Molecular Genetics and Genomics
https://doi.org/10.1007/s00438-017-1403-4

ORIGINAL ARTICLE

Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407

Yun-Zhi Huang[1] · Lan-Hai Wei[1,2] · Shi Yan[1] · Shao-Qing Wen[1] · Chuan-Chao Wang[3] · Ya-Jun Yang[1] · Ling-Xiang Wang[1] · Yan Lu[4] · Chao Zhang[4,5] · Shu-Hua Xu[1,4,5,6] · Da-Li Yao[7] · Li Jin[1] · Hui Li[1]

Received: 4 September 2017 / Accepted: 4 December 2017

Abstract

The Y-chromosome haplogroup C2c1a1a1-M407 is a predominant paternal lineage in Mongolic-speaking populations, especially in Buryats and Kalmyks. However, the origin and internal phylogeny of C2c1a1a1-M407 have not been investigated in detail. In this study, we analyzed twenty-three Y-chromosome sequences of haplogroup C2c1a1a1-M407 and its most closely related clades. We generated a high-resolution phylogenetic tree of haplogroup C2c1a1a1-M407 and its upstream clade C2c1a1-CTS2657, including 32 subclades and 144 non-private Y-chromosome polymorphisms. We discover that all available C2c1a1a1-M407 samples from Mongolic-speaking populations belong to its newly defined downstream clade C2c1a1a1b-F8465, whereas all samples of C2c1a1-CTS2657(xF8465) come from northern Han Chinese, Korean, and Japanese. Furthermore, we observe that C2c1a1a1b-F8465 and its subclade C2c1a1a1b1-F8536 expanded at approximately 0.86 and 0.44 thousand years ago, respectively. Therefore, we conclude that C2c1a1a1-M407 in Mongolic-speaking populations has originated from northeastern Asia. C2c1a1a1b1-F8536, the newly defined subclade of C2c1a1a1-M407, probably represents the genetic relationships between ancient Oyrats, modern Kalmyks, Mongolians, and Buryats.

Keywords C2c1a1a1-M407 · Mongolic-speaking · Han Chinese · Y-chromosome sequence

Introduction

The Y-chromosome single nucleotide polymorphism (Y-SNP) marker M407 was first reported by Sengupta et al. 2006. Previous studies have regarded haplogroup C2c1a1a1-M407 as a predominant paternal lineage in Mongolic-speaking populations, such as Buryats, Kalmyks, and Mongolians in northeastern Asia. However, haplogroup C2c1a1a1-M407 was rarely seen in other Eurasian populations, including Han Chinese from China (Zerjal et al. 2002; Malyarchuk et al. 2010, 2013; Zhong et al. 2010). Additionally,

Communicated by S. Hohmann.

Yun-Zhi Huang and Lan-Hai Wei contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-017-1403-4) 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 of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China

[2] Institut National des Langues et Civilisations Orientales, 75013 Paris, France

[3] Department of Anthropology and Ethnology, Xiamen University, Xiamen 361005, China

[4] 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 200031, China

[5] University of Chinese Academy of Sciences, Beijing 100049, China

[6] School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

[7] Center for Historical Geographical Studies of Fudan University, Shanghai 200433, China

Published online: 29 December 2017

C2c1a1a1-M407 has a special Y-chromosome short tandem repeat (Y-STR) cluster, characterized by allelic combination 11/11 at the DYS385a/b loci (Malyarchuk et al. 2010, 2016). C2c1a1a1-M407 has also been determined to be the Y-haplotype of Batu-Möngke Dayan Khan (1474–1517 AD, the ruler of North Yuan Khanate), who was a direct descendant of Genghis Khan (Batbayar and Sabitov 2012).

Previous studies suggested that the distribution of C2c1a1a1-M407 in North Asia resulted from the dispersal of Mongolic-speaking populations, and the expansion time with Y-STR was about 1.95 ± 1.26 ky (Malyarchuk et al. 2010, 2013; Zhong et al. 2010). Moreover, an analysis of full Y-chromosome sequences indicated that all of C2c1a1a1-M407 samples in Buryats belonged to a relatively young subclade (~ 1 kya, Karmin et al. 2015). All the above results suggested that C2c1a1a1-M407 had experienced a rapid expansion in Mongolic-speaking populations. However, previous studies indicated that all subclades of C2c-F1067 appeared in northeastern Asia, like northern China, Korea and Japan, excepting C2c1a1a1-M407 (Lippold et al. 2014; Yan et al. 2014; Karmin et al. 2015; Poznik et al. 2016; Mallick et al. 2016). Therefore, we would like to investigate whether C2c1a1a1-M407 also originated from northeastern Asia, how C2c1a1a1-M407 migrated to North Asia, and how C2c1a1a1-M407 successfully expanded in Mongolic-speaking populations.

