 4233 Figure 2 here.}

Figure 2. Schematic diagram for all sampling locations are listed in Table 1. Green diamonds indicated the approximate location while the red circle indicated the capital of some countries.

{~?~IM: insert Table 1 here.}

No.	Populations	N	Y* Others	SNPs																Ref.						
				SRY152.1 BT	M1 DE	M17.4 D	M139 C	M217 C2*	M48 C2a	M77/AB6 C2a	M407 C2b	M89 F*	M9 K*	M214 NO*	M175 O	M231 N	LLY228 N1*	P48 N1a2b	M46 N1a1		M45 P*	M242 Q	M207 R*	SRY152.2 R1a1*	M17 R1a1a	
1	Mongolian	24	0	0	0	4.2	13	-	46	-	-	8.3	13	-	13	-	-	0	0	0	0	0	4.2	1,2,3		
2	Mongolian	147	0	0	0	2.7	0	52	-	-	-	4.1	31	-	-	-	2.7	2.7	-	2.0	2.0	-	4			
3	Mongolian-Urunkhai	60	0	0	0	1.7	25	-	33	-	-	0	5	-	12	-	-	8.3	8.3	-	-	0	6.7	5		
4	Mongolian-Zakchia	60	0	0	0	3.3	17	-	30	-	-	3.3	5	-	20	-	-	3.3	5	-	-	0	13	5		
5	Mongolian-Khalkh	85	0	0	0	3.5	41	-	15	-	-	4.7	3.5	-	19	-	-	4.7	4.7	-	-	0	3.5	5		
6	Mongolian-Khalkh	48	4.2	-	2.1	-	44	-	-	-	-	-	10	15	25	-	-	-	-	-	-	-	6			
7	Mongolian-Inner Mongolia	45	0	4.4	0	-	0	38	8.9	-	-	-	4.5	0	31	-	0	0	13	0	0	0	0	7		
8	Mongolian-Mongolia	65	0	6.2	1.5	-	0	34	20	-	-	-	1.5	0	12	-	0	3.1	7.7	4.6	-	-	0	9.2	7,8	
9	Mongolian	149	0	0	0	2.7	0	34	-	18	-	4.0	26	0.7	-	-	0	6.0	2.0	0	2.7	4.0	-	9		
10	Mongolian	47	0	0	0	0	57	-	-	-	-	2.1	21	-	-	-	6.4	2.1	4.3	-	4.3	0	2.1	10,11		
11	Mongolian	46	26	-	-	-	0	39	11	-	15	-	-	-	-	-	6.5	2.2	-	-	-	-	-	12,13		
12	Mongolian	12	0	0	0	0	0	25	0	0	0	0	58	-	-	-	8.3	-	-	-	0	8.3	-	14,15,16		
13	Mongolian	23	48	-	-	-	0	35	17	-	0	-	-	-	-	-	-	-	-	-	-	-	-	14		
14	Mongolian	50	0	0	0	4	40	-	-	-	0	4	-	38	6	-	-	0	2	2	-	4	16			
15	Mongolian	22	0	0	0	0	55	-	-	-	9.1	18	-	4.6	-	-	-	4.6	0	0	9.1	17				
16	Mongolian-East	18	0	0	0	11	0	22	-	22	5.6	5.6	0	0	5.6	0	-	5.6	0	0	11	11	0	18		
17	Mongolian-Northeast	20	0	0	0	5	0	20	-	15	10	0	0	0	30	0	-	10	10	0	0	0	0	18		
18	Mongolian-Northwest	97	0	0	0	1	0	16	-	30	6.2	7.2	0	0	13	0	-	1	12	0	7.2	3.1	0	2.1	18	
19	Mongolian-Southeast	23	0	0	0	8.7	0	35	-	4.4	8.7	4.4	0	0	17	0	-	4.4	0	0	4.4	13	0	18		
20	Mongolian-Southwest	2	0	0	0	0	0	50	-	0	50	0	0	0	0	0	-	0	0	0	0	0	0	18		
21	Buryats	4	0	0	0	0	75	-	-	-	-	25	0	0	0	0	-	0	0	0	0	0	0	1		
22	Buryats	13	0	0	0	0	0	0	7.7	-	0	7.7	62	-	-	-	-	23	0	0	0	0	0	19		
23	Buryats	61	0	6.6	0	0	0	84	-	-	-	-	0	-	9.8	-	-	0	0	0	0	0	0	20		
24	Buryats	50	4	-	4	-	38	-	-	-	-	4	30	20	-	-	-	-	-	-	-	-	-	6		
25	Buryats	81	0	0	0	0	0	56	-	4.9	-	2.5	0	0	2.5	-	0	2.5	28	0	0	2.5	12	4.9		
26	Buryats	238	0	0	0	0	64	-	-	-	2.5	8.8	-	-	-	-	1.3	19	1.7	-	0.8	0	2.1	10,11		
27	Buryats	217	13	-	-	-	0	8.3	6.0	-	54	-	-	-	-	-	0.5	18	-	-	-	-	-	12,13		
28	Buryats-Barghuts	76	0	0	0	0	0	10.5	-	55	2.6	1.3	0	1.3	0	0	0	28	0	0	1.3	0	0	21		
29	Kalmyks	99	0	1.0	-	24	-	37	-	-	-	6.1	13	-	-	-	1.0	11	-	6.1	-	0	22			
30	Kalmyks	68	0	0	0	0	71	-	-	-	0	5.9	-	-	-	-	2.9	0	12	-	2.9	0	5.9	10,11		
31	Kalmyks	60	97	-	-	-	-	-	-	-	-	-	-	-	-	-	3.3	0	-	-	-	-	-	13		
32	Kalmyks	91	37	-	-	-	0	5.5	45	-	12	-	-	-	-	-	-	-	-	-	-	-	-	12		
33	Kalmyk Derbet	165	0	0	0	1.2	0	5.5	33	-	19	2.4	0	0	7.9	0	-	1.2	5.5	0	3.6	17	-	3.6	23	
34	Kalmyk Khoshut	82	0	0	0	0	0	1.2	38	-	0	0	0	0	17	2.4	-	1.2	37	0	1.2	2.4	-	0	23	
35	Kalmyk Torgut	150	0	0	0	0.7	2	0	11	46	-	8.7	3.4	0	0	17	0	-	1.3	2	0	0.7	3.4	-	4.7	23
36	Kalmyk Derbet	69	0	0	0	2.9	0	0	0	32	29	4.3	0	0	4.3	0	0	1.4	0	0	0	2.6	0	0	24	
37	Kalmyk Torgut	58	0	0	0	0	0	3.4	0	43	8.6	5.2	0	0	24	0	0	3.4	0	3.4	0	3.4	0	5.1	24	
38	Kalmyk Buzava	52	0	0	0	0	0	3.8	0	48	14	12	0	0	1.9	0	0	0	0	0	1.7	0	0	3.8	24	
39	Kalmyk Khoshut	28	0	0	0	0	0	0	0	50	3.6	0	0	0	14	0	0	0	32	0	0	0	0	0	24	
40	Kalmyk Sart	61	0	0	0	1.6	0	8.2	0	4.9	0	2.5	0	0	1.6	0	0	4.9	0	0	1.8	0	0	3.6	24	
41	Kalmyk Xinjiang	12	0	0	0	0	8.3	0	17	0	25	8.3	8.3	0	0	25	0	0	8.3	0	0	0	0	0	24	
42	Mongol Torgut	47	0	0	0	0	0	17	4.3	30	15	6.4	0	0	8.6	0	0	4.2	11	0	0	0	0	4.2	24	
43	Mongol Khoshut	18	0	0	0	0	0	0	11	67	0	5.6	0	0	5.6	0	0	5.6	0	5.6	0	5.6	0	0	24	
44	Mongol Derbet	40	0	0	0	0	0	10	2.5	53	7.5	2.5	0	0	7.5	0	0	13	0	0	5	0	0	24		
45	Khamnikan	51	45	-	-	-	0	2	0	-	53	-	-	-	-	-	-	-	-	-	-	-	-	12		
46	Yugur-East	45	0	0	2.2	-	4.4	-	-	-	18	16	-	44	-	-	0	13	-	-	0	2.2	25			
47	Daur	39	0	5.1	0	-	0	28	2.6	-	-	0	2.6	54	-	0	0	7.7	0	0	0	0	0	7		
48	Tu	50	0	0	0	8	28	-	-	-	4	2	-	40	10	-	-	0	2	4	-	2	16			
49	Dongxiang	35	0	0	0	0	2.9	-	-	-	2.9	0	0	11	5.7	-	-	0	23	0	0	54	16			
50	Bao'an	27	0	0	0	15	0	0	0	0	0	0	22	0	41	3.7	-	-	0	0	11	0	7.4	16		
	Sum	3130																								

