

Phylogeography of the Y-chromosome haplogroup C in northern Eurasia

Boris Malyarchuk^{1*}, Miroslava Derenko¹, Galina Denisova¹, Marcin Wozniak², Tomasz Grzybowski², Irina Dambueva³ and Ilia Zakharov⁴

¹*Institute of Biological Problems of the North, Russian Academy of Sciences, Portovaya str. 18, 685000 Magadan, Russia*

²*Forensic Medicine Institute, the Ludwik Rydygier Medical College, the Nicolaus Copernicus University in Torun, 85-094 Bydgoszcz, Poland*

³*Institute of General and Experimental Biology, Russian Academy of Sciences, 670047 Ulan-Ude, Russia*

⁴*N.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

*Corresponding author: Dr. Boris A. Malyarchuk, Genetics Laboratory, Institute of Biological Problems of the North, Portovaya str., 18, 685000 Magadan, Russia. Fax/Phone: +7 4132 634463; E-mail: malyarchuk@ibpn.ru

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

Population	N	Linguistic affiliation	C, in total	C3c	C3d	C3*
Altaians	89	Turkic	14 (15.7)	5 (5.6)	0	9 (10.1)
Teleuts	44	Turkic	5 (11.4)	0	1 (2.4)	4 (9.1)
Altaiian Kazakhs	36	Turkic	21 (58.3)	15 (41.7)	0	6 (16.7)
Khakassians	64	Turkic	1 (1.6)	0	0	1 (1.6)
Shors	38	Turkic	1 (2.6)	0	0	1 (2.6)
Todjins	26	Turkic	2 (7.7)	2 (7.7)	0	0
Tuvinians	108	Turkic	12 (11.1)	6 (5.6)	3 (2.8)	3 (2.8)
Yakuts	10	Turkic	2 (20.0)	2 (20.0)	0	0
Sojots	28	Turkic	15 (53.6)	0	15 (53.6)	0
Mongols	46	Mongolic	30 (65.2)	5 (10.9)	7 (15.2)	18 (39.1)
Buryats	217	Mongolic	148 (68.2)	13 (6.0)	117 (53.9)	18 (8.3)
Khamnigans	51	Mongolic	28 (54.9)	0	27 (52.9)	1 (0.5)
Kalmyks	91	Mongolic	57 (62.6)	41 (45.1)	11 (12.1)	5 (5.5)
Evenks	41	Tungusic	20 (48.8)	18 (43.9)	1 (2.4)	1 (2.4)
Evens	63	Tungusic	34 (54.0)	33 (52.4)	0	1 (1.6)
Koryaks	39	Chukotko-Kamchatkan	15 (38.5)	0	0	15 (38.5)
Koreans	52	Korean	5 (9.6)	0	0	5 (9.6)
Russians	406	Indo-European	3 (0.7)	1 (0.2)	0	2 (0.5)

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, Altaiian 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, Foster City, CA, USA). The Y-SNP haplogroup nomenclature used here is according to the recommendations of the Y Chromosome Consortium (Karafet et al., 2008).

A total of 413 samples belonging to hg C were analyzed at twelve STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437,

DYS438, DYS439) using PowerPlex[®] Y System (Promega Corporation, Madison, USA). Several 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, DYS635) using AmpFI-STR[®]Yfiler[®] PCR Amplification kit (Applied Biosystems) according to the manufacturer's instructions. Products of amplification were analyzed on ABI 3100 and ABI 3130 Genetic Analyzers (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

of a mutation rate (Zhivotovsky et al., 2004; Sengupta et al., 2006). A mutation rate of 2.5×10^{-3} per 35 years calculated in father-son pairs (Goedbloed et al., 2009) and the evolutionary effective mutation rate of 6.9×10^{-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 1000 permutation steps. χ^2 analysis of haplogroup frequencies in populations was performed by means of the program CHIRXC, which estimates the probability of homogeneity using Monte Carlo simulation (1000 runs) (Zaykin & Pudovkin, 1993).

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 was detected at a frequency of 0.7%.