In this study, we analyzed the whole Y-chromosome sequences and Y-STR data of C2c1a1a1-M407 and its most closely related clades from a broad geographical scale, and constructed a phylogeographic distribution and a revised phylogenetic tree. We also connected C2c1a1a1-M407 and its most closely related clades with Buryats, Mongolians, ancient Oyrats, and modern Kalmyks. The investigation on haplogroup C2c1a1a1-M407 and its upstream clade C2a1a1-CTS2657 will be helpful to explore the origin and diffusion process of Mongolic-speaking and other populations in Northeastern Asia.

Methods

Blood or saliva samples were collected from unrelated healthy males in East Eurasia. All individuals signed their informed consent forms before their participation. The ethics committee for biological research at the School of Life Sciences in Fudan University (Shanghai, China) approved the study.

Genomic DNA was extracted using the DP-318 Kit (Tiangen Biotechnology, Beijing, China) according to the manufacturer's protocol. The Y-chromosome marker M130 was genotyped to identify haplogroup C-M130 samples within 6348 samples from 74 populations (Table S1 "This study"). In total, 947 individuals with the derived allele at M130 were subjected to further typing of ten biallelic markers: C1a1-M8, C1b1a1-M356, C1b2a-M38, C2-M217, C2a-M93, C2c1a1a1-M407, C2b1a1a-P39, C2b1a2-M48, C1b2b-M347, and C-P53.1. Seventeen Y-short tandem repeat markers (Y-STR) (DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y-GATA H4, and DYS385a/b) were amplified using the AmpFlSTR® YFiler™ PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA). Amplification products were analyzed using the ABI 3730 and ABI 3130 Genetic Analyzers (Applied Biosystems). Electrophoresis results were analyzed using Genscan v. 3.7 and Genotyper v. 3.7 (Applied Biosystems).

The DNA of seven CTS2657+ samples was sent for next-generation sequencing on the Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA). We designed a series of bait libraries to capture the sequences of ~ 11 M region on the Y-chromosome. We used the procedure that we described previously for the other steps prior to next-generation sequencing, i.e., DNA shearing, adding an adaptor, and gel electrophoresis (Yan et al. 2014). Typical procedures were used for next-generation sequencing analyses and the determination of genotypes and haplogroups for each sample (Li et al. 2009; Li and Durbin 2010). The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (GSA) (Wang et al. 2017) in the BIG Data Center (Members BIGDC 2017), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under an accession number PRJCA000359 that is publicly accessible at http://bigd.big.ac.cn/gsa.

We followed the standard procedures (bwa + samtools) to analyze the next-generation sequencing results (Li and Durbin 2009; Li et al. 2009). We applied a series of strict filters on the original variants file, including: (1) restriction to 8.47 M bp confident region reported by Adamov et al. 2015, which is also a subset of region used by our laboratory, Poznik et al. 2016 and Karmin et al. 2015; (2) restriction to variants that are Y-SNP; (3) removal of all positions with call rate $< 90\%$ on all samples; (4) removal of position with heterozygosis call rate $> 5\%$ on all samples; (5) base coverage ≥ 3, base quality > 20, and distance between SNPs > 10 bp; (6) removal of all recurrent or triadic mutations even though they are confident on the currently known phylogenetic tree. Back imputation was also done on some Y-SNP markers according to the known phylogenetic tree, so as to infer missing genotype on some samples. A number of variants were determined after the analysis of Y-chromosome sequences. We followed the regulations proposed by the Y Chromosome Consortium (YCC) to revise the phylogenetic tree with respect to new variants in the non-recombining region of the Y chromosome (Y Chromosome Consortium 2002).

We used twenty-three Y-chromosome sequences to construct the phylogeny of haplogroup C2c1a1a1-M407. The detailed information of studied samples was listed in Supplementary Table S1. Six sequences came from this study, two from Yan et al. (2014), two from Lippold et al. (2014), nine from Karmin et al. (2015), three from 1000 Genomes Project (Y Chromosome Consortium 2002), and one from Lu et al. (2016).

