Phylogeography of the Y-chromosome haplogroup C in northern Eurasia

Boris Malyarchuk1, Miroslava Derenko1, Galina Denisova1, Marcin Wozniak2, Tomasz Grzybowski2, Irina Dambueva3 and Ilia Zakharov4

1Institute of Biological Problems of the North, Russian Academy of Sciences, Portovaya str. 18, 685000 Magadan, Russia
2Forensic Medicine Institute, the Ludwik Rydygier Medical College, the Nicolaus Copernicus University in Torun, 85-094 Bydgoszcz, Poland
3Institute of General and Experimental Biology, Russian Academy of Sciences, 670047 Ulan-Ude, Russia
4N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin Str. 3, 119991 Moscow, Russia

Summary

To reconstruct the phylogenetic structure of Y-chromosome haplogroup (hg) C in populations of northern Eurasia, we have analyzed the diversity of microsatellite (STR) loci in a total sample of 413 males from 18 ethnic groups of Siberia, Eastern Asia and Eastern Europe. Analysis of SNP markers revealed that all Y-chromosomes studied belong to hg C3 and its subhaplogroups C3c and C3d, although some populations (such as Mongols and Koryaks) demonstrate a relatively high input (more than 30%) of yet unidentified C3∗ haplotypes. Median joining network analysis of STR haplotypes demonstrates that Y-chromosome gene pools of populations studied are characterized by the presence of DNA clusters originating from a limited number of frequent founder haplotypes. These are subhaplogroup C3d characteristic for Mongolic-speaking populations, “star cluster” in C3∗ paragroup, and a set of DYS19 duplicated C3c Y-chromosomes. All these DNA clusters show relatively recent coalescent times (less than 3000 years), so it is probable that founder effects, including social selection resulting in high male fertility associated with a limited number of paternal lineages, may explain the observed distribution of hg C3 lineages.

Keywords: Y-chromosome, haplogroup, STR, South Siberia, Eastern Europe, phylogeography

Introduction

Molecular phylogeography of Y-chromosome haplogroups (hg) provides important information about the history of human populations, allowing reconstruction of ancient migrations and more recent gene flows. One of the most widespread Y-chromosome haplogroups in East Asia is hg C defined by marker RPS4Y711 and several accompanying SNPs (Karafet et al., 2008). This haplogroup is represented by several subhaplogroups (named from C1 to C6), which appear to be informative with respect to the genetic history of human populations. It has been generally accepted that the hg C ancestor is of South-East Asian or Indian origin, because haplotypes belonging to ancient C∗-branches are still present in the Indian subcontinent, Sri Lanka and in parts of South-East Asia (Underhill et al., 2001; Sengupta et al., 2006). Subhaplogroup C2 is found mainly in New Guinea, Melanesia, and Polynesia (Kayser et al., 2003; Hammer et al., 2006). C4 is found exclusively among aboriginal Australians (Kayser et al., 2006). C5 is present mainly in India (Sengupta et al., 2006), whereas the rare C1 lineage seems to be restricted to Japan (Hammer et al., 2006). The most likely place of origin of the most successful C3 lineage is south-eastern or central Asia, from where this haplogroup has spread into northern Asia and the Americas (Karafet et al., 2002; Lell et al., 2002; Katoh et al., 2005; Hammer et al., 2006; Sengupta et al., 2006; Xue et al., 2006). C3 is also present at low frequencies in populations of eastern and central Europe, where it may represent evidence of the westward expansion of the steppe Nomads in the early middle Ages (Derenko et al., 2007b). Although there are several recognizable phylogenetic branches of C3 (from C3a to C3f), the geographic pattern of its distribution in northern Asia is still unclear (Karafet et al., 2008).

In the present study, we improve our understanding of the hg C phylogeographic structure by examining STR
variation within SNP-defined subclusters in a large number of individuals of northern Asian and Eastern European origin.