Analysis of SNP markers revealed that all Y-chromosomes studied belong to hg C3 and its subhaplogroups C3c and C3d (Table 1). Phylogenetic membership of some C3-chromosomes remains unclear, so they belong to C3* (xC3a, C3b, C3c, C3d, C3e, C3f). However, the frequency of this type of Y-chromosome is low, being equal to 6.2% in the overall data set. The Mongols and Koryaks are characterized by the highest frequency of C3* (39.1% and 38.5%, respectively), but they differed substantially at the level of STR-haplotypes. As for the C3c subhaplogroup, the highest frequencies of C3c were observed in linguistically diverse set of populations: in Tungusic-speaking Evens and Evenks (52.4% and 43.9%, respectively), in Mongolic-speaking Kalmyks (45.1%) and in Turkic-speaking Altaian Kazakhs (58.3%). It has been reported previously that many aboriginal peoples of northeastern Asia and the Far East region, such as Evens, Evenks, Negidals, Itelmens, Udegey, Ulchi, Nanai, Nivkhs, are characterized by a high frequency (50%, on average) of subcluster C3c (Karafet et al., 2002; Lell

et al., 2002; Pakendorf et al., 2007). According to Wells et al. (2001) and Zerjal et al. (2002), a high frequency of C3c (about 50%) has also been observed in Central Asia (in Kazakhs) and Eastern Asia (in Mongols). Therefore, our results are in general agreement with the above data. The only exception is the Yakut sample where C3c haplotypes were found at a frequency of 20%, in contrast to about 2% reported in previous studies (Pakendorf et al., 2007; Khar'kov et al., 2008). This is most likely due to the small sample size of Yakuts in our study.

Typing of the Y-chromosome marker M407 allowed us to clarify the phylogeographic pattern of subcluster C3d in northern Eurasia. Previously, C3d haplotypes have rarely been observed in individuals from Yakutia and China (Sengupta et al., 2006), but our study demonstrates that subcluster C3d 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 Tadjins, 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 Kalmyks a high proportion of duplicated DYS19 alleles is associated with deletion of the locus DYS448. Using a 17 STR amplification kit (AmpFI-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 Tadjins 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

Table 2 Population origins and allelic spectrum of DYS19 duplications on the background of subcluster C3c

Population	DYS19 duplications						Number of C3c Y-chromosomes
	15,16	15,17	16,17	16,18	17,18	In total	
Kalmyks	0	3	21	0	1	25	41
Mongols	0	0	2	0	0	2	5
Tuvinians	0	0	2	1	0	3	6
Todjins	0	0	2	0	0	2	2
Altaian Kazakhs	1	9	3	0	1	14	15
Altaians	0	3	1	0	0	4	5

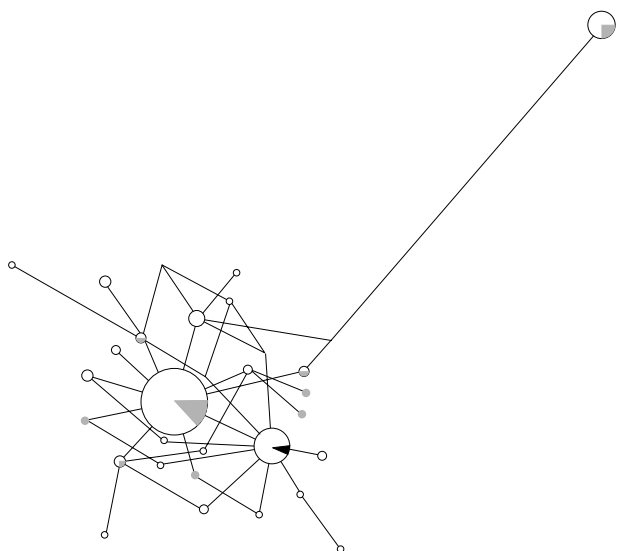


Figure 1 Median joining network of subhaplogroup C3d based on twelve STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) in populations of northern Asia. Each circle represents a haplotype, defined by a combination of STR markers. Circle size 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 Tuviniian samples form a separate branch within C3d, shown at the periphery of the median joining network.

