

The Origins of the Yakut People: Evidence from Mitochondrial DNA Diversity

Mark Zlojutro¹, Larissa A. Tarskaia², Mark Sorensen³, J. Josh Snodgrass⁴,
William R. Leonard⁵ and Michael H. Crawford¹

1. Department of Anthropology, University of Kansas, Lawrence, Kansas 66045, USA

*2. Institute of Molecular Genetics, Russian Academy of Medical Sciences,
Moscow 115478, Russia*

*3. Department of Anthropology, University of North Carolina, Chapel Hill,
North Carolina 27599, USA*

4. Department of Anthropology, University of Oregon, Eugene, Oregon 97403, USA

5. Department of Anthropology, Northwestern University, Evanston, Illinois 60208, USA

KEYWORDS Yakut origins; mitochondrial DNA; RFLP analysis; nested cladistic analysis (NCA)

ABSTRACT The Yakuts are a Turkic-speaking population of northeastern Siberia and based on archaeological and ethnohistorical evidence are believed to have originated from Turkic populations in south Siberia. To investigate this model, the HVS-I of the mitochondrial DNA control region was sequenced for 144 Yakut individuals representing seven communities from central Yakutia and compared to HVS-I data for other Asian populations. Haplogroups C and D comprise 75.7% of the Yakut sample, with only 9.7% assigned to west Eurasian lineages. The Ewens-Watterson homozygosity test revealed a significant deviation ($P = 0.045$) in the observed frequencies of common haplotypes relative to the expected values, indicating the genetic effects of a founder event. This is supported by a fragmented MJ network dominated by high-frequency haplotypes within haplogroups C and D. Nested cladistic analysis identified subhaplogroup D5a as the product of a long distance colonization event and potential founder lineage for the Yakuts, dating to approximately 1,630 years BP. SAMOVA analyses and MDS plot of genetic distances show close genetic affinities between the Yakuts and south Siberian populations, and thus affirming the south origin model.

INTRODUCTION

The Yakuts are a Siberian population that primarily lives in the Republic of Sakha (also known as Yakutia), an autonomous region within the Russian Federation. They are traditionally cattle- and horse-breeders who speak a Turkic language, which differs from the subsistence patterns and languages of neighboring populations. The origins of the Yakuts have been the subject of much speculation that has been investigated from different lines of inquiry. The Yakut language is considered an isolate within the Turkic language family, but shares features with Turkic languages spoken in south Siberia, such as Tuvan (Schönig 1990). Furthermore, aspects of Yakut culture (e.g., pastoralism, clothing, festivals) suggest ancestral ties to southern Turkic peoples (Tokarev and Gurvich 1956). Interestingly, archaeological remains of a semi-nomadic population from the Lake Baikal region, identified as the Turkic-speaking Kurykan people, exhibit close affinities to Yakut culture. Based on these similarities, it has been postulated that

a group of Kurykans migrated northwards along the Lena River to escape Mongol incursions during the 11th to 13th centuries AD and ultimately forming the Yakuts of northeastern Siberia (Alekseev 1996; Kostanstinov 1975; Okladnikov 1970).

Studies based on classical genetic polymorphisms have revealed genetic affinities between the Yakuts and various Tungusic- and Turkic-speaking populations in Siberia (Boeva 1988; Fefelova 1990; Matsumoto 1988; Novoradovsky et al. 1993; Pakendorf et al. 1999; Tarskaia et al. 2002). More recent investigations of mtDNA and Y-chromosome variation in the Yakuts have been conducted. Federova et al. (2003) sequenced the HVS-I region for a sample representing communities throughout the republic and found evidence supporting a Mongol/Central Asian origin. In a paper by Puzyrev et al. (2003), a sample collected from the Yakut settlement of Chiriktei was analyzed for mtDNA and Y-chromosome markers and found predominantly east Eurasian haplotypes. And in a study by Pakendorf et al. (2006), samples for

the Yakuts, as well as the Yukaghirs and Evenks, representing a number of villages within the Sakha Republic and neighboring districts were analyzed for mtDNA and Y-chromosome variation. The mtDNA results support a southern origin for the Yakuts, whereas the Y-chromosomal data showed evidence for a strong founder effect and unknown origins for the majority of Yakut paternal lineages. In this paper we contribute to this body of genetic information by characterizing the mtDNA variation in seven communities from central Yakutia and perform inter-population analyses to further elucidate the genetic structure of the Yakuts and their origins.

MATERIALS AND METHODS

Samples and DNA Extraction: Yakut blood samples were collected during 2000 from seven

communities located in the central Lena River watershed of the Sakha Republic: Asyma, Berdigestiakh, Dikyimdye, Khorobut, Maia, Nizhny Bestiakh, and Viljujsk (Fig. 1). All participants provided informed consent, and details such as name, gender and age were recorded. The use of these samples was approved by the University of Kansas Human Research Committee. DNA was extracted from whole blood using the Super Quik-Gene DNA isolation kit (LBA, University of Kansas), a salting-out methodology.

The following population data were compiled from mtDNA studies for comparative analyses (Fig. 1): Evenk and Yukaghir (Pakendorf et al. 2006), Even (Derenko and Shields 1997), Koryak and Itelmen (Schurr et al. 1999), Mongol (Kolman et al. 1996), Oroqen and Daur (Kong et al. 2003), Han (Yao et al. 2002), Uighur, Kazakh and Kirghiz



Fig. 1. Map of Asia showing the approximate geographic locations of the Yakut samples and comparative populations analyzed in this study.