Table 1. Frequencies of paternal Y-chromosome haplogroup in Mongolic-speaking populations. Notes: The actual number of positive individuals for each frequency number and the details of all references are listed Table S1. The dash line in the table means that this Y-SNP maker was not tested in the corresponding article.

{~?~IM: insert Table 2 here.}

No. of Lineage	Haplogroup	Positive individuals / Sampling size	Frequency	Mainly contains	Category
	DE*-YAP all	47/2642	1.80%		

1	DE-YAPxM174	3/2251	0.13%	E-M96	Admixed components
2	D-M174	38/2251	1.69%		Admixed components
	C-M130 all	1676/3074	54.52%		
	C2-M217 all	1269/2413	51.31%		
3, 4	C2*-M217 (xM48/M86,M407)	127/1399	9.08%	C2a-M504, C2a-F1756	Common components
5	C2a-M48/M86	548/2129	25.74%		Common components
6	C2b-M407	304/1475	20.61%		Common components
7	F*-M89xM9	103/2334	4.41%	G, H, J	Admixed components
8	K*-M9x(N, O, Q, R)	7/1160	0.60%	L	Admixed components
	N-M231 all	304/2806	10.83%		Common components
9	N1a2b-P43	36/1724	2.09%		
10	N1a1-M46/Tat	245/2550	9.61%		
11	O-M175	296/1849	16.01%		Common components
12	Q-M242	36/1630	2.21%		Admixed components
13	R-M207xSRY1532.2	110/2019	5.45%	R1b-M73, R1b-M269, R2a-M124	Admixed components
14	R1a1-SRY1532.2	107/2397	4.46%	R1a1a-M17	Admixed components

Table 2, The general frequencies of varied Y-chromosome haplogroup among Mongolic-speaking populations and the corresponding category.

According to Table 1 and Table 2, the overall proportion of haplogroups DE*-YAPx(M174) and D-M174 in Mongolic-speaking populations are very low, which is 0.13% and 1.69% respectively. Haplogroup D-M174 is the predominant paternal lineage of populations on the Tibet Plateau (Shi et al. 2008; Qi et al. 2013). According to historic records, with the spread of Tibetan Buddhism, there were continuous and extensive population and cultural exchanges between the people of the Tibetan Plateau and the Mongolian Plateau after the Yuan Dynasty (Wallace 2015). Furthermore, we inferred that the haplogroup DE*-YAPx(M174) belongs to haplogroup E-M96. This paternal lineage concentrated in populations from Africa and the Middle East (de Filippo et al. 2011). Therefore, we proposed that the haplogroups DE*-YAPx(M174) and D-M174 in Mongolic-speaking populations resulted from recent admixtures.

As shown in Table 2, The total frequency of macro-haplogroup C2-M217 reached 51.31% among all Mongolic-speaking populations. The proportion of this macro-haplogroup is much higher than that of all other branches, including DE*, D, F*, K*, N, O, Q, and R. Many ancient DNA studies also revealed larger number of ancient remains of C2-M217 in archaeological sites of the Mongolian Plateau and adjacent regions (de Barros Damgaard et

al. 2018; Allentoft et al. 2015; Ning et al. 2020; Wang et al. 2021; Robbeets et al. 2021) (Table S2). Therefore, we can conclude that macro-haplogroup C2-M217 is the most predominant paternal lineage of Mongolic-speaking populations.

In further study, we found that haplogroup C2*-M217 (xM48/M86,M407) can be divided into two clades, C2a-M504 (nearly equal to “Star Cluster”) and C2a-F1756(DYS448del) (Wei, Yan, et al. 2018; Wei et al. 2017). Haplogroup C2a-M504 had been found with high frequencies in many modern Mongolic-speaking populations and some Turkic-speaking populations (Abilev et al. 2012; Wei, Yan, et al. 2018; Zerjal et al. 2003). It is generally accepted that the current distribution of haplogroup C2a-M504 is directly related to the expansion of Mongol tribes since the 13th century across the Eurasia steppe.