We collected frequencies and Y-STR data for haplogroup C-M130 in 218 East Eurasian populations from previous studies (Supplementary Table S1 and S2). The data including 15 Y-STR polymorphisms (excluding DYS385a/b from 17 Y-STRs) were used to construct the reduced-median network by the program NETWORK 5.0.0.0 (Fluxus Engineering). The TMRCA of Y-STR data was estimated using average squared distances (ASD) (Zhivotovsky 2001; Ramakrishnan and Mountain 2004; Sengupta et al. 2006). The genealogical mutation rate cited from http://www.YHRD.org was utilized for time estimates.

Sixteen Y-chromosome sequences of C2c1a1a1-M407 and its most closely related clades were used for time estimates (ESM_2). The outgroup consisted of 21 samples from haplogroups C2-M48, C2*-Star Cluster, D-M174, NO* (Ust' Ishim, Fu et al. 2014), N-M231, and O1-M95. The number of variants was determined after the analysis of Y-chromosome sequences. The age of haplogroup CT-M168 (71,760 years, 95% confidence interval [CI] = 69,777–73,799 years) (Karmin et al. 2015) and the splitting time (470 ± 20 years ago) of samples (YCH508 and YCH1981) from the Aisin Gioro family (Wei et al. 2016) were used as calibration points for time estimates.

BEAST v.2.4.3 was employed to estimate the coalescent times for haplogroup C2c1a1a1-M407 and its most closely related clades (Bouckaert et al. 2014). A Bayesian skyline coalescent tree prior was selected with the bModelTest package in Beast 2.0 software to improve the substitution model. The calculation was 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

Since all available M407 samples and their close relatives have the derived state at marker CTS2657, we redefine haplogroup C2c1a1-CTS2657. The revised phylogenetic tree of haplogroup C2c1a1-CTS2657 contains 32 subclades, 144 non-private polymorphisms, and a number of private mutations (Fig. 2, Supplementary Table S3). The geographic locations of studied samples are in Supplementary Table S1, and their Y-STRs are in Supplementary Table S2. Detailed information for all Y-chromosome polymorphisms within haplogroup C2c1a1-CTS2657 is provided in Supplementary Table S4. Since all available C2c1a1a1-M407 samples from Mongolic-speaking populations belong to a new marker F8465, we define C2c1a1a1b-F8465 as a new downstream clade of C2c1a1a1-M407. We determine that two F8465 samples of special Y-STR cluster DYS385a/b = 11/11 belong to a new marker F8536 by sequencing results. Therefore, we define C2c1a1a1b1-F8536 as a new subclade of C2c1a1a1b-F8465.

The distribution of haplogroup C2c1a1a1-M407 and the sampling sites of haplogroup C2c1a1-CTS2657(xF8465) are shown in Fig. 1. Samples of haplogroup C2c1a1-CTS2657(xF8465) mainly come from northern and northeastern China, Korea, and Japan, which are indicated by red triangles on Fig. 1. Based on all available data, haplogroup C2c1a1-CTS2657(xF8465) is not present in Mongolia and Siberia. By contrast, the distribution center of C2c1a1a1-M407 (including all F8465 samples) is observed in the Transbaikal region of Siberia and northeastern Mongolia (Fig. 1).

The revised phylogenetic tree and distribution map reveal a clear phylogeographic pattern of haplogroup C2c1a1-CTS2657 (Figs. 1, 2). According to Fig. 2, all available C2c1a1a1-M407 samples from Mongolic-speaking populations belong to its downstream clade C2c1a1a1b-F8465. Both C2c1a1a1b-F8465 and its subclade C2c1a1a1b2-B97 give birth to a large number of subclades (Fig. 2), which represent a signal of successful expansion. Time estimates are also shown in Fig. 2. Haplogroup C2c-F1067 split from its upstream clade C2-M217 at around 35 kya. As a subclade of C2c-F1067, haplogroup C2c1a1-CTS2657 is separated from its most closely related clades at about seven kya. Haplogroup C2c1a1a1-M407 emerged at around five kya, and generated two subclades C2c1a1a1b-F8465 and C2c1a1a1b1-F8536 within one ky. We observe that C2c1a1a1b-F8465 has successful expansions in Mongolic-speaking populations from the Y-SNP phylogenetic tree.

