



Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan

Lan-Hai Wei^{1,2} · Shi Yan¹ · Yan Lu³ · Shao-Qing Wen¹ · Yun-Zhi Huang¹ · Ling-Xiang Wang¹ · Shi-Lin Li¹ · Ya-Jun Yang¹ · Xiao-Feng Wang¹ · Chao Zhang^{3,4} · Shu-Hua Xu^{1,3,4,5} · Da-Li Yao⁶ · Li Jin¹ · Hui Li¹

Received: 17 January 2017 / Revised: 27 July 2017 / Accepted: 23 August 2017

© European Society of Human Genetics 2018

Abstract

The Y-chromosome haplogroup C3*-Star Cluster (revised to C2*-ST in this study) was proposed to be the Y-profile of Genghis Khan. Here, we re-examined the origin of C2*-ST and its associations with Genghis Khan and Mongol populations. We analyzed 34 Y-chromosome sequences of haplogroup C2*-ST and its most closely related lineage. We redefined this paternal lineage as C2b1a3a1-F3796 and generated a highly revised phylogenetic tree of the haplogroup, including 36 sub-lineages and 265 non-private Y-chromosome variants. We performed a comprehensive analysis and age estimation of this lineage in eastern Eurasia, including 18,210 individuals from 292 populations. We discovered that the origin of populations with high frequencies of C2*-ST can be traced to either an ancient Niru'un Mongol clan or ordinary Mongol tribes. Importantly, the age of the most recent common ancestor of C2*-ST (2576 years, 95% CI = 1975–3178) and its sub-lineages, and their expansion patterns, are consistent with the diffusion of all Mongolic-speaking populations, rather than Genghis Khan himself or his close male relatives. We concluded that haplogroup C2*-ST is one of the founder paternal lineages of all Mongolic-speaking populations, and direct evidence of an association between C2*-ST and Genghis Khan has yet to be discovered.

Lan-Hai Wei and Shi Yan contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0012-3>) contains supplementary material, which is available to authorized users.

✉ Hui Li
LHCA@Fudan.edu.cn

- ¹ MOE Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, 200438 Shanghai, China
- ² Institut National des Langues et Civilisations Orientales, 75214 Paris, France
- ³ Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
- ⁴ University of Chinese Academy of Sciences, 100049 Beijing, China
- ⁵ School of Life Science and Technology, ShanghaiTech University, 201210 Shanghai, China
- ⁶ Center for Historical Geographical Studies of Fudan University, 200433 Shanghai, China

Introduction

The Y chromosome is a powerful tool for tracing the paternal history of certain ethnic groups or clans. Y-chromosome haplogroup C3*-Star Cluster (revised to C2*-ST in this study), first discovered by Zerjal et al. [1] is one of the most well-known paternal profiles. This lineage was common among the peoples (~8%) residing between the Yellow River and the Caspian Sea. The time to the most recent common ancestor (TMRCA) of the lineage (1000 ± 300 years ago) coincided with Genghis Khan (1162–1227 A.D.), and the present distribution of the lineage matches that of Genghis Khan's Mongol Empire. A possible descendant population of Genghis Khan, Hazara in Afghanistan, has a high frequency of this paternal lineage. Based on these lines of evidence, it has been proposed that Genghis Khan or his close relatives are the origin of this special Y-chromosome lineage. However, genome wide studies have revealed large scale recent expansions in Eurasia [2, 3] and it is still unclear whether or not there is a connection between the origin of C2*-ST and Genghis Khan.

As discussed by Zakharov et al. [4] some studies have opposed the connection between C2*-ST and Genghis Khan [5–10]. The highest frequencies of C2*-ST were found in two populations of Kazakhs (Kerey-Abakh and Kerey-Ashmaily, 89.3% and 55.0%, respectively) [7]. Additionally, the Y haplotype of a direct descendant of Genghis Khan, Batu-Mungke Dayan Khan (1474–1517 A. D., ruler of North Yuan Khanate) is C2c1a1a1-M407 based on the direct testing of his well-documented descendants [11]. In this study, we reanalyzed the whole sequences and Y-chromosome short tandem repeat (Y-STR) haplotypes of samples from a broader geographical scale to clarify the origin of C2*-ST and its connections with Genghis Khan and Mongol populations.

