

## ORIGINAL ARTICLE

# Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels

Boris Malyarchuk<sup>1</sup>, Miroslava Derenko<sup>1</sup>, Galina Denisova<sup>1</sup>, Sanj Khoyt<sup>2</sup>, Marcin Woźniak<sup>3</sup>, Tomasz Grzybowski<sup>3</sup> and Ilya Zakharov<sup>4</sup>

The Mongolic-speaking Kalmyks currently inhabiting the steppes of the Volga region have Central Asian ancestry and are organized into the tribal groups. The genetic relationships among these tribes and their origin have remained obscure. We analyzed 17 short tandem repeat and 44 binary polymorphisms of Y-chromosome in 426 individuals mainly from three major tribes of the Kalmyks (the Torguuds, Dörwöds and Khoshuuds). Among these tribes, the Dörwöds and Torguuds, as well as the Kalmyks collectively as an ethnic group, showed relatively close genetic affinities to each other and to the Mongols and Altaian Kazakhs, whereas the Khoshuuds were clearly separated from all of them, gathering with the Manchu, Tibetans or Evenks (depending on the algorithm used to calculate genetic distances). The genetic results also indicate that paternal gene flow from East Europeans to the Kalmyks is very little, despite their cohabitation in the North Caspian Steppe during the last 380 years. The occurrence of unique cluster of N1c-Tat haplotypes in the Khoshuuds, which dates to about 340 years and is likely to have East European ancestry, is considered as a result of interethnic contacts occurred soon after the appearance of the Kalmyk tribes in the Volga-Ural region.

*Journal of Human Genetics* (2013) 58, 804–811; doi:10.1038/jhg.2013.108; published online 17 October 2013

**Keywords:** haplogroup; phylogeography; STR; the Kalmyks; tribes; Y-chromosome

## INTRODUCTION

The Mongolic-speaking Kalmyks currently inhabiting the steppes to the west of the mouth of the Volga River close to the Caspian Sea coast are the descendants of the Oyrats (or the Oirads, Oyrads, Oyrads) originating from Western Mongolia.<sup>1,2</sup> In the 13th century, the Oyrat confederation was established consisting of the Torguud (or Torgut, Torgaut, Torgoud), Dörwöd (or Dörböd, Derbet), Khoshuud (or Khoshut, Khoshugut) and Khoyd (or Khoyt, Khoit) tribes.<sup>1</sup> In the 17th century, the largest Oyrat tribal unions of the Torguuds and Dörwöds migrated to the steppes of Western Siberia, but later they relocated to the left bank of the Volga River. Therefore, for over 380 years (from 1630), all the descendants of the Oyrats (currently classified as Kalmyks) have been living interspersed among other ethnic groups of this region. In the Republic of Kalmykia, the Torguuds and Dörwöds are the numerically dominant subdivisions of the Kalmyks. The Torguuds are located in the eastern and southeastern regions of the Republic. The Dörwöds primarily inhabit the northern and central regions of Kalmykia, and the Khoshuuds, few in number, are located in the Torguud and Dörwöd regions.

There are several genetic studies in which different Kalmyk populations were analyzed. An analysis of biochemical polymorphisms (blood groups, serum proteins and red cell enzymes) has been

performed in three tribes of the Kalmyks: the Torguuds, Dörwöds and Buzawa (or Buzavs).<sup>3</sup> The Buzawa are descendants of those Torguuds, Dörwöds and Züüngars who migrated to the Don River region and intermingled with other ethnic groups of this region during the 18th century. This genetic study revealed that close genetic affinities exist between the Dörwöds and Buzawa, but both populations differ significantly from the Torguuds. As an ethnic group, however, the Kalmyks genetically resemble the contemporary Buryats of the Baikal region of southeastern Siberia and the Mongols of Mongolia.<sup>3</sup> The mitochondrial DNA and Y-chromosome results also indicate a close relationship between the Kalmyks and Mongols<sup>4–6</sup> and Buryats.<sup>7,8</sup> More detailed analysis of mitochondrial DNA variation in Northern Asia has shown that the Kalmyks are clustered together with the Mongolic-speaking populations of the Buryats, Barghuts, Khamnigans, Mongols and Turkic-speaking Sojots, thus suggesting their origin from a common maternal ancestral gene pool.<sup>9</sup> The same trend is also evident for some of the Y-chromosome lineages—a relatively high frequency of haplogroup C3d-M407 has been revealed in all populations mentioned above, thus suggesting that this sub-haplogroup is characteristic of the Mongolic-speaking populations.<sup>10</sup>

Meanwhile, population samples studied to date using mitochondrial DNA and Y-chromosome markers represent heterogeneous

<sup>1</sup>Genetics Laboratory, Institute of Biological Problems of the North, Russian Academy of Sciences, Magadan, Russia; <sup>2</sup>Center for Promoting Social and Research Initiatives, Elista, Russia; <sup>3</sup>Department of Molecular and Forensic Genetics, Bydgoszcz, Institute of Forensic Medicine, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland and <sup>4</sup>Animal Comparative Genetics Laboratory, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia  
Correspondence: Dr BA Malyarchuk, Genetics Laboratory, Institute of Biological Problems of the North, Portovaya Street, 18, 685000 Magadan, Russia.  
E-mail: malyarchuk@ibpn.ru

Received 1 August 2013; revised 16 September 2013; accepted 27 September 2013; published online 17 October 2013

collections of individuals of different tribal ancestry living in Elista city, the capital of the Kalmyk Republic; therefore, molecular data obtained may simply result from sampling effects. Therefore, it would be reasonable to study descendants from different tribes of the Kalmyks obtained from different locations in Kalmykia. Apparently, such sets of samples may more accurately represent the overall genetic diversity in the Kalmyks. In this article, we explore the genetic polymorphisms of three tribes of the Kalmyks by characterizing the Y-chromosome variation through analysis of high-resolution biallelic markers and short tandem repeat (STR) typing. This approach allowed us to reconstruct the genetic structure of separate Kalmyk subpopulations (tribes of the Torguuds, Dörwöds and Khoshuuds) and the Kalmyks collectively as an ethnic group. The genetic relationships between the Kalmyks and other populations of Central and Eastern Asia as well as South Siberia were examined to clarify the paternal history of the Kalmyks.

## MATERIALS AND METHODS

### Subjects and DNA typing

Hair root samples of the Kalmyk males were collected from different locations in the Republic of Kalmykia, the autonomous Republic within the Russian Federation, during field expedition conducted in 2007 (Supplementary Figure S1, Supplementary Table S1). All the samples were collected with ethical approval at the Institute of Biological Problems of the North in Magadan, Russia, and informed consent was written in Russian.