**Materials and Methods**

**Subjects and DNA Typing**

A total of 1449 samples (whole blood and hair root samples) from unrelated males were collected in populations of South Siberia (Altaians, Teleuts, Altaian Kazakhs, Khakassians, Shors, Tuvinians, Todjins, Sojots, Buryats, and Khamnigans), central and eastern Siberia (Evenks, Evens, Yakuts, and Koryaks), Eastern Asia (Mongols and Koreans) and Eastern Europe (Kalmyks and Russians) (Table 1). All samples were collected with appropriate ethical approval and informed consent.

Hg C markers RPS4Y711 (for the whole hg C), M8 (for C1), M38 (for C2), M217 (for C3), M93 (for C3a), P39 (for C3b), M77 (for C3c), M407 (for C3d), P53.1 (for C3e) and P62 (for C3f) were assayed using PCR primers summarized in Karafet et al. (2008). Several SNPs were typed by means of RFLP analysis: RPS4Y711 by Bsc4I-analysis (Bergen et al., 1999), M8 by DdeI-analysis (Underhill et al., 1997), M38 by Bst4CI-analysis (Kayser et al., 2003) and M217 by Bsc4I-analysis (Shen et al., 2000). The remaining markers were analyzed by means of DNA sequencing on ABI 3130 Genetic Analyzer (Applied Biosystems). Electrophoresis results were analyzed using Genscan v. 3.7 and Genotyper v. 3.7 software (Applied Biosystems).

**Data Analysis**

Median joining (MJ) networks of hg C STR-haplotypes were constructed using the Network 4.5.1.6 program (http://www.fluxus-engineering.com). For the network construction, STR variants were weighted (with a weight assigned to a range of variance values) following the distribution of the number of mutations at character (Bandelt et al., 2000). Since C3c chromosomes carry the DYS385a,b genotypes 11,12, 12,12 and 12,13, an unambiguous assignment of the alleles to the loci DYS385a and DYS385b is impossible without separate typing of these loci (Niederstätter et al., 2004). In this connection, loci DYS385a and DYS385b were excluded from the MJ network analysis. In addition, DYS19 was omitted from analysis because C3c chromosomes include a set of DYS19 duplicated haplotypes.

The age of STR variation within hg C 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

---

**Table 1** Haplogroup C subhaplogroup distribution in populations studied (no. of individuals and% values in parenthesis)

<table>
<thead>
<tr>
<th>Population</th>
<th>N</th>
<th>Linguistic affiliation</th>
<th>C, in total</th>
<th>C3c</th>
<th>C3d</th>
<th>C3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altaians</td>
<td>89</td>
<td>Turkic</td>
<td>14 (15.7)</td>
<td>5</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Teleuts</td>
<td>44</td>
<td>Turkic</td>
<td>5 (11.4)</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Altaian Kazakhs</td>
<td>36</td>
<td>Turkic</td>
<td>21 (58.3)</td>
<td>15</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Khakassians</td>
<td>64</td>
<td>Turkic</td>
<td>1 (1.6)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Shors</td>
<td>38</td>
<td>Turkic</td>
<td>1 (2.6)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Todjins</td>
<td>26</td>
<td>Turkic</td>
<td>2 (7.7)</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tuvinians</td>
<td>108</td>
<td>Turkic</td>
<td>12 (11.1)</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Yakuts</td>
<td>10</td>
<td>Turkic</td>
<td>2 (20.0)</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sojots</td>
<td>28</td>
<td>Turkic</td>
<td>15 (53.6)</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Mongols</td>
<td>46</td>
<td>Mongolic</td>
<td>30 (65.2)</td>
<td>5</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Buryats</td>
<td>217</td>
<td>Mongolic</td>
<td>148 (68.2)</td>
<td>13</td>
<td>117</td>
<td>18</td>
</tr>
<tr>
<td>Khamnigans</td>
<td>51</td>
<td>Mongolic</td>
<td>28 (54.9)</td>
<td>0</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Kalmyks</td>
<td>91</td>
<td>Mongolic</td>
<td>57 (62.6)</td>
<td>41</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Evenks</td>
<td>41</td>
<td>Tungusic</td>
<td>20 (48.8)</td>
<td>18</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Evens</td>
<td>63</td>
<td>Tungusic</td>
<td>34 (54.0)</td>
<td>33</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Koryaks</td>
<td>39</td>
<td>Chukotko-Kamchatkan</td>
<td>15 (38.5)</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Koreans</td>
<td>52</td>
<td>Korean</td>
<td>5 (9.6)</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Russians</td>
<td>406</td>
<td>Indo-European</td>
<td>3 (0.7)</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Annals of Human Genetics (2010) 00:1–8 © 2010 The Authors**