Materials and methods

Samples

Blood or saliva samples of 6348 individuals from 74 populations in eastern Eurasia were collected (see Supplementary Table S1) from unrelated healthy males between 2005 and 2014. All individuals signed their consent forms before participating in the study. The ethics committee for biological research at the Fudan School of Life Sciences approved the study.

Molecular methods

Genomic DNA was extracted using the DP-318 Kit (Tiangen Biotechnology, Beijing, China) according to the manufacturer's protocol. First, the Y-chromosome marker M130 was genotyped to identify haplogroup C-M130 samples. Individuals with the derived allele at M130 were then subjected to further typing of ten biallelic markers: M8, M38, M217, M93, P39, M48, M407, P53.1, M347, and M356. The C2*-ST samples were M130+, M217+, M93-, P39-, M48-, M407-, and P53.1-. Seventeen STR loci of all DNA samples were genotyped using the AmpFISTR® YFiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA). Amplification products were analyzed using ABI 3730 and ABI 3130 Genetic Analyzers (Applied Biosystems). Electrophoresis results were analyzed using GeneScan v. 3.7 and Genotyper v. 3.7 (Applied Biosystems).

The DNA extracted from 17 selected C2*-ST samples was sent for next-generation sequencing using the Illumina HiSeq2000 platform (San Diego, CA, USA). A series of bait libraries were designed to capture the sequences of a region of ~11 Mb on the Y chromosome. A procedure that we described previously was used for all other steps prior to next-generation sequencing, i.e., DNA shearing,

adding an adaptor, and gel electrophoresis [12]. The raw sequence data reported in this paper have been deposited in the Genome sequence archive (GSA) [13] in BIG Data Center [14], Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers PRJCA000420 that are publicly accessible at <http://bigd.big.ac.cn/gsa>. Standard procedures (bwa + samtools) were followed to analyze the next-generation sequencing data [15, 16]. Another 17 previously published Y-chromosome sequences were also used to construct a phylogeny of haplogroup C2*-ST, including one sequence from Yan et al. [12], three from Wei et al. [17], one from Karmin et al. [18], and 11 from Lippold et al. [19]

The regulations proposed by the YCC were followed to revise the phylogenetic tree with respect to new variants in the non-recombining region of the Y chromosome [20]. According to recent phylogenetic studies, C1-M8, C2-M38, C4-M347, and C5-M356 now belong to the newly defined C1-F3393 haplogroup [18, 21]. Accordingly, previous C3-M217 has been redefined as C2-M217 [22]. Therefore, we redefined the C3*-ST lineage as C2*-ST in this study. The Y-SNP marker M401, an important variant under C2*-ST, was first reported in Di Cristofaro et al. [8] The F914 + sample from Kazakh (ID: 13752) in Karmin et al. [18] is also derived from the marker M401.

Statistical analysis

The proposed Genghis Khan Y-profile is C2*-ST (M217 +, M93-, P39-, M48-, M407-, P53.1-) [5–7]. To more comprehensively characterize this lineage, Y-chromosome haplogroup frequencies and Y-STR data were collected for haplogroup C-M130 in 218 East Eurasia populations from the literature (see Supplementary Table S1 and S2). Y-STR haplotypes of C2*-ST were identified based on the definition in the original paper [1], i.e., within one to three mutational steps from a central haplotype with a particular set of Y-STR loci (DYS389I-DYS389b-DYS390-DYS391-DYS392-DYS393-DYS388- DYS425-DYS426-DYS434-DYS435-DYS436-DYS437-DYS438-DYS439, but excluding DYS425, DYS426, DYS434, DYS435, and DYS436, which have rarely been studied). The frequencies of C2*-ST in populations were plotted on a geographic map using Surfer 7.0 (Golden Software, Inc., Golden, CO, USA). Only those haplotypes with 15 Y-STR markers (excluding DYS385a and DYS385b from the set of 17 Y-STRs) were used to construct the median-joining network using NETWORK 4.6.0.0 (Fluxus Engineering, Suffolk, UK) [23].

Coalescence dating

The TMRCA of the Y-STR haplotypes was estimated using average squared distances (ASD) [24–26] and Bayesian

analysis of trees with internal node generation (BATWING) [27, 28]. The genealogical mutation rate [29] was used for the age estimates using both ASD and BATWING. The generation time was set to 30 years.

