

BRIEF COMMUNICATION



Dual origins of the Northwest Chinese Kyrgyz: the admixture of Bronze age Siberian and Medieval Niru'un Mongolian Y chromosomes

Shao-qing Wen^{1,2,5}, Pan-xin Du^{2,5}, Chang Sun², Wei Cui³, Yi-ran Xu², Hai-liang Meng², Mei-sen Shi⁴✉, Bo-feng Zhu³✉ and Hui Li²✉

© The Author(s), under exclusive licence to The Japan Society of Human Genetics 2021

The Kyrgyz are a trans-border ethnic group, mainly living in Kyrgyzstan. Previous genetic investigations of Central Asian populations have repeatedly investigated the Central Asian Kyrgyz. However, from the standpoint of human evolution and genetic diversity, Northwest Chinese Kyrgyz is one of the more poorly studied populations. In this study, we analyzed the non-recombining portion of the Y-chromosome from 298 male Kyrgyz samples from Xinjiang Uygur Autonomous Region in northwestern China, using a high-resolution analysis of 108 biallelic markers and 17 or 24 STRs. First, via a Y-SNP-based PCA plot, Northwest Chinese Kyrgyz tended to cluster with other Kyrgyz population and are located in the West Asian and Central Asian group. Second, we found that the Northwest Chinese Kyrgyz display a high proportion of Y-lineage R1a1a1b2a2a-Z2125, related to Bronze Age Siberian, and followed by Y-lineage C2b1a3a1-F3796, related to Medieval Niru'un Mongols, such as Uissun tribe from Kazakhs. In these two dominant lineages, two unique recent descent clusters have been detected via NETWORK analysis, respectively, but they have nearly the same TMRCA ages (about 13th–14th centuries). This finding once again shows that the expansions of Mongol Empire had a striking effect on the Central Asian gene pool.

Journal of Human Genetics; <https://doi.org/10.1038/s10038-021-00979-x>

INTRODUCTION

The activities of the Eurasian Steppe Nomads, and the culture they created, have an eternal and mysterious fascination for the world, since their history is often recorded sporadically in the historical records. The Kyrgyz (or Kirghiz), a cross-border ethnic group, are mainly distributed in Kyrgyzstan, which accounts for 69.2% of its total population [1]. Others are scattered in China, Uzbekistan, Tajikistan, Kazakhstan, and Afghanistan. According to the sixth population survey of China in 2010, the Kyrgyz population totals 0.18–0.19 million (<http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm>). They live primarily in the Kizilsu Kirghiz Autonomous Prefecture in the southwestern Xinjiang Uygur Autonomous Region with some living in the Heilongjiang Province. The economic life of the Kyrgyz is primarily based on pastoralism.

Historically, ancient Kyrgyz were considered to be the Yenisei Kyrgyz that may perhaps be concerned with the Tashtyk culture [2]. They lived on the upper reaches of Yenisei River in the south of the Minusinsk Basin and dispersed among many stock-raising peoples of the Sayano-Altai from the 6th to 13th century [3, 4]. They appeared as Gekun (鬲昆) in the Chinese annals *Shiji* (*Records of the Grand Historian*, 91 B.C.), as Jiankun (坚昆) in *Hanshu* (*Book of Han*, 92 A.D.), as Hegu (纥骨) in *Beishi* (*History of Northern Dynasties*, 659 A.D.), as Qigu (契骨) in *Zhoushu* (*Book of Zhou*, 636

A.D.) and *Suishu* (*Book of Sui*, 636 A.D.), as Jiegu (结骨) in *Jiu Tangshu* (*Old Book of Tang*, 945 A.D.) and *Tang Huiyao* (*Institutional History of Tang*, 618–907 A.D.), as Xiajiashi (黠戛斯) in *Xin Tangshu* (*New Book of Tang*, 1060 A.D.), *Liaoshi* (*History of Liao*, 1344 A.D.), and *Songshi* (*History of Song*, 1345 A.D.), as Heliqisi (纥里迄斯) in *Jinshi* (*History of Jin*, 1344 A.D.), as Jilijisi (吉利吉思) in *Yuanshi* (*History of Yuan*, 1370 A.D.) and *Mingshi* (*History of Ming*, 1739 A.D.), and as Bulute (布鲁特) in *Qing Shigao* (*Draft History of Qing*, 1927 A.D.). Notably, according to the records of *Xin Tangshu*, the majority of the Jiegu people are “all tall and big and have red hair, white faces, and green eyes” and the ones with “dark hair and eyes” claimed to be the descendants of Li Ling, grandson of the famous general Li Guang during the Han Dynasty.

