



# Y-chromosome evidence confirmed the Kerei-Abakh origin of Aksay Kazakhs

Shao-Qing Wen<sup>1,2</sup> · Chang Sun<sup>1</sup> · Dan-Lu Song<sup>1</sup> · Yun-Zhi Huang<sup>1</sup> · Xin-Zhu Tong<sup>1</sup> · Hai-Liang Meng<sup>1</sup> · Hong-Bing Yao<sup>3</sup> · Pan-Xin Du<sup>1</sup> · Lan-hai Wei<sup>1,4</sup> · Ling-Xiang Wang<sup>1</sup> · Chuan-Chao Wang<sup>1,4</sup> · Mei-Sen Shi<sup>5</sup> · Yong-Mei Lan<sup>6</sup> · Jiu-Cun Wang<sup>1</sup> · Li Jin<sup>1</sup> · Maxat Zhabagin<sup>7</sup> · Xiao-Dong Xie<sup>8</sup> · Hui Li<sup>1</sup>

Received: 16 January 2020 / Revised: 14 March 2020 / Accepted: 2 April 2020  
© The Author(s), under exclusive licence to The Japan Society of Human Genetics 2020

## Abstract

Aksay Kazakhs are the easternmost branch of Kazakhs, residing in Jiuquan city, the forefront of the ancient Silk Road. However, the genetic diversity of Aksay Kazakhs and its relationships with other Kazakhs still lack attention. To clarify this issue, we analyzed the non-recombining portion of the Y-chromosome from 93 Aksay Kazakhs samples, using a high-resolution analysis of 106 biallelic markers and 17 STRs. The lowest haplogroup diversity (0.38) was observed in Aksay Kazakhs among all studied Kazakh populations. The social and cultural traditions of the Kazakhs shaped their current pattern of genetic variation. Aksay Kazakhs tended to migrate with clans and had limited paternal admixture with neighboring populations. Aksay Kazakhs had the highest frequency (80%) of haplogroup C2b1a3a1-F3796 (previous C3\*-Star Cluster) among the investigated Eurasian steppe populations, which was now seen as the genetic marker of Kerei clan. Furthermore, NETWORK analysis indicated that Aksay Kazakhs originated from sub-clan Kerei-Abakh in Kazakhstan with DYS448 = 23. TMRCA estimates of three recent descent clusters detected in C2\*-M217 (xM48) network, one of which incorporate nearly all of the C2b1a3a1-F3796 Aksay Kazakhs samples, gave the age range of 976-1405 YA for DC1, 1059-1314 YA for DC2, and 1139-1317 YA for DC3, respectively; this is coherent with the 7<sup>th</sup> to the 11<sup>th</sup> centuries Altaic-speaking pastoral nomadic population expansion.

---

These authors contributed equally: Shao-Qing Wen, Chang Sun, Dan-Lu Song

---

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s10038-020-0759-1>) contains supplementary material, which is available to authorized users.

---

✉ Shao-Qing Wen  
wenshaoqing@fudan.edu.cn

✉ Xiao-Dong Xie  
xdxie@lzu.edu.cn

✉ Hui Li  
lihui.fudan@gmail.com

<sup>1</sup> MOE Key Laboratory of Contemporary Anthropology and B&R International Joint Laboratory for Eurasian Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China

<sup>2</sup> Institute of Archaeological Science, Fudan University, Shanghai 200433, China

## Introduction

Aksay, located in Gansu province, is bounded on the west by Xinjiang Uygur Autonomous Region and the south by Qinghai province and is the only one Kazakh autonomous county in China except in Xinjiang. Aksay is under the jurisdiction of Jiuquan city, which was one of the four prefectures in Hexi Corridor, an important part of the

<sup>3</sup> Key Laboratory of Evidence Science of Gansu Province, Gansu Institute of Political Science and Law, Lanzhou 730070, China

<sup>4</sup> Department of Anthropology and Ethnology, Institute of Anthropology, Xiamen University, Xiamen 361005, China

<sup>5</sup> Institute of the Investigation, School of Criminal Justice, China University of Political Science and Law, 100088 Beijing, China

<sup>6</sup> Medical College of Northwest Minzu University, Lanzhou 730030, China

<sup>7</sup> National Center for Biotechnology, Astana 010000, Republic of Kazakhstan

<sup>8</sup> Institute of Medical Genetics, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China

ancient Silk Road in Han dynasty. Many populations migrated frequently via this region during historical periods, including Kazakhs [1].