Seventeen Y-chromosome sequences of C2*-ST were used to calculate the age of the haplogroup and its sub-lineages (Supplementary Table S3 and Supplementary Fig. S1). Additionally, another 18 samples were also included as outgroup samples, including samples of haplogroups D-M174, N-M231, O2a-M95, C2-M407, C2-M48, and C2-F8951 (Supplementary Fig. S1). Haplogroup C2-F8951, also called the “C3*-Daur clade” in our previous study [17], is the most closely related lineage to C2*-ST. A number of variants were determined after the analysis of Y-chromosome sequences (see Supplementary Table S4, Supplementary Fig. S1, and Supplementary file 5,.vcf file). To obtain a confident data set for age estimation, a series of strict filters were applied to the file containing original variants. The age of haplogroup CT-M168 (71,760 years, 95% confidence interval [CI] = 69,777–73,799) [18] and the splitting time (470 ± 20 years ago) of samples (YCH508 and YCH1981) from the Aisin Gioro family [17] were used as calibration points to estimate the coalescent times of all samples in the study. The details of the filters and the selection of calibration points are described in the Supplementary Text.

BEAST v.2.4.3 was used to estimate the coalescent times of haplogroup C2*-ST and its sub-lineages [30]. A Bayesian skyline coalescent tree prior was selected with the general time reversible (GTR) substitution model with gamma distributed rates and a strict clock. The GTR model of nucleotide substitution is a basic model for many distance-based and character-based phylogeny inference methods [31]. Besides, the bModelTest package allows the BEAST program to infer the most optimistic substitution model for input sequences [32]. Therefore, separate runs were conducted using the bModelTest package. Both runs were performed with 10 million iterations and sampling every 1000 steps. The results were visualized in Tracer v.1.6 and FigTree v1.4.2 with a burn-in of 20%, and all effective sample sizes were above 200.

Results

Among 18,210 samples from 292 populations throughout eastern Eurasia, we identified 809 samples with 685 Y-STR haplotypes as C2*-ST (see Supplementary Table S2). The distribution of C2*-ST in East Eurasian populations is shown in Fig. 1. Typing data for 15 Y-STR were only available for 476 haplotypes from 511 samples of C2*-ST and its most closely related branch C2*-Daur clade (C2-