(Comas et al. 1998), Uzbek (M. Zlojutro and R. Devor, unpublished data), and south Siberian Buryat, Tuva, Khakass and Altai (Derenko et al. 2003).

mtDNA Sequencing and RFLP Analyses: A total of 144 Yakut samples were amplified with primers L15976 and H16401, using the PCR conditions described by Vigilant et al. (1989). The HVS-I amplicons were purified using the QIAquick Purification Kit (Qiagen) and sequenced for the light strand at Integrated DNA Technologies (Coralville, Iowa) using the Big Dye Terminator ver. 3.1 Cycle Sequencing Kit on the ABI PRISM 310 Genetic Analyzer (Applied Biosystems). The sequences were aligned with BioEdit software (Ibis Therapeutics) against a standard mtDNA sequence known as the Cambridge Reference Sequence (CRS) (Anderson et al. 1981). The HVS-I output ranged from approximately np 16020 to 16400, however the sequence information for np 16020 to 16049 lacked consistent quality within the data set and was thus excluded from further analysis. Based on network analyses (see below), samples that exhibit unusual HVS-I profiles (e.g., excessive transversions and inter-haplogroup parallelisms) were re-sequenced for both strands to check for potential errors. Sequencing problems were encountered with Yakut samples exhibiting a 16189 T→C transition. This HVS-I variant, which is a mutation hotspot and common to several different haplogroups, creates a homopolymer of 10 cytosines that causes DNA polymerases to “stutter” during the sequencing PCR reaction, resulting in poor sequence data beyond this site. Therefore, samples carrying this transition were sequenced in both directions, producing overlapping data that were combined to create continuous sequences for np 16050 to 16400.

All sequences were assigned to haplogroups on the basis of their HVS-I motifs in relation to sequence data published by Kivisild et al. (2002) and Derenko et al. (2003). Assignments were confirmed through PCR-RFLP analyses of diagnostic sites for macrohaplogroups M and N and the major east Eurasian haplogroups A, C, D, F and G. Primer sequences, PCR conditions and enzyme information are provided in Table 1. Restriction reactions were performed according to manufacturer recommendations (New England Biolabs) and typed on 3% 3:1 NuSieve agarose gels (ISC BioExpress). Subhaplogroup D5, a common mtDNA lineage in East Asian

populations (Kolman et al. 1996; Kong et al. 2003; Yao et al. 2002), was identified in the Yakuts by the loss of 10394 *DdeI* and 10397 *AluI* and the concurrent gain of a 10393 *Bmrl* cut site. The intergenic 9 bp COII-tRNA^{Lys} deletion, a diagnostic polymorphism for haplogroup B (Torroni et al. 1992, 1993; Wrischnik et al. 1987), was typed by gel electrophoresis of the L8239/H8363 PCR amplicon and confirmed by direct sequencing.

Statistical Methods: Measures of sequence variation – haplotype diversity, θ_s and θ_π – were calculated for the Yakut sample and comparative population data using the software program Arlequin ver. 2.0 (Schneider et al. 2000). In addition, Nei’s intermatch-mismatch distances (D_A) and the neutrality test statistics Tajima’s D , Fu’s F_S and Ewens-Watterson homozygosity test (Watterson 1978) were calculated with this program. The significance of these values was evaluated with 1,000 permutations. Multi-dimensional scaling (MDS) analyses of pairwise D_A values were conducted using the NTSYS statistical package (Rohlf 2002). A spatial analysis of molecular variance was performed with the program SAMOVA ver. 1.0 (Dupanloup et al. 2002) to detect the presence of genetic barriers in the population HVS-I data. Median-joining (MJ) networks (Bandelt et al. 1999) were constructed for each major haplogroup using Network ver. 4.1 (Fluxus Engineering) and incorporated into a single network based on the known Eurasian mtDNA phylogeny (Kivisild et al. 2002; Macauley et al. 1999). Nested cladistic analysis (NCA) (Templeton et al. 1995) was performed with the program GeoDis ver. 2.0 (Posada et al. 2000) to identify any significant associations between haplogroups and geography and infer the effects of population structure (e.g., recurrent gene flow restricted by the Isolation by Distance model) and/or population history (e.g., contiguous range expansion, long distance colonization). Yakut haplogroups identified as founder lineages were dated by converting ρ (rho) statistics calculated in Network 4.1 into time (years BP) using different HVS-I mutation rate estimates.

RESULTS

mtDNA Analyses

The HVS-I sequences for 144 Yakuts harbor 53 different haplotypes (k) characterized by 64

segregating sites (the sequences were submitted to the online database HvrBase and are available at <http://www.hvrbase.org>). The haplotype diversity for the HVS-I data is 0.955, and 11.365 and 6.025 for the θ_s and θ_π estimators, respectively (Table 2). These diversity values are intermediate to those exhibited by northeastern Siberian populations and the elevated diversity levels displayed by Asian populations further to the south. The Itelmen and Koryaks of the Kamchatkan Peninsula and the Evenks possess the lowest values, whereas the Han have the highest.

The neutrality test statistics Tajima's D and Fu's F_s for the Yakut sequences are -1.462 and -

25.025, respectively, both of which are significant at the 0.05 level (Table 3). Like the diversity measures, the Tajima's D score is approximately intermediate to the values displayed by northeastern Siberians and Asian populations to the south, whereas the Fu's F_s for the Yakuts is similar to the large negative values exhibited by other Turkic-Mongol populations. For selectively neutral systems such as the HVS-I, negative scores for these two particular measures have been typically interpreted as evidence for the genetic effects of population growth (Fu 1997; Ramos-Onsins and Rozas 2002; Tajima 1989). However, most human populations display negative scores, with the exception of certain

Table 1: Primers and restriction enzymes used in PCR-RFLP analyses.