Linguistically, the present-day Kyrgyz language belongs to the western branch of Turkic group within the Altaic linguistic family (www.ethnologue.com). However, the Kyrgyz Fuyu language, a dialect found in Fuyun County (Heilongjiang Province, China), has common roots with the modern Khakas and Shor languages of South Siberia. These ancient Siberian populations belong to the northern branch of Turkic language group. In this context, language replacement may have occurred in the Kyrgyz language, during their migration from South Siberian to the region where they currently reside [5].

¹Institute of Archaeological Science, Fudan University, Shanghai 200433, China. ²B&R International Joint Laboratory for Eurasian Anthropology, MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China. ³Multi-Omics Innovative Research Center of Forensic Identification; Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China. ⁴Institute of the Investigation, School of Criminal Justice, China University of Political Science and Law, Beijing 100088, China. ⁵These authors contributed equally: Shao-qing Wen, Pan-xin Du. ✉email: shimeisen2000@163.com; zhubofeng7372@126.com; LHCA@fudan.edu.cn

Received: 1 July 2021 Revised: 31 August 2021 Accepted: 13 September 2021

Published online: 17 September 2021

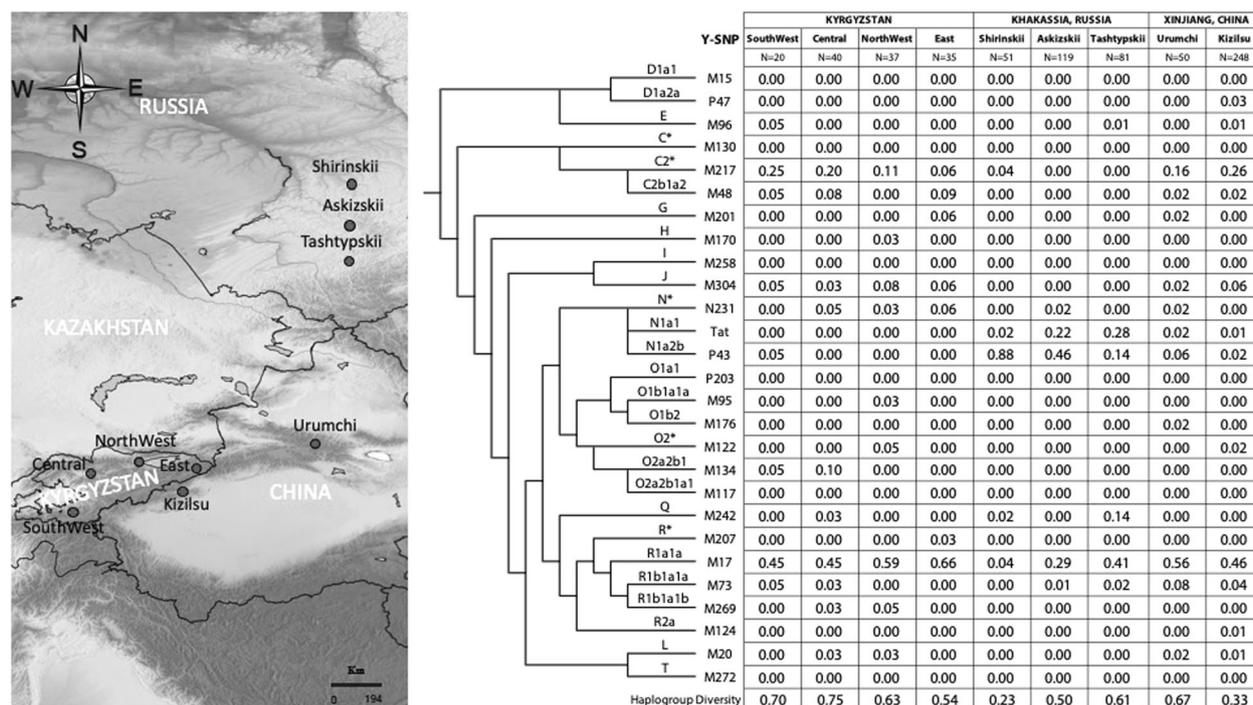


Fig. 1 Map with the sampling points of Northwest Chinese Kyrgyz and other studied Kyrgyz populations. The distributions of the Kyrgyz populations are shown in red. The phylogenetic relationship of Y-chromosome haplogroups surveyed in this study and their frequencies are exhibited on the right side of the figure. The haplogroup names are shown along the branches, and the marker names are shown according to ISOGG Y-DNA Haplogroup Tree 2018. Potentially paraphyletic undefined subgroups are distinguished from recognized haplogroups by the asterisk symbol