The Y-STR network (Fig. 3) for haplogroup C2c1a1a1-M407 consists of C2c1a1a1-M407(xF8536) and its downstream clade C2c1a1a1b1-F8536 (equal to special Y-STR cluster DYS385a/b = 11/11 in this study). The samples of C2c1a1a1-M407(xF8536) from Mongolic-speaking populations (including Buryats, Mongolians, and Kalmyks) mainly have close distances from each other, whereas samples from non-Mongolic-speaking populations (indicated by the word "Others" and the light pink color) disperse at the end of branches. The samples of C2c1a1a1b1-F8536 show a star-like expansion pattern and locate near the samples of C2c1a1a1-M407(xF8536) from Mongolic-speaking populations. This scenario indicates that C2c1a1a1-M407 had recent expansions in Mongolic-speaking populations.

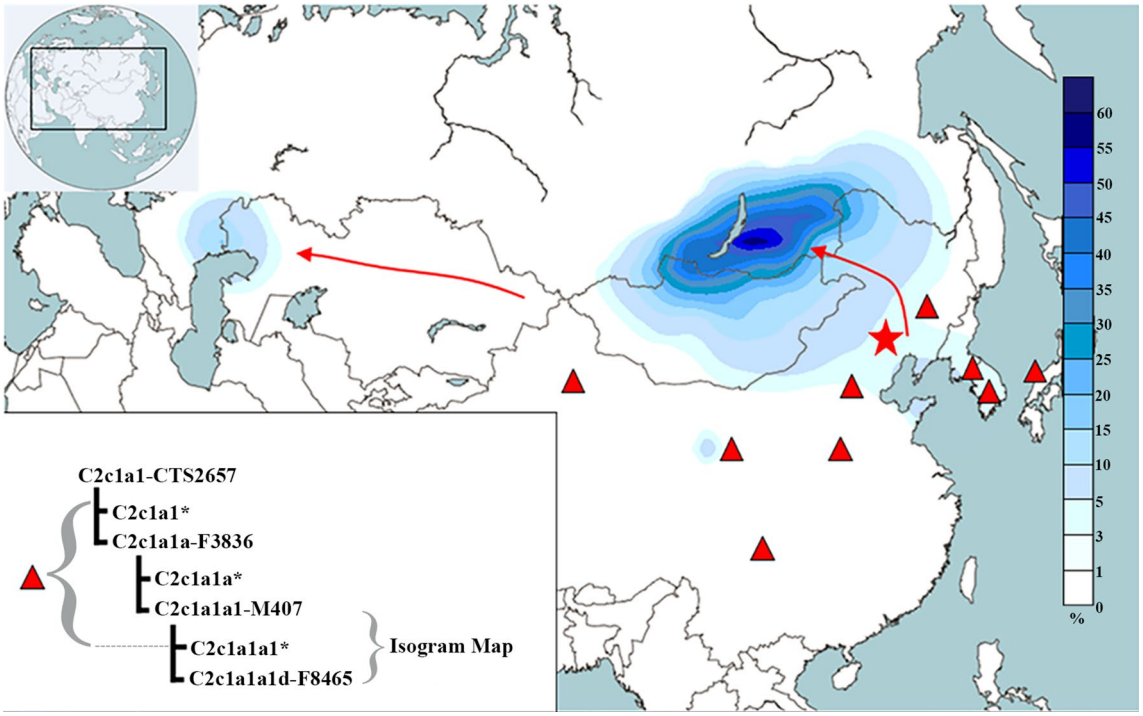

Fig. 1 Phylogeographic distributions of haplogroup C2c1a1a1-M407 and its closely related clades. We generated the maps by the Generic Mapping Tools 4.5.7 (GMT 4.5.7, https://www.soest.hawaii.edu/gmt/). We showed frequencies (%) of haplogroup C2c1a1a1-M407 by blue colour, and the red arrows indicated the migration routes. The red star represented the most likely origin of haplogroup C2c1a1a1-M407. The red triangles showed the locations of samples of haplogroup C2c1a1-CTS2657(xF8465), and the note in the bottom-left corner listed the specific subclades used in the map. (Color figure online)

Discussion

In this paper, we provide a revised phylogenetic tree of haplogroup C2c1a1a1-M407 with newly defined subclades C2c1a1a1b-F8465 and C2c1a1a1b1-F8536 (Fig. 2). Our result indicates that subclade C2c1a1a1b-F8465 is specific for Mongolic-speaking populations. The special Y-STR cluster DYS385a/b = 11, 11 of C2c1a1a1b-F8465 is defined as a new subclade C2c1a1a1b1-F8536, which was also widespread in Mongolic-speaking populations (Malyarchuk et al. 2010, 2016).