Haplogroup(s)	Primer Sequences (5'→3')	T_m (°C)	Enzyme	Size
M/N, D5	L10270: CCTCTTTACCCCTACCATGAGH10579: ATTATTCCTTCTAGGCATAGTAG	57 ^a	<i>AluI</i> , <i>Bmrl</i> , <i>DdeI</i>	310
A	L535: CCCATACCCGAACCAACCH725: GGTGAACTCACTGGAAGGGG	57 ^a	<i>HaeIII</i>	191
B	L8239: CTTTGAATAGGGCCCGTH8363: CACTGTAAAGAGGTGTTGG	50 ^a	–	125
C	L13204: ACTCTGTTTCGAGCAGH13335: GTGCAGGAATGCTAGG	55 ^b	<i>AluI</i> , <i>HincII</i>	132
D	L5147: AAACCTCCAGCACACGH5261: CAAAAAGCCGGTTAGC	55 ^b	<i>AluI</i>	121
F	L12217: CAAGAACTGCTAACTCATGCCH12595: TGGAGAAAGTAGTCTAGTTTGAA	55 ^a	<i>HpaI</i>	379
G	L4738: CCAATCAATACTATCATTATAATCH4972: CCTCAACTGCCTGCTATG	61 ^a	<i>HaeIII</i> , <i>HhaI</i>	235

^a PCR conditions: 94°C for 1 minute (1 cycle); 94°C for 40 seconds (35 cycles); annealing temperature for 30 seconds (35 cycles); 72°C for 45 seconds (35 cycles); and 72°C for 5 minutes (1 cycle) (Rubicz *et al.* 2003)

^b PCR conditions: 94°C for 1 minute (1 cycle); 94°C for 1 minute (35 cycles); 55°C for 1 minute (35 cycles); 72°C for 1 minute (35 cycles); and 72°C for 5 minutes (1 cycle) (Lalueza *et al.* 1997)

Table 2: HVS-I diversity values for the Yakuts and comparative populations.

Populations	n	Language	Haplotype Diversity	θ_s	θ_π
Yakut	144	Turkic	0.955 (0.007)	11.365 (2.912)	6.025 (3.196)
<i>Northeastern Siberia</i>					
Even	65	Tungus	0.953 (0.012)	7.378 (2.264)	5.978 (3.202)
Evenk	39	Tungus	0.923 (0.020)	6.150 (2.115)	4.821 (2.672)
Itelmen	46	Kamchatkan	0.921 (0.022)	5.233 (1.795)	4.059 (2.290)
Koryak	147	Kamchatkan	0.923 (0.011)	6.829 (1.873)	5.583 (2.985)
Yukaghir	22	Yukaghir	0.935 (0.032)	7.132 (2.685)	5.355 (2.996)
<i>South Siberia</i>					
Altai	110	Turkic	0.982 (0.004)	14.413 (3.758)	6.581 (3.470)
Buryat	91	Mongol	0.990 (0.004)	13.969 (3.767)	5.330 (2.876)
Khakass	53	Turkic	0.972 (0.009)	10.137 (3.097)	6.488 (3.460)
Tuva	90	Turkic	0.964 (0.008)	12.620 (3.442)	6.336 (3.359)
<i>East Asia</i>					
Daur	45	Mongol	0.979 (0.009)	12.807 (3.941)	6.665 (3.558)
Han	250	Sino-Tibetan	0.999 (0.001)	20.831 (4.650)	7.667 (3.968)
Mongol	103	Mongol	0.995 (0.002)	14.403 (3.796)	6.435 (3.402)
Oroqen	44	Tungus	0.948 (0.015)	9.195 (2.936)	6.151 (3.310)
<i>Central Asia</i>					
Kazakh	55	Turkic	0.990 (0.006)	13.769 (4.057)	6.540 (3.482)
Kirghiz	94	Turkic	0.990 (0.004)	15.444 (4.107)	6.234 (3.309)
Uighur	55	Turkic	0.993 (0.005)	13.769 (4.057)	5.848 (3.148)
Uzbek	47	Turkic	0.996 (0.005)	12.679 (3.873)	5.876 (3.172)

Standard deviation provided in parantheses

Table 3: Neutrality test statistics for HVS-I data for the Yakuts and comparative populations.

Populations	Tajima's <i>D</i>	Fu's <i>F_s</i>	Σp^2
Yakut	-1.462*	-25.025**	0.051*
<i>Northeastern Siberia</i>			
Even	-0.617	-9.342**	0.061
Evenk	-0.738	-3.097	0.101
Itelmen	-0.737	-4.266	0.099
Koryak	-0.545	-11.105*	0.083
Yukaghir	-0.947	-2.838	0.107
<i>South Siberia</i>			
Altai	-1.752*	-24.997**	0.027
Buryat	-2.026**	-25.391**	0.021
Khakass	-1.227	-12.010**	0.047
Tuva	-1.624*	-22.599**	0.046
<i>East Asia</i>			
Daur	-1.690*	-16.050**	0.043
Han	-1.941**	-24.419**	0.005**
Mongol	-1.794**	-25.054**	0.015
Oroqen	-1.147	-5.141*	0.073
<i>Central Asia</i>			
Kazakh	-1.812*	-25.149**	0.028
Kirghiz	-1.957**	-25.129**	0.021*
Uighur	-1.985**	-25.310**	0.025
Uzbek	-1.879**	-25.304**	0.025

Significance levels: * $P < 0.05$; ** $P < 0.01$

hunter-gatherer groups. Excoffier and Schneider (1999) have interpreted this global pattern as a signal of a Pleistocene expansion that has been erased in some hunter-gatherers due to post-Neolithic bottlenecks. But in a recent simulation

study conducted by Ray et al. (2003), it was shown that the degree of gene flow experienced by a population can significantly impact the scores of Tajima's *D* and Fu's *F_s*, which suggests that the generally lower densities and migration rates of hunter-gatherers is the source for their non-significant negative scores for mtDNA diversity. Hence, the evolutionary meaning of these two neutrality tests is rather ambiguous.