---
of a mutation rate (Zhivotovsky et al., 2004; Sengupta et al., 2006). A mutation rate of $2.5 \times 10^{-3}$ per 35 years calculated in father-son pairs (Goedbloed et al., 2009) and the evolutionary effective mutation rate of $6.9 \times 10^{-4}$ per 25 years based on STR variation within Y chromosome haplogroups in the populations with documented short-term histories (Zhivotovsky et al., 2004) were used. The upper bound for divergence time of two groups of haplotypes was calculated as $T_D$, assuming STR variance in repeat number at the beginning of population subdivision ($V_0$) equal to zero (Zhivotovsky, 2001). In the network construction and the age calculation procedures, the allele sizes for DYS389II were determined with the subtraction of DYS389I.

The software ARLEQUIN 3.5 (Excoffier et al., 2005) was used to estimate genetic distances (as pairwise values of $F_{ST}$ and $R_{ST}$ between STR haplotypes) and to test for correlations between linguistic and genetic distances through a Mantel test with RST between STR haplotypes) and to test for correlations between STR-haplotypes. As for the C3c subhaplogroup, the high-frequencies of C3c (53.6%) was also found in Sojots from the Baikal region. Although Sojots speak the Turkic language, they are genetically related to Mongolic-speaking Buryats, as follows from the analysis of maternal mtDNA lineages (Derenko et al., 2003). In that study, analysis of the population structure of mtDNA sequences in South Siberia showed that between-population differences (in terms of the pairwise $F_{ST}$ values) were statistically non-significant only for three population pairs, namely for Altaians and Khakassians, Tuvinians and Tujdins, and Buryats and Sojots.

To obtain a better resolution of the phylogenetic relationships between Y-chromosomes revealed in individuals carrying C3 lineages, we have analyzed haplotypes at 12 STR loci (Table S1). As a result, we have found that about 35% of individuals belonging to subcluster C3c are characterized by duplication of DYS19 (Table S1). This duplication on the background of subcluster C3c has been previously observed in Kalmyk, Mongolian, Kazakh, Kyrgyz and Tajik Y-chromosomes (Nasidze et al., 2005; Roewer et al., 2007; Balaresque et al., 2009). In addition, Roewer et al. (2007) demonstrated that in Kalmyk a high proportion of duplicated DYS19 alleles is associated with deletion of the locus DYS448. Using a 17 STR amplification kit (AmpFl-STR®YFiler®, Applied Biosystems) we have also analyzed the DYS448 locus in the Kalmyk population and found that three out of nine samples with duplicated alleles 16 and 17 at DYS19 are characterized by deletion of the locus DYS448. We should note however that the duplication in Kalmyks, Mongols, Tuvinians and Tujdins involves mostly alleles 16 and 17, whereas both Altaian Kazakhs and Altaians exhibit alleles 15 and 17 (Table 2).

Another interesting case can be seen in the median joining network of subcluster C3d (Fig. 1). What is notable is a separate branch of identical Kalmyk, Mongolian, Tuvinian and

