



# Y-chromosomal analysis of clan structure of Kalmyks, the only European Mongol people, and their relationship to Oirat-Mongols of Inner Asia

Natalia Balinova<sup>1</sup> · Helen Post<sup>2,3</sup> · Alena Kushniarevich<sup>2</sup> · Rodrigo Flores<sup>2</sup> · Monika Karmin<sup>2</sup> · Hovhannes Sahakyan<sup>2,4</sup> · Maere Reidla<sup>2</sup> · Ene Metspalu<sup>2</sup> · Sergey Litvinov<sup>5</sup> · Murat Dzhaubermezov<sup>6</sup> · Vita Akhmetova<sup>5</sup> · Rita Khusainova<sup>5,6</sup> · Phillip Endicott<sup>2,7</sup> · Elza Khusnutdinova<sup>5</sup> · Keemya Orlova<sup>8</sup> · Elza Bakaeva<sup>9</sup> · Irina Khomyakova<sup>10</sup> · Nailya Spitsina<sup>11</sup> · Rena Zinchenko<sup>1</sup> · Richard Villems<sup>2,3</sup> · Siiri Rootsi<sup>2</sup>

Received: 2 October 2018 / Revised: 8 March 2019 / Accepted: 26 March 2019  
© The Author(s) 2019. This article is published with open access

## Abstract

Kalmyks, the only Mongolic-speaking population in Europe, live in the southeast of the European Plain, in Russia. They adhere to Buddhism and speak a dialect of the Mongolian language. Historical and linguistic evidence, as well as shared clan names, suggests a common origin with Oirats of western Mongolia; yet, only a limited number of genetic studies have focused on this topic. Here we compare the paternal genetic relationship of Kalmyk clans with ethnographically related groups from Mongolia, Kyrgyzstan and China, within the context of their neighbouring populations. A phylogeny of 37 high-coverage Y-chromosome sequences, together with further genotyping of larger sample sets, reveals that all the Oirat-speaking populations studied here, including Kalmyks, share, as a dominant paternal lineage, Y-chromosomal haplogroup C3c1-M77, which is also present in several geographically distant native Siberian populations. We identify a subset of this clade, C3c1b-F6379, specifically enriched in Kalmyks as well as in Oirat-speaking clans in Inner Asia. This sub-clade coalesces at around 1500 years before present, before the Genghis Khan era, and significantly earlier than the split between Kalmyks and other Oirat speakers about 400 years ago. We also show that split between the dominant hg C variant among Buryats—C3-M407—and that of C3-F6379, took place in the Early Upper Palaeolithic, suggesting an extremely long duration for the dissipation of hg C3-M217 carriers across northern Eurasia, which cuts through today's major linguistic phyla.

These authors contributed equally: Natalia Balinova, Helen Post and Siiri Rootsi

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41431-019-0399-0>) contains supplementary material, which is available to authorized users.

✉ Helen Post  
helen2po@ut.ee

<sup>1</sup> Federal State Budgetary Institution Research Centre for Medical Genetics, Moscow 115522, Russia

<sup>2</sup> Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia

<sup>3</sup> Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia

<sup>4</sup> Laboratory of Ethnogenomics, Institute of Molecular Biology of National Academy of Sciences, Yerevan 0014, Armenia

<sup>5</sup> Institute of Biochemistry and Genetics - Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa 450054, Russia

## Introduction

Kalmyks are the only Mongolic-speaking people living in Europe, residing in the easternmost part of the European Plain. According to historical evidence, the ancestors of

<sup>6</sup> Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa 450076, Russia

<sup>7</sup> The UMR 7206, Muséum National d'Histoire Naturelle, Site du Musée de l'Homme, Paris 75116, France

<sup>8</sup> Institute of Oriental Studies, Russian Academy of Sciences, Moscow 107031, Russia

<sup>9</sup> Kalmyk Scientific Center, Russian Academy of Sciences, 358000 Elista, Russia

<sup>10</sup> Anuchin Institute and Museum of Anthropology, Lomonosov Moscow State University, Moscow 125009, Russia

<sup>11</sup> Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow 119991, Russia