Genetic analysis of the Kyrgyz populations initially focused on mitochondrial DNA and limited Y chromosomal markers to assess levels of diversity and the relationships between them and other Eurasian populations. The frequencies of Eastern Asian-specific and European-specific mtDNA haplogroups were observed at 72.9% and 25.0% in Talas Kyrgyz (lowland Kyrgyz) and 59.6% and 31.9% in Sary-Tash Kyrgyz (highland Kyrgyz), respectively [6–9]. Among them, Eastern Asian-specific Haplogroup D and European-specific Haplogroup HV both present the highest frequency (25.5%) in the Kyrgyz [8]. Therefore, the Kyrgyz are an admixed population between the East and the West. Different patterns have been observed in the patrilineal gene pool of the Kyrgyz. Extremely low Y-diversity and the presence of a high-frequency (63% [10], 54.5% [11], or 68.9% [12]) Y-chromosome haplogroup R1a1-M17 (a diagnostic Indo-Iranian marker [10]) are striking features of Kyrgyz populations in central Asia. It appears that a serial of founder effects or strong bottlenecks have occurred in Kyrgyz populations. Moreover, recent genome-wide SNP study on Central Asian Kyrgyz by Petr Triska et al. [13] suggested that high levels of shared IBD blocks in Central Asian Kyrgyz and other Altaic-speaking populations from Southern Siberia (Tuva, Buryat) and North Asia (Yakut), support their recently formed common genetic core in Southern Siberia.

However, much of the previous genetic research into this issue has focused mainly on Siberian and Central Asian Kyrgyz populations, leaving our understanding of the genetic diversity of the Kyrgyz in China incomplete. In this study, we collected and tested 298 male Kyrgyz samples from Xinjiang Uygur Autonomous Region in northwestern China. This paper attempts to investigate Y-chromosomal variation of Northwest Chinese Kyrgyz and its genetic relationships with other Kyrgyz populations and understand the population history of Northwest Chinese Kyrgyz by placing them into the genetic context of the Eurasian Steppes.

MATERIALS AND METHODS

With the informed consent from the Ethics Committee of Fudan University of Life Sciences, 50 saliva samples and 248 bloodstain samples were collected from healthy Kyrgyz males in Urumchi City and Kizilsu Kyrgyz Autonomous Prefecture, respectively (Fig. 1). Care was taken to avoid related individuals. Genomic DNA was extracted through the QIAamp DNA Blood Mini Kit (QIAGEN, Germany). In all DNA samples, 108 Y-SNP markers were hierarchically genotyped by SNaPshot (ABI SNaPshot® Multiplex Kit) as described in previous studies [14, 15]. Primers were based on ones previously published with additional primers based on published sequences [16]. Urumchi Kyrgyz samples were typed with 17 microsatellites by using AmpFISTR® Yfiler™ PCR Amplification kit (Applied Biosystems, Carlsbad, CA) and Kizilsu Kyrgyz samples were typed with 24 microsatellites by using AGCU Y24 STR amplification system (AGCU ScienTech Incorporation, Wuxi, China). AGCU Y24 STR amplification system includes Y-filer 17 STRs and 7 additional Y-STR loci, five single-copy loci (DYS449, DYS447, DYS522, DYS388, and DYS444) and a multicopy locus (DYS527a, b). Data were submitted to the YHRD (Y-chromosomal haplotype reference database) and received the accession number YA004490.

Haplotype diversity and haplogroup diversity were calculated using the Arlequin package. 7855 samples from 116 Eurasian populations, especially central Asian/Altaic-speaking populations, were selected for population comparisons at a similar level of haplogroup resolution (Supplementary Tables 2 and 3). Principal Component Analysis (PCA) was constructed using R software version 3.6.3. Median-joining networks were calculated by use of the program NETWORK v. 5.0.0.1 (<http://www.fluxus-engineering.com>) for the common haplogroups in the Northwest Chinese Kyrgyz. For use in these analyses, DYS389II was calculated by subtracting the DYS389I allele size. The time to the most recent

common ancestor (TMRCA) of each cluster detected in common haplogroups was determined by using the average squared distance (ASD) estimator as described previously [14, 15, 17]. ASD method is based on the assumption that median or modal STR haplotype in a lineage is the founder haplotype. A generation time of 25 years was used to produce time estimates in years [18].

We called Y-chromosomal variants in published ancient samples using bcftools (<http://www.htslib.org/doc/bcftools.html>) [19] mpileup and bcftools call emitting all sites within mappable Y-chromosomal regions [20]. Haplogroup determination was done with the script Yleaf.py with Yleaf [21], which outputs the allele counts for ancestral and derived SNPs along a path of branches of the Y-chromosome tree. At last, we re-checked the SNPs by visual inspection with IGV software [22].