DNA was extracted from the hair roots as described elsewhere.<sup>11</sup> The Y-chromosome single-nucleotide polymorphism was characterized using PCR primers summarized in Karafet *et al.*<sup>12</sup> Markers M25, M40, M48, M73, M120, M124, M128, M174, M217, M231, M267, M269, M314, M356, M407, M410, M458, MEH2, P31, P36.2 and P43 were analyzed through direct sequencing. All these polymorphisms were typed on ABI 3130 and ABI 3500xL Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). Markers Lly22g, M9, M17, M77, M89, M122, M130, M170, M172, M173, M201, M207, M214, M242, M253, P37.2, RPS4Y, SRY1532, Tat and 92R7 were detected by PCR-restriction fragment length polymorphism analysis.<sup>13–21</sup> An insertion/deletion polymorphism at DYS287 (YAP), 12f2 and MSY2.2 was typed as described elsewhere.<sup>13,22,23</sup>

Samples were additionally typed for 17 Y-STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, GATA-H4, DYS448, DYS456, DYS458 and DYS635) using AmpF<sup>®</sup>-STR YFiler PCR Amplification kit (Applied Biosystems) according to the manufacturer's instructions. Products of amplification were analyzed on ABI 3500xL Genetic Analyzer. Electrophoresis results were analyzed using GeneMapper Software v. 4.1 (Applied Biosystems).

### Statistical analysis

Loci DYS385a and DYS385b were excluded from statistical analysis of population data sets because an unambiguous assignment of the alleles to these loci is impossible without their separate typing. In addition, we excluded DYS19 because it is duplicated in some haplogroups, particularly in haplogroup C3c1-M77. The allele sizes for DYS389II were determined with the subtraction of DYS389I, and both loci were included in the calculations.

Summary statistics were calculated using ARLEQUIN 3.5.<sup>24</sup> Genetic differentiation between populations was estimated by means of distance methods based on the number of different alleles of the microsatellite (FST) and the sum of the squared number of repeat differences between two alleles (RST). The measure RST differs from FST in taking explicit account of the mutation process at microsatellite loci, for which a generalized stepwise mutation model appears appropriate.<sup>25</sup> The software ARLEQUIN 3.5 was used to perform the analysis of molecular variation. The statistical significance tests for pairwise FST and RST values were performed at 1000 permutations and for analysis of molecular variation at 10 000 permutations.

Our population data sets were compared with previously described populations of Central and Eastern Asia and South Siberia belonging to

different groups of Altaic family of languages, for which data for 12 Y-STR loci were available. These populations included Buryats,<sup>26</sup> Tibetans,<sup>27</sup> Manchu<sup>28</sup> and Altaian Kazakhs.<sup>29</sup> In addition, the population samples of the Mongols, Khamnigans, Sojots, Southern Altaians, Evenks, Khakassians and Tuvinians (a total of 462 individuals) typed for 12 Y-STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439) using PowerPlex Y System (Promega, Madison, WI, USA) were also analyzed. These population samples have been partially published by us,<sup>9,10,30,31</sup> but the complete set of Y-STR profiles is included here, in Supplementary Table S2.

Population pairwise FST values were also calculated from the Y-chromosome haplogroup frequencies to assess the genetic similarity of the comparative Kalmyk tribes, the Kalmyks collectively and South Siberian and Central Asian ethnic groups. As there is a scarcity of high-resolution haplogroup information in Central and Eastern Asian populations, we have used only the following South Siberian and Central Asian populations in comparative analysis: Tuvinians, Khakassians, Southern Altaians, Buryats, Yakuts, Altaian Kazakhs and Kyrgyzs from the study by Kharkov<sup>32</sup> and additional sample of Altaian Kazakhs from Dulik *et al.*<sup>29</sup>

Between-population distances (in the form of FST or RST values) were illustrated by constructing a neighbor-joining tree using the program MEGA 5.05<sup>33</sup> and by creating a multidimensional scaling plot using the software package STATISTICA v. 7.1 (StatSoft, Tulsa, OK, USA).

The age of STR variation within haplogroups was estimated as the average squared difference in the number of repeats between all current chromosomes and the founder haplotype (formed by the median values of the repeat scores at each STR locus within the haplogroup), averaged over STR loci and divided by means of a mutation rate.<sup>34</sup> The genealogical mutation rate equal to  $2.5 \times 10^{-3}$  per 35 years<sup>35</sup> was used in the calculations.

### Median network analysis

Median-joining networks of the Kalmyk Y-STR haplotypes were constructed using the Network 4.6 program (<http://www.fluxus-engineering.com>). Ambiguous loci DYS385a and DYS385b were excluded from median network analysis of individual haplogroups (with the exception of the C3d-M407 haplogroup). In addition, DYS19 locus was excluded from analysis of the C3c1 haplogroup. For the network construction, the weight of each character within the network was assigned to be inversely proportional to the number of mutations at character.<sup>36</sup> The weights varied from the default value of 10 at characters with 1 mutation per character to the value of 1 at fast-mutating characters.

For reconstruction of haplogroup R1a1a-M17 phylogeny, published 12-loci STR haplotypes for South Siberians (Southern Altaians, Buryats, Sojots, Evenks, Khakassians, Tuvinians, Shors and Tofalars),<sup>7</sup> Central Asians (Mongols,<sup>7</sup> Tajiks<sup>37</sup> and Pathans from Afghanistan<sup>38</sup>), East Europeans (Russians),<sup>7</sup> Western Asians (Persians and Kurds of Iran)<sup>37</sup> and South Asians (Indians)<sup>39</sup> were included in the analysis. For median network analysis of haplogroup R2a, Y-STR haplotypes for Buryats,<sup>31</sup> Persians, Kurds, Tajiks,<sup>37</sup> Pathans<sup>38</sup> and Indians<sup>39</sup> were used. For haplogroup N1c-Tat analysis, the following haplotypes for South Siberians and Central Asians (the Mongols, Southern Altaians, Buryats, Sojots, Evenks, Khakassians, Tuvinians, Shors, Tofalars, Yakuts, Evens and Koryaks)<sup>30</sup> and East Europeans (Russians,<sup>30</sup> Udmurts, Maris, Bashkirs, Chuvashes,<sup>40</sup> Komi,<sup>41,42</sup> Khanty and Mansi<sup>43</sup>) were also used.

## RESULTS

In total, 23 Y-chromosome haplogroups were identified in the Kalmyks (Table 1). However, more than half (56.1%) of the Kalmyk Y-chromosomes belonged to three C3-M217-derived haplogroups (C3\*-M217, C3c1-M77 and C3d-M407). N1c-Tat, O2-P31, O3-M122 and R2a-M124 were also common haplogroups, encompassing about 30% of the total male population. The Kalmyk tribes of the Torguuds, Dörwöds and Khoshuuds are characterized by the presence of haplogroups C3\*-M217, C3c1-M77, N1c-Tat, O3-M122 and R2a-M124 haplogroups, but the Khoshuuds had no haplogroup C3d-M407 that is typical of many Mongolic-speaking populations.<sup>10</sup>

In addition, the Khoshuuds are characterized by a high frequency of haplogroups N1c-Tat and O3-M122, whereas the Dörwöds and Torguuds had a high frequency of haplogroups R2a-M124 and O2-P31, respectively. The fourth tribal group that can be retrieved from

our data is the Buzawa (Supplementary Table S1) but the sample size of this group is small (23 individuals). Nevertheless, we can report that this subpopulation also had high frequencies of C3-haplogroups (8.7% of C3\*-M217, 39.1% of C3c1-M77 and 8.7% of C3d-M407), N1-haplogroups (8.7% of N1a-M128, 4.3% of N1b-P43 and 4.3% of N1c-Tat) and haplogroup R2a-M124 (17.4%). Haplogroups R1a1a-M17 and G-M201, each having a frequency of 4.3%, were also found in the Buzawa.