---

**Results**

SNP analysis of 1449 Y-chromosomes from 18 ethnic groups representing populations of northern Asia and eastern Europe demonstrates that hg C is frequent in some Siberian populations studied (Table 1). The highest frequencies of haplogroup C (more than 50%) were revealed in Mongolic-speaking Buryats (68.2%), Mongols (65.2%), Kalmyks (62.6%) and Khamnigans (54.9%) as well as in Turkic-speaking Altaian Kazakhs (58.3%) and Sojots (53.6%) and Tungusic-speaking Evens (54.0%). In other Siberian populations hg C was found at low or moderate frequencies. In Russians, this haplogroup is very frequent in Mongolic-speaking populations, such as Buryats (53.9%), Khamnigans (52.9%), Mongols (15.2%) and Kalmyks (12.1%). A high frequency of this subcluster (53.6%) was also found in Sojots from the Baikal region. Although Sojots speak the Turkic language, they are genetically related to Mongolic-speaking Buryats, as follows from the analysis of maternal mtDNA lineages (Derenko et al., 2003). In that study, analysis of the population structure of mtDNA sequences in South Siberia showed that between-population differences (in terms of the pairwise $F_{ST}$ values) were statistically non-significant only for three population pairs, namely for Altaians and Khakassians, Tuvinians and Tujdins, and Buryats and Sojots.

To obtain a better resolution of the phylogenetic relationships between Y-chromosomes revealed in individuals carrying C3 lineages, we have analyzed haplotypes at 12 STR loci (Table S1). As a result, we have found that about 35% of individuals belonging to subcluster C3c are characterized by duplication of DYS19 (Table S1). This duplication on the background of subcluster C3c has been previously observed in Kalmyk, Mongolian, Kazakh, Kyrgyz and Tajik Y-chromosomes (Nasidze et al., 2005; Roewer et al., 2007; Balaresque et al., 2009). In addition, Roewer et al. (2007) demonstrated that in Kalmyk a high proportion of duplicated DYS19 alleles is associated with deletion of the locus DYS448. Using a 17 STR amplification kit (AmpFl-STR®YFiler®, Applied Biosystems) we have also analyzed the DYS448 locus in the Kalmyk population and found that three out of nine samples with duplicated alleles 16 and 17 at DYS19 are characterized by deletion of the locus DYS448. We should note however that the duplication in Kalmyks, Mongols, Tuvinians and Tujdins involves mostly alleles 16 and 17, whereas both Altaian Kazakhs and Altaians exhibit alleles 15 and 17 (Table 2).

Another interesting case can be seen in the median joining network of subcluster C3d (Fig. 1). What is notable is a separate branch of identical Kalmyk, Mongolian, Tuvinian and

---

© 2010 The Authors
Annals of Human Genetics © 2010 Blackwell Publishing Ltd/University College London

Annals of Human Genetics (2010) 00,1–8
Table 2 Population origins and allelic spectrum of DYS19 duplications on the background of subcluster C3c

<table>
<thead>
<tr>
<th>Population</th>
<th>DYS19 duplications</th>
<th>Number of C3c Y-chromosomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15,16</td>
<td>15,17</td>
</tr>
<tr>
<td>Kalmyks</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Mongols</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tuvinnians</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Todjins</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Altaian Kazaks</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Altaians</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 1 Median joining network of subhaplogroup C3d based on twelve STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) in populations of northern Asia. Each circle represents a haplotype, defined by a combination of 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 the linguistic subdivisions: white for Mongolic-speaking populations, black for Tungusic-speaking populations, and grey for Turkic-speaking populations (according to data presented in Table 1). Fourteen identical Kalmyk, Mongolian, Altaian and Tuvinnian samples form a separate branch within C3d, shown at the periphery of the median joining network.

Altai samples characterized by allelic combination 11,11 at the DYS385 locus, whereas the overwhelming majority of C3d haplotypes are defined by allelic combination 11,18 (or sometimes 11,17 or 11,19) (Table S1). Unfortunately, we do not know whether the appearance of this 11,11 variant is a consequence of a deletion of any one of the two DYS385a and b loci lying on different arms of the Y-chromosome (Kittler et al., 2003) or a consequence of mutation in the primer-binding sites. Therefore, due to uncertainties with loci DYS19 and DYS385a,b, the evolutionary ages of the hg C3 subclusters were calculated for all the markers except for these loci (Table 3).

The age of accumulated STR variation within hg C3, estimated using the method of Zhivotovsky et al. (2004), is about 14.9 ky or 4.1 ky depending on the mutation rate values selected for calculations (Table 3). The older time estimate is most compatible with the view that hg C3 haplotypes were present in Siberia during the Last Glacial Maximum from where the ancestors of C3b Native Americans migrated to the Beringia (Karafet et al., 2002; Zegura et al., 2004).