RESULTS

We typed the 298 males from Northwest Chinese Kyrgyz with 108 binary markers, to identify the Y haplogroups according to the nomenclature of Y Chromosome Consortium (YCC) and ISOGG Y-DNA Haplogroup Tree 2018, and with 17 or 24 microsatellites, to define more detailed haplotypes (Supplementary Table 1). The haplotype diversity of Northwest Chinese Kyrgyz was 0.96 when using 17 microsatellites, among which Urumchi Kyrgyz and Kizilsu Kyrgyz have haplotype diversity as 0.98 and 0.96, respectively. Expectedly higher haplotype diversity (0.98) would be observed in Kizilsu Kyrgyz when testing 24 microsatellites. In Northwest Chinese Kyrgyz, Urumchi Kyrgyz had a moderate haplogroup diversity (0.67), whereas Kizilsu Kyrgyz have a haplogroup diversity with extremely low value (0.33), only little higher than Shirinskii Khakass (0.23; another name of Kyrgyz given by Russian ethnologists) among all studied Kyrgyz populations, including Kyrgyz populations from southwestern, central, northwestern, and eastern Kyrgyzstan [11] and Khakass populations from Askizskii, Tashtypskii, and Shirinskii district of the Republic of Khakassia [3] (Fig. 1).

Haplogroup R1a1a*-M17+, M198+, M458- and C2*-M217+, M48- together account for 72% of the studied Y chromosomes (Fig. 1). As reported in the other Kyrgyz populations, the highest frequencies of Haplogroup R1a1a*-M17+, M198+, M458- were also present in the Northwest Chinese Kyrgyz (Urumchi Kyrgyz: 56% and Kizilsu Kyrgyz: 46%). This haplogroup generally was frequent in a wide geographic area extending from South Asia to Central East Europe and South Siberia [16]. Zerjal et al. [10] postulated that it could be the most evident male genetic legacy of the "Kurgan Culture" population expansion. We further tested the diagnostic markers R1a1a1b1a-Z282 and R1a1a1b2-Z93 [16]. Of the 143 Kyrgyz R1a1a*-M17+, M198+, M458- samples, more than 90% were assigned to Central Asian lineage R1a1a1b2-Z93 whereas the rest belonged to European lineage R1a1a1b1a-Z282.

Haplogroup C2*-M217+, M48- appeared with relatively high frequency (Urumchi Kyrgyz: 16% and Kizilsu Kyrgyz: 26%) in Northwest Chinese Kyrgyz. However, this haplogroup is frequently observed among Kyrgyz populations from Kyrgyzstan (6–25%), but rather rare in Khakass. Four paternal lineages in haplogroup C2*-M217 are common in Altaic-speaking populations, especially most Mongolic-speaking populations, that is haplogroup C2b1a3a1-F3796 [23], C2a1a1b-M48 [24], C2b1a1a1a-M407 [25], and C2a1a1a2a-F1756 [26]. In Northwest Chinese Kyrgyz, haplogroup C2b1a3a1-F3796 accounts for about 23% of the studied Y chromosomes, whereas haplogroup C2a1a1b-M48 and C2a1a1a2a-F1756 are sporadic (about 2 and 1%, respectively) and haplogroup C2b1a1a1a-M407 is absent.

Western Eurasian lineages, haplogroup J-M304 and haplogroup R1b1a1a-M73, have the low or moderate frequencies among Urumchi Kyrgyz (2 and 8%) and Kizilsu Kyrgyz (6 and 4%). North Asian lineage N1a2b-P43 is observed in Urumchi Kyrgyz (6%) and in Kizilsu Kyrgyz (2%), whereas it occurs at a high frequency in

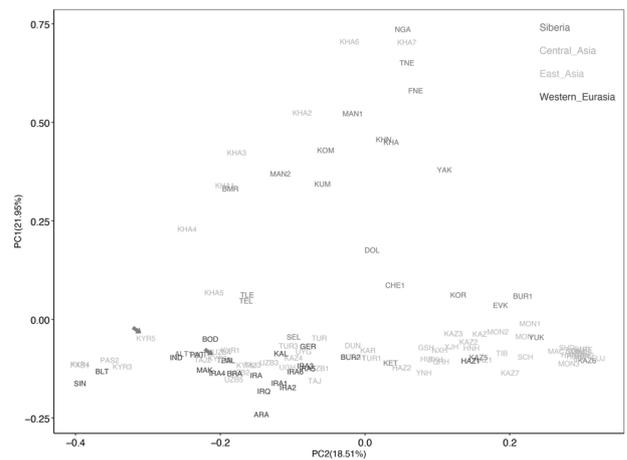


Fig. 2 Y-SNP-based PCA plot of Northwest Chinese Kyrgyz samples with samples from other Eurasian populations. Detailed information and codes of the studied populations are described in Supplementary Table 2

Khakass, especially Shirinskii Khakass (88%). The rest detected haplogroups in Northwest Chinese Kyrgyz were found at very low frequencies (<5%, Fig. 1) and can be classified into eastern Eurasian lineages (C*-M130, C2b-F1067, D1a1-M15, D1a2a-P47, N1a1a-M178, N1b-F2930, O1a1-P203, O1b2-M176, O2a1b-002611, O2a2b1a1-M117, and O2a2b1a2-F444) and western Eurasian lineages (E-M96, G-M201, H-M69, I-M170, L-M20, R1b1a1b-M269, and R2a-M124) based on their putative origin. Notably, haplogroup Q-M242 has a Central Asian and Southern Siberian origin and reached a high level in North Asian population [27], such as Tashtypskii Khakass (14%), but it was not seen in our dataset of Northwest Chinese Kyrgyz.