The Kazakhs are the second largest Muslim group of Central Asia. Most of the Kazakhs live in Kazakhstan. About 18.6 million Kazaks are scattered throughout Central Asia, China, Russia, and as far away as Turkey and Germany. Approximately 1.5 million Kazaks were counted in China according to the 2010 census. They are centered primarily in the Ili Prefecture in northern Xinjiang. Moreover, about 2712 Kazaks live in Aksay Kazakh autonomous county, the northwest part of Gansu Province. Kazakh language belongs to the Western group of the Turkic language family and in China it can be roughly divided into two dialects: Northeastern Kazakh and Southwestern Kazakh (<http://www.ethnologue.com>). The Alban and Suwan tribes speak the southwest dialect; Kerei, Naiman, and Kazay tribes speak the northeastern dialect [2]. Traditionally, as their nomadic pastoral lifestyle, the Kazakhs lived in dome shaped felt tents called yurts. However, present Kazakhs took up a semi-nomadic or sedentary lifestyle, with only some Kazakhs maintaining the seasonal migrations. The Kazakhs' marriage system is monogamy in a patrilocal and exogamous way.

From the 13<sup>th</sup> to 15<sup>th</sup> century, Kazakhs emerged as an ethnic group on the Kazakh Steppe in Central Asia, and set up the Kazakh Khanate in now southern Kazakhstan [3, 4]. During the 16<sup>th</sup> century, Kazakhs split into three groups, separately called the Senior, Middle, and Junior *Zhüz*, or Hordes [3]. With the attack of Kalmyks in the 18<sup>th</sup> century, Kazakhs had to submit themselves to Russians. Then some Kazakhs, who mainly came from Kerei clan of the Middle *Zhüz*, spread to northwestern China, the Altai Mountains in western Mongolia and southern Russia [3–5]. Kerei originated from Keraites who were under the rule of the Genghis Khan's empire [6]. In alliance with the Keraites, Genghis Khan destroyed the Naimans, and then redirected all the force against his former ally. Some Keraites migrated from Mongolia to Central Asia, and assimilated some Turkic populations, such as Kazakhs, Uzbeks, and Kyrgyz [6]. In this context, the population incorporating Keraites and Kazakhs formed the Kerei clan. Kerei clan comprises two sub-clans, Kerei-Abakh and Kerei-Ashmaily [6]. Kerei in China mainly belong to its sub-clan Kerei-Abakh [5]. During the early 20<sup>th</sup> century, Kerei-Abakh mainly stayed in Xinjiang and partially migrated to Gansu [3]. In 1954, the Aksay Kazakh autonomous county was established and most Kazakhs in Gansu province gathered in this place [5].

Previous studies found Kazakhs shared a common paternal source and showed the genetic influence by the 13<sup>th</sup> century Mongol Empire [3, 7, 8]. However, these studies mainly focused on Kazakhs in Kazakhstan, the Altai Republic of Russia, and the Xinjiang Uygur Autonomous