On the other hand, the significant Yakut value computed for the Ewens-Watterson homozygosity test is more informative. Along with the Yakuts, only the Han and Kazakhs exhibit significant values. This statistical method tests for deviations in the observed haplotype frequency spectrum relative to the Ewens sampling distribution (Ewens 1972) by comparing the observed homozygosity (Σp^2) to the expected homozygosity of the Ewens distribution. Two main types of deviations can be observed: (1) the most common haplotypes are at higher frequencies than expected; and (2) common haplotypes have lower frequencies, with intermediate frequency haplotypes at higher levels (Halliburton 2004). Both deviations may be due to natural selection or demographic processes. In Figure 2, the observed haplotype frequency spectrum for the Yakut sequences is

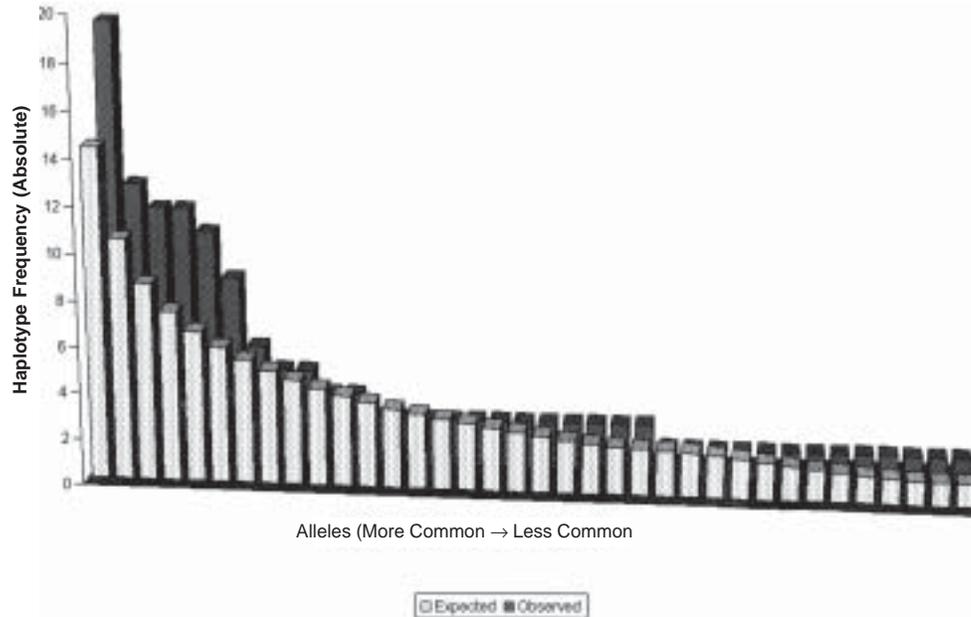


Fig. 2. The expected Ewens haplotype frequency spectrum plotted against the observed haplotype frequency spectrum for Yakut HVS-I sequences.

plotted against the expected frequencies (given $k = 53$ and $n = 144$), revealing six of the most common Yakut haplotypes with frequencies greater than expected. The expected homozygosity is 0.0384, which is significantly less than the observed value of 0.0513 ($P = 0.045$) and due to the type 1 deviation for the common haplotypes. Based on the current understanding of Yakut origins, this deviation likely stems from a founder effect associated with the ancestral migration to northeastern Siberia that may have been dominated by a limited number of maternal lineages.

The mismatch distribution for the Yakut sequences is primarily unimodal with a peak of six pairwise differences (Fig. 3), which is considered a signature of population expansion (Rogers and Harpending 1992). The majority of the comparative populations in this study also exhibit unimodal mismatch distributions (results not shown). The coalescent date for the Yakut HVS-I data derived from Rogers and Harpending's mismatch model (based on the τ parameter calculated by a non-linear least-squares approach in Arlequin and using a conservative mutation rate of 16.5% per nucleotide per million years) is approximately 58,800 years BP (95% CI: 36,000 to 75,700). This result is generally consistent with previous studies that have dated HVS-I data for Eurasian populations using the mismatch model, indicating a widespread population expansion of anatomically modern humans some time during the Paleolithic period (Rogers 1995; Sherry et al. 1994). Thus, the unimodality of the Yakut mismatch

distribution is a product of ancient demographic processes that is not population-specific.

The MJ network for Yakut HVS-I sequences is presented in Figure 4. Fifteen haplotypes, representing 67 individuals or 46.5% of the Yakut sample, belong to haplogroup C based on the absence of 13259 *HhaI*, the presence of 13262 *AluI*, and the HVS-I motif of 16223T, 16298C and 16327T. The most common haplotype in the sample is the root for subhaplogroup C4a that is exhibited by 19 individuals (13.2%), which is defined by the transitions 16171G, 16344T and 16357C. This haplotype is shared with many Asian populations, including the neighboring Evenks and Yukaghirs, as well as various Mongol-Turkic populations from south Siberia.

Seventeen haplotypes, representing 42 Yakut individuals (29.1%), were assigned to haplogroup D on the absence of 5176 *AluI* and HVS-I transitions 16223T and 16362C. Three of the haplotypes belong to a fragmented subhaplogroup D5a (9.7%) based on the loss of two diagnostic restriction sites for macrohaplogroup M, 10394 *DdeI* and 10397 *AluI* (both of which are due to a 10397G transition that concomitantly creates a *Bmrl* restriction site), and transitions 16189C and 16266T. Subhaplogroup D5a is predominantly found among south Siberians and East Asian groups such as the Han, although at low frequencies (about 1-3%). The second most common HVS-I sequence (12 Yakut individuals or 8.3%) is the D5a lineage exhibiting the transitions 16092C and 16172C and is evolutionarily removed from the other two D5a