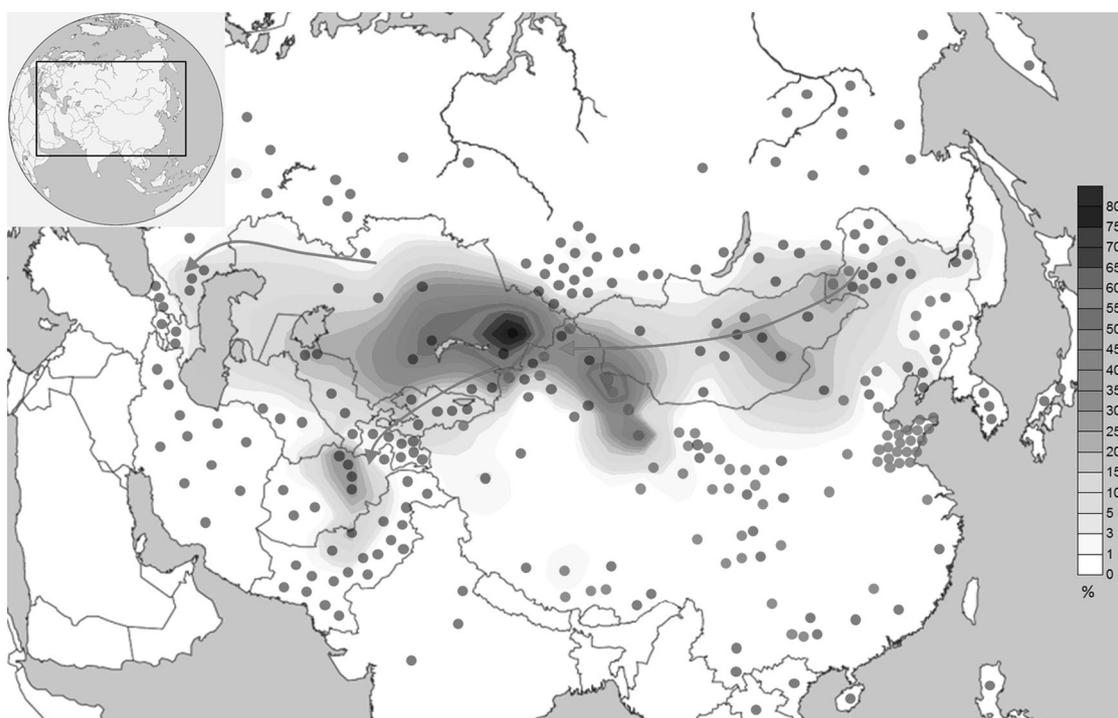
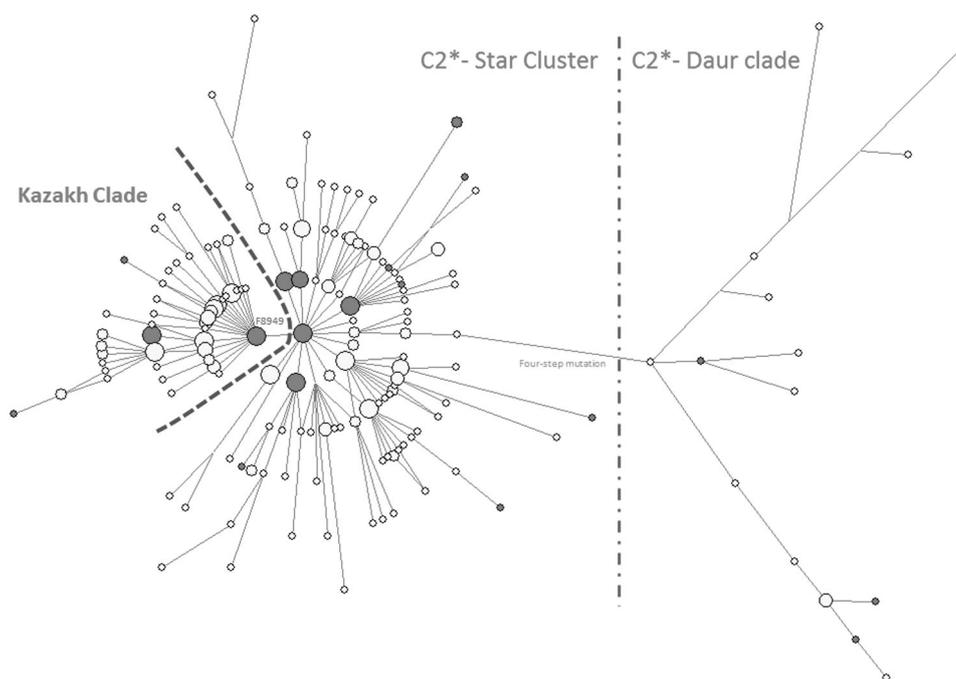


Fig. 1 Distribution of the Y-chromosome lineage C2*-Star Cluster across Eurasia. Note: black dots indicate populations taken from the literature and red dots indicate populations reported for the first time in this study

Fig. 2 Y-STR network of C2*-ST based on 15 Y-STRs. Note: The Y-STR haplotype of the sequenced sample is indicated by red circles on the network. Sample YCH509 and HLB-179 share the same Y-STR haplotype. Sample HLB-095, even though negative on F8949, shares the same value with the central haplotype of the Kazakh clade



F8951). We used these haplotypes in our subsequent analyses (see Supplementary Table S2). Y-STR networks based on 15 markers showed a star-like expansion of haplogroup C2*-ST and its sub-branches (Fig. 2). Since all of the sequenced C2*-ST samples had the derived state at marker F3796, we redefined this haplogroup as C2b1a3a1-F3796. The revised phylogenetic tree of haplogroup C2b1a3a1-F3796 contained 36 sub-clades, 265 non-private variants, and a number of private variants (Fig. 3, also see Supplementary Table S3).

We observed the highest frequencies of C2*-ST in several Kazakh populations in Southeast Kazakhstan and Northwest China, followed by Mongolian, Buryat-Bargut, and Uzbeks. According to sampling information, the Kazakh populations with C2*-ST frequencies of greater than 50% were collected in the Great Jüz of Kazakhs or Kazakh-Kerey. By contrast, C2*-ST was absent or found at very low frequencies in non-Altaic populations. The only exceptions were Hazara. Hazara is a Persian-speaking population with a Mongol origin.