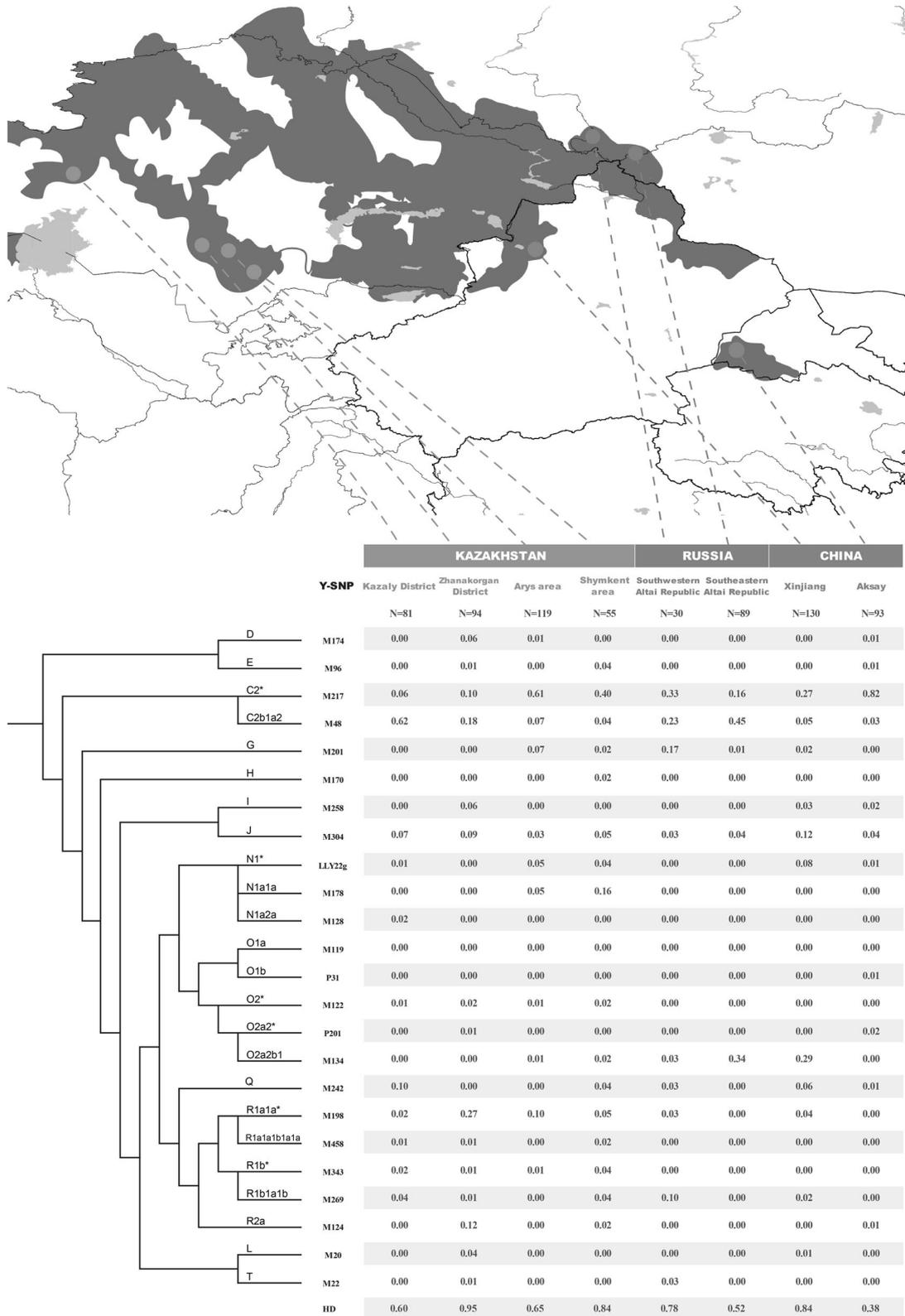
Region of China [3, 7–10]. The genetic relationships of Aksay Kazakhs with other Kazakhs still lack attention. To address this issue, we collected and tested 93 male Kazakh samples from Aksay Kazakh autonomous county in northwestern China.