Haplogroup C2a-F1756(DYS448del) is widely distributed in Mongolic-speaking populations and several Turkic-speaking populations, but generally had a low frequency (Wei et al. 2017). The ancient DNA evidence revealed that C2a-F1756(DYS448del) is the predominant paternal lineage of ancient Donghu, Xian-bei, and Rouran populations (Zhang et al. 2018; Li et al. 2018a; Wang et al. 2012) (Table S2). Historians considered that these ancient people might be classified as “proto-Mongols” (Zhang 1998). Therefore, we conclude that this paternal lineage has played an important role in studying the origin and expansion of the Mongols. Although the frequency of this haplogroup is generally low among modern Mongolic-speaking populations, the available ancient DNA and modern DNA data demonstrated that this haplogroup had played key role during the origin and formation of ancient “proto-Mongols” and modern Mongolic-speaking populations. Therefore, the haplogroup was also classified as one of common components in ancestor groups of Mongolic-speaking populations.

High frequency of C2a-M48 was found in Mongolic-speaking populations (especially the Oriat-speaking groups) (Di Cristofaro et al. 2013; Balinova et al. 2019; Malyarchuk et al. 2013) and some Turkic-speaking populations, like Kazakhs (Dulik, Osipova, and Schurr 2011) and Kyrgyz (Zerjal et al. 2002). Further analyses of Y-chromosome sequences confirmed that most samples from these populations belong to lineage C2a1a2a2-F6170, the western branch of C2a-M48 (Liu et al. 2020; Zhabagin et al. 2021). The other major sub-branch of C2a-M48, C2a1a2a1-F5484, is the only founding paternal lineage of Tungusic-speaking populations (Liu et al. 2020). We suggested that haplotype C2a1a2-F6170 may originate from northwest Mongolia and southern Siberia. After the 13th century, the activities of Mongols in the Western Mongolian Plateau and Central Asia may contribute to the distribution of this branch among modern populations.

High frequency of this haplogroup was also found in Buryats and Kalmyks, but generally low in other Mongolic-speaking populations (Malyarchuk et al. 2010; Balinova et al. 2019; Malyarchuk et al. 2013). Apart from Mongolic-speaking populations, haplogroup C2b-M407 was found with very low frequencies in population of other language families (Huang et al. 2018). Thus, this haplogroup can be considered as a characteristic paternal lineage of Mongolic-speaking populations and their ancestor group (Huang et al. 2018). Except for C2b-M407, almost all downstream branches of C2b-F1067 are mainly distributed in East Asia (Wu et al. 2020).

Haplogroup F*-M89x(M9) contains the subclades G-M201, H-L901, I-M170, and J-M304 (Karafet et al. 2008). The total frequency of haplogroup F*-M89x(M9) is about 4.41% in all Mongolic-speaking populations. Haplogroup K*-M9x(N, O, Q, R) mainly contains lineage L-M20. According to previous studies, these paternal lineages originated in South Asia, Middle East, the Caucasus, and Europe (Rootsi et al. 2012; Sengupta et al. 2006; Fechner et al. 2008; Rootsi et al. 2004; Semino et al. 2004). Therefore, we considered that these paternal lineages are the reflection of populations admixture from western Eurasia.

The total frequency of N-M231 reached 10.83% among all Mongolic-speaking populations (Table 2). High frequency of this haplogroup was found in Buryats, Kalmyk, and some samplings of Mongolian. In addition, varied sub-branches of N-M231 were founding paternal lineages of Uralic-speaking populations (Ilumäe et al. 2016; Karmin et al. 2015; Rootsi et al. 2007). Large proportions of N-M231 were also found with high frequency in many Turkic-speaking populations (Rootsi et al. 2007; Khar'kov et al. 2008). Ancient DNA indicated that N-M231 is one of predominant paternal lineages in ancient people on the Mongolian Plateau (de Barros Damgaard et al. 2018; Wang et al. 2021; Yu et al. 2020; Jeong et al. 2018; Jeong et al. 2019) (Table S2). Most of these people were integrated as part of Mongols after the 13th century. Therefore, we can conclude that haplogroup N-M231 is also one of common components in ancestor group of modern Mongolic-speaking populations.

The frequencies of haplogroup O-M175 reached 15.98% among all Mongolic-speaking populations (Table 2). Among the fourteen paternal components listed in Table 2, the proportion of haplogroup O-M175 is the third highest. It appeared with high frequency (11.7%~38%) in nearly all samplings of modern Mongolian (Table 1, No.1~No.19). As shown in Table 1, high frequency of haplogroup O-M175 can be found in samples from very remote regions, like those of Uriankhai and Zakhchin (Kato et al. 2005). Therefore, we concluded that this haplogroup is one of major components in ancestor group of modern Mongolic-speaking populations. In general, the appearance of haplogroup O-M175 in

Mongolic-speaking populations can be considered as the result of intense exchange and interaction in the past five thousand years between ancient people on the Mongolian Plateau and people from other parts of East Asia.

The frequency of haplogroup Q-M242, R-M207xSRY1532.2, and R1a1-SRY1532.2 are 2.21%, 5.45%, and 4.46% respectively. Ancient DNA analyses suggested that haplogroup Q is the dominant paternal lineage among populations in Northern Asia 4000 years ago (de Barros Damgaard et al. 2018; Wang et al. 2021; Yu et al. 2020; Jeong et al. 2018; Jeong et al. 2019; Raghavan et al. 2014). The appearance of Q-M242 is likely the result of admixture during the expansion of ancient Mongolians. Some sub-branches of R1b-M343, like M73 (Malyarchuk et al. 2011) and PH155 (Cinnioglu et al. 2004) may be lineages of indigenous populations in western Siberia and central Asia. However, most of R1b-M269 (Balaesque et al. 2010) and R1a1-SRY1532.2 (Underhill et al. 2015) are likely the results of admixture with populations who expanded from western Eurasia. The total frequency of R-M207 exceeded ten percent. Therefore, although we believe that the present of R-M207 is the result of admixture, the contribution of ancient people with dominant lineage R-M207 to the genetic structure and cultural traditions of modern Mongolic-speaking populations cannot be underestimated.

In summary, we provided a comprehensive frequency list for fourteen kinds of paternal components among all Mongolic-speaking populations. We thoroughly discussed the distribution and possible source of these lineages. Based on the discussion above, we propose that the following paternal lineages can be considered as common components in ancestor group of Mongolic-speaking populations, including C2a-F3796, C2a-F1756, C2a-M48/M77, C2b-M407, N-M231 and O-M175 (Table 2). We also propose that the following paternal components are the results of the continuous mixing of Mongolic-speaking populations with surrounding populations during the recent historical period, including various sub-branches of E, D, I, J, Q, and R (Table 2).