Population pairwise FST values calculated from the Y-chromosome haplogroup frequencies among different tribes of the Kalmyks and between them and other comparative groups are summarized in Supplementary Table S3. The relative affinities of populations studied are shown in Supplementary Figure S2. This neighbor-joining tree indicates the Dörwöds and Torguuds, as well as the Kalmyks collectively, are much more similar genetically to each other compared with the Khoshuuds. Comparing the Kalmyks with other Siberian populations, it is surprising that the Kalmyks cluster with Altaian Kazakhs. On the basis of Y-chromosome haplogroup frequencies, the Kalmyks differ from the other comparative groups, especially from the Mongolic-speaking Buryats.

Analysis of 17 fast evolving Y-STRs provided additional details of the relative affinities of the Kalmyk tribes. We reduced the 17-STR profile to a 9-STR profile (loci DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439) to compare the Kalmyk data with published data sets. Population pairwise FST and RST estimates were computed from the Y-STR data (Table 2) and illustrated by creating neighbor-joining trees (Supplementary Figures S3 and S4).

The neighbor-joining tree based on the FST values (Supplementary Figure S3) shows that the Kalmyks, as well as the Dörwöds and Torguuds, cluster together with Altaian Kazakhs and Mongols, whereas the Khoshuuds were clearly separated from all of them, gathering with the Manchu and Tibetans. The results of the multi-dimensional scaling of the FST values presented in Supplementary Figure S4 illustrate the close genetic affinities between the Kalmyks (and the Dörwöds) and Altaian Kazakhs, whereas the Torguuds were separated from all other populations and the Khoshuuds were close to the Mongols. Somewhat different picture emerged from results of the

**Table 1 Y-chromosome haplogroup frequencies (in percentages) in the Kalmyks**

| Haplogroups               | Tribes of the Kalmyks |                |                | Kalmyks, in total (426) |
|---------------------------|-----------------------|----------------|----------------|-------------------------|
|                           | Dörwöds (165)         | Khoshuuds (82) | Torguuds (150) |                         |
| C3*-M217                  | 5.5 (9)               | 1.2 (1)        | 10.7 (16)      | 6.6 (28)                |
| C3c1-M77                  | 33.3 (55)             | 37.8 (31)      | 46.0 (69)      | 38.7 (165)              |
| C3d-M407                  | 18.8 (31)             |                | 8.7 (13)       | 10.8 (46)               |
| D-M174                    | 1.2 (2)               |                | 2.0 (3)        | 1.2 (5)                 |
| E-M40                     |                       |                | 0.7 (1)        | 0.2 (1)                 |
| G-M201                    |                       |                |                | 0.2 (1)                 |
| I1-M253                   |                       |                |                | 0.2 (1)                 |
| I2a-P37.2                 |                       |                |                | 0.2 (1)                 |
| J1-M267                   | 0.6 (1)               |                | 0.7 (1)        | 0.5 (2)                 |
| J2a-M410                  | 1.8 (3)               |                | 2.7 (4)        | 1.6 (7)                 |
| J2b-M314                  |                       |                |                | 0.5 (2)                 |
| N*-M231                   |                       | 2.4 (2)        |                | 0.5 (2)                 |
| N1a-M128                  |                       |                |                | 0.5 (2)                 |
| N1b-P43                   | 1.2 (2)               | 1.2 (1)        | 1.3 (2)        | 1.4 (6)                 |
| N1c-TatC                  | 5.5 (9)               | 36.6 (30)      | 2.0 (3)        | 10.1 (43)               |
| O1-MSY2.2                 |                       |                | 0.7 (1)        | 0.2 (1)                 |
| O2-P31                    | 1.8 (3)               |                | 12.7 (19)      | 5.2 (22)                |
| O3-M122                   | 6.1 (10)              | 17.1 (14)      | 3.3 (5)        | 6.8 (29)                |
| Q1a1-M120                 | 2.4 (4)               |                |                | 0.9 (4)                 |
| Q1a2-M25                  |                       | 1.2 (1)        |                | 0.2 (1)                 |
| Q1a3-M346                 | 1.2 (2)               |                | 0.7 (1)        | 0.7 (3)                 |
| R1a1a-M17                 | 3.6 (6)               |                | 4.7 (7)        | 3.3 (14)                |
| R1b1b1-M73                | 1.8 (3)               | 1.2 (1)        | 0.7 (1)        | 1.4 (6)                 |
| R1b1b2-M269               |                       |                | 0.7 (1)        | 0.2 (1)                 |
| R2a-M124                  | 15.2 (25)             | 1.2 (1)        | 2.0 (3)        | 7.7 (33)                |
| <i>h</i> , gene diversity | 0.82 ± 0.02           | 0.70 ± 0.03    | 0.75 ± 0.03    | 0.81 ± 0.02             |

**Table 2 Pairwise comparisons among the Kalmyks and other populations of Siberia, Central and Eastern Asia based on Y-chromosome STR variation**

