



# Phylogenetic analysis of the Y-chromosome haplogroup C2b-F1067, a dominant paternal lineage in Eastern Eurasia

Qiao Wu<sup>1</sup> · Hui-Zhen Cheng<sup>2</sup> · Na Sun<sup>3</sup> · Peng-Cheng Ma<sup>4</sup> · Jin Sun<sup>2</sup> · Hong-Bing Yao<sup>5</sup> · Yong-Mei Xie<sup>6</sup> · Yong-Lan Li<sup>6</sup> · Song-Lin Meng<sup>7</sup> · Maxat Zhabagin<sup>8,9</sup> · Yan-Huan Cai<sup>2</sup> · Da-Ru Lu<sup>1</sup> · Shi Yan<sup>10</sup> · Lan-Hai Wei<sup>2,9</sup>

Received: 29 December 2019 / Revised: 29 April 2020 / Accepted: 30 April 2020 / Published online: 19 May 2020  
© The Author(s), under exclusive licence to The Japan Society of Human Genetics 2020

## Abstract

Human Y-chromosome haplogroup C2b-F1067 is one of the dominant paternal lineages of populations in Eastern Eurasia. In order to explore the origin, diversification, and expansion of this haplogroup, we generated 206 new Y-chromosome sequences from C2b-F1067 males and coanalyzed 220 Y-chromosome sequences of this haplogroup. BEAST software was used to reconstruct a revised phylogenetic tree of haplogroup C2b-F1067 with age estimates. The revised phylogeny of C2b-F1067 included 155 sublineages, 1986 non-private variants, and >6000 private variants. The age estimation suggested that the initial splitting of C2b-F1067 happened at about 32.8 thousand years ago (kya) and the major sublineages of this haplogroup experienced continuous expansion in the most recent 10,000 years. We identified numerous sublineages that were nearly specific for Korean, Mongolian, Chinese, and other ethnic minorities in China. In particular, we evaluated the candidate-specific lineage for the Dayan Khan family and the Confucius family, the descendants of the ruling family of the Chinese Shang dynasty. These findings suggest that ancient populations with varied C2b-F1067 sublineages played an important role during the formation of most modern populations in Eastern Eurasia, and thus eventually became the founding paternal lineages of these populations.

## Introduction

Distribution of the paternal Y-chromosome haplogroup C-M130 in global populations is considered strong

---

These authors contributed equally: Qiao Wu, Hui-Zhen Cheng

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s10038-020-0775-1>) contains supplementary material, which is available to authorized users.

✉ Lan-Hai Wei  
Ryan.lh.wei@gmail.com

<sup>1</sup> State Key Laboratory of Genetic Engineering, School of Life Science, Fudan University, 2005 Songhu Road, 200438 Shanghai, China

<sup>2</sup> Department of Anthropology and Ethnology, Institute of Anthropology, Xiamen University, 361005 Xiamen, China

<sup>3</sup> College of Foreign Languages, Huaqiao University, 362021 Quanzhou, China

<sup>4</sup> School of Life Sciences, Jilin University, 130012 Changchun, China

<sup>5</sup> Key Laboratory of Evidence Science of Gansu Province, Gansu

supporting evidence for the initial coastal settlement of modern humans from the Middle East to South Asia, East Asia, and Australia [1]. Specific sublineages of C-M130 are present among populations from different geographic regions, such as C1a1-M8 in Japan, C1a2-V20 in Europe, C1b1a1-M356 in South Asia, C1b1a2-B65 in South China and Southeast Asia, C1b3a-M38 in Southeast Asia and Islands, C1b3b-M347 in Australia, and C2-M217 in East Asia, Northern Asia, and America. Recent studies indicate that these local sublineages have been separated for more

University of Political Science and Law, Lanzhou, China

<sup>6</sup> Laboratory for Human Biology and Human Genetics, Institute of Ethnology and Anthropology, School of Ethnology and Anthropology, Inner Mongolia Normal University, 010022 Hohhot, China

<sup>7</sup> School of History and Ethnic Culture, Hulunbuir University, 021008 Hulunbuir, China

<sup>8</sup> National Center for Biotechnology, Astana 010000, Kazakhstan

<sup>9</sup> B&R International Joint Laboratory for Eurasian Anthropology, Fudan University, 20043 Shanghai, China

<sup>10</sup> Human Phenome Institute, Fudan University, Shanghai, China

than 45,000 years [2]. Haplogroup C2-M217 is present at the highest frequency among populations in East Asia and Northern Asia (Zhong et al., 2010).