Possible sources of six major components

Combining the achievements of studies of ethnology, history, archaeology and linguistics, we discussed the possible origins of six major paternal components in the ancestor group of Mongolic-speaking populations and the evolution history of related ancient populations (Figure 3).

In one of our early studies (Wei, Wang, et al. 2018), we generated a phylogenetic tree

with divergence age and distribution map for haplogroup C2a-F1756(DYS448del). More and more ancient DNA data of this lineage in recent years provided a lot of insights about this paternal lineage (Mao et al. 2021; Sikora et al. 2019; Li et al. 2018b; de Barros Damgaard et al. 2018) (Table S2). We provided a list with relative ancient DNA data in the supplementary file of Sun et al. 2020 (Sun et al. 2020) and updated it in this study (Table S2). Summarizing all available evidence, we proposed that haplogroup C2a-F1756(DYS448del) may be the predominant paternal lineage in Neolithic populations from the southeastern Lake Baikal to the eastern part of the Mongolian Plateau from 5.5 to 3.3 thousand years ago (kya, Figure 3.A). About 4.3 thousand years ago, haplogroup C2a-F1756(DYS448del) differentiated into two sub-branches. The western branch, C2a-F8497, mainly distributed in populations originated from the Altai Mountain areas. The eastern branch, C2a-F3889, experienced continuous expansion between 3.3 kya and 1.6 kya. As a result, this haplogroup became the predominant paternal lineage of Donghu, Wuhuan, Xianbei and their ancestor groups. These ancient populations were considered as “proto-Mongols.” The westward diffusion of C2a-F3899 may had led to the appearance of this lineages in Dingling, Tiele, and Uyghur tribes who were considered as ancestor of modern Turkic-speaking populations (Figure 3A). Many Xianbei tribes once lived in northern and western China (Lin 2007). They integrated into the Han populations at later age. In summary, although the frequency of haplogroup C2a-F1756(DYS448del) is generally low among modern Mongolic- and Turkic-speaking populations, the current available evidence is sufficient to demonstrate that ancient populations with this lineage was one of the most powerful groups in the eastern Eurasian grassland before the rising of Mongols in the 13th century.

In another research of our team, we generated a phylogenetic tree with divergence age and distribution map for haplogroup C2a-M504 (Wei, Yan, et al. 2018). Ancient DNA research has found several ancient remains belonging to this branch during the Mongolian Empire (Jeong et al. 2020; Wang et al. 2021) (Table S2). Due to the lack of ancient DNA data, the origin of this lineage before 4 kya remains unclear. Haplogroup C2a-M504 diverged into two major sub-branches (F3796 and F8951) at about 4 kya. Sub-branch F8951 concentrated in Daur people (Wang et al. 2019) while F3796 spread widely with high frequency in many other Mongolic-speaking populations (Wei, Yan, et al. 2018). According to research of History and Ethnology, the common ancestor of Mongolian and Daur is Shiwei tribes who live in the northern part of Greater Khingan mountain region between 4 kya and 1 kya (Lin 2007; Zhang 1998). Therefore, we suggested that this region might be the source of expansion of lineage C2a-M504 at later age. According to previous studies, the continuous

branching of sub-lineage C2a-F12308-F9700 between 3ky and 1kya may correspond to the in-situ population growth of Mongols tribes in the Greater Khingan mountain region and the subsequent westward diffusion of them onto the steppe after the collapse of Uyghur Khannate at around 9th century (Zerjal et al. 2003; Wei, Yan, et al. 2018). Furthermore, the significant expansion of sub-branch C2a-F5481 may directly related to the dispersal of Mongols across the Eurasian steppe since the 13th century (Abilev et al. 2012).

Ancient DNA data about haplogroup C2b-M407 is limited (only one individual in Table S2). Nevertheless, the phylogeographics of this lineage and studies of archaeology and history provide some clues for the origin and diffusion process of this lineage (Balinova et al. 2019; Huang et al. 2018; Malyarchuk et al. 2016; Malyarchuk et al. 2010). The ancient populations with paternal lineage C2b-M407 may live in the southern part of northeast China during the Neolithic age. Divergence time of phylogeny of this lineage suggested a continuous differentiation process during the Bronze Age (about 4 kya to 2 kya in northeast China). Xianbei-Shiwei tribes rose from 2000 to 1500 years ago, and established a series of political power in north China and northeast China (Lin 2007). Ancient populations with sub-branches of lineage C2b-M407, like C2b-F8465 and C2b-F3916, may became parts of Shiwei tribe during this period. The spread of lineage C2b-M407 in eastern part of Mongolian Plateau and Cis-Baikal region may be related to Wu-gu (乌古), Yu-jue-li (羽厥里), and Khongirad (弘吉剌惕) (Feng 2002). The admixture between immigrants and local people in the south Baikal Lake gave birth to the varied tribes of “Forest people” (林中百姓), which is the direct ancestor of the later Oirat Mongols (Onon 2001). We proposed that these demographic processes resulted in high frequency of C2b-M407 in Mongolic-speaking populations in the west, especially Kalmyks.

Summarizing all available evidence, we proposed that the origin of C2a-M48/M86 in Mongolic-speaking populations can be summarized into several steps (Liu et al. 2020; Zhabagin et al. 2021) (three individuals in Table S2). Firstly, the lineage M48/M86 differentiated into two sub-branches about 3.4 thousand years ago. The eastern branch, C2a-F5484, became the founding paternal lineage of Tungusic-speaking populations. Most of M48/M86 individuals in Mongolic-speaking populations belong to the western sub-branch C2a-F7171 (Balinova et al. 2019; Malyarchuk et al. 2013). Therefore, the southern Baikal Lake region may be the place where the initial differentiation of haplogroup C2a-F7171 took place. Haplogroup C2a-F7171 may be one of predominant paternal lineages of diverse tribes who live in the boundary region of Siberia and modern Mongolia, like “Forest people”,

Kerait (克烈部), Naiman (乃蛮部), and ancestor of modern Tuvian (Onon 2001). After the rising of Mongols, some of these tribes were integrated as parts of Mongols and involved in the dispersal of Mongols at later age. The rest of them may involve in the formation of Modern Kazakhs and Tuvian. We proposed that the demographic process might be the source of high frequency of C2a-F7171 in Mongolic-speaking populations and some Turkic-speaking populations.