| Populations     | 1            | 2            | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12           | 13    | 14    | 15    |
|-----------------|--------------|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|-------|-------|-------|
| Kalmyks         | 0            | <b>0.001</b> | 0.085 | 0.019 | 0.080 | 0.164 | 0.142 | 0.135 | 0.018 | 0.185 | 0.207 | 0.078        | 0.233 | 0.138 | 0.045 |
| Dörwöds         | <b>0.001</b> | 0            | 0.113 | 0.025 | 0.064 | 0.190 | 0.131 | 0.117 | 0.021 | 0.208 | 0.143 | 0.097        | 0.268 | 0.153 | 0.050 |
| Khoshuuds       | 0.026        | 0.035        | 0     | 0.173 | 0.152 | 0.070 | 0.158 | 0.218 | 0.093 | 0.069 | 0.294 | <b>0.025</b> | 0.105 | 0.164 | 0.117 |
| Torguuds        | 0.028        | 0.041        | 0.101 | 0     | 0.196 | 0.295 | 0.267 | 0.235 | 0.055 | 0.251 | 0.173 | 0.160        | 0.314 | 0.144 | 0.099 |
| Sojots          | 0.147        | 0.142        | 0.144 | 0.264 | 0     | 0.196 | 0.061 | 0.051 | 0.091 | 0.197 | 0.155 | 0.114        | 0.248 | 0.200 | 0.146 |
| Khakassians     | 0.127        | 0.141        | 0.087 | 0.240 | 0.161 | 0     | 0.241 | 0.321 | 0.180 | 0.133 | 0.320 | <b>0.018</b> | 0.089 | 0.140 | 0.229 |
| Khamnigans      | 0.205        | 0.201        | 0.204 | 0.350 | 0.115 | 0.286 | 0     | 0.026 | 0.158 | 0.220 | 0.263 | 0.152        | 0.283 | 0.304 | 0.242 |
| Buryats         | 0.176        | 0.174        | 0.187 | 0.301 | 0.050 | 0.250 | 0.019 | 0     | 0.144 | 0.256 | 0.232 | 0.221        | 0.358 | 0.348 | 0.218 |
| Mongols         | 0.039        | 0.035        | 0.045 | 0.108 | 0.109 | 0.085 | 0.184 | 0.135 | 0     | 0.155 | 0.158 | 0.062        | 0.251 | 0.139 | 0.075 |
| Manchu          | 0.118        | 0.124        | 0.070 | 0.198 | 0.173 | 0.156 | 0.204 | 0.198 | 0.107 | 0     | 0.331 | 0.117        | 0.119 | 0.220 | 0.160 |
| Tibetans        | 0.132        | 0.136        | 0.097 | 0.208 | 0.197 | 0.151 | 0.258 | 0.247 | 0.120 | 0.063 | 0     | 0.266        | 0.386 | 0.228 | 0.170 |
| Evenks          | 0.074        | 0.071        | 0.073 | 0.152 | 0.143 | 0.050 | 0.268 | 0.220 | 0.040 | 0.162 | 0.166 | 0            | 0.119 | 0.103 | 0.146 |
| Tuvinians       | 0.170        | 0.181        | 0.116 | 0.273 | 0.153 | 0.074 | 0.261 | 0.249 | 0.148 | 0.156 | 0.143 | 0.141        | 0     | 0.159 | 0.271 |
| Altaians        | 0.136        | 0.140        | 0.123 | 0.216 | 0.168 | 0.107 | 0.306 | 0.287 | 0.139 | 0.177 | 0.116 | 0.126        | 0.077 | 0     | 0.199 |
| Altaian Kazakhs | 0.065        | 0.078        | 0.086 | 0.097 | 0.195 | 0.164 | 0.295 | 0.247 | 0.056 | 0.101 | 0.139 | 0.121        | 0.204 | 0.190 | 0     |

Abbreviation: STR, short tandem repeat.

FST values are shown below the diagonal, RST values are shown above the diagonal. Nonsignificant values ( $P > 0.05$ ) are indicated in bold print.

RST values analysis. As shown in the phylogenetic tree (Supplementary Figure S5), populations formed two clusters, with most Mongolic-speaking populations (the Mongols, Kalmyks and their tribes Dörwöds and Torguuds, Buryats, Khamnigans) and genetically close to them samples of Altaian Kazakhs, Tibetans and Sojots, being separated from the other reference populations, such as the Altaians, Evenks, Manchu, Khakassians and Tuvinians, as well as the Khoshuuds. In the multidimensional scaling plot (Figure 1), the RST estimates indicated that the Khoshuuds had genetic affinities with the Evenks, whereas the Kalmyks collectively, as well as their tribes of the Dörwöds and Torguuds, clustered together with the Mongols and Altaian Kazakhs. Therefore, the Kalmyk tribes did not constitute a separate cluster and the Khoshuuds were distinctive among them. Meanwhile, one should note that based on the population pairwise FST and RST values, the Kalmyks shared a smaller distances with the Mongols (Table 2). However, only three pairs of populations exhibited nonsignificant genetic distances: Khoshuuds-Evenks and Evenks-Khakassians for the RST values and Dörwöds-Kalmyks for the FST and RST values (Table 2).

The analysis of molecular variation performed using the 12 populations of South Siberia, Central and Eastern Asia showed that a 15 and 20.2% of genetic differentiation (for FST and RST distances, respectively) are because of differences among ethnic groups. This is consistent with previous studies pointing to a high level of Y-chromosomal differentiation in this geographical region.<sup>7,43-46</sup> However, our study shows that genetic differentiation is strong even within the Kalmyks as an ethnic group (Table 3). The level of genetic differentiation is higher at the inter-tribal group level than that at the interethnic group level for such closely related ethnic groups as the Kalmyks, Mongols and Altaian Kazakhs (Table 3). In addition, genetic diversity at the tribal level is comparable to that at the level of ethnic group (Table 4). It is noteworthy that this is in agreement with previous studies of Y-chromosome variation in nomadic populations of Central Asia demonstrating that there is no common ancestry at the tribal level, in contrast to the clan and lineage levels.<sup>45</sup>

In order to compare patterns of distribution of Y-STR haplotypes in different tribes of the Kalmyks, we performed median network analysis of STR haplotypes within their major haplogroups, such as C3c1-M77, C3d-M407, C3\*-M217, N1c-Tat, R-M207, O2-P31 and O3-M122 (Supplementary Figure S6). Haplogroup C3c1-M77 has a

star-like phylogenetic structure, with the most frequent haplotype found in 69 individuals from different tribes of the Kalmyks. About one-third of the Kalmyk Y-chromosomes (27.9% in the Dörwöds, 26.8% in the Khoshuuds and 37.3% in the Torguuds) are characterized by duplication of DYS19 locus on the background of haplogroup C3c1-M77 (Table 5). The same duplication has been previously observed in different populations of Central Asia and South Siberia.<sup>4,5,10,47</sup>

C3\*-M217 paragroup consists of two major haplotype clusters (Supplementary Figure S6). One of them is associated with deletion of the locus DYS448 and has been found at a frequency of 1.2% in the Khoshuuds, 2.4% in the Dörwöds and 7.3% in the Torguuds. Second haplotype cluster corresponded to the ‘star cluster’, which was thought to be carried by likely male-line descendants of Genghis Khan<sup>48</sup> and found at the highest frequency (76.5%) in the Kerey clan from Kazakhstan.<sup>49</sup> ‘Star cluster’ haplotypes represented only 3% of the Dörwöds and 2.7% of the Torguuds. Another interesting haplogroup is C3d-M407 (Supplementary Figure S6). Two separate haplotype subclusters are seen in the MJ network of this clade. One of them is characterized by allelic combination 11,18 at the DYS385 locus, whereas the overwhelming majority of C3d-M407 haplotypes in the Kalmyks are defined by allelic combination 11,11, which is probably a consequence of a deletion of allele 18 in any one of the two DYS385a and b loci lying on different arms of the Y-chromosome or gene conversion creating identical copies at the two loci during intrachromosomal recombination.<sup>50-52</sup> Quite the contrary, the overwhelming majority of C3d-M407 haplotypes in populations of South Siberia and Central Asia are defined by allelic combination 11,18 (or sometimes 11,17 and 11,19), and the DYS385-11,11 branch is present in the Kalmyks, Mongols, Tuvinians and Altaians.<sup>10</sup>

**Table 3 AMOVA results (in percentages of variation) based on the Y-chromosome STR diversity in the Kalmyk tribes and closely related ethnic groups**