A Y-SNP-based PCA plot (Fig. 2) at population level compares our Kyrgyz data with 116 Eurasian populations covering Central Asia, East Asia, Siberia and West Asia (Supplementary Table 2). The PCA plot appears well structured, and it seems to mirror the historical and geographical background of the studied population groups (Fig. 2). Three separated groups radiate in three directions away from the center of the plot. West Asian and Central Asian populations, except Khakass, cluster together in the lower left direction, East Asian populations cluster tightly with each other in the lower right direction, and the Khakass have close genetic affinities with the Siberian populations in the upper direction. The northwest Chinese Kyrgyz were close to other Kyrgyz populations and located in the West Asian and Central Asian group, thus showing a general similarity in geographically neighboring populations.

To reveal the detail relationship at individual level between Northwest Chinese Kyrgyz and other reference Eurasian populations, we further investigated the two dominant lineages, R1a1a1b2-Z93 and C2b1a3a1-F3796, by analyzing the reduced median networks (Figs. 3, 4, and Supplementary Table 3). In the R1a1a1b2-Z93 network, most Northwest Chinese Kyrgyz samples (red) tend to cluster with haplogroup R1a1a1b2a2a-Z2125 samples (magenta) and they therefore should be assigned to the identical lineages (Fig. 3). Among the Asian R1a1a1b2-Z93 lineages, R1a1a1b2a2a-Z2125 is quite common in Kyrgyzstan and Afghan Pashtuns (>40%), and less frequent in other Afghan ethnic groups and some Caucasus and Iran populations (>10%). Notably, the basal lineage R1a1a1b2-Z93* is commonly distributed in the South Siberian Altai region of Russia. In the upper-right corner of Fig. 3, a marked recent descent cluster (we used the criterion that haplotypes linked to the modal haplotype fewer than 5 mutational steps in the shaded area [17]) can be easily observed. Except two R1a1a1b2-Z93* samples (a Khakassian and an Altaian, dark blue),

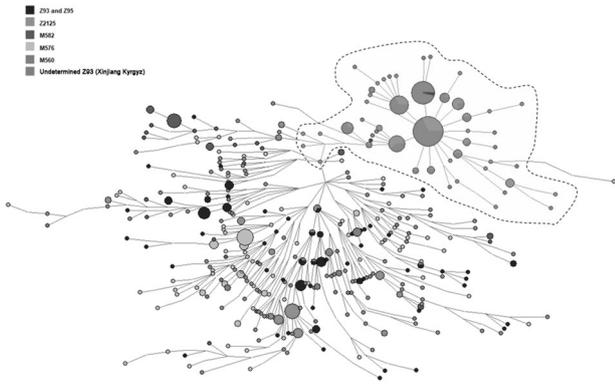


Fig. 3 Reduced median joining network of haplogroup R1a1a1b2-Z93. Haplotypes are represented by circles with area proportional to the number of individual. The lines between circles represent mutational distance, the shortest distance being a single mutational step. Colors indicate different sub-haplogroups: mazarin for the basal Z93 and Z95 samples, carmine for Z2125 samples, loden for M582 samples, light green for M576 samples, gray for M560 samples, and red for undetermined Z93 Xinjiang Kyrgyz samples

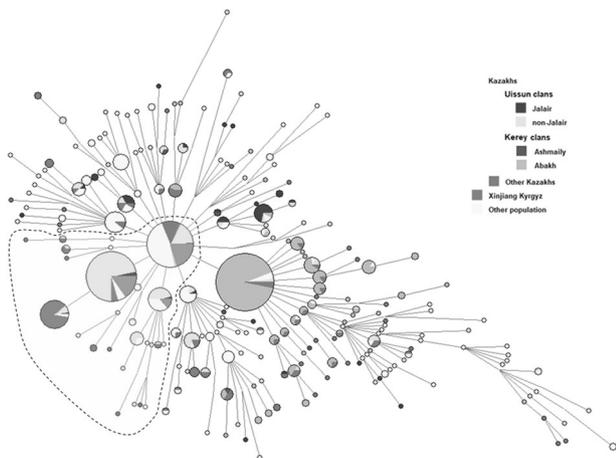


Fig. 4 Reduced Median joining network of haplogroup C2b1a3a1-F3796. Haplotypes are represented by circles with area proportional to the number of individual. The lines between circles represent mutational distance, the shortest distance being a single mutational step. Colors indicate different groups of populations: blue for Uissun clans from Kazakhs (dark blue for Jalair Uissun clan samples and brilliant blue for non-Jalair Uissun clan samples), green for Kerey clans from Kazakhs (loden for Ashmally Kerey clan samples, light green for Abakh Kerey clan samples), gray for other clans from Kazakhs, red for Xinjiang Kyrgyz samples, and yellow for the remaining populations