It is worth noting the distribution of haplogroup C2b1a3a2-F8951, the closest branch to C2*-ST. We defined this clade as C2*-Daur clade in our previous study [17]. The Y-STR C2b1a3a2-F8951 haplotype has been randomly detected in the Daur, Manchu, Mongolian, Mongolian Buryats, Oroqen, and Xibe populations [17]. This haplogroup is the founding paternal lineage of the Daur and Aisin Gioro clan in the Manchu populations [17]. We tended to detect C2b1a3a2-F8951 in populations in the eastern Greater Khingan Mountains, while we mainly detected C2b1a3a1-F3796 in populations in the western

Greater Khingan Mountains, i.e., the Mongolia Plateau, Central Asia, western Asia, and Northern Caucasus region (Supplementary Table S2).

Estimates of the TMRCA and expansion time can provide insights into the history of a particular lineage, including its origin and diffusion. As shown in the Y-STR network, a special branch with $DYS448 = 23$ (defined as the C2*-ST-Kazakh clade, Fig. 2 and Supplementary Table S2) emerged in the Kazakh-Kerey-Abakh population [7]. The C2*-ST-Kazakh clade showed a similar star-like expansion to that of C2*-ST. Y-chromosome sequence data also indicated that samples with $DYS448 = 23$ form a special sub-lineage, C2b1a3a1c2a-F8949 (Fig. 3 and Supplementary Table S3). Furthermore, we detected a unique downstream sub-lineage, C2b1a3a1c2-F5481, mainly in Kazakh, Kirghiz, and Hazara populations. We estimated the ages of the C2b1a3a2-F8951 (C2*-Daur clade), C2*-ST, C2b1a3a1c2-F5481, and C2*-ST-Kazakh clades. The TMRCA estimates obtained using ASD, BATWING, and BEAST are shown in Table 1. We obtained similar age estimates using BEAST with two modes and ASD. However, there were two exceptions. The ages of C2b1a3-F1918 and C2b1a3a1-F3796 were younger using ASD than BEAST. Additionally, the total age of C2*-ST calculated by ASD was less than the age of its sub-clade C2b1a3a2-F8951 (Table 1). ASD may underestimate the ages owing to large number of identical Y-STR haplotypes in haplogroup C2*-ST (see Supplementary Table S2). We obtained older age estimates using BATWING than the other three methods. Therefore, we preferred the age estimates of BEAST, which are based on whole Y-chromosome sequences. The