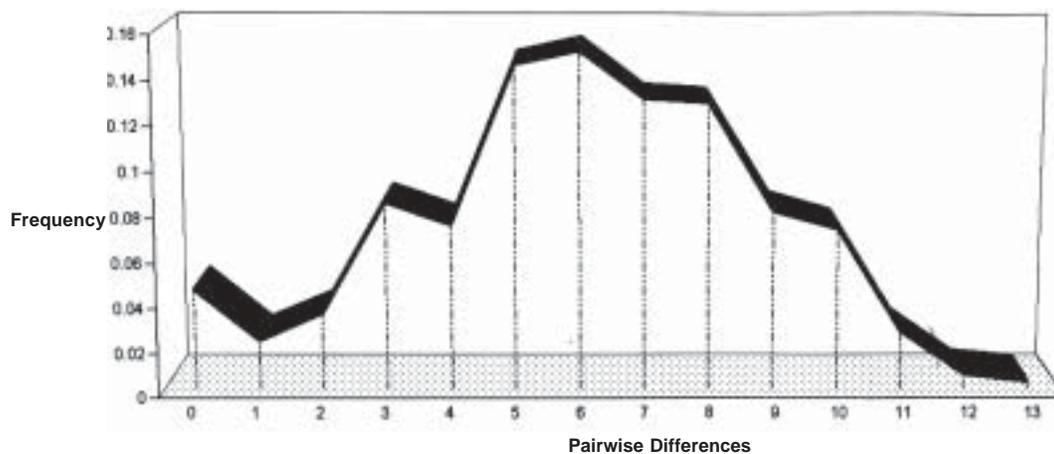


Fig. 3. Mismatch distribution for Yakut HVS-I sequences.

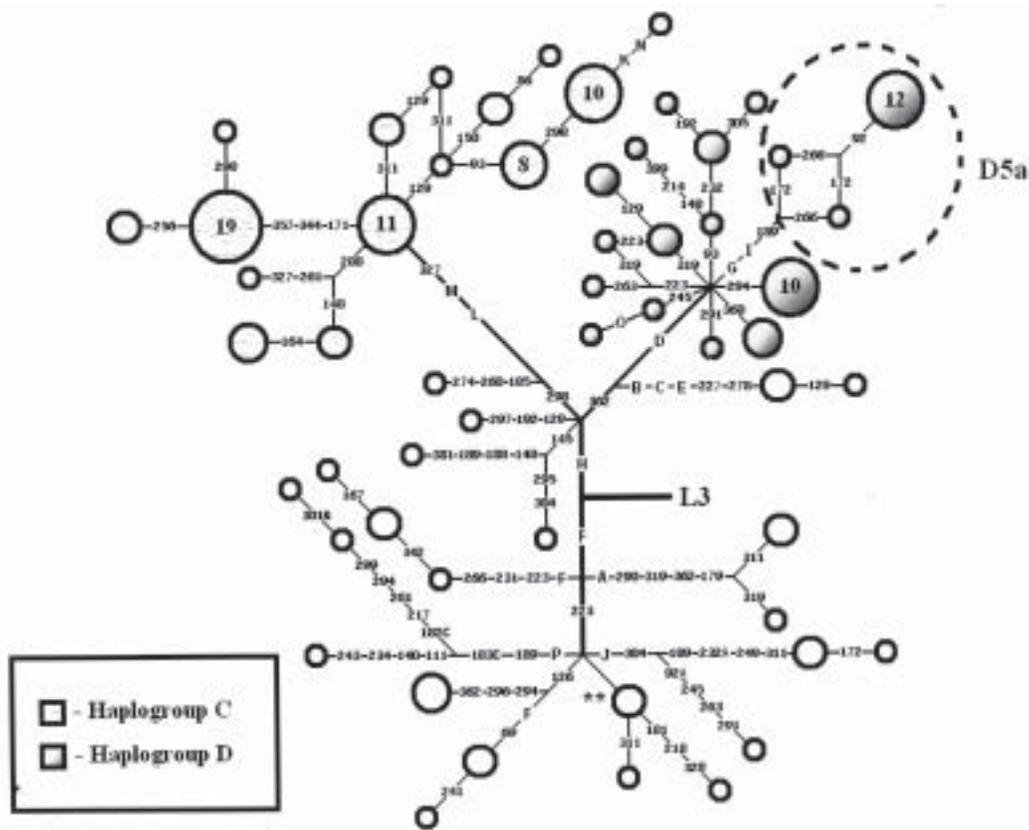


Fig. 4. Median-joining (MJ) network for Yakut HVS-I sequences. Each circle represents an observed haplotype, with its area proportional to its frequency, and the connecting lines describe mutation pathways. Transitions are abbreviated to the last three digits of the nucleotide position (np 16050-16400) and followed by a letter for transversion events to indicate the result of the nucleotide change. The haplotype with the double asterisk (**) represents the Cambridge Reference Sequence. Diagnostic RFLPs were characterized to resolve some of the haplogroup assignments. The RFLPs are represented by the following letters: A (+) 663 *HaeIII*; B (+) 4830 *HaeII*; C (+) 4831 *HhaI*; D (-) 5176 *AluI*; E (-) 7598 *HhaI*; F (+) 10394 *DdeI*; G (-) 10394 *DdeI*; H (+) 10397 *AluI*; I (-) 10397 *AluI*; J (-) 12406 *HincII*; K (+) 13259 *HincI*; L (-) 13259 *HincI*; M (+) 13262 *AluI*; and N (-) 13262 *AluI*. The presence (+) or absence (-) of a cut site is indicated in parentheses. The intergenic 9 bp insertion and deletion of the COII-tRNA^{Lys} region are coded by the letters O and P, respectively.

singletons by these two mutations. This common haplotype is only shared with the Yukaghir and Mongol samples (each with only a single sequence) and given its relative isolation within the Yakut network suggests that it is the genetic consequence of a population bottleneck.