The geographic distribution of C2c1a1a1-M407 is dramatically different from that of C2c1a1-CTS2657x(F8465) (Fig. 1). C2c1a1a1-M407 (all belong to C2c1a1a1b-F8465 in this study) mainly locates in southern Siberia and Mongolia, while C2c1a1-CTS2657x(F8465) focuses on northeastern Asia, including northern and northeastern China, Korea and Japan (Fig. 2). The time estimates also show that C2c1a1a1-M407 underwent a rapid expansion in Mongolic-speaking populations at approximately 0.86 kya, while C2c1a1-CTS2657(xF8465) dispersed in northeastern Asia between 5.83 and 0.86 kya (Fig. 2). Therefore, we propose that C2c1a1a1-M407 in Mongolic-speaking populations originated from northeastern Asia.

Furthermore, the distribution of C2c1a1a1-M407 provides insight into the internal differentiation of Mongolic-speaking populations, because all C2c1a1a1-M407 samples belong to C2c1a1a1b-F8465 in this study. C2c1a1a1-M407 has high frequencies (> 50%, Supplementary Table S1) in northeastern Asian populations: Buryats (Mongolic-speaking), Sojots and Khamnigans (Turkic-speaking), who live near to Lake Baikal (Malyarchuk et al. 2010; Zhong et al. 2010). Besides, moderate frequencies (around 10%, Supplementary Table S1) of C2c1a1a1-M407 are also detected in Mongolian and Kalmyks (Mongolic-speaking). By contrast, the frequencies of C2c1a1a1-M407 in other Mongolic-speaking populations are relatively low or absent (Supplementary Table S1). However, previous studies showed that C2*-Star Cluster was the most predominant paternal lineage of all Mongolic-speaking populations, not C2c1a1a1-M407 (Derenko et al. 2007; Malyarchuk et al. 2010; Zhong et al. 2010; Di Cristofaro et al. 2013). Thus, our results suggest that Buryats assimilated most C2c1a1a1-M407 samples and brought this haplogroup to other Mongolic-speaking populations.

The expansion pattern of C2c1a1a1-M407 is consistent with the history of Oyrats to some degree. Previous studies showed that Kalmyks were the descendants of Oyrats, originating from Western Mongolia (Jungaria). The Oyrat