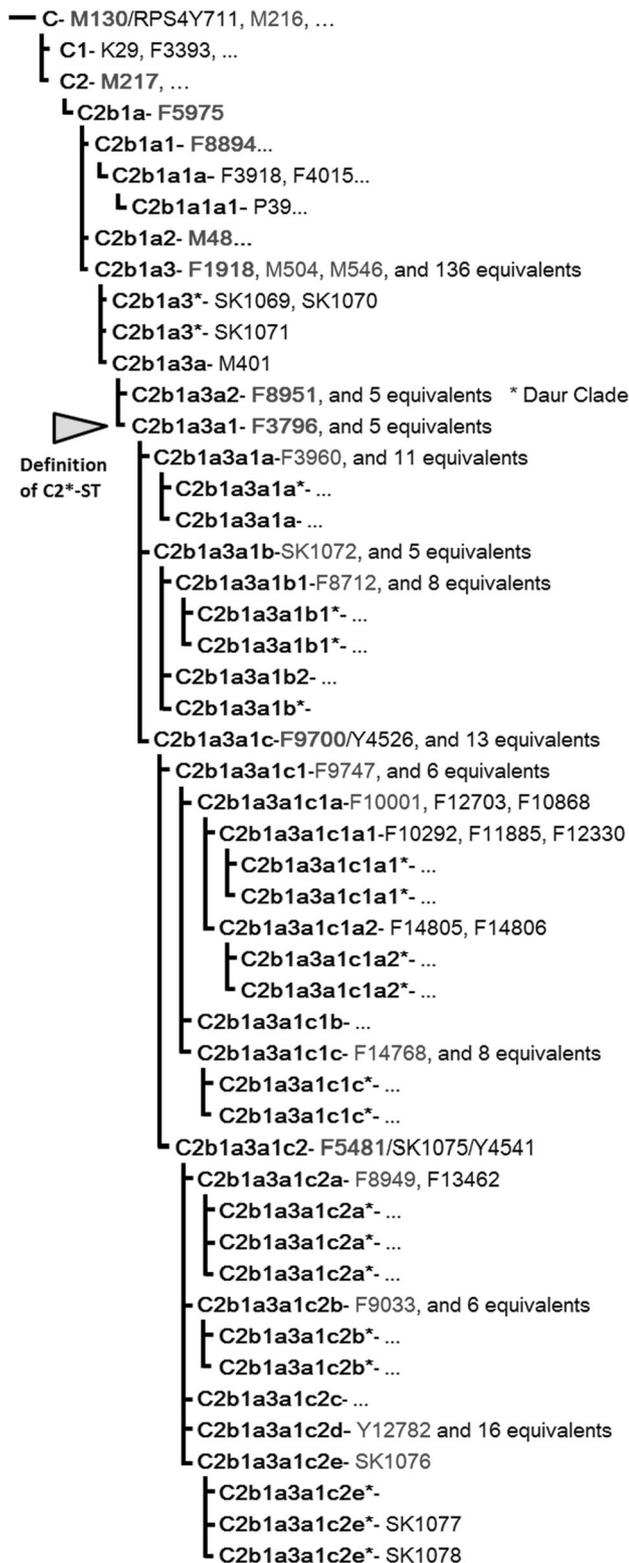


Fig. 3 Revised phylogeny of the Y-chromosome lineage C2*-Star Cluster

TMRCAs of C2*-ST was 2 576 years (95% CI = 1975–3178) as calculated by BEAST with bModelTest mode.

Discussion

Combined with the results of historical studies, we found that several modern populations with high frequencies of C2*-ST can be traced back to either an ancient Mongol Niru'un clan or ordinary Mongol tribes, including the Manghit tribe in Uzbekistan and Nogay populations, the Keneges tribe from Uzbekistan, the Hazara population from Afghanistan, the Daur population from China, and the Dulat, Uysun, and Kerey tribe in Kazakh populations. The details of the origin and migration history of these tribes or populations are discussed in the Supplementary Text. The Niru'un Mongols (which translates to “the pure Mongols”) were believed to be the descendants of Alan Quo'a. Genghis Khan belonged to the Kiyan clan in the Niru'un tribe (Supplementary Fig. S2) [33].

Hazara people were considered as direct descendants of Genghis Khan, and hence become a strong evidence to support the connection between C2*-ST and Genghis Khan in Zerjal et al. [1] However, the original material used in this study indicated that the Hazara were derived from ten military detachments sent by Genghis Khan [34]. According to available historical records, those military detachments, amounting to about 20,000 soldiers in total, were ordinary people of Mongol tribes [34, 35]. There is no evidence that they were direct descendants of Genghis Khan, whose descendants were well-documented during that era [33, 35].

According to the revised phylogenetic tree, C2b1a3a2-F8951 is the most closely related lineage to C2*-ST (Fig. 3). We observed C2b1a3a2-F8951 mainly in Daur and Manchu populations in the eastern Greater Khingan Mountains [17], while we observed other C2*-ST lineages in populations in the Mongolia Plateau, Central Asia, or even western regions (Supplementary Table S1 and Supplementary Table S2). Interestingly, according to historical records, the northern region of the Greater Khingan Mountains is the home of the Shi-Wei tribes, ancestors of Mongols [36]. The Daur population lived in the middle reaches of the Amur River at the end of 16th century. There is no earlier historical record of the Daur. Additionally, there is no evidence that they were part of the Mongol tribes when Genghis Khan and his descendants started to unify the Mongol populations and subsequently establish a vast empire across Eurasia [33]. Therefore, we concluded that both Hazara and Daur originated from ordinary ancient Mongolic-speaking populations, rather than from Genghis Khan or his close male relatives.