## Materials and methods

We collected blood samples of 93 Kazakh male individuals in Aksay Kazakh Autonomous County, Gansu province of China (Fig. 1). Genealogical relationships were recorded prior to sample collection, confirming that all participants were unrelated within at least the last three generations. All individuals were adequately informed and signed informed consent forms before their participation. The ethics committee for biological research at the School of Life Sciences in Fudan University approved the study. Genomic DNA was extracted using QIAamp DNA Blood Mini Kit (QIAGEN, Germany). About 100 Y-SNP markers and 17 STR loci were tested in all DNA samples as described in previous studies [11–13]. Data were submitted to the YHRD (Y-chromosomal haplotype reference database) and received the accession number YA003979. Our Kazakh sample was compared with other Central Asian/Altaic-speaking populations and to surrounding areas using data from previous studies performed at a similar level of haplogroup resolution (Supplementary Tables 2 and 3). Furthermore, we analyzed Y-chromosomal STR data for haplogroups C2\*-M217 (xM48) and C2b1a3a1-F3796 by NETWORK v. 5.0.0.1 (<http://www.fluxus-engineering.com>). For use in the analyses, the locus DYS389b was obtained after subtracting the DYS389I from DYS389II. The loci DYS385a and DYS385b were excluded from calculations. The time to the most recent common ancestor (TMRCA) of each DC was determined by using the average squared distance (ASD) estimator as described previously [10–12]. A generation time of 25 years was used to produce time estimates in years [14].

## Results and discussion

We typed 17 Y-chromosomal microsatellites and 106 binary markers in 93 male Aksay Kazakh samples, and defined 31 different haplotypes and 12 haplogroups according to the nomenclature of Y-Chromosome Consortium and ISOGG Y-DNA Haplogroup Tree 2018, respectively (Supplementary Table 1). The Y-chromosomal haplogroup distribution in Aksay Kazakhs was compared with the other Kazakh populations in Xinjiang of China [15], southeastern and southwestern Altai Republic of Russia [3], and Kazaly, Zhanakorgan, Arys, and Shymkent area of Kazakhstan [10]



**Fig. 1** Map with the sampling points of Aksay Kazakhs and other studied Kazakhs. The distributions of the Kazakhs are shown in gray. The phylogenetic relationship of Y-chromosome haplogroups surveyed in this study and their frequencies in the studied Kazakhs is exhibited in the lower half of the figure. The marker names are shown

along the branches, and haplogroup names are shown on the right side according to ISOGG Y-DNA Haplogroup Tree 2018. Potentially paraphyletic undefined subgroups are distinguished from recognized haplogroups by the asterisk symbol

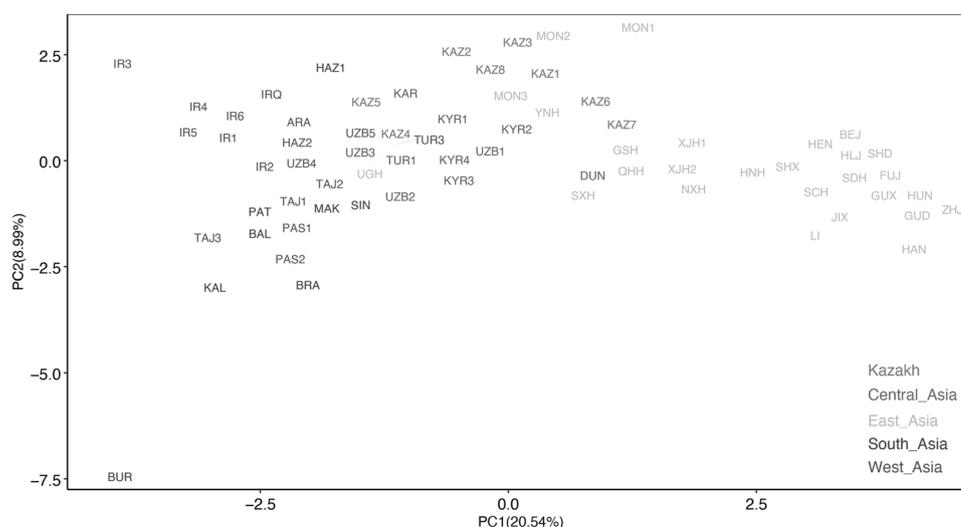
(Fig. 1). Haplogroup diversity was lowest for the Aksay Kazakhs (0.38) in all Kazakh populations. The most prominent clade, C2\*-M217 (xM48) as a whole, represents 82% of Y chromosomes among the Aksay Kazakhs, 27% among the Xinjiang Kazakhs, 6%–61% among the Kazakh populations in Kazakhstan, and 16–33% among the Kazakh populations in Russia. In Altaic-speaking populations, especially most Mongolic-speaking populations, there are four predominant paternal lineages of C2\*-M217 clade, including C2b1a3a1-F3796 (previous C3\*-Star Cluster) [16], C2a1a1b-M48 [17], C2b1a1a1a-M407 [18] and C2a1a1a2a-F1756 (with deletion on DYS448) [19]. In this study, C2b1a3a1-F3796 accounts for about 80% of the Aksay Kazakhs and reaches the highest frequency among reported Eurasian steppe populations (previous population screening revealed its highest frequency (76.5%) in Kerei clan from Kazakhstan [6]). In Aksay Kazakhs, Haplogroup C2a1a1b-M48 and C2a1a1a2a-F1756 are found at relatively low levels (3% and 2%, respectively) and Haplogroup C2b1a1a1a-M407 is absent. The rest detected haplogroups including present-day western Eurasian lineages (E-M96, I-M170, J-M304 and R2a-M124) and eastern Eurasian lineages (D1a2a-P47, N1a2b-P43, O1b1a1a-M95, O2a2\*-P201+, M7-,P164-, and Q1a\*-MEH2+,M120-,M25-,M346-) were found at low frequencies (<5%, Fig. 1).