Prior studies evaluated the histories of some sublineages of haplogroup C2-M217. Most of the C2-M217 samples from Northern Asia belong to C2a-F1396, referred to as the “Northern branch of C2-M217” [3–6]. Most of the C2-M217 samples from East Asia belong to C2b-F1067, referred to as the “Southern branch of C2-M217” [7, 8]. However, haplogroup C2b1a1a1a1-F3850, a sublineage of C2b-F1067 and C2b1a1a1-M407, is one of the founding paternal lineages of Mongolic-speaking populations [9]. Further, numerous samples are downstream of C2b-F1067, such as those from the Human Genome Diversity Project [10] and the 1000 Genomes Project (Supplementary Data S1) [11].

Several studies revealed the origin and internal phylogeny of C2a-F1396 sublineages. Haplogroup C2a1a3-M504 is one of the founding paternal lineages of Mongolic-speaking populations [6]. Haplogroup C2a1a1a2a-F1756 is the dominant lineage of the ancient Donghu and Xianbei tribes [3, 12]. C2a1a2-M48 is also dominant paternal lineage in Tungusic-speaking populations [1, 5]. In general, many researches have been done to reveal the demographic history of C2a-F1396. However, the internal phylogenetic structure of haplogroup C2b-F1067, the other major sub-branches of C2-M217, remains unclear. Furthermore, the origin, diversification, and expansion of this paternal lineage are currently uncertain.

In the present study, we generated a large number of C2b-F1067 male sample sequences. The objectives were (1) to reconstruct a phylogeny with age estimation of this haplogroup, and (2) to explore the origin, diversification, and expansion of this paternal lineage. Specifically, we sought to investigate the contributions of C2b-F1067 sublineages to the formation of modern populations in Eastern Eurasia. In this study, the term ‘East Asia’ refers to China, North Korea, South Korea, Mongolia, and Japan. The term ‘Southwest China’ refers to those provinces that lie to the west or the south of Hubei Province. The term ‘Northern Asia’ refers to the region of Asia that lies to the north of the Yangtze River. The term ‘Eastern Eurasia’ refers to East Asia, Southeast Asia, and eastern part of Siberia.

## Materials and methods

### Materials

Blood or saliva samples were collected from unrelated healthy males in East Asian populations (Supplementary Table S1). All participants provided signed consent prior to participating in the study. Relatedness was checked

according to the information of signed consent and biologically relevant samples were excluded from this study. The ethics committee for biological research at the Fudan School of Life Sciences approved the study. A series of Y-SNP markers were genotyped to identify C2b-F1067 individuals, including M130, P54, M105, M48, M208, M407, P33, M93, P39, P92, P53.1, M217, M38, M210, M356, P55, M347, M86, and F1067. DNA extracted from 206 C2b-F1067 samples was sent for next-generation sequencing using the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA). A series of bait libraries were designed to capture the sequences of an ~11 Mbp region on the Y chromosome. DNA shearing, adaptor addition, and gel electrophoresis prior to next-generation sequencing were performed as described previously [8]. The raw sequence data reported in this study were deposited in the Genome Sequence Archive [13] at the BIG Data Center, Beijing Institute of Genomics (BIG) [14], Chinese Academy of Sciences, under accession number HRA000106, and are publicly accessible at <http://bigd.big.ac.cn/gsa>.