The median joining network of subcluster C3c appears to be complex, with several common haplotypes present in different populations (Fig. 2). Our analysis revealed that the age of this subcluster is about 5.9 ky or 1.6 ky, whereas the age of subcluster C3d appears to be younger — about 2.0 ky or 0.5 ky, depending on the mutation rate values selected. The age of STR variation within a subset of subcluster C3c formed by DYS19 duplicated alleles is about 3.0 ky or 0.8 ky depending on the mutation rate selected. Note that the latter value, calculated using the genealogical mutation rate, is lower than the time estimates from the study by Balaresque et al. (2009) (1.80 ± 0.63 ky) and Nasidze et al. (2005) (0.86 ky, with 95% confidence interval equal to 0.56–1.16 ky). In any case, all these estimations are largely biased because Y-chromosomes presenting only a single peak (allele) in the electrophoregram can nonetheless be duplicated for the STR (Balaresque et al., 2009). To differentiate non-duplicated from duplicated chromosomes with identical DYS19 alleles, a further search of the subhaplogroup-defining SNPs is required.

STR haplotypes belonging to a heterogeneous group C3* are represented by members of different subclusters with still unknown SNP markers. The best known among them corresponds to the “star cluster”, which was thought to be carried by likely male-line descendants of Genghis Khan, as described...
in Zerjal et al. (2003). It is suggested that this subcluster appears to have originated in Mongolia about 1 ky ago, taking into account the genealogical mutation rate (Zerjal et al., 2003). Our present and previous (Derenko et al., 2007b) studies have shown that the highest frequency of the “star cluster” in C3* is observed in Mongols (35%), whereas in Siberia it varies from 8% in Altai Kazakhs and 6.5% in Buryats to less than 3% in Tuvinians, Altaians and Shors (Table S2). According to our data, the age of the “star cluster” in C3* is 2.8 ± 1.0 or 0.78 ± 0.27 ky, based on the evolutionary and genealogical mutation rates, respectively.

Since the frequency of subhaplogroup C3d is significantly higher in the Mongolic-speaking populations than in Turkic or Tungusic (P < 0.001) and the frequency of subhaplogroup C3c is higher in Tungusic-speaking populations than in Turkic or Mongolic (P < 0.005) (Table 1) and taking into consideration that a statistically significant correlation between linguistic and genetic distances has been found in a previous Y-chromosome haplogroup study of native Siberians (Karafet et al., 2002), we have performed correlation analysis (Mantel test) between genetic and linguistic distances in the populations studied. Genetic distances were based on FST and RST-differences between hg C3 STR haplotypes. The Mantel test analysis did not reveal any significant correlation between the FST-matrix and the matrix of linguistic distances (r = 0.11, P = 0.23). The correlation between RST-genetic and linguistic distances was significant (r = 0.34, P = 0.047), although it is at the border of significance.

### Discussion

Overall, the data obtained indicate that in populations of northern Asia hg C3 is represented mostly by subclusters C3c and C3d, although the frequency of unidentified C3* lineages is also high in some populations, e.g. in Mongols or Koryaks (Table 1). Despite the fact that we were not able to find a significant correlation between linguistic and genetic distances in Siberian populations, a correspondence between genetic and linguistic affiliation is observed in the median joining networks for some sets of C3 haplotypes. For instance, C3d haplotypes are widespread in Mongolic-speaking populations (Fig. 1). A relatively high frequency of subhaplogroup C3d was also found in the Turkic-speaking Sojots (53.6%), so in this respect the Sojots are situated closer to their Mongolic-speaking neighbors, the Buryats. The same trend is also evident for maternal mtDNA lineages (Derenko et al., 2003). However, the Sojots are characterized by a relatively high frequency of the Y-chromosome haplogroup R1a1 (about 25%), which is typical for the Turkic-speaking populations such as Altaians, Teleuts and Shors, all characterized by the highest frequencies of R1A1 (about 50%) in Siberia (Derenko et al., 2006). Therefore, it seems that the Turk males might have contributed genetically to the formation of Sojots, imposing a language of the Turkic group. In this scenario, most likely an elite dominance process should be assumed (Renfrew, 1994).