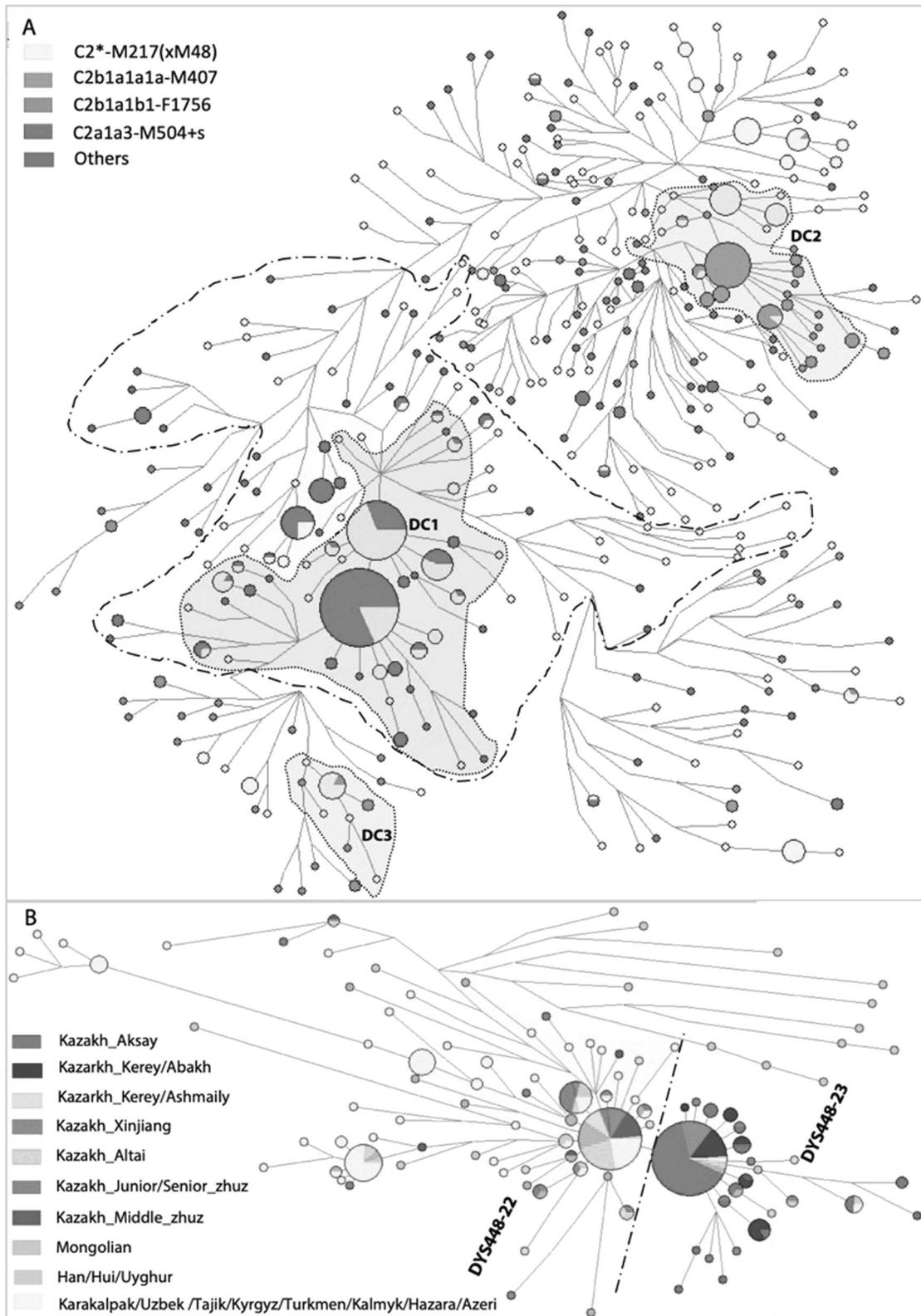
We then compared our data to published data from 5115 males from 70 Eurasian populations covering Central Asia, East Asia, South Asia and West Asia (Supplementary Table 2 and Fig. 2). Results of the Y-SNP-based PCA are presented by the plots of the first two principal components, which together account for 29.53% of the studied Y-chromosome variation in these populations. Populations from a particular region or ethnic group tend to cluster together. For example, rather than dispersed throughout the plots, most East Asian populations are seen in the right-