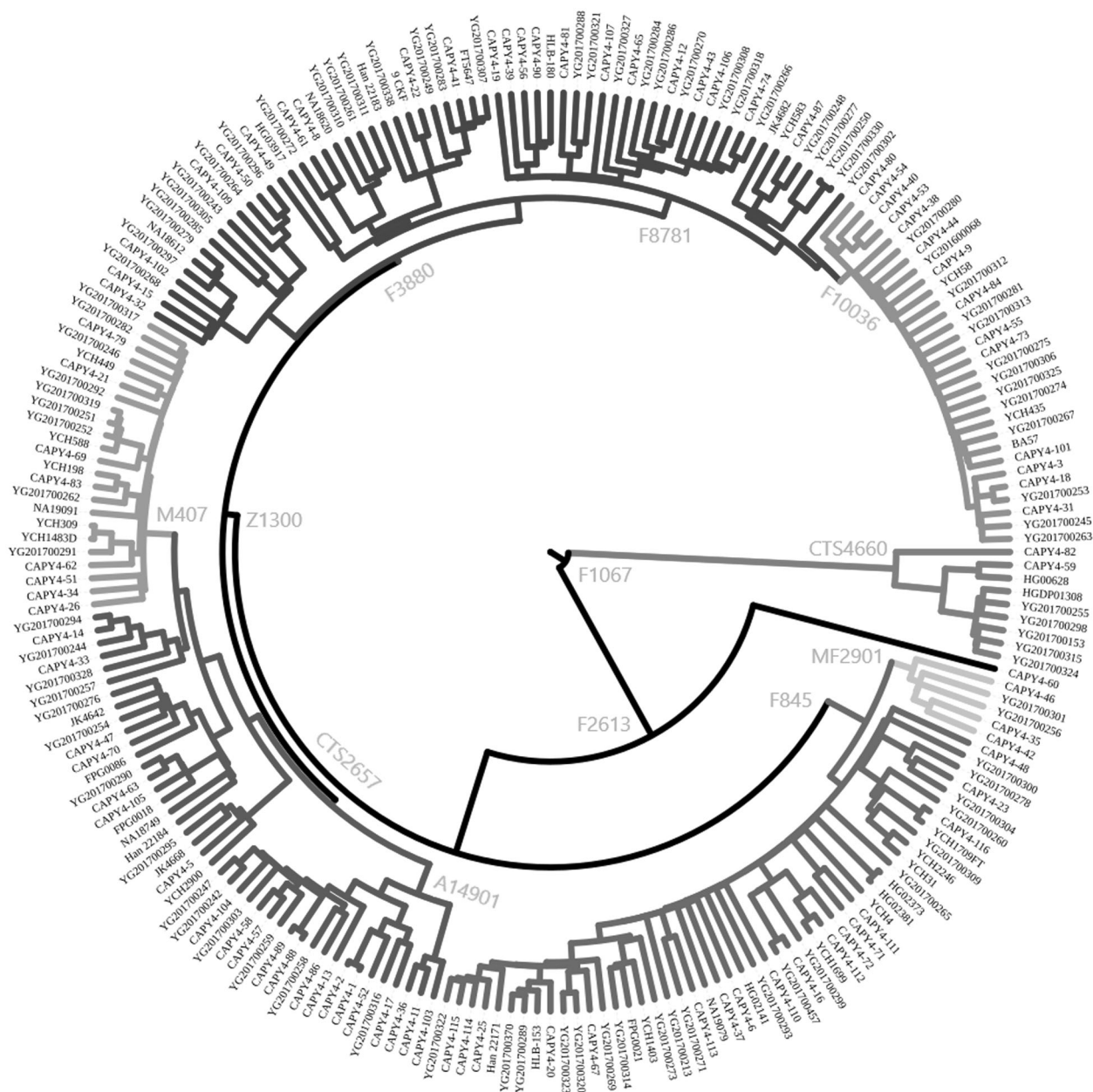
### Methods

Read mapping and SNP calling of the next-generation sequencing data were conducted with the human reference hg38 and standard procedures (BWA and SAMtools) [15, 16]. Another 14 previously published C2b-F1067 sequences were also used to construct the phylogeny of this haplogroup, including three sequences from Yan et al. [8], ten sequences from the 1000genome project [11], and one sequence from the SGDP project [17] (Supplementary Table S1). Regulations proposed by the YCC were used to revise the phylogenetic tree with respect to new variants in the non-recombining region of the Y chromosome [18].

Bayesian evolutionary analyses were conducted using BEAST software (v2.4.3) [19]. To calculate the splitting time of the phylogenetic tree, we applied the point mutation rate of  $0.74 \times 10^{-9}$  per site per year, which was inferred from the genome of the Anzick-1 Boy [20]. The calculation was performed with 10 million iterations and sampling every 1000 steps. Results were visualized in Tracer v.1.6 [21] and FigTree v1.4.2 [22], with a burn-in of 20% and over 200 effective sample sizes.