Northwest Chinese Kyrgyz samples specifically shared the haplotypes with Kyrgyz from Central Asia, forming a “Kyrgyz” cluster. Based on Y-STR profiles, the TMRCA (time to the most recent common ancestor) of all haplogroup R1a1a1b2a2a-Z2125 samples is estimated to be 2585–3107 YA and that of the Kyrgyz cluster is 628–721 YA.

Haplogroup C2b1a3a1-F3796 (dating 1296–1624 YA) has a classic star-like pattern with a central founder haplotype and a few subfounders (Fig. 4). In the center of Fig. 4, the central haplotype is mainly occupied by the Kazakhs from Central and North Asian (60%) and some Mongolian-speaking populations (30%), such as Buryats, Daur, Kalmyk, Khalkh, Oirats, and other Mongols. In the network, previous reported clusters are clearly visible, such as kerey cluster, Jalair cluster, and a Hazara cluster, a non-Jalair

Uissun cluster [28]. Significantly, Northwest Chinese Kyrgyz samples (red) tend to cluster with non-Jalair Uissun clan samples (brilliant blue). This cluster (dating 696–736 YA) consisted of the majority of the Northwest Chinese Kyrgyz C2b1a3a1-F3796 samples (61%), and the representatives of the Uissun clans, such as Dulat ($n = 56$), Alban ($n = 21$), Suan ($n = 15$), Shaprashty ($n = 6$), Shanyshkly ($n = 3$), SaryUissun ($n = 2$), and Oshakty ($n = 1$).

DISCUSSION

Three competing hypotheses have been debated regarding the origins of the Kyrgyz: an upper Yenisei River (Minusinsk basin) origin, advocated by Russian academician G.F. Miller (1705–1783) in his *History of Siberia*, a Tenir Too Mountains origin, which holds that the Kyrgyz were the indigenous Central Asian, and a multiple independent origin, suggesting that the Kyrgyz consisted of at least two ethnic groups of Central Asian [29]. In this study, a set of 108 Y chromosome SNPs and 17 or 24 Y chromosome microsatellites was employed to trace the genetic components of the Northwest Chinese Kyrgyz paternal gene pool. We found that the Northwest Chinese Kyrgyz were characterized by the presence of two major Y chromosome haplogroups (R1a1a1b2a2a-Z2125 and C2b1a3a1-F3796).

Haplogroup R1a1a1b2a2a-Z2125 was the most common lineage (39%) in Northwest Chinese Kyrgyz. It was frequent in several central Asian populations. The oldest specimen (SVP27, Utyevka VI, kurgan 7, grave 1 [2200–1900 BCE]) of this lineage originated from the Early Bronze Age Potapovka culture (closely related to the Sintashta culture) site from south of the Sok River in the Samara oblast, Russia [30]. According to the published ancient DNA data, we found that, in Middle Bronze Age, Haplogroup R1a1a1b2a2a-Z2125 was mainly found in Sintashta culture population from Kamennyi Ambar 5 cemetery, western Siberia, in Fedorovo type of the Andronovo culture or Karasuk culture population from Minusinsk Basin, southern Siberia, and in Andronovo culture populations from Maitan, Ak-Moustafa, Aktogai, Kazakh Mys, Satan, Oy-Dzhaylau III, Karagash 2, Dali, and Zevakinskiy stone fence, Kazakhstan [30]. Subsequently, from Late Bronze Age to Medieval Age, this lineage was observed among Mongolia [31–33], Kyrgyzstan [34], Uzbekistan [30], Tajikistan [30], Hungary [35, 36] and Moldova [37], except for Russia [30, 34, 37–39] and Kazakhstan [30, 34, 38, 40, 41]. Notably, in Tianshan area of Kyrgyzstan, the lineage was seen in Scythian_Saka and Hun people at 259–93 BCE and 286–406 CE, respectively [34]. Therefore, on basis of the spatial and temporal distribution of haplogroup R1a1a1b2a2a-Z2125, some Kyrgyz might have a Siberian origin.