However, additional studies are required to clarify the relationships between the Sojots and Buryats due to the small sample size of Sojots studied.

On the other hand, the results of subhaplogroup C3c phylogenetic analysis indicate that C3c-haplotypes widespread in Tungusic-speaking populations are clustered together. However, the Evenks, which are equally represented in our study by Western and Eastern Evenks, do not share any haplotypes with Evenks represented here by Eastern Evenks inhabiting the Okhotsk Sea coast. The median joining network (Fig. 2) demonstrates that haplotype 1 is present in Evenks as well as in Buryats, Mongols, Kalmuks, Tuvinians and Yakuts, whereas derived haplotypes 2, 3, 4 and 5 are characteristic only for Evenks. Using the T1b-estimator and different mutation rates, divergence time between the founder Evenk haplotype 1 and derived Even haplotypes is 1.40 ± 1.06 or 3.39 ± 0.29 ky for evolutionary or genealogical mutation rate, respectively. We should note that according to ethnological data the split between the ancestors of the northern Tungus (Evenks and Evens) occurred either about 1.5 ky ago (Vasilevich, 1969) or much earlier, beginning from the 12th and 13th century AD (Tugolukov, 1980). Therefore, due to the uncertainty of mutation rates in STRs of the Y-chromosome, molecular dating results are in accord with both hypotheses on the origin.
of the northern Tungusic groups. Meanwhile, median joining network analysis of C3c haplotypes demonstrates that the Evenk gene pool represents the more ancient genetic variant in comparison to the Evens. However, more detailed Y-STR analyses are necessary for investigation of genetic relationships between populations of north-eastern Asia, taking into account their small sizes, isolation and the effect of genetic drift.

An interesting feature of Y-chromosome gene pools of the populations studied is the presence of DNA clusters originating from a limited number of frequent founder haplotypes. These are the C3∗ “star cluster” that may correspond to the descendants of Genghis Khan (Zerjal et al., 2003), subhaplogroup C3d characteristic for Mongolic-speaking populations, and the C3c cluster with several founding haplotypes, including a remarkable set of DYS19 duplicated haplotypes. Another lineage, the “Manchu cluster”, associated probably with Qing Dynasty nobility, has occurred at a relatively high frequency (about 3.3%) in northeastern China and Mongolia and has arisen about 600 years ago (Xue et al., 2005). Although there are questions about the accuracy of molecular dating of these lineages and their historical assignments, it is nonetheless clear that such Y-chromosome lineages have increased dramatically in frequency, and hence social selection remains the most probable explanation (Zerjal et al., 2003). It has been suggested that social selection may result in high male fertility associated with one paternal lineage, which can potentially have a large impact on the Y-chromosomal gene pool at the population level (Zerjal et al., 2003; Stoneking & Delfin, 2010). However, indigenous groups of Siberia are characterized by small effective population sizes and isolation for long periods of time (Karafet et al., 2002), thus founder effects may explain the observed distribution of hg C3. In addition, one should note that other Y-chromosome haplogroups (e.g., N and Q) also often show evidence of founder effect/social selection in boreal Eurasian populations (Zegura et al., 2004; Derenko et al., 2007a; Rootsi et al., 2007), so this phenomenon requires further examination.

Acknowledgements

The authors would like to thank the reviewers for very constructive comments. We are grateful to all the voluntary donors of DNA samples used in this study. This study was supported by the Program of Presidium of Russian Academy of Sciences “Biodiversity and Gene Pools”, the Russian Foundation for Basic Research (07-04-00445) and the Far-East Branch of Russian Academy of Sciences (10-III-B-06-134).

References


Boris Malyarchuk et al.


**Supporting Information**

Additional supporting information may be found in the online version of this article:

**Table S1** Haplogroup C STR haplotypes in populations of northern Eurasia

**Table S2** STR haplotypes belonging to the “star cluster” in C3

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 15 March 2010
Accepted: 9 June 2010