### Results

A schematic phylogenetic tree and the locations of the studied samples are shown in Fig. 1, Fig. 2, Supplementary Table S1, and Supplementary Table S2. Detailed sample information and a detailed phylogenetic tree with markers of important clades and age estimate were provided in



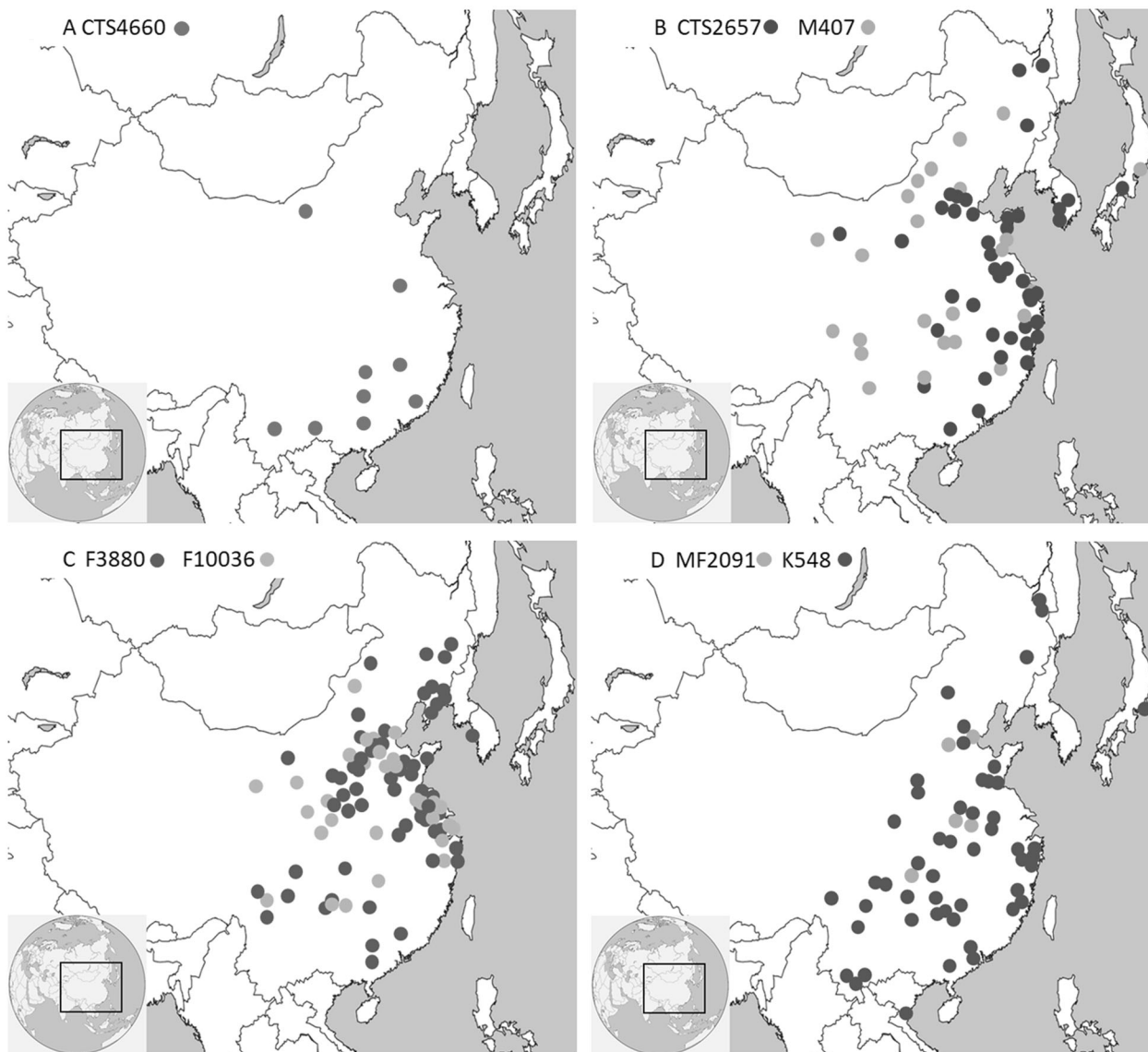
**Fig. 1** Brief phylogenetic tree of haplogroup C2b-F1067