Overall, haplogroups C and D comprise 75.7% of the Yakut sample, which is consistent with the range of frequencies exhibited by other Siberian groups (from 34.5% in the Altai to 92.3% in the Evenks). The seven most common haplotypes all belong to haplogroups C and D. Of the

remaining Yakut sequences, 14.6% belong to east Eurasian haplogroups A, B, F, G, M and Y. The rest (9.7%) were assigned to west Eurasian haplogroups such as J and T.

Nested cladistic analyses (NCA) were performed for the major Yakut mtDNA lineages of haplogroups C and D and included closely related HVS-I sequences observed in the comparative population data (note, some D5a sequences for the Han, Daur and Oroqen were excluded from the analyses due to the underlying uncertainty of transversions 16182C and 16183C).

This phylogeographic method calculates statistics for each clade in order to identify significant geographic differentiation: D_c (the clade distance), measures the geographical range of a particular clade or haplotype; and D_n (the nested clade distance), compares the geographical range of a particular clade to that of its sister clades (Jobling et al. 2004). The nested cladograms are shown in Figure 5 and were produced following the methodology outlined by Templeton and Sing (1993). From a total of 28 nested clades, seven exhibited significant associations with geography. With the aid of an inference key provided with the GeoDis program, three main competing explanations are provided for the significant results: restricted gene flow (i.e., isolation by distance); population

fragmentation; and range expansion (either continuous range expansion or long distance colonization). Only one of the nested clades that included Yakut sequences was identified as the product of a long distance colonization event: clade II-4 of haplogroup D that exhibits a significantly large D_c difference between internal and tip clades, with the major Yakut D5a haplotype represented by clade I-7.

Therefore, subhaplogroup D5a appears to be a founder lineage for the Yakuts based on the significant deviation of the common D5a haplotype from the Ewens frequency spectrum, its fragmented structure within the MJ network, and the results of the NCA. This is further supported by ancient mtDNA studies, which have characterized five Yakut skeletal specimens

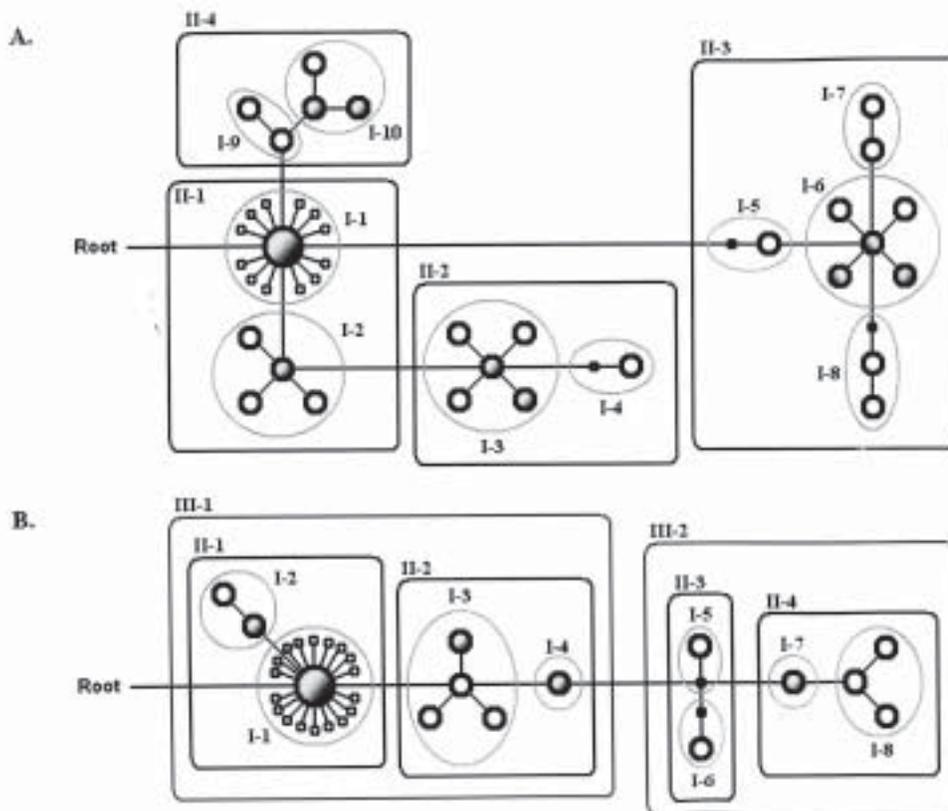


Fig. 5. Nested cladistic designs for haplogroups C (part A) and D (part B). Each line represents a HVS-I mutation, and the open circles indicate observed haplotypes (shaded ones include Yakut sequences), whereas the black circles are missing haplotypes. Ovals contain one-step clades (I-1 to I-10), black rectangles designate two-step clades (II-1 to II-4), and grey rectangles contain three-step clades (III-1 and III-2).

dating to about 300-400 years BP to this particular subhaplogroup (Ricaud et al. 2004, 2006). The age of D5a was estimated with the ρ (rho) statistic in order to obtain the age of the Yakut founder event. Using the conservative mutation rate of 20,180 years per mutation (Forster et al. 1996), a date of 5,771 years BP (\pm 3,532 years) was calculated. However, the faster rate of 0.5 mutations per million years per nucleotide based on pedigree estimates (Howell et al. 2003) yielded an age of 1,630 years BP (\pm 997 years), which is more congruent with archaeological evidence.