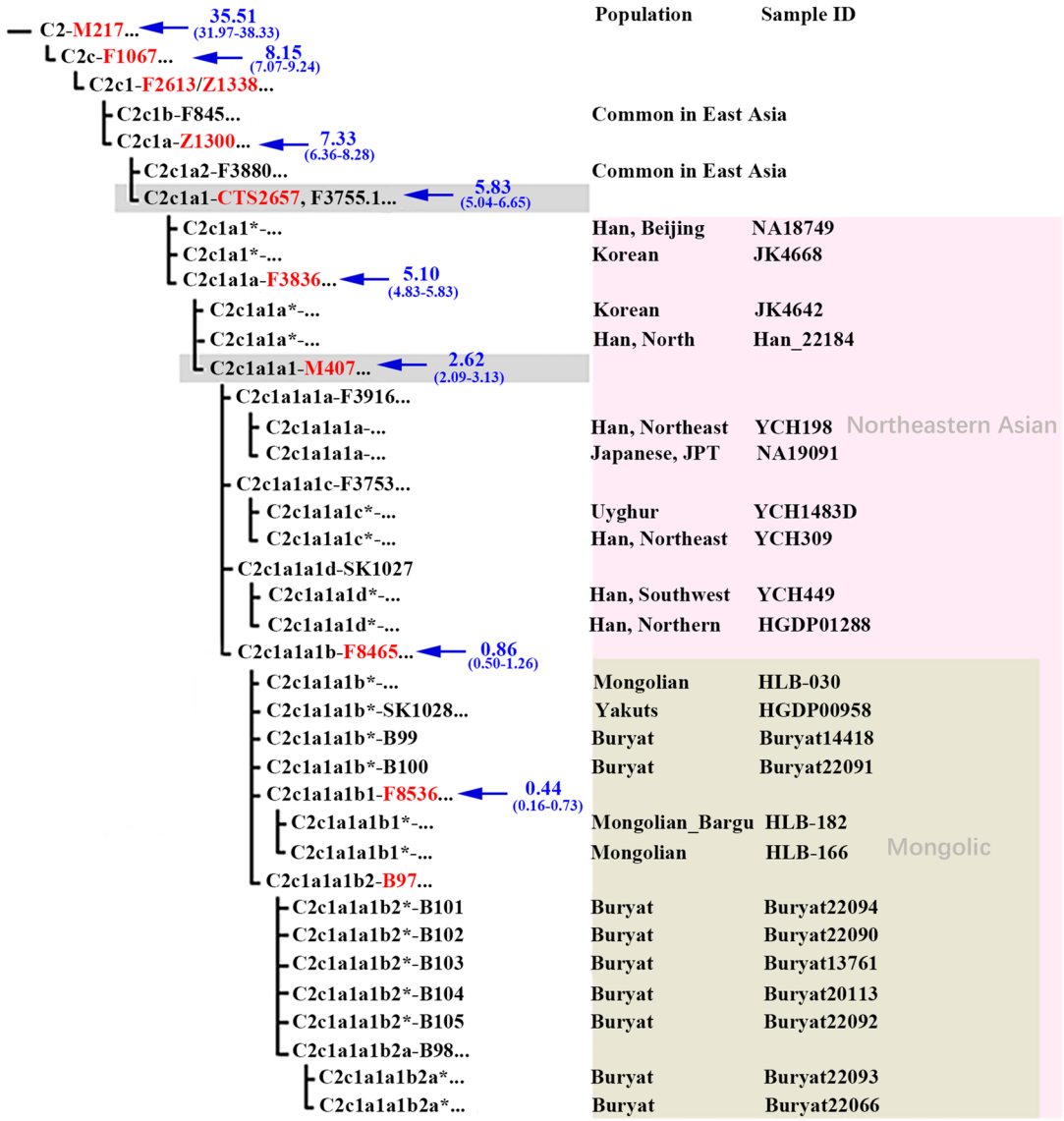

Fig. 2 Revised phylogenetic tree and time estimates for haplogroup C2c1a1a1-M407 and its upstream clade C2c1a1-CTS2657. The red and blue words represented the important subclades of haplogroup C2-M217 and their time estimates, respectively. The two studied subclades M407 and CTS2657 were shown by grey highlights. The light pink part indicated that most samples of M407 and CTS2657 come from northeastern Asia. The light green part showed that all samples of haplogroup C2a1a1a1b-F8465 come from Mongolic-speaking populations. (Color figure online)

confederation was established in the thirteenth century. During the late 16th and early seventeenth centuries, most Oyrats firstly migrated to the steppes of Western Siberia, and then relocated to the left bank of the Volga River. In the late seventeenth century, Oyrats formed the Kalmyk Khanate on the Lower Volga and became a new Mongolic-speaking ethnic group, Kalmyks (Nasidze et al. 2005; Malyarchuk et al. 2013). Our results show that C2c1a1a1-M407 is present in the Volga River (Fig. 1). Additionally, haplogroup C2c1a1a1-M407 has many samples in Kalmyks, especially in its subclade C2c1a1a1b1-F8536 (Fig. 3). The TMRCA of C2c1a1a1b1-F8536 is 0.44 kya (Fig. 2), which mainly corresponds with the history of Oyrats. Thus, we suggest that Oyrats had brought C2c1a1a1-M407 to western Eurasia and generated C2c1a1a1b1-F8536 in the modern Kalmyks in Russia. What is more, Mongolian and Buryats also have several F8536 samples (Fig. 3). Therefore, we propose that haplogroup C2c1a1a1b1-F8536 probably represents the genetic relationships between ancient Oyrat, and modern Kalmyk, Mongolian, and Buryat populations.

In conclusion, our studies provide a high-resolution phylogenetic tree and a clear phylogeographic pattern of haplogroup C2c1a1a1-M407. We propose that C2c1a1a1-M407 in Mongolic-speaking populations has originated