Supplementary Table S1 and Fig. S1, respectively. The revised phylogenetic tree of haplogroup C2b-F1067 contained 154 sublineages, 1986 non-private variants, and >6000 private variants (Fig. S1). The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) [23] in Big Data Center [14], Beijing Institute of Genomics (BIG), Chinese Academy of Science, under accession numbers GVM000050 that are publicly accessible at <https://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000050>.

As shown in Fig. 1 and Fig. S1, C2b-F1067 exhibited a number of sublineages, and the initial split time between

and C2b-F2613/Z1338 and C2b-CTS4660 was ~32.77 kya (95% confidence interval [CI] = 31.94–36.62 kya). After this split time, we observed only one branching event for C2b-F2613/Z1338 and C2b-CTS4660 before their expansion after 11 kya. Most C2b-CTS4660-F10027 samples come from South China or Southeast Asia. However, the age of this lineage is young. The C2b-Z1318 sublineages are comprised of three sub-branches, including C2b-CTS2657, C2b-F3880, and C2b-F845.

Among the 50 sub-branches of C2b-CTS2657, lineage C2b-F845 is specific to the Mongolic-speaking population, and the most recent common ancestor of this lineage is very



**Fig. 2** Geographic distribution of samples of sub-branches of C2b-F1067

young (0.86 kya). Most samples from the other sub-branches of C2b-CTS2657 come from Korean and Han populations. Previous studies found high frequencies (~15%) of C2b-CTS2657 in most Korean samples from different locations [7]. Three samples of Korean in the present study revealed the sub-branches of C2b-CTS2657 in Korean. Furthermore, samples of M407 were found randomly throughout China. Two samples of Mongolian and one Manchu individual formed a special sub-branch of M407 (F8536). The continuous expansion pattern of C2b-CTS2657 after 6 kya suggested that ancient populations with this paternal lineage may have once lived in a vast geographic region and subsequently became part of modern Korean, Han, and Mongolic-speaking populations in a later historical age.

Haplogroup C2b-F3880 was comprised of 61 sub-lineages, most of which were samples from Han populations. However, C2b-F3880 samples were also randomly found in other populations in East Asia. Interestingly, three such samples came from varied populations of South Asia. These samples can be explained by the influence of Tibeto–Burman populations. Most importantly, continuous expansion and rapid differentiation were observed in the phylogeny of C2b-F10036 sublineages between 3.6 kya and 2.2 kya. This process led to six layers of splitting structure, which were only defined by one or two SNPs.

In total, 33 sublineages were present within lineage C2b-F845. Lineage C2b-F845 experienced significant expansion approximately 6.5 kya, and most C2b-F845 sublineages appeared after this expansion. According to the background

information of F845-positive individuals, Southwest China is likely the diffusion center of this lineage. Accordingly, the surnames of many individuals are originally connected to Tujia populations or their ancient relatives.

## Discussion

In this study, we found more than one hundred sub-branches of C2b-F1067, which is widely distributed in East Eurasia. The current available ancient DNA is limited and only several major show clear connections with modern populations. Therefore, we are only able to discuss briefly about the demographic history of haplogroup C2b-F1067 and explore the possible role of five major sub-branches during the formation of modern East Eurasian, including C2b-F10036, C2b-F8465, C2b-CTS2657, and C2b-F845.

### Holocene differentiation of C2b-F1067

Previous studies suggested that haplogroup C2a-F1396, the most closely related lineage of lineage C2b-F1067, originated in Northern Asia [5, 24]. The revised phylogeny and age estimations in the present study were suggestive of different diffusion centers in East Asia for different sub-branches of C2b-F1067. For example, C2b-CTS2657 originated in Northeast China, C2b-F3880 originated in Northern China, and C2b-F845 originated in Southwest China. We propose that Northern China may be the distribution center of C2b-F1067 prior to 11 kya.