A MDS plot of a pairwise D_A distance matrix is presented in Figure 6. The stress value for the plot is 0.0975, which indicates a strong goodness of fit to the original matrix (Kruskal 1964). Along the first dimension, the Evenki and Itelmen represent the extreme positions, with the rest of the populations clustering in the center. The second dimension differentiates the Yakuts and other Mongol-Turkic groups, along with the Oroqen and Han, from the indigenous populations

of northeastern Siberia positioned at the bottom of this axis. The main finding from the SAMOVA analyses is also shown in Figure 6. When the number of groupings (k) was set within the range of 2 to 5, the Mongol-Turkic groups, the Oroqen, and the Han consistently clustered together with significant ϕ_{CT} values ranging from 0.0537 to 0.0799 ($P < 0.05$). Overall, these results show a close genetic relationship between the Yakuts and south Siberian populations, and thus supporting a southern origin for the Yakut people.

CONCLUSION

The goal of this study was to examine the origins of the Yakut people based on mtDNA variation. The findings are generally consistent with the results of previous mtDNA studies on the Yakuts that show evidence of a close genetic relationship to southern Turkic populations (Federova et al. 2003; Pakendorf et al. 2006; Puzyrev et al. 2003). The majority of the Yakut

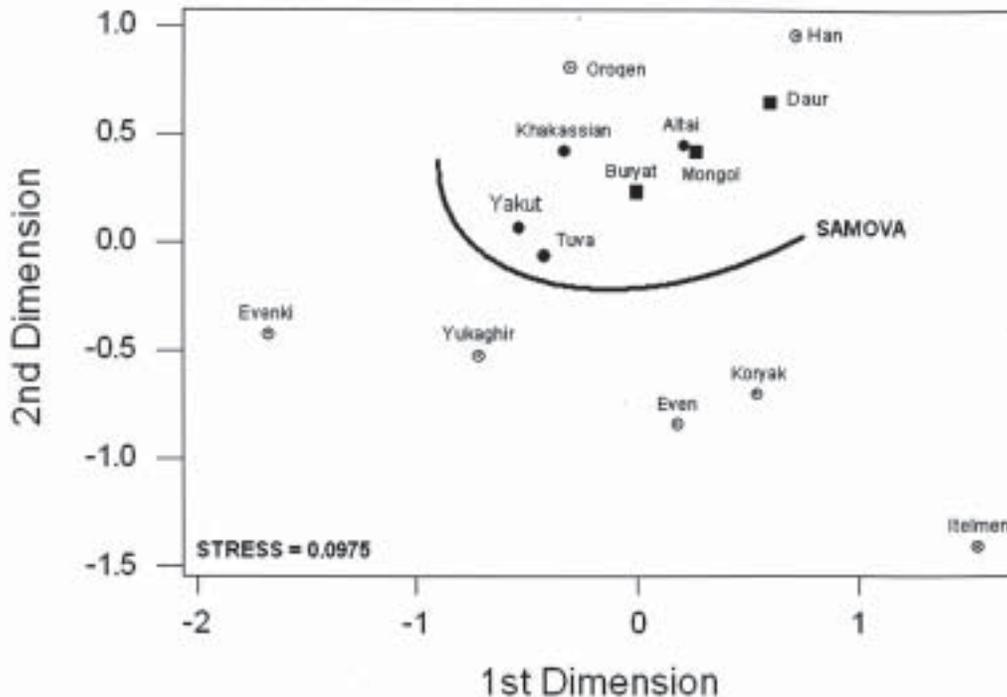


Fig. 6. MDS plot of a D_A distance matrix and the primary grouping identified in SAMOVA. Black circles and squares represent Turkic- and Mongol-speaking populations, respectively, and open circles are non-Mongol-Turkic groups. The curved line indicates those populations that were found to consistently cluster together in SAMOVA analyses (for $k = 2$ to 5 groups).

haplotypes belong to East Eurasian lineages that include haplogroups commonly found in south Siberian and Central Asian populations, such as B, D5a, F, and G2a. This is reflected in the MDS plots and SAMOVA analyses that show close genetic affinities between the Yakuts and south Siberian populations. However, the Yakuts and Tuva are somewhat removed from the other populations in the Mongol-Turkic MDS cluster in the direction of the Evenki and Yukaghirs. And with Yakut diversity levels that are intermediate to northeastern Siberians and Mongol-Turkic groups from south Siberia, these results may be indicative of some degree of admixture with indigenous groups and/or differentiation after the Yakut migration northwards to the Lena river basin. The genetic effects of this postulated founder event are evident from the significant Ewens-Watterson homozygosity test that is the product of a type 1 deviation in the Yakut HVS-I data. Subhaplogroup D5a is identified as one of the founder lineages and was dated to approximately 1,630 years BP, which is earlier than the founder dates based on the archaeological record. Overall, the results from this study are congruent with other investigations and lines of evidence that point to a southern origin for the Yakut people that involved a founder event dominated by a limited number mtDNA lineages within haplogroups C and D.

ACKNOWLEDGEMENTS

We are grateful to all the participants from Yakutia for providing blood samples. And we would like to thank Ric Devor, Kristin Wilson and Mike Grose for their assistance in sequencing the samples. This research was funded in part by a grant from the National Science Foundation.