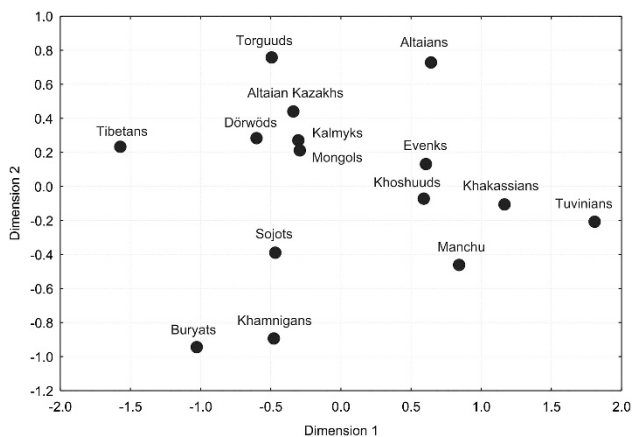
| Source of variation | The Kalmyk tribes<br>(the Dörwöds, Torguuds<br>and Khoshuuds) |       | Ethnic groups<br>(the Kalmyks, Mongols<br>and Altaian Kazakhs) |       |
|---------------------|---|-------|--|-------|
|                     | FST   | RST   | FST  | RST   |
| Among populations   | 5.77  | 9.72  | 5.57   | 3.92  |
| Within populations  | 94.23   | 90.28 | 94.43  | 96.08 |

Abbreviation: AMOVA, analysis of molecular variation; STR, short tandem repeat.

**Table 4 Molecular diversity indices calculated from the Y-chromosome STR haplotypes in the Kalmyks**

| Molecular diversity indices         | Tribes of the Kalmyks |                   |                   |                            |
|-------------------------------------|-----------------------|-------------------|-------------------|----------------------------|
|                                     | Dörwöds<br>(165)      | Khoshuuds<br>(82) | Torguuds<br>(150) | Kalmyks, in total<br>(426) |
| Mean number of pairwise differences | 4.78 ± 2.35           | 4.94 ± 2.43       | 3.78 ± 1.92       | 4.83 ± 2.36                |
| Average gene diversity over loci    | 0.53 ± 0.29           | 0.55 ± 0.3        | 0.42 ± 0.24       | 0.54 ± 0.29                |

Abbreviation: STR, short tandem repeat.



**Figure 1** Multidimensional scaling (MDS) plot of pairwise RST genetic distances using Y-chromosome STR haplotypes data for the Kalmyks and other populations of Siberia, Central and Eastern Asia. The stress value for the MDS plot is 0.112.

**Table 5** Frequencies (in percentages) and the age of STR variation of some Y-chromosome haplogroups in the Kalmyks

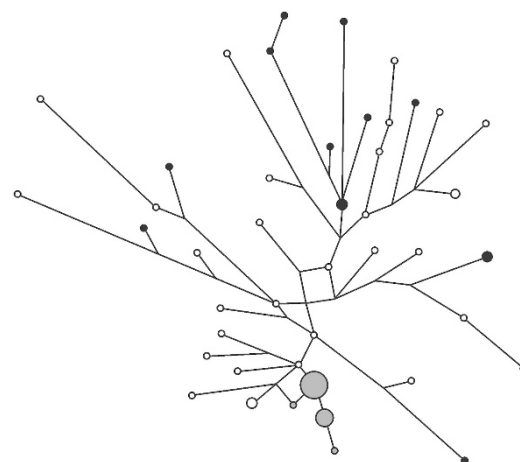
| Tribes                              | Haplogroups   |   |   |                 |
|-------------------------------------|---|---|---|-----------------|
|                                     | <i>C3c1-M77</i> ,<br><i>DYS19</i><br><i>duplication</i> | <i>C3*-M217</i> ,<br><i>DYS448</i><br><i>deletion</i> | <i>C3d-M407</i> ,<br><i>DYS385a</i> ,<br><i>b-11,11</i> | <i>R2a-M124</i> |
| Dörwöds (165)                       | 27.9 (46)   | 2.4 (4)   | 18.8 (31)   | 15.2 (25)       |
| Khoshuuds (82)                      | 26.8 (22)   | 1.2 (1)   | 0   | 1.2 (1)         |
| Torguuds (150)                      | 37.3 (56)   | 7.3 (11)  | 4.0 (6)   | 2.0 (3)         |
| The age of STR variation (in years) | 666 ± 202   | 1367 ± 379  | 511 ± 197   | 411 ± 168       |

Abbreviation: STR, short tandem repeat.

High frequency of such unusual mutations as the *DYS19* duplication, *DYS448* deletion and *DYS385-11,11* variant may be related to the founder effects due to the recent migration of the Kalmyk ancestors and their isolation in the North Caspian Steppe. Therefore, one can expect age-synchronous pattern in these haplogroups. However, their ages differed significantly, varying from 1300 years for *DYS448-del* haplotypes to about 400 years for *R2a-M124* haplogroup (Table 5). The latter haplogroup, rare in South Siberian and Central Asian populations, was found at high frequency in the Dörwöds (15.2%). It is noteworthy that the age of haplogroup *R2a-M124* shows the best correspondence to the historical chronology of the Kalmyks.

Figure 2 illustrates a median network of *R2a-M124* haplotypes derived from 12 Y-STR loci in the Kalmyks and Buryats and populations of Southwestern and South Asia, in which this haplogroup is present at a relatively high frequency.<sup>37–39</sup> Common haplotypes are observed in the Kalmyks and Buryats, whereas there is no haplotype sharing between these and other comparative populations. However, it is important that the *R2a-M124* haplotypes found in the Kalmyks and Buryats are derivatives of the Indian haplotypes. This suggests a recent common ancestry and/or expansion of the Kalmyk *R2a-M124* lineages from India.

Previous study of the Y-chromosome variability in the Kalmyks indicated that there has been no substantial admixture with Russians during the last 380 years of their neighborhood because the paternal gene pool of the Kalmyks hardly included East European-specific haplogroups, such as *R1a1a-M17*, *N1c-Tat* and *I-M170*.<sup>4</sup> In our study, we have found haplotypes belonging to the *R1a1a-M17* and *N1c-Tat* haplogroups at perceptible frequencies (3.3% and 10.1%, respectively); hence, their East European origin is quite possible. To study this question, we reconstructed median networks of haplogroups *R1a1a-M17* and *N1c-Tat* based on STR variation in the Kalmyks, Russians and populations of South Siberia and Central Asia. The latter were taken into analysis because haplogroups *R1a1a-M17* and *N1c-Tat* are typical not only of East Europeans but also of some populations of Siberia and Central Asia.<sup>7,30,32,40,44</sup> In addition, haplogroup *R1a1a-M17* is specific for populations of Southwestern and South Asia; therefore, we have also analyzed Y-STR haplotypes for Pathans from Afghanistan,<sup>38</sup> Persians and Kurds from Iran,<sup>37</sup> Tajiks<sup>37</sup> and South Indians.<sup>39</sup> Phylogenetic network of haplogroup *R1a1a-M17* (Supplementary Figure S7) demonstrates that two haplotypes of the Kalmyks coincide with the haplotypes common among different populations of South Siberia (Tuvinians, Altaians