The second major haplogroup observed in Northwest Chinese Kyrgyz is C2b1a3a1-F3796. This lineage has the highest diversity in Mongols and the highest frequency in Kazakhs [28]. Previous studies have shown that the origin of this lineage is related to the Medieval Niru’un Mongols [23, 28]. At present, only two ancient specimens unquestionably belonging to the haplogroup C2b1a3a1-F3796 have been found [31]. They (MRI001 and SHG003) were from typical Late Medieval (ca. 1200–1400 CE) Mongol burials from Mori Baridag, Khuvsugul and Sharga Uul, Sukhbaatar, respectively, in Mongolia. Notably, except I1193 (from Primorsky Krai, Boisman-2 [4989–4794 calBCE] [33]), most ancient samples (SHA001, TSA003, and DAS001 [32]) belonged to its upstream SNP, haplogroup C2b1a3a1-F1918, also originated from medieval Mongolia. Therefore, compared to haplogroup R1a1a1b2a2a-Z2125, this lineage should have been integrated into Kyrgyz from Niru’un Mongols at a later stage. The pattern of dual origins of the Northwest Chinese Kyrgyz (the admixture of Bronze Age Siberian and Medieval Mongolian) has also been observed in central Asian Kyrgyz by using genome-wide IBD sharing analysis [13]. Of note, within haplogroup C2b1a3a1-F3796, Northwest Chinese Kyrgyz and non-Jalair Uissun clan samples have a close paternal genetic relationship, forming a descent

cluster (dating 696–736 YA). Within haplogroup R1a1a1b2a2a-Z2125, TMRCA (time to the most recent common ancestor) of the “Kyrgyz” cluster is estimated to be 628–721YA. The nearby coincidence of ages of these two clusters indicates the rapid population growth of Northwest Chinese Kyrgyz in 13th–14th centuries, which coincides with the expansion period of the Mongol Empire.

Based on the above dual origins model, Kyrgyz language changed from the northern branch to the western branch of the Turkic group, which can be readily understood. Like modern Khakas and Shor, the ancient Kyrgyz language might belong to northern branch of Turkic group. However, during the Late Medieval period, the replacement of ancient Kyrgyz language might be accompanied by the western branch Turkic-speaking populations, such as Uissun tribe from Kazakhs.