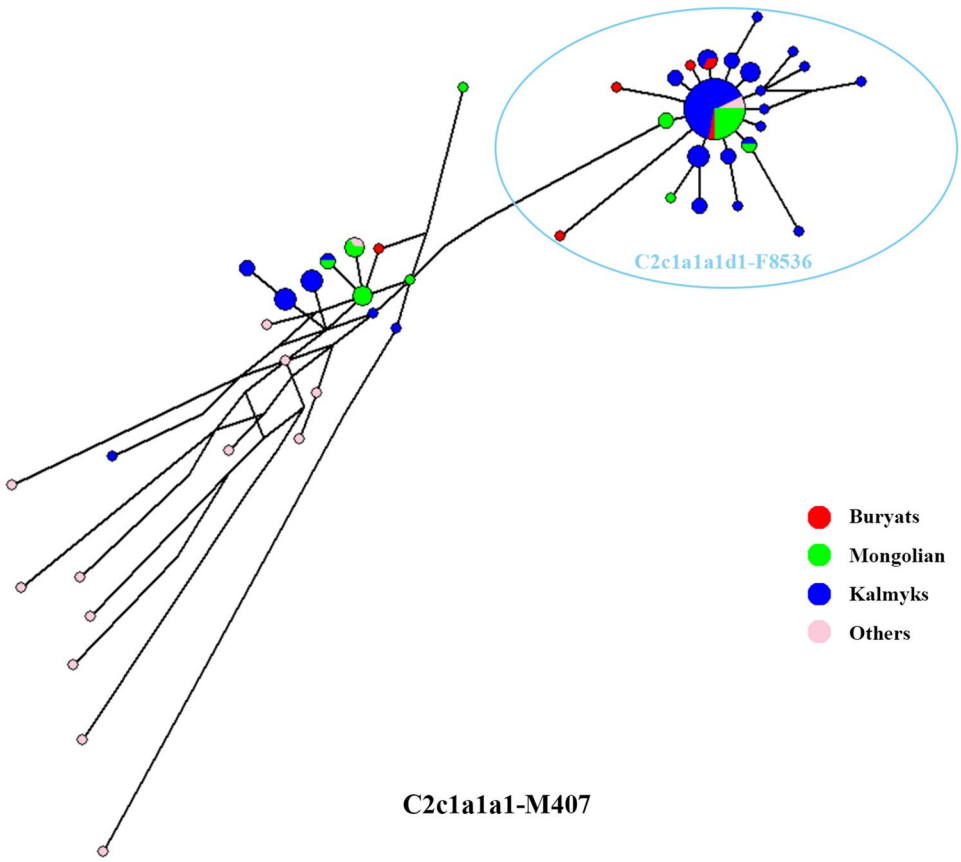

Fig. 3 Y-STR network for haplogroup C2c1a1a1-M407. The points on the network were colored according to the color code presented in the bottom-right color illustration. The light blue circle represented samples of haplogroup C2c1a1a1b1-F8536 (the downstream clade of C2c1a1a1-M407, equal to special Y-STR cluster DYS385a/b = 11/11 in this study). (Color figure online)

from northeastern Asia. Moreover, C2c1a1a1b1-F8536, the newly defined subclade of C2c1a1a1-M407, probably represents the genetic relationships between ancient Oyrats, modern Kalmyks, Mongolians, and Buryats. However, we need more ancient DNA, extensive Y-SNP genotyping and whole Y-chromosome sequencing to reveal the origin and migration of Mongolic-speaking populations in the future.

Acknowledgements This study was supported by the National Natural Science Foundation of China (31222030 and 31671297 to H.L., 31401060 to Y.S., 31271338 to L.J., 31501011 to Y.L., 91331204 to S.H.X.), MOE Scientific Research Project (113022A to H.L.), Shanghai Shuguang Project (14SG05 to H.L.), Ministry of Science and Technology of China (MOST) (2016YFC0900300 to H.L.) and Science and Technology Commission of Shanghai Municipality (STCSM) (14YF1406800 to Y.L.). L.H.W. was supported by Future Scientists Project of China Scholarship Council. C.C.W was supported by the Nanqiang Outstanding Young Talents Program of Xiamen University. We also thank all the sample donors and the laboratory assistants.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Research involving human participants and/or animals All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Adamov D, Guryanov V, Karzhavin S, Tagankin V, Urasin V (2015) Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. Russ J Genet Geneal 1:3–36

Batbayar K, Sabitov ZM (2012) 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 4:1–8

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Comput Biol 10:1–6. https://doi.org/10.1371/journal.pcbi.1003537

Derenko M, Maliarchuk B, Wozniak M, Denisova G, Dambueva I, Dorzhu C, Grzybowski T, Zakharov I (2007) Distribution of the male lineages of Genghis Khan's descendants in northern Eurasian populations. Genetika 43:422–426

Di Cristofaro J, Pennarun E, Mazieres S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA, Villems R, Chiaroni J (2013) Afghan Hindu Kush: Where Eurasian sub-continent gene flows converge. PLoS One. https://doi.org/10.1371/journal.pone.0076748