Altaian 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 Balaesque 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 (Balaesque 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

Table 3 The age of STR variation within hg C3 in northern Asia based on STR diversity

Haplogroups	Time estimates with evolutionary mutation rate ^a	Time estimates with genealogical mutation rate ^b
C3	14.92 ± 3.83	4.12 ± 1.06
C3c	5.94 ± 2.90	1.63 ± 0.80
C3d	1.95 ± 1.26	0.54 ± 0.35
C3c (DYS19 duplicated)	2.98 ± 1.21	0.82 ± 0.33

^aAges of haplogroups (in ky) calculated by the method of Zhivotovsky et al. (2004) with the evolutionary effective mutation rate equal to 6.9×10^{-4} per 25 years (Zhivotovsky et al., 2004) and genealogical mutation rate equal to 2.5×10^{-3} per 35 years (Goedbloed et al., 2009). STR loci DYS19, DYS385a and DYS385b were excluded from analysis.

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 Altaian 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 F_{ST} and R_{ST} -differences between hg C3 STR haplotypes. The Mantel test analysis did not reveal any significant correlation between the F_{ST} -matrix and the matrix of linguistic distances ($r = 0.11$, $P = 0.23$). The correlation between R_{ST} -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 Evens represented here by Eastern Evens 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, Kalmyks, Tuvinians and Yakuts, whereas derived haplotypes 2, 3, 4 and 5 are characteristic only for Evens. Using the T_D -estimator and different mutation rates, divergence time between the founder Evenk haplotype 1 and derived Even haplotypes is 1.40 ± 1.06 or 0.39 ± 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

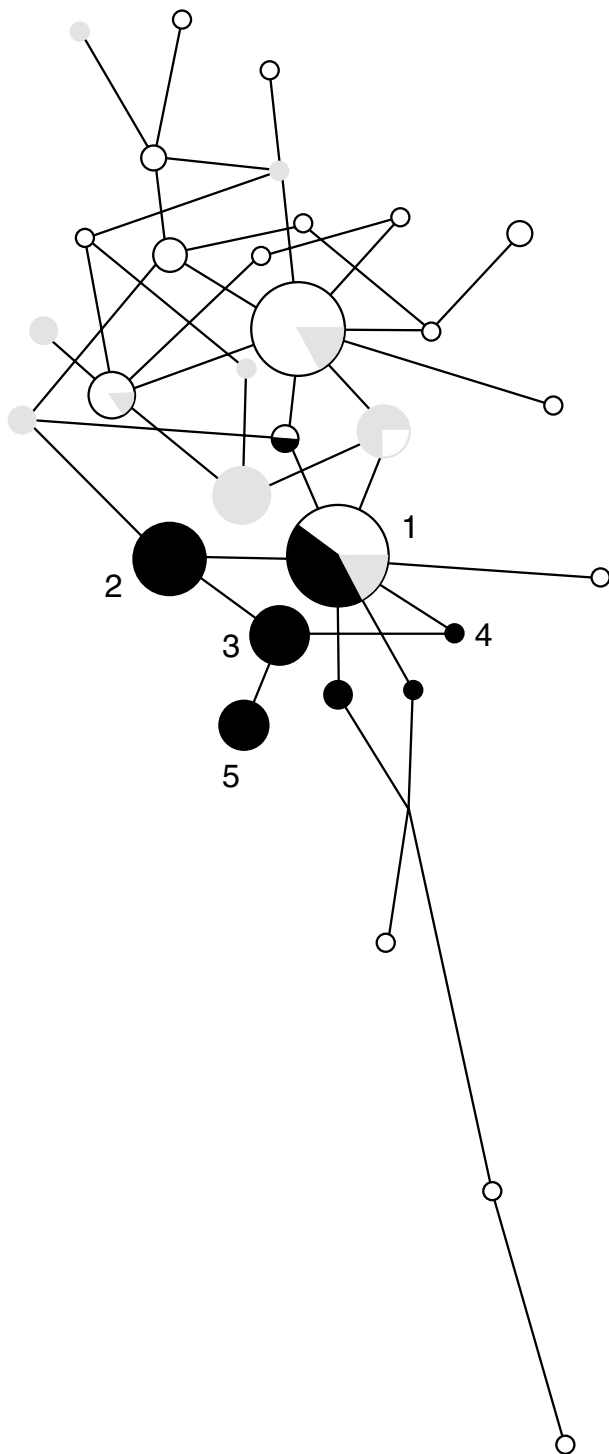


Figure 2 Median joining network of subhaplogroup C3c based on nine STR loci (DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) in populations of northern Asia. The numbers attached to some circles are haplotypes discussed in the text. Remaining designations are as in Figure 1.