hand part, while South Asian and West Asian populations are shown in the left-hand part of the plots. Kazakh populations and other Central Asian populations tend to lie between the right-hand part (East) and left-hand part (West) cluster. Aksay Kazakhs (KAZ8) close to other Kazakh populations and Mongolian populations.

To reveal the detail relationship at individual level between the prominent haplogroup C2b1a3a1-F3796 in Aksay Kazakhs and other related populations, a reduced median network was constructed based on Y-STR haplotypes of haplogroup C2\*-M217 (xM48) (Fig. 3 and Supplementary Table 3a). Although the tested Y-SNP vary a lot in different studies, we merged the data into the following haplogroups with different colors: C2b1a1a1a-M407 (sky-blue), C2b1a1b1-F1756 (green), C2a1a3-M504 + s (+s means certain haplogroup and its sublineages (red), including C2a1a3-M504/F1918/M546/F914 and its sublineages C2b1a3a-M401 and C2a1a3a-F3796), other minor lineages (gray) and not further typed C2\*-M217 (xM48) samples (yellow) (Supplementary Table 2). The C2\*-M217 (xM48) haplotypes belonging to the identical lineages (colors) tend to cluster together; the undetermined samples (yellow) therefore should be assigned to certain haplogroups. The C2\*-M217 (xM48) network has three marked star-like shape, the “skyblue cluster” of C2b1a1a1a-M407 haplotypes, the “green cluster” of C2b1a1b1-F1756 haplotypes, and the “red cluster” of C2a1a3-M504+s haplotypes. Recent descent clusters (DCs) are characterised by a high-frequency Y-microsatellite haplotype and a set of close mutational neighbors, which means the signals of continued transmission of success over generations. In Fig. 3a, at least three DCs patterns (we used the criterion that haplotypes linked to the modal haplotype fewer than 5 mutational steps in the shaded area [10]) can be detected in C2\*-M217 (xM48) network. All Aksay Kazakhs haplotypes of

**Fig. 2** Phylogenetic relationship between Aksay Kazakhs and reference populations analyzed by PCA with the frequencies of haplogroups