Previous studies suggested that haplogroup N-M231 spread widely in nearly all populations in northern Eurasia (Rootsi et al. 2007). Further analyses demonstrated that varied sub-branches of N-M231 are founding paternal of Uralic-speaking populations (Ilumäe et al. 2016; Karmin et al. 2015). This haplogroup was also found with high frequency among some Mongolic- and Turkic-speaking populations (Table 1) (Keyser et al. 2015). Ancient DNA data suggested that this haplogroup is dominant in ancient populations on the Mongolian Plateau before the rise of Mongols since the 13th century (de Barros Damgaard et al. 2018; Yu et al. 2020; Jeong et al. 2018; Jeong et al. 2019) (Table S2). High frequency of N-M231 was found in Buryats (Table 1) and Jalair tribe of Kazakhs (Zhabagin et al. 2020) while these two populations can be traced to the oldest parts of ancient Mongol tribes. In general, we suggested that haplogroup N-M231 is one of important and major paternal lineages of ancestor groups of Mongolic-speaking populations.

Haplogroup O-M175 spread widely with high frequency in some Kazakhs (Dulik, Osipova, and Schurr 2011) and near all populations from East Asia, such as Han, Korean, Japanese, and Tibeto-Burman speaking populations (Zhong et al. 2011) (also see the ancient DNA information in Table S2). It is worth noting that O1b-M268, a sub-branch of O-M175, was found in Kalmyk-Torgut, Mongolian-Uriankhai, and Mongolian-Zakhchin (Kato et al. 2005; Balinova et al. 2019). Haplogroup O1b-M268 is a special paternal lineage. It is major component among Korean and also spread widely with low frequency among populations from northeast China (Xue et al. 2006). Therefore, the appearance of O1b-M268 in Mongolic-speaking populations may be the result of admixture with ancient populations in northeast part of China. Since the beginning of Chalcolithic Age and Bronze Age, there were many communication channels between the Mongolian Plateau and North and Northwest China in the past four thousand years (Liu et al. 2016; Eng 2016). These channels have been used since ancient time, promoting the close and profound exchange and integration of people, materials, and cultures in two regions. These demographic processes may have contributed the appearance of other sub-branches of O-M175, like O1a-M119 and O2-M122,

in modern Mongolic-speaking populations.

{~?~IM: insert 4233 Figure 3 here.}

Figure 3. The possible origin and diffusion pattern of ancient populations with the six major paternal lineages of modern Mongolic-speaking populations. Dash line indicated that the details of diffusion during the corresponding period remain ambiguous.

Implications for multidisciplinary research in the future

After 20 years' research, scholars have basically revealed the complete picture of the paternal genetic structure of Mongolic-speaking populations. In the above sections, we had discussed possible origin of fourteen components of the Mongolic-speaking populations. We also identified six possible major paternal lineages in ancestor groups of Mongolic-speaking populations. There have also been many publications about the maternal mtDNA of Mongolic-speaking populations. So far, more than ten ancient DNA studies have involved ancient human remains that might be related to the ancestors of Mongolic-speaking populations. Several recent articles based on ancient genomic and genome-wide chip data have provided the fine mixing and evolution history of populations in northern Asia, including Mongolic-speaking populations and their ancient relatives. In conclusion, we supposed that genetic studies have revealed a very detailed evolution history of the Mongolic-speaking populations. These research results will be helpful to the future multidisciplinary research involving ethnology, history, archaeology and linguistics.

Theory of molecular anthropology and some critiques

The main work of studies of molecular anthropology is to reveal some ambiguous parts of human evolutionary history through DNA data. However, related research has caused a lot of controversy and concern in other disciplines. People are generally worried that DNA markers will be used as new labels and symbols by the idea of racial discrimination that has existed for a long time in human history. We believe that, like all other types of scientific research, "seeking truth" is the original and only purpose of molecular anthropology. To a greater extent, how people interpret the data displayed by molecular anthropology and how they use it to construct their own cognition and theory is determined by the orientation of contemporary scholars of Humanities and Social Sciences and the public themselves. These topics are crucial to the development of human society, so scholars from different disciplines need to carry out more and more in-depth dialogues.

For example, in this study, the ancestral groups of modern Mongolian people have six

major paternal components. The “ancient Mongols” were formed after Genghis Khan established the Mongol Empire. The mixing of people from all directions is the key attribute of this newly formed “ancient Mongols” group. It is worth noting that all of these six components also appear in many other populations and they are not purely unique for modern Mongolic-speaking populations. Here in this study, we only discuss the part of these six components among modern Mongolic-speaking populations. The other parts of these six components may be related to evolutionary history of other populations. We believe that in the more ancient prehistoric period, the population with these components as the major paternal line cannot be called “Mongols.” These ancient people either have their own names in historic records, or they are ancient people with unknown ethnic status, and their ethnic status may also change many times in history. We propose that an ethnic name in a special historical period cannot be used on their ancestor groups in older age. We believe that this criterion helps to avoid the confusion caused by relevant research among scholars and the public.

Research of History

Studies of genetics have revealed differences between the evolution process of Mongolic-speaking populations and other populations in eastern Asia. As discussed in the Introduction section, there are multiple groups of people in northern Asia, including Japanese, Korean, Paleosiberians, Tungusic-speaking populations, Turkic-speaking populations, and Uralic-speaking populations. It can be seen from the above sections that each of those groups has its own major paternal lineages, some of which are shared within the Mongolic-speaking populations. Further analyses are needed to explain the emergence of Mongolic-speaking populations from the highly diverse northern Asian populations. Also, the detailed process of evolution and mixing in prehistoric time requires more ancient genomic research.

The current available genetic data also have demonstrated the differences between sub-groups of Mongolic-speaking populations. Mongolic-speaking populations in the East, such as the Mongols and Daur people in Hulunbuir, share many recently mixed genetic components with Tungusic-speaking populations (Table 1) (Liu et al. 2020). This represents the exchange and intimate relationship of these ethnic groups in recent historical period. There is a high diversity in the genetic structure of Mongolian in different regions of Inner Mongolia Autonomous Region (Table 1). The history of exchange and integration between these populations and populations in North China demands further research. Mongolic-

speaking populations in the West share many genetic components (like R-M207) with Turkic-speaking populations and other populations in central Eurasia, which represents the complex evolution of the Mongolic-speaking populations in the western part of the Mongolia plateau and central Asia after the 13th century. Along the northern border, Mongolic-speaking populations (such as Buryats) experienced long-term admixture with Siberian populations. All of these issues need further study. In short, genetic data can provide new evidence and clues for future ethnological and historical research in the future.