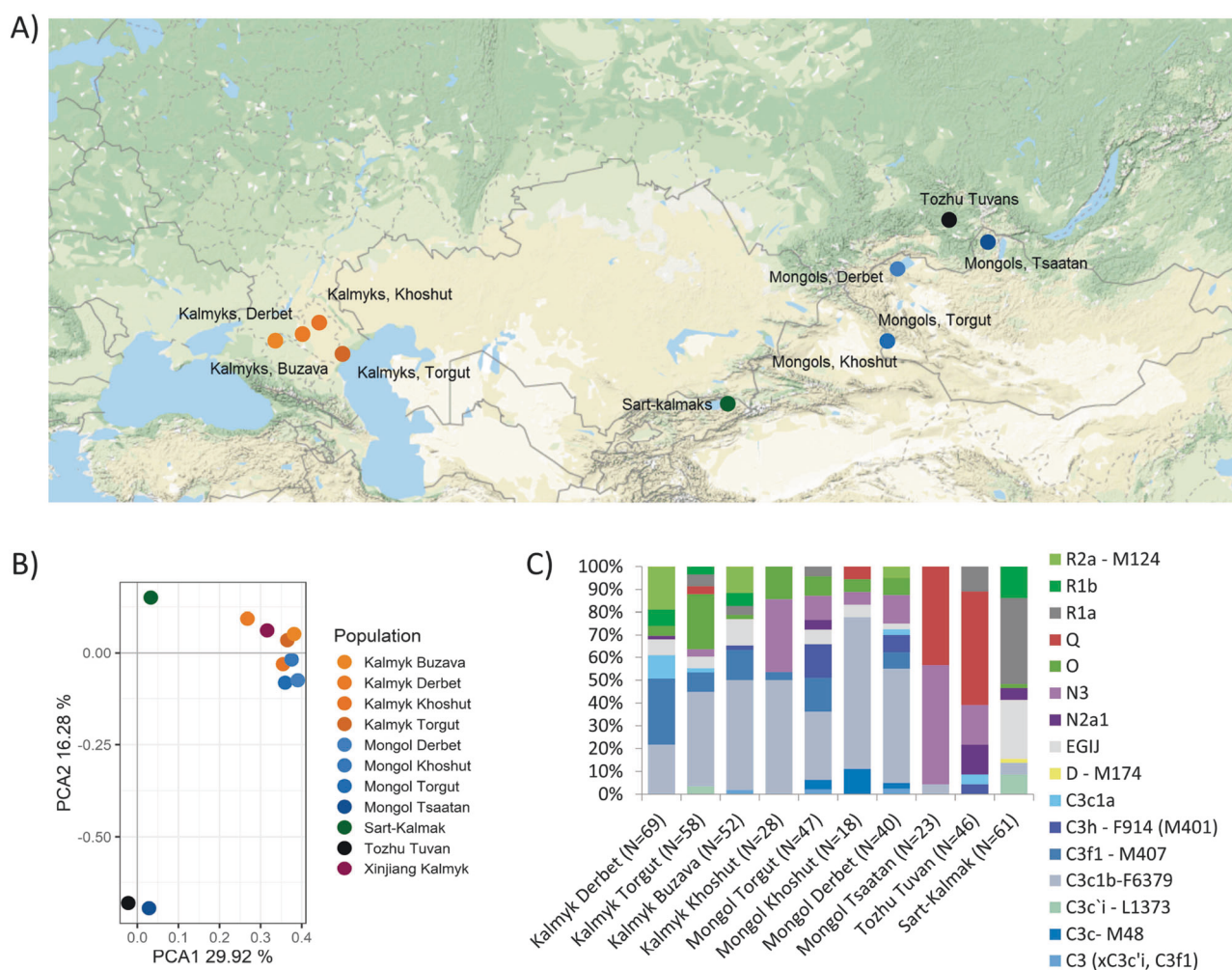
Kalmyks were nomadic groups of Oirat-speaking people, who migrated from Western Mongolia to Eastern Europe about four centuries ago [1, 2]. Kalmyks settled down in the dry steppe area, west of the lower Volga River basin on the Northwest shore of the Caspian Sea.

Oirat dialects belong to the western branch of the Mongolian language family [3], whose speakers include numerous sub-ethnic groups (Derbet, Torgut, Khoshut, Olot, Dzungar (Zunghar), Bayad, Zakhchin, Khoton, Myangad, Buzava) across a wide geographical area ofUvs and Khovd provinces (aimags) of Western Mongolia ( $N = 209,412$  [4]), and in Xinjiang Uygur Autonomous Region, China ( $N = 194,891$  [5]). Ethnic groups of Oirat speakers in the Republic of Kalmykia, Russia ( $N = 162,740$  ([http://www.gks.ru/free\\_doc/new\\_site/perepis2010/croc/perepis\\_itogi1612.htm](http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm))) include Torguts, Derbets and Buzavas (Fig. 1a), together with a smaller group called Khoshuts, who live in just two villages of Kalmykia. Up until today the Kalmyks have retained their

distinguished sub-ethnic groups, being quite separated from their geographical neighbours in Russia and northeast Caucasus.

Within the Torguts of Kalmykia the Tsaatan people form a sub-group, whereas in Mongolia Tsaatans are considered a distinct group on their own. Historically, Mongolian Tsaatans lived together with Tozhu Tuvans who speak a dialect of Tuvan, which is a Turkic language. They also shared a subsistence pattern, as both were reindeer herders. Researchers often erroneously consider the two Tsaatan groups together, relying on the similarity of ethnonyms and ethno-cultural characteristics [6]. The genetic relatedness of the two Tsaatan groups is therefore of great interest and has not been characterized thus far.

Kalmaks are another population of Oirat origin residing in small groups on the migration routes of Oirats through Altai and Central Asia. The largest group of Kalmaks is the Sart-Kalmak people living in Kyrgyzstan [7], who speak the



**Fig. 1 a** Map with the sampling points of the studied populations. Kalmyk students from Xinjiang province in China were left out due to small sample size ( $<15$ ). Mongol Torguts and Mongol Khoshuts live in the same location and are indicated with one dot. The map was created

using ggmap package in software environment R [50]. Map data is from c 2018 Google Maps. **b** PC plot based on haplogroup frequencies in studied populations. **c** Bar chart of Y-chromosomal haplogroup distribution among the studied populations

Oirat dialect of the Mongolian language and profess adherence to the Islam faith. Relating their paternal gene-pool to that of the Kalmyk people could reveal the extent of potential common paternal genetic legacy.