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, Meyer M, Zwyns N, Salazar-García DC, Kuzmin YV, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov NV et al (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514:445–449. https://doi.org/10.1038/nature13810

Karmin M, Saag L, Vicente M, Sayres MAW, Järve M, Talas UG, Rootsi S, Ilumäe A, Mägi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G et al (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res 25:459–466. https://doi.org/10.1101/gr.186684.114

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M (2014) Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5:13. https://doi.org/10.1186/2041-2223-5-13

Lu D, Lou H, Yuan K, Wang X, Wang Y, Zhang C, Lu Y, Yang X, Deng L, Zhou Y, Feng Q, Hu Y, Ding Q, Yang Y, Li S, Jin L, Guan Y, Su B, Kang L et al (2016) Ancestral origins and genetic history of Tibetan Highlanders. Am J Hum Genet 99:580–594. https://doi.org/10.1016/j.ajhg.2016.07.002

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C et al (2016) The simons genome diversity project: 300 genomes from 142 diverse populations. Nature 538:201–206. https://doi.org/10.1038/nature18964

Malyarchuk B, Derenko M, Denisova G, Wozniak M, Grzybowski T, Dambueva I, Zakharov I (2010) Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. Ann Hum Genet 74:539–546. https://doi.org/10.1111/j.1469-1809.200.00601.1x

Malyarchuk B, Derenko M, Denisova G, Khoyt S, Woźniak M, Grzybowski T, Zakharov I (2013) Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels. J Hum Genet 58:804–811. https://doi.org/10.1038/jhg.2013.108

Malyarchuk BA, Derenko M, Denisova G, Woźniak M, Rogalla U, Dambueva I, Grzybowski T (2016) Y chromosome haplotype diversity in Mongolic-speaking populations and gene conversion at the duplicated STR DYS385a,b in haplogroup C3-M407. J Hum Genet 61:491–496. https://doi.org/10.1038/jhg.2016.14

Members BIGDC. (2017) The BIG Data Center: from deposition to integration to translation. Nucleic Acids Res 45:D18–D24. https://doi.org/10.1093/nar/gkw1060

Nasidze I, Quinque D, Dupanloup I, Cordaux R, Kokshunova L, Stoneking M (2005) Genetic evidence for the Mongolian ancestry of Kalmyks. Am J Phys Anthropol 128:846–854. https://doi.org/10.1002/ajpa.20159

Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q, McCarthy SA, Narechania A, Kashin S, Chen Y, Banerjee R, Rodriguez-Flores JL, Cerezo M, Shao H, Gymrek M, Malhotra A, Louzada S, Desalle R et al (2016) Punctuated bursts in human male demography inferred from 1244 worldwide Y-chromosome sequences. Nat Genet 48:593–599. https://doi.org/10.1038/ng.3559

Ramakrishnan U, Mountain JL (2004) Precision and accuracy of divergence time estimates from STR and SNPSTR variation. Mol Biol Evol 21:1960–1971. https://doi.org/10.1093/molbev/msh212

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow C-ET, Lin AA, Mitra M, Sil SK, Ramesh A, Rani MVU, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA (2006) Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am J Hum Genet 78:202–221

Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, Bai Z, Dong X, Chen H, Sun M, Zhai S, Sun Y, Yu L, Lan L, Xiao J et al (2017) GSA: Genome Sequence Archive *. Genom Proteom Bioinform 15:14–18. https://doi.org/10.1016/j.gpb.2017.01.001

Wei L-H, Yan S, Yu G, Huang Y-Z, Yao D-L, Li S-L, Jin L, Li H (2016) Genetic trail for the early migrations of Aisin Gioro, the imperial house of the Qing dynasty. J Hum Genet 1–5. https://doi.org/10.1038/jhg.2016.142

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348. https://doi.org/10.1101/gr.217602

Yan S, Wang C-C, Zheng H-X, Wang W, Qin Z-D, Wei L-H, Wang Y, Pan X-D, Fu W-Q, He Y-G, Xiong L-J, Jin W-F, Li S-L, An Y, Li H, Jin L (2014) Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. PLoS One 9:e105691. https://doi.org/10.1371/journal.pone.0105691

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. Am J Hum Genet 71:466–482. https://doi.org/10.1086/342096

Zhivotovsky L (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. Mol Biol Evol 18:700

Zhong H, Shi H, Qi X-B, Xiao C-J, Jin L, Ma RZ, Su B (2010) Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. J Hum Genet 5540:428–435