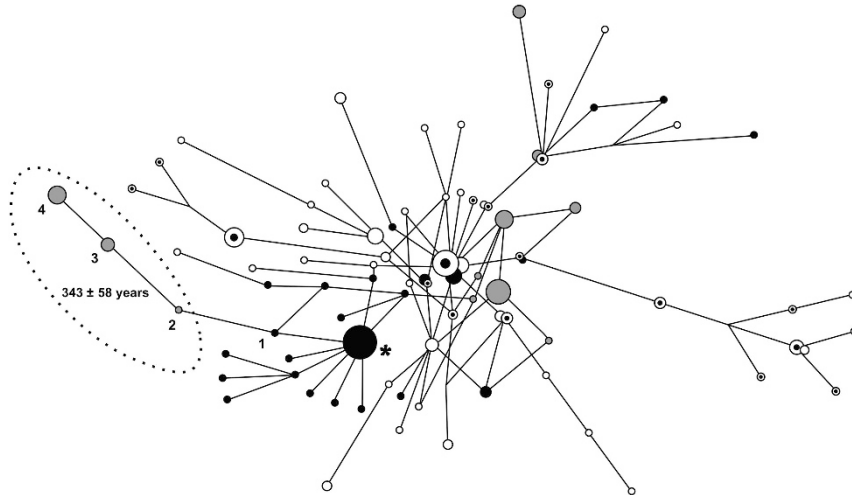


**Figure 2** A median-joining network based on 12 Y-STR loci within the *R2a-M124* haplogroup. Each circle represents a haplotype, defined by a combination of short tandem repeat (STR) markers. Circle size is shown proportional to haplotype frequency and the smallest circle represents one haplotype. The lines between circles represent mutational distance, the shortest distance being a single mutational step. Colors indicate different groups of populations: gray for the Kalmyks and Buryats,<sup>31</sup> black for populations of Central and West Asia (Pathans from Afghanistan, Persians and Kurds from Iran and Tajiks)<sup>37,38</sup> and white for South Indians.<sup>39</sup>

and Sojots) and Central Asia (Tajiks and Pathans), two haplotypes with populations of Pathans, Persians and Indians, and one haplotype with South Siberians (Shors and Tuvinians). Only one common haplotype was detected in the Kalmyks and Russians (Supplementary Figure S7).

Median network analysis of *N1c-Tat* haplotypes shows the bipartite distribution of the Kalmyk haplotypes (Supplementary Figure S6, Figure 3). Majority of them can be found among a dozen haplotypes that were detected in South Siberians and East Europeans, and the origin of these paternal lineages cannot be established undoubtedly. The Kalmyk haplotypes are in general unique, with only one haplotype revealed in common between the Kalmyks and Tuvinians. Thus, it is possible that all these haplotypes may have originated in South Siberia and Eastern Asia. Second part of the Kalmyk *N1c-Tat* haplotype network is represented by a separate branch consisting of three haplotypes that were found only in the Khoshuuds—haplotypes 2, 3 and 4 in Figure 3.

A search of these STR haplotypes in the YHRD 3.0 database (<http://www.yhrd.org>; release 44 built on 15 July 2013; 67 430 haplotypes within 503 world populations) does not find matching haplotypes (for loci *DYS19*, *DYS385a*, *DYS385b*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS437*, *DYS438* and *DYS439*) in any population. It is noteworthy, however, that earlier haplotype (haplotype 1 in Figure 3) preceding the branch consisting of haplotypes 2–4 is the haplotype revealed in Chuvashes from the Volga region.<sup>40</sup> This finding points to possibility of East European origin of the Khoshuud *N1c-Tat* branch. Rare haplotype 1 also does not occur in the YHRD 3.0 database, but earlier haplotype, indicated by a star (\*) in Figure 3, was found frequently only in populations of the Ural region (in the Komi, Mansi and Khanty).<sup>41,42,53</sup> Coalescence age of the Khoshuud *N1c-Tat* branch is about  $343 \pm 58$  years, pointing to a possibility that inflow of this East European paternal lineage into the gene pool of the Khoshuuds occurred soon after their appearance in the Volga-Ural region (after 1645 according to historical data<sup>2</sup>).



**Figure 3** A median-joining network based on 12 Y-STR loci within the N1c-Tat haplogroup in the Kalmyks, East Europeans and South Siberians. Each circle represents a haplotype, defined by a combination of short tandem repeat (STR) markers. Circle size is shown proportional to haplotype frequency and the smallest circle represents one haplotype. The lines between circles represent mutational distance, the shortest distance being a single mutational step. Colors indicate different groups of populations: gray for the Kalmyks, black for the Volga-Ural populations (Komi, Udmurts, Chuvashes and Bashkirs),<sup>40–43</sup> white for the Russians<sup>30</sup> and white circle with black dot inside for aboriginal South Siberian populations.<sup>30</sup> Explanations for haplotypes \*, 1, 2, 3 and 4 are in the main text.

## DISCUSSION

The pattern of Y-chromosome diversity in a population can provide a clear overview of its origin.<sup>4,29,48</sup> Haplogroup C3c1-M77 is the most frequent in all Kalmyk tribes studied. It is known that, in addition to the Kalmyks, haplogroup C3c1-M77 is widespread in Tungusic-speaking people of North-Eastern Asia, such as Evens and Evenks, and Turkic-speaking Altaian Kazakhs.<sup>54</sup> However, the three Kalmyk tribes studied here exhibited quite different patterns of Y-chromosome haplogroup variation. In the Khoshuuds, N1c-Tat and O3-M122 were the dominant haplogroups and C3d-M407 was absent, although it appeared as a genetic marker for the Mongolic-speaking populations.<sup>10</sup> Such composition is interesting because haplogroup O3-M122 is the dominant paternal haplogroup in Sino-Tibetan populations,<sup>55</sup> but haplogroup N1c-Tat occurs at high frequency in different populations of Eastern and Central Siberia.<sup>30,40,56</sup> However, the Torguuds showed another pattern, having high frequency of haplogroup O2-P31 and low frequency of O3-M122. The Dörwöds are characterized by high frequencies of C3d-M407 and R2a-M124. Haplogroup D-M174, albeit rarely found in the Dörwöds and Torguuds (about 2%), reveals genetic affinity to the Tibetan populations where this haplogroup was present probably before the Last Glacial Age.<sup>57</sup> Rare N\*-M231 haplotype was found in the Khoshuuds. It is suggested that this haplogroup has likely originated in Southern China and now is mostly present in southern Eastern Asian populations, such as Daic, southern Han Chinese, Tibeto-Burman and Hmong-Mien.<sup>58</sup> Another rare N1a-M128 haplogroup, which is frequently occurred only among Kazakhs (8.1%) in Central Asia, was also detected in the Kalmyks (in the Buzawa). Haplogroup R2a-M124 is also interesting, notwithstanding it accounted relatively high frequency only in the Dörwöds (15.2%). This haplogroup is very rare in Central Asia and Siberia, being found only in the Buryats (0.3–2.7%), Kyrgyzs (1.3%), Tajiks (12.5–16.7%)<sup>31,32,37,59</sup> and also in Chinese populations (in Uygurs, Han, Hui at a frequency of 2–7%).<sup>60</sup> Haplogroup R2a-M124 has a predominant distribution in South Asia (in India and Pakistan);<sup>38,39,60,61</sup> thus, its presence in the Kalmyks and other

Central Asians is likely a result of male influence from the south of Asia, possibly from India.