Table 1 Age estimations of Y-chromosome lineage C2*-ST and its sub-clades

Haplogroup	Y-Sequences			Y-STR		
	<i>N</i>	bModelTest Median, 95% CI	GTR + G + I Median, 95% CI	<i>N</i>	ASD $\rho \pm \sigma$	Batwing Median, 95% CI
C2b1a3-F1918	20	2896 [2256–3602]	2868 [2261–3566]	476	1790 ± 1133	6327 ± 1736
C2b1a3a2- F8951	3	2378 [1734–3110]	2365 [1722–3059]	17	2260 ± 1198	4653 ± 1392
(C2*-Daur Clade)						
C2b1a3a1- F3796	17	2576 [1975–3178]	2560 [2032–3183]	460	1590 ± 1217	3381 ± 982
(C2*-ST)						
C2b1a3a1c2- F5481	6	1069 [686–1470]	1059 [676–1441]	N/A	N/A	N/A
C2b1a3a1c2a- F8949	3	644 [346–982]	635 [353–979]	230	489 ± 233	2813 ± 1063
(C2*-ST-Kazakh Clade)						

ASD average squared distance, BATWING Bayesian analysis of trees with internal node generation, CI confidence interval, *bModelTest* applying the “bModelTest” package in Beast, GTR + G + I general time reversible + gamma distribution + proportion of invariable sites

According to historical studies and legends, ancestors of Mongol tribes lived in the northern region of the Greater Khingan Mountains before they moved westward onto the Mongolia Plateau [33, 35, 36]. Subsequently, they expanded to Central Asia and Europe after the establishment of the Mongol Empire. The estimated age of haplogroup C2*-ST in this study (~2600 years) was much older than the earliest record of Mongol tribes in Chinese historical materials (~1300 years ago) [36]. However, the Xian-Bei and Shi-Wei tribes appeared in Chinese historical materials at about 1900 years ago [36]. It is generally accepted that the Xian-Bei and Shi-Wei tribes are the direct ancestors of modern Mongolic-speaking populations. In the context of these historical records, we propose that haplogroup C2*-ST originated in the northern region of the Greater Khingan Mountains, and the genetic expansion of this lineage corresponds to the differentiation of ancient Mongolic-speaking populations.

According to the revised phylogenetic tree of C2*-ST in this study, we propose that sub-lineage C2b1a3a1c2-F5481/SK1075 is an important clade of C2*-ST in Central Asia and adjacent region (Fig. 3 and Supplementary Table S3). Haplogroup C2b1a3a1c2-F5481 only had one definitive marker and gave birth to five different sub-branches determined by samples from Mongolian populations and populations from Central Asia, such as Hazara, Kazakhs, and Kirgiz (Fig. 3 and Supplementary Table S3). As discussed in the Supplementary Text, Manghit, Keneges, Dulat, and Hazara can be clearly traced to Niru'un Mongols. They are all descendants of the armies sent to various regions of the

Mongol Empire by Genghis Khan. Therefore, genetic evidence in this study suggests that sub-lineage C2b1a3a1c2-F5481/SK1075 is one of the most important lineages in ancient Mongols, which eventually spread to the west of the Altai Mountain region.

We did not find evidence to support the previous proposition that this lineage has direct connections with Genghis Khan himself [1]. Since Genghis Khan was a member of the Niru'un clan (see Supplementary Fig. S2), it is possible that the Great Conqueror carried the Y-STR profile of C2*-ST. However, based on all available data in the literature and the results reported in this study [4–10], none of the donors who carried C2*-ST can trace their genealogy to Genghis Khan, and none of the self-claimed direct descendants of Genghis Khan carry C2*-ST. As shown in Batbayar et al. [11], a direct descendant of Genghis Khan, Dayan Khan, had a different Y-chromosome haplogroup (C2c1a1a1-M407).

In conclusion, our genetic analyses of C2*-ST provide a clear account of the dispersal and expansion patterns of the lineage throughout the steppe zone of Eurasia in the last millennium. We propose that C2*-ST is a predominant paternal profile in ordinary Mongol tribes, where Genghis Khan's paternal family came from. However, direct genotyping of more of his well-documented male descendants from a wider geographic region is needed to definitively characterize the Y chromosome of Genghis Khan in the future.

Acknowledgements We are grateful to all sample donors. LHW was supported by Future Scientists Project of China Scholarship Council.