In the Great Mongol Empire, the ruling dynasties of Oirats entered into matrimonial alliances with the dynasty of Genghis Khan [8]. Having a privileged position, they retained their tribal structure and were released from the need to pay tribute. The collapse of the Mongol Empire was followed by the formation of the Durben-Oirat alliance, which existed from the fourteenth to the eighteenth century [9]. The genealogical connection of Kalmyk Khoshut rulers to the younger brother of Genghis Khan (Habutu Hasar) [10] is mentioned in various documents (e.g., *Tale of the Derben-Oirats* composed by Buddhist monk Gaban Sharab (1737)) and ideal for investigation using the Y chromosome.

Previous studies of the paternal genetic legacy of Genghis Khan and his male descendants [11–16] identified a C3\*- Star Cluster [16]. However, a recent study by Wei et al. [17] suggests that the age of the most recent common ancestor (TMRCA) of the Star Cluster (proposed to be the Y-profile of Genghis Khan) and its sub-lineages, together with their expansion patterns, are more consistent as having resulted from the diffusion of Mongolic-speaking populations. Hence, there is a need for direct genotyping of additional, well-documented, male descendants from a wider geographic region, to investigate the paternal legacy of Genghis Khan and his descendants with greater precision.

Previous studies of the Y chromosome diversity among Kalmyks are limited by a lack of sub-ethnic differentiation among samples [18], or being limited to just one of the smallest sub-groups, the Khoshut [19], and the large sub-ethnic group of Buzava [20, 21] was not sampled in any of the genetic studies thus far. Here we study 454 Oirat-speaking individuals from Kalmykia (all 4 ethnic groups), Mongolia (4 groups including Torguts and Tsaatans) and Kyrgyzstan (Sart Kalmaks), together with 28 Tozhu Tuvans from the Russian Federation. A combination of genotyping, microsatellite analysis and Y chromosome sequencing provides high-resolution comparative data between these geographically distant Oirat-speaking groups.

We have set the following objectives in our study: (a) to characterize the patrilineal genetic structure of the sub-ethnic groups of Kalmyks; (b) to determine the relation of Kalmyks to Oirat groups in Western Mongolia; (c) to clarify the controversial issues of ethnogenesis of the groups of Sart Kalmaks of Kyrgyzstan and the Tsaatans of Mongolia; (d) to show whether the members of the Khoshut dynasty in Kalmykia, whose genealogy traces back to Habutu Hasar (Genghis Khan's younger brother) belong to the putative Genghis Khan Y-chromosomal lineage.

## Materials and methods

### Samples

We studied a total of 454 unrelated male individuals for Y-chromosome polymorphisms. Included samples were from four sub-populations of Kalmyks from Russian Federation: Torgut (58), Derbet (69), Buzava (52), Khoshut (28); 4 populations from Western Mongolia: Torgut (47), Derbet (40), Khoshut (18), Tsaatan (23); Sart Kalmaks from Kyrgyzstan (61); Kalmyk students from Xinjiang province in China (12); and Tozhu Tuvans from Russian Federation (46) (Fig. 1a). The study was approved by the Federal State Budgetary Institution Research Centre for Medical Genetics Moscow, Russia. All donors provided informed consent forms that were translated into the native languages of the donors. All experiments were performed in accordance with the relevant guidelines and regulations of the collaborating institutions. DNA from peripheral blood leukocytes was isolated using standard phenol chloroform method.

### Genotyping and data analyses

The following Y-chromosomal biallelic markers were genotyped: M9, M130, M217, M77, F6379, B469, B80, B90, L1373, M407, M217, M48, M401, M207, Page07, Z93, Z95, Z2125, M343, M269, M412, Z2105, M458, M558, M478, M124, M175, M122, M119, M268, M95, P203, P201, M7, M231, P43, TAT, M128, F2930, F4205, CTS6967, B478, B525, B187, M2118, YAP, M174, M35, M201, P15, M170, M423, 12f2, M410, Page08, M242, M25 and M346. Genotyping was performed using PCR and subsequent direct sequencing or restriction fragment length polymorphism analysis. For genotyping new markers (F6379, B80 and B469) discovered from the refined phylogenetic tree inside C3c1 sub-clade, we designed primers with Primer3 software [22, 23]. Primer specificity was first assessed with Primer-BLAST [24] and GenomeTester v.1.3 software [25], and verified by Sanger sequencing. The specifications for the new markers inside C3c1 sub-clade can be found in Supplementary Table S1.

We used R to perform the principal component analysis (PCA) [26] (Fig. 1b, Supplementary Figs. S1 and S2) and correspondence analysis (CA) [27] (Supplementary Fig. S3 and Supplementary Table S2). To obtain the pairwise  $F_{st}$  genetic distances between groups and to perform the analysis of molecular variation (AMOVA), we used Arlequin 3.5.1.2 [28] (Supplementary Tables S3 and S4).

### Microsatellites and phylogenetic network

The microsatellite analysis was performed for 78 samples using 23 Short Tandem Repeats (STRs) from PowerPlex

Kit: DYS19, DYS385a, DYS385b, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643 and YGATAH4 (Supplementary Table S5).