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

Balaresque, P., Parkin, E. J., Roewer, L., Carvalho-Silva, D. R., Mitchell, R. J., van Oorschot, R. A., Henke, J., Stoneking, M.,

- Nasidze, I., Wetton, J., de Knijff, P., Tyler-Smith, C. & Jobling, M. A. (2009) Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications. *Int. J. Legal Med.* **123**, 15–23.
- Bandelt, H.-J., Macaulay, V. & Richards, M. (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16**, 8–28.
- Bergen, A. W., Wang, C. Y., Tsai, J., Jefferson, K., Dey, C., Smith, K. D., Park, S. C., Tsai, S. J. & Goldman, D. (1999) An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* **63**, 63–80.
- Derenko, M. V., Grzybowski, T., Malyarchuk, B. A., Dambueva, I. K., Denisova, G. A., Czarny, J., Dorzhu, C. M., Kakpakov, V. T., Miścicka-Sliwka, D., Woźniak, M. & Zakharov, I. A. (2003) Diversity of mitochondrial DNA lineages in South Siberia. *Ann. Hum. Genet.* **67**, 391–411.
- Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Dambueva, I., Dorzhu, C., Luzina, F., Miścicka-Sliwka, D. & Zakharov, I. (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altay-Sayan regions. *Hum. Genet.* **118**, 591–604.
- Derenko, M., Malyarchuk, B., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I. & Zakharov, I. (2007a) Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J. Hum. Genet.* **52**, 763–770.
- Derenko, M. V., Malyarchuk, B. A., Wozniak, M., Denisova, G. A., Dambueva, I. K., Dorzhu, C. M., Grzybowski, T. & Zakharov, I. A. (2007b) Distribution of the male lineages of Genghis Khan's descendants in northern Eurasian populations. *Russ. J. Genet.* **43**, 422–426.
- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.
- Goedbloed, M., Vermeulen, M., Fang, R. N., Lembring, M., Wollstein, A., Ballantyne, K., Lao, O., Brauer, S., Krüger, C., Roewer, L., Lessig, R., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M. R. & Kayser, M. (2009) 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.
- Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. & Horai, S. (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* **51**, 47–58.
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. (2008) New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. *Genome Res.* **18**, 830–838.
- Karafet, T. M., Osipova, L. P., Gubina, M. A., Posukh, O. L., Zegura, S. L. & Hammer, M. F. (2002) 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.
- Katoh, T., Munkhbat, B., Tounai, K., Mano, S., Ando, H., Oyungerel, G., Chae, G. T., Han, H., Jia, G. J., Tokunaga, K., Munkhtuvshin, N., Tamiya, G. & Inoko, H. (2005) Genetic feature of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene* **346**, 63–70.
- Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovskiy, L. A., Moysse-Faurie, C., Rutledge, R. B., Schiefenhövel, W., Gil, D., Lin, A. A., Underhill, P. A., Oefner, P. J., Trent, R. J. & Stoneking, M. (2006) Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol. Biol. Evol.* **23**, 2234–2244.
- Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M. & Stoneking, M. (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302.
- Khar'kov, V. N., Stepanov, V. A., Medvedev, O. F., Spiridonova, M. G., Maksimova, N. R., Nogovitsyna, A. N. & Puzyrev, V. P. (2008) The origin of Yakuts: analysis of Y-chromosome haplotypes. *Mol. Biol. (Mosk)*. **42**, 226–237.
- Kittler, R., Erler, A., Brauer, S., Stoneking, M. & Kayser, M. (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur. J. Hum. Genet.* **11**, 304–314.
- Lell, J. T., Sukernik, R. I., Starikovskaya, Y. B., Su, B., Jin, L., Schurr, T. G., Underhill, P. A. & Wallace, D. C. (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* **70**, 192–206.
- Nasidze, I., Quinque, D., Dupanloup, I., Cordaux, R., Kokshunova, L. & Stoneking, M. (2005) Genetic evidence for the Mongolian ancestry of Kalmyks. *Am. J. Phys. Anthropol.* **128**, 846–854.
- Niederstätter, H., Berger, B., Oberacher, H., Brandstätter, A., Huber, C. G. & Parson W. (2004) Separate analysis of DYS385a and b versus conventional DYS385 typing: is there forensic relevance? *Int. J. Legal Med.* **119**, 1–9.
- Pakendorf, B., Novgorodov, I. N., Osakovskij, V. L. & Stoneking, M. (2007) Mating patterns amongst Siberian reindeer herders: inferences from mtDNA and Y-chromosomal analyses. *Am. J. Phys. Anthropol.* **133**, 1013–1027.
- Renfrew, C. (1994) World linguistic diversity. *Sci. Am.* **270**, 104–110.
- Roewer, L., Krüger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L., Rothämel, T., Kravchenko, S., Jobling, M. A., Stoneking, M. & Nasidze, I. (2007) Y-chromosomal STR haplotypes in Kalmyk population samples. *Forensic Sci. Int.* **173**, 204–209.
- Rootsi, S., Zhivotovskiy, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R., Bermisheva, M. A., Gubina, M., Fedorova, S. A., Ilumaa, A. M., Khusnutdinova, E. K., Voevoda, M. I., Osipova, L. P., Stoneking, M., Lin, A. A., Ferak, V., Parik, J., Kivisild, T., Underhill, P. A. & Villems, R. (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* **15**, 204–211.
- Sengupta, S., Zhivotovskiy, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E., Lin, A. A., Mitra, M., Sil, S. K., Ramesh, A., Usha Rani, M. V., Thakur, C. M., Cavalli-Sforza, L. L., Majumder, P. P. & Underhill, P. A. (2006) Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am. J. Hum. Genet.* **78**, 202–221.
- Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W. H., Roxas, A., Sung, R., Lin, A. A., Hyman, R. W., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000) Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **97**, 7354–7359.
- Stoneking, M. & Delfin, F. (2010) The human genetic history of East Asia: weaving a complex tapestry. *Curr. Biol.* **20**, R188–193.