Analysis of the Y-chromosome STR diversity has shown that the Kalmyks collectively as an ethnic group have close genetic affinities to the Mongols and Altaian Kazakhs, but the individual Kalmyk tribes, especially the Khoshuuds, have differentiated genetically from each other. Recent studies based on Y-STR haplotypes have also demonstrated that the Kalmyks, Mongols, Altaian Kazakhs and Altaians-Kizhi showed a similar haplotype composition, resulting probably from common origin of populations.<sup>6,62</sup> Anthropologists have suggested that in the last 200–300 years, the Kalmyk tribes have experienced some admixture with neighboring East European populations.<sup>63</sup> However, comparison of haplotypes belonging to haplogroups R1a1a-M17 and N1c-Tat, which are frequently occurred in East European populations, shows a low level of East European admixture in the Kalmyks. In addition, the lack or very low frequency of European-specific lineages R1a1a7-M458, I1-M253 and I2a-P37.2 in the Kalmyks suggests that paternal gene flow from East Europeans to the Kalmyks is very little. Therefore, intra-ethnic heterogeneity of the Kalmyks is likely maintained by inter-tribal differences because of extensive endogamous practices of the tribal groups. Previous molecular genetic studies have shown that Y-chromosome results are in agreement with ethnological data suggesting that nomadic populations of Central Asia, especially Turkic-speaking ones, are exogamous at the clan level but endogamous at the tribe level.<sup>29,43,45,46,49</sup> Our study shows that there are more differences between the tribes of the Kalmyks than between some ethnic groups, such as the Kalmyks, Mongols and Altaian Kazakhs. This is also consistent with previous studies of the Y-chromosome variation in the Turkic nomadic groups of Central Asia.<sup>45,46</sup>

Future research should be conducted with high-resolution Y-chromosome analysis of Mongolian populations throughout Mongolia, especially of the Oyrat groups living in Western Mongolia, which is necessary to answer the questions about genetic affinities between different tribes of the Mongols and Kalmyks. Previous studies of Mongolian populations were performed using low-resolution

Y-chromosome analysis and reported conflicting results.<sup>64,65</sup> In addition, analysis of genome-wide autosomal single-nucleotide polymorphism data in the Kalmyks would be a helpful approach to reveal their genetic affinity to populations of Central Asia and Siberia and to estimate the extent of their recent admixture with East Europeans.

## ACKNOWLEDGEMENTS

This research was financially supported by the Program of Fundamental Studies of Russian Academy of Sciences 'Living Nature: Modern State and Problems of Development', subprogram 'Dynamics and Conservation of Gene Pools' (Project 12-I-P30-12) and the Russian Foundation for Basic Research (Projects 11-04-00620 and 11-04-98508-vostok-a).