We combined the data of 17 Y-STRs from 50 hg C3c1-M77 samples from this study (Supplementary Table S5) and hg C3c1-M77 samples published previously [19], to construct the phylogenetic network for hg C3c1 (Supplementary Fig. S4). We applied the median joining algorithm in the Network 4.6.1.1 software [29]. The expansion time calculations of the central node in the network were performed for 17 STRs using the rho-statistic and the genealogical mutation rate  $\sim 8.6 \times 10^{-5}$  per locus per year for the Y-filer (discussed in ref. [30]).

### Whole Y chromosome sequencing

To reconstruct the phylogeny of haplogroup C3 we combined the sequences of 10 novel Y chromosomes with 28 published by us earlier [31] (Supplementary Table S6). The ten whole Y chromosome sequences reported in this study are deposited in the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under the accession number PRJEB27561. Variants that did not have previously reported rs numbers were submitted to NCBI dbSNP database (<https://www.ncbi.nlm.nih.gov/projects/SNP/>) (dbSNP official release version 153).

All novel samples were sequenced using the Illumina HiSeq 2500 platform following Y chromosome capture with a proprietary capture protocol available at Gene by Gene (Family Tree DNA) using the commercially available “BigY” service (<https://www.familytreedna.com/>), a targeted enrichment design utilizing 67,000 capture probes for sequencing at least 10 Mbp at  $> 60 \times$  coverage. The targeted regions lie within the non-recombining male-specific parts of the Y chromosome. Published genomes are generated with Complete Genomics (San Jose, California) technology at a mean coverage of  $40 \times$ .

### Mapping and calling the Y chromosome variants

The fastq files of the newly sequenced genomes were mapped with BWA-MEM (v0.7.12) [32]. Read duplicates were removed using Picard (v2.0.1) (<http://broadinstitute.github.io/picard/>). GATK (v3.5) [33] was used to perform local realignment around known indels and for base quality score recalibration. Variant calling was performed with GATK tools HaplotypeCaller and GenotypeGVCF, which were asked to report variant and non-variant calls over the whole Y chromosome.

### Filtering

Individual genotypes were filtered with bcftools (v1.4) [34] in the raw VCF files that included all Y-chromosomal sites. The filtered Illumina and Complete Genomics data sets were merged using CombineVariants from GATK (v3.8) [33]. All the positions with  $> 5\%$  of missing genotypes in the combined dataset were masked out resulting in the final effective overlap between the sets. In addition, regions with poor mappability as described in ref. [31] were also masked out. This resulted in a final total of 9.2 Mb of usable sequence.

### Phylogeny reconstruction

We reconstructed the phylogeny for Y chromosome haplogroup C3 with hg C7a as an outgroup and estimated coalescent times using the software package BEAST v.1.7.5 [35] (Fig. 2). We chose Bayesian skyline coalescent model as the tree model [36] the general time reversible substitution model [37] with  $\gamma$ -distributed rates [38] and a relaxed lognormal clock model [39]. We used previously published [31] age of hg C (50,865 years, 95% confidence interval = 49,191–52,699) as the calibration point to get coalescent times for the inner structure. The run was performed with a piecewise-constant coalescent model with seven groups. The number of groups was obtained by dividing the sample size by 5. Markov Chain Monte Carlo method had 30 million iterations with a sampling made every 2000 steps. We visualized the BEAST run in Tracer v1.5 (<http://beast.bio.ed.ac.uk>) and confirmed that effective sample size was above 200. The tree was visualized using FigTree 1.4.2. (<http://tree.bio.ed.ac.uk/software/figtree/>). One newly sequenced sample was left out of BEAST analysis due to quality issues but can be seen on annotated tree indicated with dashed line (GRC171156839) (Supplementary Fig. S5).

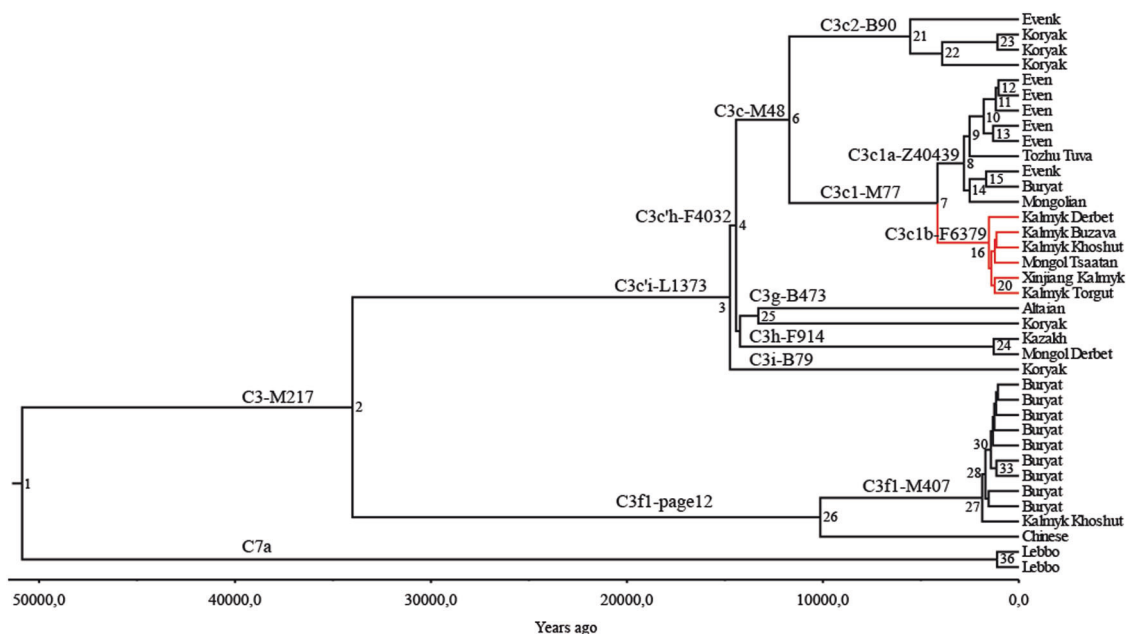
Throughout the study, nomenclature of Karmin et al. [31] and its updates were followed.