Research of Archaeology

Up till now, genetic studies have revealed the detailed evolution history and unique genetic lineages of modern Mongolic-speaking populations and their ancient relatives. For an archaeological site, the DNA test results of the remains will help to determine the kinship and social status of the corresponding ancient populations. Comparing the newly tested ancient DNA with the previously accumulated data of modern populations, ancient DNA is expected to shed light on the origin and flow direction of the ancient populations at this archaeological site. When the ethnic specific genetic lineages of modern populations have been identified, the discovery of these lineages in the remains of ancient sites is helpful to explore the connection between ancient people and modern populations. The ancient DNA test of the remains of ancient sites in eastern part of Mongolia Plateau is expected to reveal the formation and evolution of multiple ancestral groups of Mongolic-speaking populations in prehistory age. Such research will greatly help archaeologists to interpret archaeological culture and archaeological sites, and will also promote the development of archaeological anthropology to a large extent (also known as archaeology of ethnicity or ethnoarchaeology).

Research of linguistics

Due to the lack of written records in ancient times, it is difficult for linguists to study the origin and early transition process of a languages group (such as language families, language groups, and language branches). Firstly, studies of population genetics can reveal a very detailed process of population evolution, thus providing a very clear and solid background data for the transition of cultural traditions (including language). For example, ancient DNA studies have shown that C2a-F1756 is the major paternal lineage of the “proto-Mongols,” such as Donghu, Xianbei, and Shiwei populations (Wang et al. 2012; Zhang et al. 2018; Li et al. 2018a). However, C2a-M504 is the predominant paternal lineage of Mongols in the 13th

century. These two paternal lineages have been separated for more than 14 000 years (Sun et al. 2020). Why do two groups of people with different dominant paternal lineages share similar languages? Is it possible that some language replacement events took place in the early stage of proto-Mongolian language? Secondly, previous studies demonstrated that many elements of ancient Mongolian language in the 13th century can be found in modern Daur language while these elements have disappeared from other modern Mongolic languages (Eerdengtai, Wuyundalai, and Asalatu 1980). Further multidisciplinary research is needed to reveal the demographic background for this linguistic phenomenon. Thirdly, genetic studies suggested that the Mongolic-speaking populations in the Gansu-Qinghai area in northwest China has fewer components shared with other Mongolic-speaking populations, but has more mixed components with the Han, Tibetan, and Central Asian populations (Table 1). These Mongolian languages were categorized as a separate language branch. Further study is needed to reveal the demographic background of the related language evolution history. Fourthly, the language of the western Mongolians, namely, the Oirat language, is somewhat different from that of the eastern Mongolian. Genetic studies had revealed some differences between these two population groups. How does the related demographic history affect the similarities and differences between the eastern and western Mongolian language branches? All of these issues need to be further studied through multidisciplinary research involving linguistics, genetics, etc.

In conclusion, we reviewed all available publications about the paternal gene pool of Mongolic-speaking populations and their ancestor groups. We summarized a comprehensive frequency table the variant paternal lineages of modern Mongolic-speaking populations. We discussed the role of each lineage during the formation of Mongolic-speaking populations and proposed six major components in paternal gene pool of ancestor of this population group. Also, we discussed a series of possible topics that may be of interest to future interdisciplinary research. The data we summarized and the prospect we proposed in this study will be insightful for exploring more detailed history of Mongolic-speaking population in perspective of multidisciplinary study in the future.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

Preprint version. Visit <http://digitalcommons.wayne.edu/humbiol/> after publication to acquire the final version.

This study was supported by the Key projects of strategic international scientific and technological innovation cooperation of the Chinese Ministry of Science and Technology (2020YFE0201600), the National Natural Science Foundation of China (31900406 to LHW), the National Social Science Fund of China (20BY195 to NS), the Scientific and Technology Committee of Shanghai Municipality (18490750300), and the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP08855823 to MZ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received 8 December 2021; accepted for publication 28 May 2022.

Literature Cited

- Abilev, S., B. Malyarchuk, M. Derenko et al. 2012. The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey Clan from Kazakhstan. *Hum. Biol.* 84:79–89.
- Allentoft, M. E., M. Sikora, K. G. Sjogren et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.
- Amitai, R. 2007. *The Mongols in the Islamic Lands: Studies in the History of the Ilkhanate*. Aldershot, UK and Burlington, VT: Ashgate.
- Balanovsky, O., M. Zhabagin, A. Agdzhoyan et al. 2015. Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranic speakers. *PLoS One* 10:e0122968.
- Balaresque, P., G. R. Bowden, S. M. Adams et al. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8:e1000285.
- Balinova, N., H. Post, A. Kushniarevich et al. 2019. Y-chromosomal analysis of clan structure of Kalmyks, the only European Mongol people, and their relationship to Oirat-Mongols of Inner Asia. *Eur. J. Hum. Genet.* 27:1466–1474.
- Cinnioglu, C., R. King, T. Kivisild et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–148.
- Cui, Y.-Q, F. Zhang, P.-C. Ma et al. 2020. Bioarchaeological perspective on the expansion of Transeurasian languages in Neolithic Amur River basin. *Evol. Hum. Sci.* 2:E15.
- Damgaard, P. B., N. Marchi, S. Rasmussen et al. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557:369–374.
- de Barros Damgaard, P., R. Martiniano, J. Kamm et al. 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360:eaar7711.

Preprint version. Visit <http://digitalcommons.wayne.edu/humbiol/> after publication to acquire the final version.

Commented [MC3]: I couldn't find this one either. This is the original, I couldn't put it right next to how I changed it bc this paper has a linked citation manager.
Feng, E.-X. 2002. Rock paintings around Baikal Lake and the Yu-jue-li tribe in Liao Dynasty (in Chinese), *Northern Cultural Relics*:38-42.