**Fig. 3** Reduced Median joining network of Y-chromosome haplogroups C3\*-M217(xM48) (a) and haplogroup C2a1a3-M504+s (b). Haplotypes are represented by circles with area proportional to the number of individual

haplogroup C2b1a3a1-F3796 belong to DC1. The TMRCA estimates for DC 1-3 obtained using ASD are shown in Supplementary Table 4. The age of TMRCA of DC1 (975.8–1404.5 YA), DC2 (1058.9–1314.2 YA) and DC3 (1138.7–1316.8 YA) is almost identical to the published estimates [20], which corresponds to the time of the rise of Altaic-speaking pastoral nomadic populations, such as the Khitan (Great Liao), Tangut Xia, Jurchin, Kara-Khitan and the Mongol Empires [21].

Furthermore, in the network of C2a1a3-M504+s (haplotypes enclosed by dash dotted line in Fig. 3a), the basic information (geography, tribe or clan) of samples has been shown in Fig. 3b (Supplementary Table 3b). The samples from China (Han and Hui) and South/West Asia are likely derived haplotypes, as they are mostly at the terminals. There are a series of central haplotypes in the right lower part of network. This network displays a bipartite structure, defined by the difference at DYS448 locus as reported by previous analyses [6]. The left central haplotype is determined by allele DYS448-22, and right central haplotype by allele DYS448-23. These central haplotypes are occupied by the Eurasian steppe populations, such as Altai Kazakhs (green) and Kalmyks (yellow) from Russia, Mongols from Mongolia (orange), Xinjiang Kazakhs (magenta) and Aksay Kazakh (red) from China, Middle Zhüz Kazakhs (slate blue) from Kazakhstan, especially subclane Kerei-Abakh (blue) and Kerei-Ashmaily (cyan). The Kerei clan samples have been detected in left central haplotypes and right central haplotypes at 18.37% and 13.79%, respectively. Serikbai Abilev et al. found the Kerei clan appears to be descending from a common Y-chromosome ancestor, C2b1a3a1-F3796 and all members of the Kerei-Abakh belong to DYS448-23 subcluster, while the Kerei-Ashmaily belong to DYS448-22 subcluster [6]. In this context, Aksay Kazakhs are in or closer to the right central haplotype (DYS448-23), which matches the history records that its origin can trace back to “Twelve lineages of sub-clan Kerei-Abakh”, such as ZHantekej (坚铁开), ZHadik (加迪克) and Iteli (依铁里) [22]. Aksay Kazakh samples share the haplotypes with the Kerei-Abakhs ( $n = 17$ , blue), Xinjiang Kazakhs and Uyghurs ( $n = 14$ , magenta, and  $n = 2$ , pink), Mongols from northwestern Mongolia ( $n = 3$ , orange), Altai Kazakhs ( $n = 3$ , green), and one Kalmyk (buzawa tribe) from Kalmykia Republic and one Karakalpak from Uzbekistan (yellow), suggesting their potential paternal genetic relationships. Xinjiang Kazakh, by contrast, have a more complex origin, which is also consistent with historical records [5].

In this study, we find that Aksay Kazakhs had a very low paternal genetic diversity, implying that they have experienced a serial of founder effects or strong bottlenecks. Aksay is located in a border region of trans-province; hence Aksay Kazakhs were surrounded by many other populations, including Han Chinese, Uyghurs, Huis, Tibetans, and

Mongolians [5]. However, Aksay Kazakhs were mostly maintained by their tribal affiliations, they rarely absorbed male members of neighboring populations [3, 23]. Figure 1 and Supplementary Table 1 showed that Aksay Kazakhs had the highest frequency of haplogroup C2b1a3a1-F3796 among investigated Eurasian steppe populations, which was once seen as the lineage of Genghis Khan and his male descendants [24]. Subsequent studies revealed that the haplogroup C2b1a3a1-F3796 is a genetic marker of the Kerei clan, rather than Genghis Khan’s descendants. NETWORK analysis indicated that Aksay Kazakhs originated from sub-clan Kerei-Abakh in Kazakhstan with DYS448 = 23. TMRCA estimates of DCs detected in C2\*-M217 (xM48) network showed the genetic influence by the 7<sup>th</sup> to the 11<sup>th</sup> centuries steppe nomadic Empire [20].