This work was supported by NSFC for Excellent Young Scholar (nos 31222030, 31671297, 31271338, 31401060, 31501011, 91331204), MOE Scientific Research Project (113022A), Ministry of Science and Technology of China (MOST) (2016YFC0900300), Shanghai Shuguang Project (14SG05), and Science and Technology Commission of Shanghai Municipality (STCSM) (14YF1406800). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Author contributions LHW and SY carried out the molecular genetic studies and drafted the manuscript. YL, SY, SQW, YJY, SLL, XFW, CZ, and SHX collected the samples. YL, CZ, SHX, SQW, and LHW participated in the experiments. YZH and LXW participated in the statistical analysis. DLY helped to analyze the history records. JL and HL designed the study and revised the manuscript. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- 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.
- Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R, et al. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet.* 2015;11:e1005068.
- Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. *Nature.* 2015;522:167–72.
- Zakharov IA. A search for a “Genghis Khan chromosome”. *Genetika.* 2010;46:1276–7.
- Derenko MV, Malyarchuk BA, Wozniak M, Denisova GA, Dambueva IK, Dorzhu CM, et al. Distribution of the male lineages of Genghis Khan’s descendants in northern Eurasian populations. *Russ J Genet.* 2007;43:334–7.
- 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.
- 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.
- Di Cristofaro J, Pennarun E, Mazieres S, Myres NM, Lin AA, Temori SA, et al. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PLoS ONE.* 2013;8:e76748.
- Balaresque 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.
- Lkhagvasuren G, Shin H, Lee SE, Tumen D, Kim JH, Kim KY, et al. Molecular genealogy of a Mongol Queen’s family and her possible kinship with Genghis Khan. *PLoS ONE.* 2016;11:e0161622.
- Batbayar K, Sabitov ZM. The genetic origin of the Turko-Mongols and review of the genetic legacy of the Mongols. Part 1: The Y-chromosomal lineages of Chinggis Khan. *Russ J Genet Geneal.* 2012;4:1–8.
- Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, et al. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS ONE.* 2014;9:e105691.
- Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome sequence archive. *Genomics Proteomics Bioinformatics.* 2017;15:14–18.
- Members BIGDC. The BIG Data Center: from deposition to integration to translation. *Nucleic Acids Res.* 2017;45:D18–D24.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
- Wei LH, Yan S, Yu G, Huang YZ, Yao DL, Li SL, et al. Genetic trail for the early migrations of Aisin Gioro, the imperial house of the Qing dynasty. *J Hum Genet.* 2016;62:407–11.
- Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, Talas UG, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 2015;25:459–66.
- Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Invest Genet.* 2014;5:13.
- Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 2002;12:339–48.
- Bergstrom A, Nagle N, Chen Y, McCarthy S, Pollard MO, Ayub Q, et al. Deep roots for aboriginal Australian Y Chromosomes. *Curr Biol: CB.* 2016;26:809–13.
- Perez-Benedico D, La Salvia J, Zeng Z, Herrera GA, Garcia-Bertrand R, Herrera RJ. Mayans: a Y chromosome perspective. *Eur J Hum Genet.* 2016;24:1352–8.
- Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16:37–48.
- 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.
- Ramakrishnan U, Mountain JL. Precision and accuracy of divergence time estimates from STR and SNPSTR variation. *Mol Biol Evol.* 2004;21:1960–71.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, et al. 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.* 2006;78:202–21.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, et al. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics.* 2006;172:2431–9.
- Wilson IJ, Weale ME, Balding DJ. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc A Stat.* 2003;166:155–88.
- Burgarella C, Navascues M. Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet.* 2011;19:70–5.
- Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10:e1003537.
- Lanave C, Preparata G, Sacone C, Serio G. A new method for calculating evolutionary substitution rates. *J Mol Evol.* 1984;20:86–93.
- Bouckaert RR, Drummond AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol.* 2017;17:42.

33. Onon U. The secret history of the Mongols: The life and times of Chinggis Khan. 3rd ed. London and New York: Routledge; 2001.
34. Bellew H. The races of Afghanistan. New Delhi: Asian Educational Services; 1880.
35. Al-din R, Yu D, Zhou J. Jámí'u-t Tawárikh(Shiji). Beijing: Commercial Press; 1997.
36. Lin G. A history of Donghu (in Chinese). Hohhot: Inner Mongolian People's Publishing House; 2007.