- de Filippo, C., C. Barbieri, M. Whitten et al. 2011. Y-chromosomal variation in sub-Saharan Africa: Insights into the history of Niger-Congo groups. *Mol. Biol. Evol.* 28:1255–1269.
- Derenko, M., B. A. Malyarchuk, M. Wozniak et al. 2006. The diversity of Y-chromosome lineages in indigenous population of South Siberia. *Dokl. Biol. Sci.* 411:466–470.
- Derenko, M., B. Malyarchuk, G. A. Denisova et al. 2006. Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum. Genet.* 118:591–604.
- Derenko, M., B. Malyarchuk, G. Denisova et al. 2007. Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J. Hum. Genet.* 52:763–770.
- Di Cosmo, N., A. J. Frank, and P. B. Golden. 2009. *The Cambridge History of Inner Asia: The Chinggisid Age*. Cambridge, UK: Cambridge University Press.
- Di Cristofaro, J., E. Pennarun, S. Mazieres et al. 2013. Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. *PLoS One* 8:e76748.
- Dulik, M. C., L. P. Osipova, and T. G. Schurr. 2011. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One* 6:e17548.
- Dulik, M. C., S. I. Zhadanov, L. P. Osipova et al. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* 90:229–246.
- Eerdengtai, Wuyundalai, and Asalatu. 1980. *An Interpretation of Selected Words in 'The Secret History of Mongols.'* Hohhot, CN: Inner Mongolia People's Publishing House.
- Eng, J. T. 2016. A bioarchaeological study of osteoarthritis among populations of Northern China and Mongolia during the Bronze Age to Iron Age transition to nomadic pastoralism. *Quat. Int.* 405:172–185.
- Fechner, A., D. Quinque, S. Rychkov et al. 2008. Boundaries and clines in the West Eurasian Y-chromosome landscape: Insights from the European part of Russia. *Am. J. Phys. Anthropol.* 137:41–47.
- Feng, E.-X. 2002. Rock paintings around Baikal Lake and the Yu-jue-li Tribe in Liao Dynasty. *North. Cult. Relics* 38–42.
- Flegontov, P., N. E. Altinisik, P. Changmai et al. 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570:236–240.
- Flegontov, P., P. Changmai, A. Zidkova et al. 2016. Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci.*

Rep. 6:20768.

- Fu, Q., C. Posth, M. Hajdinjak et al. 2016. The genetic history of Ice Age Europe. *Nature* 534:200–205.
- Gneccchi-Ruscone, G. A., E. Khussainova, N. Kahbatkyzy et al. 2021. Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci. Adv.* 7:eabe4414.
- Hammer, M. F., T. M. Karafet, H. Park et al. 2006. Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* 51:47–58.
- Hong, S. B., H. J. Jin, K. D. Kwak et al. 2006. Y-chromosome haplogroup O3-M122 variation in East Asia and its implications for the peopling of Korea. *Korean J. Genet.* 28:1–8.
- Howorth, H. H. 1880. *History of the Mongols, from the 9th to the 19th Century*. London, UK: Longmans, Green, and Company.
- Huang, Y. Z., L. H. Wei, S. Yan et al. 2018. Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407 *Mol. Genet. Genomics* 293:657–663.
- Illumäe, A. M., M. Reidla, M. Chukhryaeva et al. 2016. Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *Am. J. Hum. Genet.* 99:163–173.
- Jeong, C., O. Balanovsky, E. Lukianova et al. 2019. The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* 3:966–976.
- Jeong, C., K. Wang, S. Wilkin et al. 2020. A dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *Cell* 183:890–904.e29.
- Jeong, C., S. Wilkin, T. Amgalantugs et al. 2018. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl. Acad. Sci. U. S. A.* 115:E11248–E11255.
- Jobling, M. A., and C. Tyler-Smith. 2003. The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* 4:598–612.
- Karafet, T. M., F. L. Mendez, M. B. Meilerman et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.
- Karafet, T. M., L. P. Osipova, M. A. Gubina et al. 2002. High levels of Y-chromosome differentiation among Native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.* 74:761–789.

- Karafet, T. M., L. Xu, R. Du et al. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 69:615–628.
- Karmin, M., L. Saag, M. Vicente et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25:459–466.
- Katoh, T., B. Munkhbat, K. Tounai et al. 2005. Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene* 346:63–70.
- Keyser, C., C. Hollard, A. Gonzalez et al. 2015. The ancient Yakuts: A population genetic enigma. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20130385.
- Khar'kov, V. N., V. A. Stepanov, O. F. Medvedev et al. 2008. The origin of Yakuts: Analysis of Y-chromosome haplotypes. *Mol. Biol. (Mosk.)*. 42:226–237.
- Kilinc, G. M., N. Kashuba, D. Koptekin et al. 2021. Human population dynamics and *Yersinia pestis* in ancient Northeast Asia. *Sci. Adv.* 7:eabc4587.
- Lamnidis, T. C., K. Majander, C. Jeong et al. 2018. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* 9:5018.
- Lell, J. T., R. I. Sukernik, Y. B. Starikovskaya et al. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* 70:192–206.
- Li, J., Y. Zhang, Y. Zhao et al. 2018. The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *Am. J. Phys. Anthropol.* 166:895–905.
- Lin, G. 2007. *A History of Donghu (in Chinese)*. Hohhot, CN: Inner Mongolian People's Publishing House.
- Liu, B.-L., P.-C. Ma, C.-Z. Wang et al. 2020. Paternal origin of Tungusic-speaking populations: Insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86. *Am. J. Hum. Biol.* 33:e23462.
- Liu, Li, N. A. Duncan, X. Chen et al. 2016. Changing patterns of plant-based food production during the Neolithic and early Bronze Age in Central-South Inner Mongolia, China: An interdisciplinary approach. *Quat. Int.* 419:36–53.
- Ma, P., X. Yang, S. Yan et al. 2021. Ancient Y-DNA with reconstructed phylogeny provides insights into the demographic history of paternal haplogroup N1a2-F1360. *J. Genet. Genomics* 48:1130–1133.
- Malyarchuk, B. A., M. Derenko, G. Denisova et al. 2013. Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels. *J. Hum. Genet.* 58:804–811.
- Malyarchuk, B. A., M. Derenko, G. Denisova et al. 2011. Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J.*

Hum. Genet. 56:583–588.