**Acknowledgements** We are grateful for the trust of the sample donors. This work was supported by National Social Science Foundation of China (19VJX074), Scientific and Technology Committee of Shanghai Municipality (18490750300), National Natural Science Foundation of China (91731303, 81671874, 31760309, 31771325, 91631105, and 31801040), Innovation Team Project of Fundamental Research Funds for the Central Universities (31920190030), and 111 Project (B13016).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

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

## References

1. 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.
2. Geng SM. On Kazak spoken and written language. *N W Ethno-Natl Stud.* 2006;2:56–63.
3. Dulik MC, Osipova LP, Schurr TG. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One.* 2011;6:e17548.
4. Gokcumen O, Dulik MC, Pai AA, Zhadanov SI, Rubinstein S, Osipova LP, et al. Genetic variation in the enigmatic Altaian Kazakhs of South-Central Russia: insights into Turkic population history. *Am J Phys Anthropol.* 2008;136:278–93.
5. Xie, GX, Wang, XP. History of the Kazakhs in the Gansu province of China. Lanzhou: Gansu Culture Press (2009).
6. Abilev S, Malyarchuk B, Derenko M, Wozniak M, Grzybowski T, Zakharov I. 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.* 2012;84:79–89.
7. Tarylkov PV, Zholdybayeva EV, R Akilzhanova A, M Nurkina Z, M Sabitov Z, Rakhypbekov TK, et al. Mitochondrial and Y-chromosomal profile of the Kazakh population from East Kazakhstan. *Croat Med J.* 2013;54:17–24.
8. 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.

9. Zhabagin M, Sarkytbayeva A, Tazhigulova I, Yerezhepov D, Li S, Akilzhanov R, et al. Development of the Kazakhstan Y-chromosome haplotype reference database: analysis of 27 Y-STR in Kazakh population. *Int J Leg Med.* 2019;133:1029–32.
10. Zhabagin M, Balanovska E, Sabitov Z, Kuznetsova M, Agdzho-yan A, Balaganskaya O. The connection of the genetic, cultural and geographic landscapes of transoxiana. *Sci Rep.* 2017;7:3085.
11. 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.
12. Wen SQ, Xu D, Yao HB, Li H. Present Y Chromosomes Refute the Roma/Gypsy Origin of the Xuejiawan People in Northwest China. In: XU Dan, LI Hui (eds) *Languages and Genes in Northwestern China and Adjacent Regions.* Singapore: Springer, 107–20 (2017).
13. Hu K, Yan S, Liu K, Ning C, Wei LH, Li SL, et al. The dichotomy structure of Y chromosome Haplogroup N. [arXiv:1504.06463](https://arxiv.org/abs/1504.06463) (2015). <https://arxiv.org/abs/1504.06463v1>
14. Zhivotovsky LA, Underhill PA, Cinnioglu 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.
15. Zhang D, Cao G, Xie M, Cui X, Xiao L, Tian C, et al. Y Chromosomal STR haplotypes in Chinese Uyghur, Kazakh and Hui ethnic groups and genetic features of DYS448 null allele and DYS19 duplicated allele. *Int J Legal Med.* <https://doi.org/10.1007/s00414-019-02049-6> (2019).
16. 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.
17. 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.
18. 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.
19. 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.
20. Balaesque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, et al. Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. *Eur J Hum Genet.* 2015;23:1413–22.
21. Rogers JD. Inner Asian states and empires: theories and synthesis. *J Archaeol Res.* 2012;20:205–56.
22. He H. *Annals of the autonomous prefecture of Aksai Kazaks.* Lanzhou: Gansu People's Publishing House (1993).
23. Zhang C. Simple study of the national characteristics Kazak Islamic religion. *J Changchun Univ Sci Technol (Soc Sci Ed).* 2010;23:39–42.
24. Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. The genetic legacy of the Mongols. *Am J Hum Genet.* 2003;72:717–21.