REFERENCES

- Heliguli Niyazi. *National conditions of the five Central Asian states*. Xi 'an: Xiaotong University Press, 110–1 (2016).
- Maenchen-Helfen, O. Manichaeans in Siberia. In: Fischel, WJ (eds) *Semitic and Oriental Studies: A Volume Presented to William Popper, Professor of Semitic Languages, Emeritus, on the Occasion of His Seventy-Fifth Birthday, October 29, 1949*. Auckland: University of California Press, 311–26 (1951).
- Khar'kov VN, Khamina KV, Medvedeva OF, Shtygasheva OV, Stepanov VA. Genetic diversity of Khakassian gene pool: subethnic differentiation and the structure of Y-chromosome haplogroups. *Mol Biol (Mosk)*. 2011;45:446–58.
- Lavryashina MB, Ul'yanova MV, Balaganskaya OA, Balanovska EV. Genetic structure of the Khakass subethnic groups from autosomal DNA markers and surnames. *Sci Evolut*. 2016;1:78–84.
- Wei, LH. Genetic Perspective on Language Replacement in Siberia. In: Xu, D, and Li, H (eds) *Languages and Genes in Northwestern China and Adjacent Regions*. Singapore: Springer, 107–20 (2017).
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, et al. Trading genes along the Silk Road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet*. 1998;63:1824–38.
- Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol Biol Evol*. 2004;21:2265–80.
- Yang L, Tan S, Yu H, Zheng B, Qiao E, Dong Y, et al. Gene admixture in ethnic populations in upper part of Silk Road revealed by mtDNA polymorphism. *Sci China C Life Sci*. 2008;51:435–44.
- Peng MS, Xu W, Song JJ, Chen X, Sulaiman X, Cai L, et al. Mitochondrial genomes uncover the maternal history of the Pamir populations. *Eur J Hum Genet*. 2018;26:124–36.
- Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet*. 2002;71:466–82.
- Di CJ, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA, et al. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PLoS One*. 2013;8:e76748.
- Shou WH, Qiao EF, Wei CY, Dong YL, Tan SJ, Shi H, et al. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. *J Hum Genet*. 2010;55:314–22.
- Triska P, Chekanov N, Stepanov V, Khusnutdinova EK, Kumar GPA, Akhmetova V, et al. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. *BMC Genet*. 2017;18(Suppl 1):110.
- Wang CC, Wang LX, Shrestha R, Zhang M, Huang XY, Hu K, et al. Genetic structure of Qiangic populations residing in the western Sichuan corridor. *PLoS One*. 2014;9:e103772.
- Wen SQ, Sun C, Song DL, Huang YZ, Tong XZ, Meng HL, et al. Y-chromosome evidence confirmed the Kerei-Abakh origin of Aksay Kazakhs. *J Hum Genet*. 2020;65:797–803.
- Underhill PA, Poznik GD, Roots S, Järve M, Lin AA, Wang J, et al. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet*. 2015;23:124–31.
- Zhabagin M, Balanovska E, Sabitov Z, Kuznetsova M, Agdzhoyan A, Balaganskaya O, et al. The connection of the genetic, cultural and geographic landscapes of Transoxiana. *Sci Rep*. 2017;7:3085.
- Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, et al. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet*. 2004;74:50–61.
- Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987–93.
- Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science*. 2013;341:562–5.
- Ralf A, Montiel GD, Zhong K, Kayser M. Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol Biol Evol*. 2018;35:1291–4.
- Helga T, James TR, Jill PM. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinforma*. 2013;14:178–192.
- Wei LH, Yan S, Lu Y, Wen SQ, Huang YZ, Wang LX, et al. Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan. *Eur J Hum Genet*. 2018;26:230–7.
- Malyarchuk B, Derenko M, Denisova G, Wozniak M, Grzybowski T, Dambueva I, et al. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Ann Hum Genet*. 2010;74:539–46.
- Huang YZ, Wei LH, Yan S, Wen SQ, Wang CC, Yang YJ, et al. Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407. *Mol Genet Genomics*. 2018;293:657–63.
- Wei LH, Huang YZ, Yan S, Wen SQ, Wang LX, Du PX, et al. Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. *J Hum Genet*. 2017;62:915–8.
- Huang YZ, Pamjav H, Flegontov P, Stenzl V, Wen SQ, Tong XZ, et al. Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia. *Mol Genet Genomics*. 2018;293:107–17.
- Zhabagin M, Sabitov Z, Tarlykov P, Tazhigulova I, Junissova Z, Yezhezhepov D, et al. The medieval Mongolian roots of Y-chromosomal lineages from South Kazakhstan. *BMC Genet*. 2020;21(Suppl 1):87.
- Osmonov, OJ, Turdalieva, CJ. *A History of Kyrgyzstan (From Stone Age to the Present): School and university textbook*. Bishkek: Sarybaev TT, 137–9 (2016).
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. *Science*. 2019;365:eaat7487.
- Keyser C, Zvěniĝorosky V, Gonzalez A, Fausser JL, Jagorel F, Gérard P, et al. Genetic evidence suggests a sense of family, parity and conquest in the Xiongnu Iron Age nomads of Mongolia. *Hum Genet*. 2021;140:349–59.
- Jeong C, Wang K, Wilkin S, Taylor WTT, Miller BK, Bemmann JH, et al. A Dynamic 6000-Year Genetic History of Eurasia's Eastern Steppe. *Cell*. 2020;183:890–904.
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, et al. Genomic insights into the formation of human populations in East Asia. *Nature*. 2021;591:413–9.
- Damgaard PB, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, et al. 137 ancient human genomes from across the Eurasian steppes. *Nature*. 2018;557:369–74.
- Amorim CEG, Vai S, Posth C, Modi A, Koncz I, Hakenbeck S, et al. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat Commun*. 2018;9:3547.
- Nagy PL, Olasz J, Neparáczki E, Rouse N, Kapuria K, Cano S, et al. Determination of the phylogenetic origins of the Árpád dynasty based on Y chromosome sequencing of Béla the Third. *Eur J Hum Genet*. 2021;29:164–72.
- Krzewińska M, Kiliç GM, Juras A, Koptekin D, Chyleński M, Nikitin AG, et al. Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci Adv*. 2018;4:eaat4457.
- Unterländer M, Palstra F, Lazaridis I, Piliipenko A, Hofmanová Z, Groß M, et al. Ancestry and demography and descendants of Iron age nomads of the Eurasian Steppe. *Nat Commun*. 2017;8:14615.
- Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015;522:167–72.
- Chikisheva TA, Gubina MA, Kulikov IV, Karafet TM, Voevoda MI, Romaschenko AG. A paleogenetic study of the prehistoric populations of the Altai. *Archeol Ethnol Anthropol Eurasia*. 2007;32:130–42.
- Gnecchi-Ruscone GA, Khussainova E, Kahbatkyzy N, Musralina L, Spyrou MA, Bianco RA, et al. Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. *Sci Adv*. 2021;7:eabe4414.

ACKNOWLEDGEMENTS

We are grateful for the trust of the sample donors. This work was supported by National Social Science Foundation of China (19VJX074), Major Project of National Social Science Foundation of China (20&ZD212), the National Natural Science Foundation of China (32070576), the Scientific and Technology Committee of Shanghai Municipality (18490750300), the National Key Research and Development Program (2020YFE0201600), Shanghai Municipal Science and Technology Major Project (2017SHZDX01), and the 111 Project (B13016).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s10038-021-00979-x>.

Correspondence and requests for materials should be addressed to Mei-sen Shi, Bo-feng Zhu or Hui Li.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.