### Results and discussion

When considering all the studied populations together,  $> 50\%$  of the samples belong to different branches of haplogroup C3. If patrilineally dissimilar groups of Turkic-speaking Tsaatans and Tozhu Tuvans, and Oirat-speaking Sart Kalmaks are excluded, then the share of C3 becomes 62.5% among the sub-populations of Kalmyk and Oirats from Mongolia.

The new phylogeny for haplogroup C3c1-M77 based on Y-chromosome whole sequences (Fig. 2) reveals two





**Fig. 2** Detailed phylogenetic tree of hg C3. A phylogenetic tree of hg C3 based on 37 high-coverage Y chr sequences. As an outgroup, two hg C7a sequences were used. The calibrated tree was constructed using BEAST v.1.7.5 software package. Internal nodes, sub-clade names and population names on the tips are indicated. Internal nodes with

posterior probabilities <0.73 are not shown. Newly characterized C3c1b-F6379 branch is indicated in red. Age estimates can be found in Supplementary Table S7. All the sub-clade (node) defining variants and marker names are reported in Supplementary Table S8

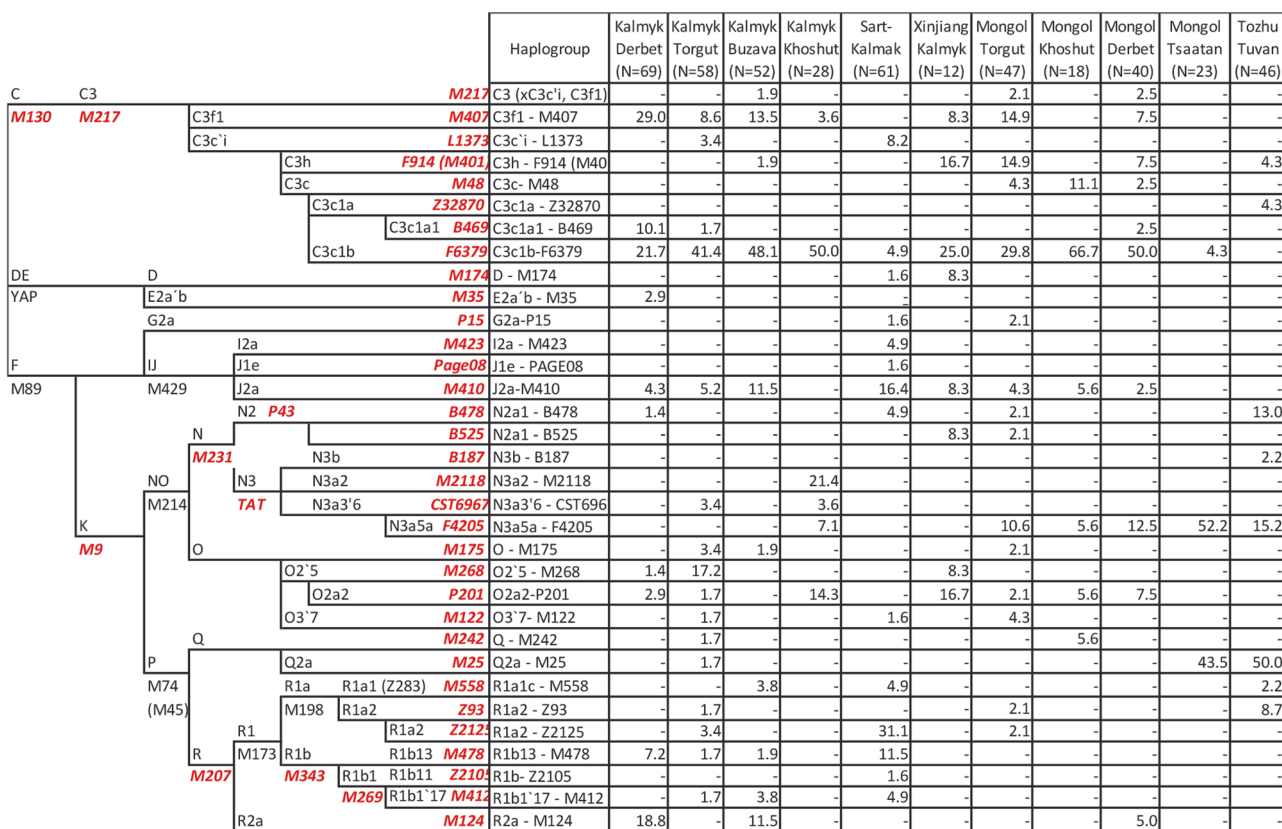
sub-branches: (i) previously characterized C3c1a-Z40439, which is common among Tungusic-speaking Evens and Evenks, Mongolic-speaking Buryats and Mongols [31], and (ii) a novel sub-branch that we name C3c1b-F6379, after 1 of 15 shared variants. The Bayesian estimate of the time to TMRCA of C3c1b is ~1.5 thousand years ago (KYA) with the 95% highest posterior density limit interval of 1.1–2.1 KYA (Table S7). The genotyping of the F6379 marker demonstrates that C3c1b is the most common paternal sub-clade among Kalmyks and Oirats from Mongolia, comprising 40.3% of male lineages (Fig. 1c and Fig. 3). It occurs at minor frequencies among Mongol Tsaatans and Sart Kalmaks.