Notably, we observed only one branching event for C2b-F2613/Z1338 and C2b-CTS4660, two major sub-branches of C2b-F1067, before their expansion in the Holocene. The period between the first splitting event at 32.8 kya and the initial expansion after 11 kya may correspond to the period of constant population size with hunter-gatherer lifestyle of ancient populations with haplogroup C2b-F1067 in Paleolithic Age. The initial expansion time of haplogroup C2b-F1067, ~11 kya, is consistent with the early domestication period of foxtail millet (*Setaria italica*) in North China [25]. The continuous expansion of C2b-F1067 since the Holocene is also consistent with the flourishing of Neolithic agriculture in Northern China [26]. In the modern age, most sublineages of C2b-F1067 are part of the paternal gene pool of East Asian populations with agricultural traditions, especially Han populations. These patterns indicate that haplogroup C2b-F1067 may be one of major paternal lineages in ancient populations who contributed to the domestication of foxtail millet. Therefore, we suggested that the Paleolithic hunter-gatherer populations with paternal lineage C2b-F1067 in Northern China may have been involved in the formation of Neolithic communities in this region, such that C2b-F1067 sublineages ultimately become

part of the paternal gene pool of modern East Asian populations.

### Founding paternal lineages of Eastern Eurasia

According to the phylogeny and detailed sample information in the present study (Table S1), we were able to identify a specific sublineage for different populations in Eastern Eurasia. As we discussed in previous sections, the C2b-F8465 is one of founding paternal lineages of Mongolic-speaking populations. The source of this paternal lineage may be ancient populations in the central part of Northeast China present 3.0 kya [9]. Furthermore, numerous specific C2b-CTS2657 sublineages were identified among Korean samples. Previous studies revealed that haplogroup C2b-CTS2657 contributed a large portion of the Korean paternal gene pool [7]. The sources of this haplogroup in Korea might have been ancient populations in the southern part of Northeast China between the Late Neolithic Age and the Bronze Age, ~6.0–3.0 kya. In general, we propose that the sub-branches of C2b-CTS2657 are founding paternal lineages of modern Koreans.

The large number of C2b-F1067 sub-branches in Han populations is one of the most important findings of the present study. According to the phylogenetic data, we suggest that numerous sub-branches of C2b-CTS2657, C2b-F3880, and C2b-F845 are founding paternal lineages of the modern Han population. Importantly, sublineage C2b-F3797 contains samples from Han and Tibeto-Burman populations. Therefore, lineage C2b-F3797 may be one of the paternal lineages in the common ancestors of Sino-Tibetan populations.

### C2b-F10036 as a candidate for the Confucius family Y-profile

Confucius (551–479 BC, one of the greatest philosophers of ancient China, was a direct descendant of the Royal family of the Chinese Shang dynasty (~1600–1027 BC [27]). In a previous study of our team, 1118 males of Kong Family from Qufu city were tested and 515 individuals (46.06%) belong to haplogroup C2-M217 [28]. Importantly, the Y-STR values of most C2-M217 individuals are close to each other. In this study, numerous C2-M217 individuals of the Kong Clan from Shandong and other provinces were sent for sequencing; include the direct descendant of clan leaders of Kong family since one thousand years ago. In the present study, numerous individuals of the Kong Clan from Shandong province and other provinces were included. These individuals are recorded as the descendants of Confucius [29]. The specific lineage of the Kong Clan (C2b-F10036–F16605) continuously expanded and formed six layers of splitting structures between 3.6 and 2.2 kya. This

interval fits well with the Shang dynasty period (~1600–1027 BC) and its successor, the Song Kingdom (~1025–286 BC [30]). Thus, we propose that lineage C2b-F10036–F16605 could be considered a candidate paternal profile of the royal house of the Shang dynasty, the ruling family of Song Kingdom, and the Confucius family. However, lineage Q-M120 also contributed with high frequency to the Kong Clan from Shandong [28]. Ancient DNA evidence is needed to validate or refute this hypothesis.