- Malyarchuk, B. A., M. Derenko, G. Denisova et al. 2010. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Ann. Hum. Genet.* 74:539–546.
- Malyarchuk, B. A., M. Derenko, G. Denisova et al. 2016. Y chromosome haplotype diversity in Mongolic-speaking populations and gene conversion at the duplicated STR DYS385a,b in haplogroup C3-M407. *J. Hum. Genet.* 61:491–496.
- Mao, X., H. Zhang, S. Qiao et al. 2021. The deep population history of Northern East Asia from the Late Pleistocene to the Holocene. *Cell* 184:3256–3266.e13.
- Narasimhan, V. M., N. Patterson, P. Moorjani et al. 2019. The formation of human populations in South and Central Asia. *Science* 365:eaat7487.
- Ning, C., T. Li, K. Wang et al. 2020. Ancient genomes from Northern China suggest links between subsistence changes and human migration. *Nat. Commun.* 11:2700.
- Onon, U. 2001. *The Secret History of the Mongols: The Life and Times of Chinggis Khan* London, UK and New York, NY: Routledge.
- Pakendorf, B., I. N. Novgorodov, V. L. Osakovskij et al. 2006. Investigating the effects of prehistoric migrations in Siberia: Genetic variation and the origins of Yakuts. *Hum. Genet.* 120:334–353.
- Qi, X., C. Cui, Y. Peng et al. 2013. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol. Biol. Evol.* 30:1761–1778.
- Raghavan, M., P. Skoglund, K. E. Graf et al. 2014. Upper Paleolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505:87–91.
- Robbeets, M., R. Bouckaert, M. Conte et al. 2021. Triangulation supports agricultural spread of the Transeurasian languages. *Nature* 599:616–621.
- Rootsi, S., C. Magri, T. Kivisild et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* 75:128–137.
- Rootsi, S., N. M. Myres, A. A. Lin et al. 2012. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J. Hum. Genet.* 20:1275–1282.
- Rootsi, S., L. A. Zhivotovskiy, M. Baldovic et al. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* 15:204–211.
- Seguin-Orlando, A., T. S. Korneliussen, M. Sikora et al. 2014. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346:1113–1118.

- Semino, O., C. Magri, G. Benuzzi et al. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* 74:1023–1034.
- Sengupta, S., L. A. Zhivotovsky, R. King et al. 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. *Am. J. Hum. Genet.* 78:202–221.
- Shi, H., H. Zhong, Y. Peng et al. 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6:45.
- Sikora, M., V. V. Pitulko, V. C. Sousa et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570:182–188.
- Su, B., J. Xiao, P. Underhill et al. 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* 65:1718–1724.
- Sun, J., P. C. Ma, H. Z. Cheng et al. 2020. Post-last glacial maximum expansion of Y-chromosome haplogroup C2a-L1373 in Northern Asia and its implications for the origin of Native Americans. *Am. J. Phys. Anthropol.* 174:363–374.
- Tanaka, M., V. M. Cabrera, A. M. Gonzalez et al. 2004. Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Res.* 14:1832–1850.
- Underhill, P. A. 2003. Inferring human history: Clues from Y-chromosome haplotypes. *Cold Spring Harb. Symp. Quant. Biol.* 68:487–493.
- Underhill, P. A., L. Jin, A. A. Lin et al. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 7:996–1005.
- Underhill, P. A., G. D. Poznik, S. Rootsi et al. 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* 23:124–131.
- Underhill, P. A., P. Shen, A. A. Lin et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26:358–361.
- Wallace, V. A. 2015. *Buddhism in Mongolian History, Culture, and Society*. Oxford, UK: Oxford University Press.
- Wang, C. C., H. Y. Yeh, A. N. Popov et al. 2021. Genomic insights into the formation of human populations in East Asia. *Nature* 591:413–419.
- Wang, C. Z., L. H. Wei, L. X. Wang et al. 2019. Relating Clans Ao and Aisin Gioro from

- Northeast China by whole Y-chromosome sequencing. *J. Hum. Genet.* 64:775–780.
- Wang, H., L. Chen, B. Ge et al. 2012. Genetic data suggests that the Jinggouzi people are associated with the Donghu, an ancient nomadic group of North China. *Hum. Biol.* 84:365–378.
- Wei, L. H., Y. Z. Huang, S. Yan et al. 2017. Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. *J. Hum. Genet.* 62:915–918.
- Wei, L. H., L. X. Wang, S. Q. Wen et al. 2018. Paternal origin of Paleo-Indians in Siberia: Insights from Y-chromosome sequences. *Eur. J. Hum. Genet.* 26:1687–1696.
- Wei, L. H., S. Yan, Y. Lu et al. 2018. Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan. *Eur. J. Hum. Genet.* 26:230–237.
- Wells, R. S., N. Yuldasheva, R. Ruzibakiev et al. 2001. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. U. S. A.* 98:10244–10249.
- Wu, Q., H. Z. Cheng, N. Sun et al. 2020. Phylogenetic analysis of the Y-chromosome haplogroup C2b-F1067, a dominant paternal lineage in Eastern Eurasia. *J. Hum. Genet.* 65:823–829.
- Xue, Y., T. Zerjal, W. Bao et al. 2006. Male demography in East Asia: A north-south contrast in human population expansion times. *Genetics* 172:2431–2439.
- Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.
- Yang, M. A., X. Fan, B. Sun et al. 2020. Ancient DNA indicates human population shifts and admixture in Northern and Southern China. *Science* 369:282–288.
- Yu, H., M. A. Spyrou, M. Karapetian et al. 2020. Paleolithic to Bronze Age Siberians reveal connections with First Americans and across Eurasia. *Cell* 181:1232–1245.
- Zerjal, T., R. S. Wells, N. Yuldasheva et al. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* 71:466–482.
- Zerjal, T., Y. Xue, G. Bertorelle et al. 2003. The genetic legacy of the Mongols. *Am. J. Hum. Genet.* 72:717–721.
- Zhabagin, M. K., Z. Sabitov, P. Tarlykov et al. 2020. The medieval Mongolian roots of Y-chromosomal lineages from South Kazakhstan. *BMC Genet.* 21:87.
- Zhabagin, M. K., Z. Sabitov, I. Tazhigulova et al. 2021. Medieval super-grandfather founder of Western Kazakh Clans from haplogroup C2a1a2-M48. *J. Hum. Genet.* 66:707–716.

- Zhang, J.-H. 1998. *History of Proto-Mongols: Research on Shiwei-Tatar Tribes (in Chinese)*. Beijing, CN: Higher Education Press.
- Zhang, Y., X. Wu, J. Li et al. 2018. The Y-chromosome haplogroup C3*-F3918, likely attributed to the Mongol Empire, can be traced to a 2500-year-old nomadic group. *J. Hum. Genet.* 63:231–238.
- Zhong, H., H. Shi, X. B. Qi et al. 2011. Extended Y chromosome investigation suggests post-glacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28:717–727.
- Zhong, H., H. Shi, X. B. Qi et al. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J. Hum. Genet.* 55:428–435.