The genotyped samples from this study belong to 33 Y-chromosomal haplogroups. The frequency distribution among the studied populations is shown in Fig. 1c and Fig. 3. Genotyping results indicate that Kalmyk sub-groups are quite diverse. This could be because different Kalmyk sub-groups experienced varying amounts of admixture after their migration from Mongolia. However, the evidence from paternal lineages suggests that Kalmyks have existed in relative isolation since arriving in Russia and maintained their common genetic heritage due to religious and cultural differences with their geographic neighbours in Eastern Europe. This scenario is also consistent with ethnographic evidence indicating that their current sub-ethnic groups were formed through the merging of different smaller tribal

groups in the past [9, 40]. In contrast, Tozhu Tuvans and Mongol Tsaatans have only a limited number of haplogroups, typical of a founder event followed by isolation and small census size (Fig. 1c).

Although sub-lineages of hg C3 form the major shared component among the various sub-populations of Kalmyks and Mongolian Oirat sub-groups, the frequency and composition of other haplogroups differ (Fig. 3 and Fig. 1b). In addition to C3, the other Y chromosome haplogroups present in Kalmyks are primarily of Siberian or Eastern Asian origin (hg N, O, Q, R1a2-Z93 derivatives, R2). For example, haplogroup N occurs in Kalmyk Khoshuts more frequently (32.1%) than in the other studied populations. N3a2-M2118 (Fig. 3), which is common in Yakutia (Central Siberia) and less frequent in Khanty and Mansi [41], is carried by 21.4% of Kalmyk Khoshuts (Fig. 3), who also have 7.1% of N3a5a-F4205, usually found in Buryats and Mongols [41], and 14% of O2a2-P201, which occurs at a high frequency in China and Southeast Asia [42].

Sart Kalmaks, Tsaatans of Mongolia and Tozhu Tuvans diverge from the general haplogroup distribution pattern (Fig. 1b, Fig. 1c and Fig. 3). Sart Kalmaks deviate due to >50% frequency of hg R. The most frequent (>30%) sub-group of hg R among Sart Kalmaks is R1a2-Z2125, which is at high frequencies in Kyrgyzstan and is also present in numbers among the Afghan Pashtuns [43]. This strongly suggests male-mediated gene flow between Kyrgyz and Sart



**Fig. 3** Schematic phylogenetic tree of detected Y SNPs and distribution of Y chromosome haplogroup frequencies (%) in the studied populations. Markers typed are indicated in red

Kalmaks reflected in small cultural and phenotypic differences between these neighbouring populations. The Tsaatans of Mongolia, with two major Y-chromosome lineages (N3a5a-F4205 and Q1a1b-M25) constituting ~96% of the paternal genepool, are another group with evidence for a founder effect followed by genetic isolation, typical of a small population of taiga reindeer herders. The same two lineages also dominate in the Tozhu Tuvans (N3a5a-F4205 (15.2%) and Q1a1b-M25 (50.0%)), providing genetic confirmation of inter-marriage between the two groups before the establishment of a political border between the Tuva Republic and the Russian Federation in 1944) [44]. According to the distribution of previously published Y chromosome haplogroup frequencies, Mongolian Tsaatans and Tozhu Tuvans are also similar to Tuvans [45]; hence, all three likely share a degree of common paternal origin.

Despite isolation by distance for four centuries, the distribution of Y-chromosomal haplogroups among Kalmyk sub-groups and Oirats of Western Mongolia is quite similar (Fig. 1c and Fig. 3). This similarity in paternal genepool composition between Kalmyks and Oirats from Mongolia is clearly visible in the PCA plot (Fig. 1b) where they cluster tightly together, whereas Kalmyks and their near geographical neighbours (NE Caucasus, Central and Southern

Russians, Tatars and Kazakhs) [46–48] are far apart (Supplementary Fig. S1). According to pairwise genetic distances (Fst), Mongol Tsaatan, Tozhu Tuvan and Sart Kalmak exhibit statistically significant divergence from the remaining studied populations (Supplementary Table S3). Repeating the PCA without these three outliers highlights the displacement of Kalmyk Derbet (Supplementary Fig. S2), which is driven by the significant level of haplogroup R2a (CA Dim1 = 21.67) (Fig. 3, Supplementary Fig. S3 and Supplementary Table S2). According to CA, other strong drivers of differentiation are N3a2-M2118 (Dim2 = 19.22) (Supplementary Fig. S3 and Supplementary Table S2), only present in Kalmyk Khoshuts, and O2'5-M268 (Dim2 = 33.9) (Supplementary Figure S3 and Supplementary Table S2) with the highest frequency among Kalmyk Torguts. In addition to Kalmyk Derbet, the genetic distances separating Kalmyk Buzava from rest of the groups are also statistically significant (Table S3). The AMOVA results confirm the pattern seen on the PC plot ( $F_{CT} = 0.15670$ ,  $P = 0.00293$ ) (Supplementary Table S4). The AMOVA results also show a low ( $F_{CT} = 0.01771$ ) and insignificant ( $P = 0.13196$ ) level of differentiation among the geographically distant groups (Supplementary Table S4). The parsimonious explanation for this concordance is that Y-chromosome genetic structure of