- Erdniev, U. E. *Kalmyks* (Nauka, Moscow, 1985) in Russian.
- Birtalan, A. in *Oirad and Kalmyk Linguistic Essays* (ed Birtalan, A.) 11–24 (Elite Eotvos Kiado, Budapest, 2012).
- Galushkin, S. K., Spitsyn, V. A. & Crawford, M. H. Genetic structure of Mongolic-speaking Kalmyks. *Hum. Biol.* **73**, 823–834 (2001).
- Nasidze, I., Quinque, D., Dupanloup, I., Cordaux, R., Kokshunova, L. & Stoneking, M. Genetic evidence for the Mongolian ancestry of Kalmyks. *Am. J. Phys. Anthropol.* **128**, 846–854 (2005).
- Roewer, L., Krüger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L. *et al.* Y-chromosomal STR haplotypes in Kalmyk population samples. *Forensic Sci. Int.* **173**, 204–209 (2007).
- Dulik, M. C., Zhadanov, S. I., Osipova, L. P., Askapuli, A., Gau, L., Gokcumen, O. *et al.* Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* **90**, 229–246 (2012).
- Derenko, M., Malyarchuk, B., Denisova, G. A., Woźniak, M., Dambueva, I., Dorzhu, C. *et al.* Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum. Genet.* **118**, 591–604 (2006).
- Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M. *et al.* Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am. J. Hum. Genet.* **81**, 1025–1041 (2007).
- Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Rogalla, U., Grzybowski, T. *et al.* Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. *PLoS One* **7**, e32179 (2012).
- Malyarchuk, B., Derenko, M., Denisova, G., Woźniak, M., Grzybowski, T., Dambueva, I. *et al.* Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Ann. Hum. Genet.* **74**, 539–546 (2010).
- Walsh, P. S., Metzger, D. A. & Higuchi, R. Research report: Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506–513 (1991).
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. *Genome Res.* **18**, 830–838 (2008).
- Rosser, Z. H., Zerjal, T., Hurler, M. E., Adojaan, M., Alavantic, D., Amorim, A. *et al.* Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**, 1526–1543 (2000).
- Flores, C., Maca-Meyer, N., Perez, J. A., Gonzalez, A. M., Larruga, J. M. & Cabrera, V. M. A predominant European ancestry of paternal lineages from Canary Islanders. *Ann. Hum. Genet.* **67**, 38–52 (2003).
- Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J. *et al.* Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**, 817–831 (1999).
- Ke, Y., Su, B., Song, X., Lu, D., Chen, L., Li, H. *et al.* African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* **292**, 1151–1153 (2001).
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
- Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am. J. Hum. Genet.* **71**, 466–482 (2002).
- Bergen, A. W., Wang, C. Y., Tsai, J., Jefferson, K., Dey, C., Smith, K. D. *et al.* An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* **63**, 63–80 (1999).
- Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W. H., Roxas, A. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).
- Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P., Shen, P. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
- Hammer, M. F. & Horai, S. Y-chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**, 951–962 (1995).
- Hammer, M. F., Karafet, T. M. & Redd, A. J. Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
- Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
- Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
- Woźniak, M., Derenko, M., Malyarchuk, B., Dambueva, I., Grzybowski, T. & Miścicka-Śliwka, D. Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. *Forensic Sci. Int.* **164**, 271–275 (2006).
- Zhu, B., Wu, Y., Shen, C., Yang, T., Deng, Y., Xun, X. *et al.* Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. *Forensic Sci. Int.* **175**, 238–243 (2008).
- He, J. & Guo, F. Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning Province, Northeast China. *Forensic Sci. Int. Genet.* **7**, e84–e85 (2013).
- Dulik, M. C., Osipova, L. P. & Schurr, T. G. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One* **6**, e17548 (2011).
- Derenko, M., Malyarchuk, B., Denisova, G. A., Woźniak, M., Grzybowski, T., Dambueva, I. *et al.* Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J. Hum. Genet.* **52**, 763–770 (2007).
- Malyarchuk, B., Derenko, M., Denisova, G., Maksimov, A., Woźniak, M., Grzybowski, T. *et al.* Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J. Hum. Genet.* **56**, 583–588 (2011).
- Kharkov, V. N. *Structure and Phylogeography of Gene Pools of Aboriginal Peoples of Siberia Based on Y-Chromosomal Markers* Dr Sci Biol thesis (Research Institute of Medical Genetics, Tomsk, 2012) in Russian.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
- Zhivotovskiy, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T. *et al.* The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **74**, 50–61 (2004).
- Goedbloed, M., Vermeulen, M., Fang, R. N., Lembring, M., Wollstein, A., Ballantyne, K. *et al.* Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit. *Int. J. Legal Med.* **123**, 471–482 (2009).
- Bandelt, H.-J., Macaulay, V. & Richards, M. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16**, 8–28 (2000).
- Malyarchuk, B., Derenko, M., Woźniak, M. & Grzybowski, T. Y-chromosome variation in Tajiks and Iranians. *Ann. Hum. Biol.* **40**, 48–54 (2013).
- Lacau, H., Gayden, T., Regueiro, M., Chennakrishnaiah, S., Bukhari, A., Underhill, P. A. *et al.* Afghanistan from a Y-chromosome perspective. *Eur. J. Hum. Genet.* **20**, 1063–1070 (2012).
- Chennakrishnaiah, S., Perez, D., Gayden, T., Rivera, L., Regueiro, M. & Herrera, R. J. Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India. *Gene* **526**, 96–106 (2013).
- Roots, S., Zhivotovskiy, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R. *et al.* A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* **15**, 204–211 (2007).
- Mirabal, S., Regueiro, M., Cadenas, A. M., Cavallini-Sforza, L. L., Underhill, P. A., Verbenko, D. A. *et al.* Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur. J. Hum. Genet.* **17**, 1260–1273 (2009).
- Pimenoff, V. N., Comas, D., Palo, J. U., Vershubsky, G., Kozlov, A. & Sajantila, A. Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur. J. Hum. Genet.* **16**, 1254–1264 (2008).
- Perez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martinez-Arias, R. *et al.* Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**, 208–219 (1999).
- Karafet, T. M., Osipova, L. P., Gubina, M. A., Posukh, O. L., Zegura, S. L. & Hammer, M. F. High levels of Y chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.* **74**, 761–789 (2002).
- Chaix, R., Austerlitz, F., Khgay, T., Jacquesson, S., Hammer, M. F., Heyer, E. *et al.* The genetic or mythical ancestry of descent groups: lessons from the Y chromosome. *Am. J. Hum. Genet.* **75**, 1113–1116 (2004).
- Heyer, E., Balaresque, P., Jobling, M. A., Quintana-Murci, L., Chaix, R., Segurel, L. *et al.* Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genet.* **10**, 49 (2009).
- Balaresque, P., Parkin, E. J., Roewer, L., Carvalho-Silva, D. R., Mitchell, R. J., van Oorschot, R. A. *et al.* Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications. *Int. J. Legal Med.* **123**, 15–23 (2009).
- Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S. *et al.* The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72**, 717–721 (2003).
- Abilev, S., Malyarchuk, B., Derenko, M., Woźniak, 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.* **84**, 79–89 (2012).

- 50 Kittler, R., Erlen, A., Brauer, S., Stoneking, M. & Kayser, M. Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur. J. Hum. Genet.* **11**, 304–314 (2003).
- 51 Park, M. J., Shin, K.-J., Kim, N. Y., Yang, W. I., Cho, S.-H. & Lee, H. Y. Characterization of deletions in the DYS385 flanking region and null alleles associated with AZFc microdeletions in Koreans. *J. Forensic Sci.* **53**, 331–334 (2008).
- 52 Hallast, P., Balaresque, P., Bowden, G. R., Ballereau, S. & Jobling, M. A. Recombination dynamics of a human Y-chromosomal palindrome: Rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. *PLoS Genet.* **9**, e1003666 (2013).
- 53 Trynova, E. G., Tsitovich, T. N., Vylegzhaniina, E. Y., Bandurenko, N. A. & Parson, W. Presentation of 17 Y-chromosomal STRs in the population of the Sverdlovsk region. *Forensic Sci. Int. Genet.* **5**, e101–e104 (2011).
- 54 Malyarchuk, B., Derenko, M. & Denisova, G. On the Y-chromosome haplogroup C3c classification. *J. Hum. Genet.* **57**, 685–686 (2012).
- 55 Shi, H., Dong, Y. L., Wen, B., Xiao, C. J., Underhill, P. A., Shen, P. D. *et al.* Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am. J. Hum. Genet.* **77**, 408–419 (2005).
- 56 Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R. *et al.* Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**, 1174–1183 (1997).
- 57 Shi, H., Zhong, H., Peng, Y., Dong, Y. L., Qi, X. B., Zhang, F. *et al.* Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* **6**, 45 (2008).
- 58 Shi, H., Qi, X., Zhong, H., Peng, Y., Zhang, X., Ma, R. Z. *et al.* Genetic evidence of an East Asian origin and Paleolithic northward migration of Y-chromosome haplogroup N. *PLoS One* **8**, e66102 (2013).
- 59 Kharkov, V. N. *Y Chromosome Structure in Populations of Siberia* PhD thesis (Research Institute of Medical Genetics, Tomsk, 2005) in Russian.
- 60 Zhong, H., Shi, H., Qi, X. B., Duan, Z. Y., Tan, P. P., Jin, L. *et al.* Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* **28**, 717–727 (2011).
- 61 Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E. *et al.* Polarity and temporality of high-resolution Y chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am. J. Hum. Genet.* **78**, 202–221 (2006).
- 62 Tarlykov, P. V., Zholdybayeva, E. V., Akilzhanova, A. R., Nurkina, Z. M., Sabitov, Z. M., Rakhypbekov, T. K. *et al.* Mitochondrial and Y-chromosomal profile of the Kazakh population from East Kazakhstan. *Croat. Med. J.* **54**, 17–24 (2013).
- 63 Ashilova, D. O. *The Anthropological Composition of the Kalmyk People* (Nauka, Moscow, 1972) in Russian.
- 64 Katoh, T., Munkhbat, B., Tounai, K., Mano, S., Ando, H., Oyungerel, G. *et al.* Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene* **346**, 63–70 (2005).
- 65 Keyser-Tracqui, C., Crubézy, E., Pamzav, H., Varga, T. & Ludes, B. Population origins in Mongolia: genetic structure analysis of ancient and modern DNA. *Am. J. Phys. Anthropol.* **131**, 272–281 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)