### C2b-F8465 in Mongolians and the Y-profile of Dayan Khan

In a previous study, we revealed that haplogroup C2b-F8465 is a sub-clade of M407, and one of the founding lineages of Mongolic-speaking populations [9]. We proposed a recent origin of haplogroup C2b-F8465 in the southern part of Northeast China. On the other hand, previous studies also indicated that the Y-profiles of descendants of Dayan Khan, the famous Khan of 16<sup>th</sup>-century Mongolia, formed a special Y-STR cluster of M407 [31]. In the present study, we sequenced several samples with similar Y-STR profiles and identified a special sub-branch of M407 and F8465 (F8536). Thus, we propose that F8536 may be the Y-chromosome lineage of the Dayan Khan family, and that this lineage is a sub-branch of M407, which is a dominant paternal lineage of Oirat populations [5, 32, 33]. How haplogroup F8465 became important paternal lineages of modern Mongolic-speaking populations remains unclear.

Here, we suggest that ancient populations with dominant C2b-M407 lineage might have been indigenous peoples in the southern part of Northeast China before they were assimilated as part of the proto-Mongols tribes between 4.0 and 2.0 kya. Sub-branches of C2b-F1067 were mainly distributed in East Asia [1], and M407 is the only sublineage that reached high frequencies in Northern Asian populations [9, 33]. From the perspective of archaeology, the Xiaolaha Culture (~4.0–3.4 kya) and the Baijingbao Culture (~3.4–2.5 kya) are successors of the local Neolithic culture in the southern part of Northeast China [34, 35]. However, their successors, such as the Hanshu II Culture (~2.5–2.0 kya) and the Hongmashan culture (~2.0–1.8 kya), show increasing similarity with other remains of the Donghu and Xianbei [35, 36]. The territory of these ancient archaeological cultures was showed in Fig. S2. In later historical periods, the southern and western parts of Northeast China were recorded as territories of the Xian-Bei and Shi-Wei tribes, the direct ancestors of modern Mongolic-speaking populations [37]. In general, we propose that the archaeological transition in the southern part of Northeast China between 4.0 and 1.5 kya may correspond with the

appearance of M407 sub-clades in proto-Mongols. Ancient DNA evidence is required to validate or refute this hypothesis.

The demographic processes of ancient populations with dominant C2b-F845 paternal lineage are more complicated. The initial expansion of this lineage may be related to the diffusion of ancient Tibeto–Burman populations in Southwest China at approximately 6 kya [38]. However, large numbers of this lineage are also present in other populations of Southwest China, which are not speakers of Tibeto–Burman languages [39]. This can be explained by intense admixture in this region in the past 6000 years. Detailed information from F845 samples indicated that this haplogroup may be the paternal profile of many major families, clans, and Tusi houses in Southwest China (Table S1 and unpublished data of surname of studied individuals). The long-term regional ruling of these families in Southwest China [38] might have contributed markedly to the predominance of C2b-F845 in Southwest China.

In conclusion, a large number of sequences were used to construct a highly revised phylogenetic tree of paternal lineage C2b-F1067, which is one of the dominant lineages of Eastern Eurasian populations. We identified more than 150 sublineages of C2b-F1067 and >8000 variants under this lineage. The age estimation in the study allowed us to explore the possible origin and diffusion of major sublineages of this haplogroup. We propose a series of founding paternal lineages for modern populations of Eastern Eurasia. The sublineages and variants identified in the present study will contribute to future exploration of the demographic history of ancient populations and modern ethnic groups in Eastern Eurasia.

**Acknowledgements** We thank all donors for providing DNA samples and/or DNA sequences. This work was supported by the National Natural Science Foundation of China (31900406 to LHW). LHW was also supported Fundamental Research Funds for the Central Universities (20720191047). This study was also supported by the Scientific and Technology Committee of Shanghai Municipality (18490750300).