sub-groups of Oirat-speaking Mongols developed in the territory of Western Mongolia and did not undergo significant changes after the split, neither among Kalmyk people nor in Oirat branches in Inner Asia.

Returning to the C31b lineage, the STR network (Supplementary Fig. S4) shows a supercluster with a strongly centred star-like pattern and men from different sub-populations of Kalmyks and Oirats from Mongolia share the central modal haplotype, suggesting a recent shared ancestry and strong founder effect. The STR haplotype of putative descendant of Genghis Khan's male lineage is two mutational steps away from the central node. The expansion time for the central node of the network, calculated with the rho-statistic and the mutation rate for STRs (discussed in ref. [30]), is 667 ( $\pm 155$ ) years, arguably giving support to the idea of this haplotype expanding with the Mongol conquests. This time estimate, however, is significantly younger than the estimation based on sequence data at 1076–2011 years (Supplementary Table S7 and Fig. 2, node 16).

The differences in TMRCA estimates based on network analysis of Y-STRs vs. BEAST analysis of sequences may be caused by using different methods and different datasets. Non-random selection of samples for sequences to cover as much variation as possible provides the maximum TMRCA estimate for the clade and rho-statistic calculations based on STR network give us the time estimate for the pronounced founder event in the populations. It should be noted that the mutation process of microsatellites is less well understood and repeat length changes can occur in both directions, leading to underestimation of times.

## Conclusions

Although it has been previously shown by Nasidze et al. [31] that both Y chromosome and mtDNA hgs display a close contact between Kalmyks and Mongolians, our higher resolution analyses of patrilineal population structure of Mongolian Oirats and Kalmyks in the Russian Federation, to our knowledge, show for the first time the unity and integrity of the paternal gene pools of Kalmyk and Mongolian Oirat sub-groups. In particular, the Derbets and Torgut ethnic groups have similar genetic profiles, despite being separated by thousands of kilometres for around 400 years, whereas there exists a clear genetic and cultural continuity between the various Kalmyk groups in the Russian Federation.

However, the ethnogenesis of the Sart Kalmaks of Kyrgyzstan and the Tsaatans of Mongolia, appears to have a somewhat different history. The Sart Kalmaks show a more limited relatedness to the Oirat-speaking groups (having only ~13% of hg C3c) together with evidence of significant

paternal gene flow, likely from neighbouring Kyrgyzs, whereas the Tsaatans of Mongolia demonstrate closer genetic affinities with Tuvinians and Tozhu Tuvans.

The evidence presented here shows that the lineage of the putative Koshut rulers with possible genealogical links to the brother of Genghis belongs to haplogroup C3-M77. Based on the aDNA findings of the putative members of Genghis Khan dynasty, Zhang et al. [49] speculate that their paternal lineage may be within C3-L1373 lineage. However, it is widespread all over northern Siberia, from Beringian Koryaks to southern Siberian Altaians and Kazakhs (Fig. 2). Sub-clades, such as M77, seem to be present more typically among Mongolic- and Tungusic-speaking populations. The present study highlights a much more recent derivative of it, C3-F6379 that is typical for Kalmyks and Oirats from Mongolia. However, it is also clear that the patrilineal identity of Genghis Khan clan can be reliably identified by ancient DNA study of the archaeologically authentic remains belonging to this imperial family that so far have not been found.

**Acknowledgements** This research was supported by the European Union through the European Regional Development Fund (Project No. 2014–2020.4.01.16–0125), Project No. 2014–2020.4.01.16–0771 and Project No. 2014–2020.4.01.15–0012. This work was supported by the Estonian Research Council grant (PUT1339 and PRG243). This work was supported by institutional research funding IUT (IUT24–1) of the Estonian Ministry of Education and Research. This work was supported by the EU European Regional Development Fund through the Centre of Excellence in Genomics to the Estonian Biocentre and by the Estonian Institutional Research grant IUT24–1.

## 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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ashilova DO. Ethnic anthropology of the Kalmyks (in Russian). Elista: Kalmyk Book Publishing House; 1976.