- Tugolukov, V. A. (1980) Tungus ethnic roots. In: *Ethnogenesis of peoples of the North* (ed. I. S. Gurvich), pp. 152–176. Moscow: Nauka (in Russian).
- Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005.
- Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., Oefner, P. J. & Cavalli-Sforza, L. L. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62.
- Vasilevich, G. M. (1969) *The Evenks. Historico-ethnographical studies (XVIII-beginning of XX centuries)*. Leningrad: Nauka (in Russian).
- Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamar-jashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A. & Bodmer, W. F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* **98**, 10244–10249.
- Xue, Y., Zerjal, T., Bao, W., Zhu, S., Lim, S. K., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., et al. (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia. *Am. J. Hum. Genet.* **77**, 1112–1116.
- Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., Hurles, M. E., Yang, H. & Tyler-Smith, C. (2006) Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* **172**, 2431–2439.
- Zaykin, D. V. & Pudovkin, A. J. (1993) Two programs to estimate significance of Chi-square values using pseudo-probability test. *J. Heredity* **84**, 152.
- Zegura, S. L., Karafet, T. M., Zhivotovsky, L. A. & Hammer, M. F. (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol. Biol. Evol.* **21**, 164–175.
- Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* **71**, 466–482.
- Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., Li, P., Yuldasheva, N., Ruzibakiev, R., Xu, J., Shu, Q., Du, R., Yang, H., Hurles, M. E., Robinson, E., Gerelsaikhan, T., Dashnyam, B., Mehdi, S. Q. & Tyler-Smith, C. (2003) The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72**, 717–721.
- Zhivotovsky, L. A. (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol. Biol. Evol.* **18**, 700–709.
- Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G. K., Herrera, R. J., Yong, K. K., Gresham, D., Tournev, I., Feldman, M. W. & Kalaydjieva, L. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **74**, 50–61.

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