**Author contributions** The authors are responsible for the content and writing of the paper.

### 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. Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, et al. Global distribution of Y-chromosome haplogroup C reveals the

- prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet.* 2010;55:428–35.
2. 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.
  3. 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.
  4. Wei LH, Wang LX, Wen SQ, Yan S, Canada R, Gurianov V, et al. Paternal origin of Paleo-Indians in Siberia: insights from Y-chromosome sequences. *Eur J Hum Genet.* 2018;26:1687–96.
  5. 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.
  6. 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.
  7. Kwon SY, Lee HY, Lee EY, Yang WI, Shin KJ. Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups. *Forensic science international. Genetics.* 2015;19:42–6.
  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.
  9. 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 Genom.* 2018;293:657–63.
  10. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science.* 2020;367. <https://science.sciencemag.org/content/367/6484/eaay5012.full>.
  11. Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet.* 2016;48:593–9.
  12. Zhang Y, Wu X, Li J, Li H, Zhao Y, Zhou H. 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.* 2018;63:231–8.
  13. Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. *Genom Proteom Bioinforma.* 2017;15:14–8.
  14. Members BIGDC. The BIG Data Center: from deposition to integration to translation. *Nucleic acids Res.* 2017;45:D18–24.
  15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
  16. 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.
  17. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature.* 2016;538:201–6.
  18. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 2002;12:339–48.
  19. 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.
  20. Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature.* 2014;506:225–9.
  21. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67:901–04.
  22. Rambaut A. <http://tree.bio.ed.ac.uk/software/figtree/> (2018). Accessed 1 Feb 2018.
  23. Song S, Tian D, Li C, Tang B, Dong L, Xiao J, et al. Genome Variation Map: a data repository of genome variations in BIG Data Center. *Nucleic acids Res.* 2018;46:D944–9.
  24. Pinotti T, Bergstrom A, Geppert M, Bawn M, Ohasi D, Shi W, et al. Y chromosome sequences reveal a short beringian standstill, rapid expansion, and early population structure of Native American Founders. *Curr Biol.* 2019;29:149–57.e3.
  25. Yang X, Wan Z, Perry L, Lu H, Wang Q, Zhao C, et al. Early millet use in northern China. *Proc Natl Acad Sci USA.* 2012;109:3726–30.
  26. Tong ZC. *Studies on Neolithic Age of China.* Chengdu: Bashu Publishing House; 1998.
  27. Lowe M, Shaughnessy EL. *The Cambridge History of Ancient China: From the Origins of Civilization to 22 B.C.* England: Cambridge University Press; 1999.
  28. Hou W-G, Wang C-C, Jiang S-H, Liu H-D, Li H. Genetic diversity of seventeen Y-STR loci among the people with the surname Kong from Qufu Prefecture. *Acta Anthropologica Sin.* 2015;35:125–31.
  29. Kaizuka S. *Confucius.* Mineola: Dover Publications; 2002.
  30. Lowe M, Shaughnessy EL. *The Cambridge History of Ancient China: From the Origins of Civilization to 221 BC.* Cambridge: Cambridge University Press; 1999.
  31. Batbayar K, Sabitov ZM. The Genetic Origin of the Turko-Mongols and Review of The Genetic Legacy of the Mongols. Part 1: Y-chromosomal Lineages Chinggis Khan Russian *J Genet Geneal.* 2012;4:1–8.
  32. Balinova N, Post H, Kushniarevich A, Flores R, Karmin M, Sahakyan H, et al. 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.* 2019;27:1466–74.
  33. Malyarchuk BA, Derenko M, Denisova G, Wozniak M, Rogalla U, Dambueva I, et al. Y chromosome haplotype diversity in Mongolic-speaking populations and gene conversion at the duplicated STR DYS385a,b in haplogroup C3-M407. *J Hum Genet.* 2016;61:491–6.
  34. Zhao BF. *Archaeology of the stone age in Northeast China.* Changchuan: Jilin University Press; 2003.
  35. Zhao BF. *Study on Archaeological Culture of Xia Dynasty to Warring States in Northeast Region of China.* Beijing: Science Press; 2009.
  36. Zhang W. *Study on the Hongmashan Culture (in Chinese).* Northern Cultural Relics; 2007.
  37. Lin G. *A history of Donghu (in Chinese).* Hohhot: Inner Mongolian People's Publishing House; 2007.
  38. Wang CY, Luo W. *A brief history of Tusi families of Tujia population (in Chinese).* Beijing: Minzu University of China Press; 1991.
  39. Xia Z-Y, Yan S, Wang C-C, Zheng H-X, Zhang F, Liu Y-C, et al. Inland-coastal bifurcation of southern East Asians revealed by Hmong-Mien genomic history. 2019. <https://www.biorxiv.org/content/10.1101/730903v1>.