2. Khomyakova IA, Balinova NV. Anthropological features of Torghts and Derbets of Kalmykia and Western Mongolia: a comparative analysis. In: *Anthropology Bulletin*. Moscow, Russia: Moscow University; 2017, p. 15–32.
3. Sanchirov VP. Formation of the ethnic community of the Oirats, the ancestors of the Kalmyks [in Russian]. In: Bakaeva EP, Zhukovskaya NL, editors. Moscow, Russia: Kalmyks Hayka; 2010, p. 12–35.
4. Ulaanbaatar. Mongolian Statistical Yearbook 'Population for 2010: Findings'. Mongolia: Ulaanbaatar; 2011.
5. Bitkejeva AN. Kalmyk language in the modern world (sociolinguistic aspect) [in Russian]. Moscow, Russia: Nauka; 2006.
6. Bakaeva EP. Tsaatan Kalmyks: towards one problem of origin of the ethnic group and etymology of the ethnonim [in Russian]. *Bull Kalmyk Inst Humanit RAS*. 2011;2:68–74.
7. Balinova NV, Khoninov VN. Studying the Issyk-Kul Kalmyk Ethnic Group. *Anthropol Archeol Eurasia*. 2014;53:47–55.
8. Juvaini A-M. Genghis Khan: the history of the world conqueror. Manchester, UK: Manchester Univ. Press; 1997.
9. Mitirov AG. Oirats – Kalmyks: Ages and generations [in Russian]. Elista, Russia: Kalmyk Publishing House; 1998.
10. Bakaeva EP. Genealogical legends of the Kalmyks and the problem of Genghis Khan's gene [in Russian]. In: Minaev VV, editor. Collection: stories of Siberia. Moscow, Russia: Russian State University for the Humanities; 2009, p. 59–73.
11. 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 General*. 2012;4:1–8.
12. Huang Y-Z, Wei L-H, Yan S, Wen S-Q, Wang C-C, Yang Y-J, et al. Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407. *Mol Genet Genomics*. 2018;293:657–63.
13. 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.
14. Zakharov IA. A search for a "Genghis Khan" chromosome. *Russ J Genet*. 2010;46:1130–1.
15. 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 [in Russian]. *Genetika*. 2007;43:422–6.
16. Zerjal T, Xue Y, Bertorelle G, Wells RSS, Bao W, Zhu S, et al. The genetic legacy of the Mongols. *Am J Hum Genet*. 2003; 72:717–21.
17. Wei L-H, Yan S, Lu Y, Wen S-Q, Huang Y-Z, Wang L-X, 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.
18. Nasidze I, Quinque D, Dupanloup I, Cordaux R, Kokshunova L, Stoneking M. Genetic evidence for the Mongolian ancestry of Kalmyks. *Am J Phys Anthropol*. 2005;128:846–54.
19. Malyarchuk B, Derenko M, Denisova G, Khoyt S, Wozniak M, Grzybowski T, et al. Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels. *J Hum Genet*. 2013;58:804–11.
20. Balinova NV. Anthropometric study of sub-ethnic groups of the Kalmyks [in Russian]. *Bull Kalmyk Inst Humanit RAS*. 2015;21:93–101.
21. Balinova NV. The Kalmyks: anthropogenetic portrait [in Russian]. Elista, Russia: ZAO NPP 'Dzhangar'; 2010.
22. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res*. 2012;40:e115.
23. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007;23:1289–91.
24. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13:134.
25. Andreson R, Reppo E, Kaplinski L, Remm M. GENOME-MASKER package for designing unique genomic PCR primers. *BMC Bioinformatics*. 2006;7:1–11.
26. Wickham H. *ggplot2: Elegant graphics for data analysis*. New York, USA: Springer-Verlag; 2016. <http://ggplot2.org>.
27. Nenadic O, Greenacre M. Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw*. 2007;20:1–13.
28. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564–7.
29. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999; 16:37–48.
30. Balanovsky O. Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Hum Genet*. 2017; 136:575–90.
31. Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve 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.
32. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. <https://arxiv.org/abs/1303.3997>.
33. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 2017. doi.org/10.1101/201178.
34. 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.
35. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29:1969–73.
36. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005;22:1185–92.
37. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Am Math Soc Lect Math Life Sci*. 1986;17:57–86.
38. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol*. 1994;39:306–14.
39. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4:699–710.
40. Erdniev UE. The Kalmyks (late 19th – early 20th cc.) [in Russian]. Elista, Russia: Kalmyk Book Publishing House; 1970.
41. Ilumäe AM, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, et al. Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *Am J Hum Genet*. 2016;99:163–73.
42. Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, et al. Major east-west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol*. 2010;27:1833–44.
43. Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, et al. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet*. 2015;23:124–31.
44. Suvandii ND, Kuular EM. Ethnic Tuvans in Tsagaan-Nuur Sumon, Mongolia: research issues [in Russian]. *New Res Tuva*. 2017;1:59–73.
45. Kharkov VN, Khamina KV, Medvedeva OF, Simonova KV, Khitrinskaya IY, Stepanov VA. Gene-pool structure of Tuvians



- inferred from Y-chromosome marker data [in Russian]. *Genetika*. 2013;49:1416–25.
46. Balanovsky O, Rootsi S, Pshenichnov A, Kivisild T, Churnosov M, Evseeva I, et al. Two sources of the Russian patrilineal heritage in their Eurasian context. *Am J Hum Genet*. 2008;82:236–50.
47. Yunusbayev B, Metspalu M, Ja M, Kutuev I, Rootsi S, Metspalu E, et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations research article. *Mol Biol Evol*. 2012;29:359–65.
48. Kharkov VN. Structure Of Y-Chromosomal Lineages In Siberian Populations [in Russian]. Tomsk, Russia: Siberian Division of Russian Academy of Medical Sciences; 2005.
49. 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.
50. Kahle D, Wickham H. ggmap: Spatial visualization with. *R J*. 2013;5:144–61.