REFERENCES

- Alekseev AN 1996. *Ancient Yakutia: The Iron Age and the Medieval Epoch*. Novosibirsk: Izdatel'stvo Instituta Arkheologii i Etnografii.
- Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG 1981. Sequence and organization of the human mitochondrial genome. *Nature*, **290**: 457-465.
- Bandelt H-J, Forster P, Röhl A 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**: 37-48.
- Boeva SB 1988. *Anthropogenetic Study of the Indigenous Populations of the Yakut ASSR in Connection with the Ethnogenesis of the Yakuts*. Ph.D. thesis submitted to Moscow State University, Moscow, Russia.
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J 1998. Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *American Journal of Human Genetics*, **63**: 1824-1838.
- Derenko MV, Shields GF 1997. Mitochondrial DNA sequence diversity in three north Asian aboriginal population groups. *Molecular Biology*, **31**: 665-669.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Micecicka-Eliwka D, Woźniak M, Zakharov IA 2003. Diversity of mitochondrial DNA lineages in south Siberia. *Annals of Human Genetics*, **67**: 391-411.
- Dupanloup I, Schneider S, Excoffier L 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**: 2571-2581.
- Ewens WJ 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**: 87-112.
- Excoffier L, Schneider S 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA*, **96**: 10597-10602.
- Federova SA, Bermisheva MA, Villems R, Maksimova NR, Khusnutdinova EK 2003. Analysis of mitochondrial DNA lineages in Yakuts. *Molecular Biology*, **37**: 544-553.
- Fefelova VV 1990. Participation of Indo-European tribes in ethnogeny of the Mongoloid population of Siberia: analysis of the HLA antigen distribution in Mongoloids of Siberia. *American Journal of Human Genetics*, **47**: 294-301.
- Forster P, Harding R, Torroni A, Bandelt H-J 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genetics*, **59**: 935-945.
- Fu XY 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics*, **147**: 915-925.
- Goebel T 1999. Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. *Evolutionary Anthropology*, **8**: 208-227.
- Halliburton R 2004. *Introduction to Population Genetics*. New Jersey: Pearson Prentice Hall.
- Howell C, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Hernstadt C 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, **72**: 659-670.
- Jobling MA, Hurles ME, Tyler-Smith C 2004. *Human Evolutionary Genetics*. New York: Garland Publishing.
- Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution*, **19**: 1737-1751.
- Kolman CJ, Sambuughin N, Bermingham E 1996.

- Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics*, **142**: 1321-1334.
- Kong Q-P, Yao Y-G, Liu M, Shen S-P, Chen C, Zhu C-L, Palanichamy MG, Zhang Y-P 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Human Genetics*, **113**: 391-405.
- Konstantinov IV 1975. *The Origins of the Yakut People and their Culture*. Yakutsk: Publications of the Prilenskaya Archaeological Expedition, Yakutskiy Filial SO AN SSSR pp. 106-173.
- Kruskal JB 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**: 1-27.
- Lalueza C, Perez-Perez A, Prats E, Cornudella L, Turbon D 1997. Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. *Human Molecular Genetics*, **6**: 41-46.
- Macaulay VA, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonn -Tamir B, Sykes B, Torroni A 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control region sequences and RFLPs. *American Journal of Human Genetics*, **64**: 232-249.
- Matsumoto H 1988. Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. *Human Genetics*, **80**: 207-218.
- Meyer S, Weiss G, Haeseler A 1999. Pattern of nucleotide substitution and rate heterogeneity in hypervariable regions I and II of human mtDNA. *Genetics*, **152**: 1103-1110.
- Novoradovsky AG, Spitsyn VA, Duggirala R, Crawford MH 1993. Population genetics and structure of Buryats from the Lake Baikal region of Siberia. *Human Biology*, **65**: 689-710.
- Okladnikov AP 1956. Ancient population of Siberia and its culture. In: Levin MG, Potapov LP (Eds): *Narody Sibiri*. Moscow: Russian Academy of Science pp. 13-98.
- Okladnikov AP 1970. *Yakutia: Before Its Incorporation into the Russian State*. Montreal: McGill-Queen's University Press.
- Pakendorf B, Spitsyn VA, Rodewald A 1999. Genetic structure of a Sakha population from Siberia and ethnic affinities. *Human Biology*, **71**: 231-244.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Human Genetics*, **120**: 334-353.
- Posada D, Crandall KA, Templeton AR 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**: 487-488.
- Puzryev VP, Stepanov VA, Golubenko MV, Puzryev KV, Maximova NR, Kharkov VN, Spiridonova MG, Nogovitsina AN 2003. mtDNA and Y-chromosomal lineages in the Yakut population. *Russian Journal of Genetics*, **39**: 816-822.
- Ramos-Onsins SE, Rozas J 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**: 2092-2100.
- Ray N, Mathias C, Excoffier L 2003. Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**: 76-86.
- Ricaud F-X, Kolodetsnikov S, Keyser-Tracqui C, Alekseev AN, Crub zy E, Ludes B 2004. Genetic analysis of human remains found in two eighteenth century Yakut graves at At-Dabaan. *International Journal of Legal Medicine*, **118**: 24-31.
- Ricaud F-X, Kolodetsnikov S, Keyser-Tracqui C, Alekseev AN, Crub zy E, Ludes B 2006. Molecular genetic analysis of 400-year-old human remains found in two Yakut burial sites. *American Journal of Physical Anthropology*, **129**: 55-63.
- Rogers AR 1995. Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**: 608-615.
- Rogers AR, Harpending H 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**: 552-69.
- Rohlf FJ 2002. *NTSYSpc: Numerical Taxonomy*. Setauket, New York: Exeter Publishing Ltd.
- Rubicz RC, Schurr TG, Babb PL, Crawford MH 2003. Mitochondrial DNA variation and the origins of the Aleuts. *Human Biology*, **75**: 809-835.
- Schneider S, Roessli D, Excoffier L 2000. *Arlequin ver. 2.00: A Software for Population Genetic Data Analysis*. Switzerland: Genetics and Biometry Laboratory, University of Geneva.
- Sch nig C 1990. Classification problems of Yakut. In : Dor R (Ed.): *L'Asie Centrale et ses Voisins*. Paris: INALCO pp. 91-102.
- Schurr TA, Sukernik RI, Starikovskaya YB, Wallace DC 1999. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *American Journal of Physical Anthropology*, **108**: 1-39.
- Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Human Biology*, **66**: 761-775.
- Tajima F 1989. The effect of change in population size on DNA polymorphism. *Genetics*, **123**: 597-601.
- Tarskaia LA, Bychkovskaia LS, Pai GV, Makarov SB, Pakendorf B, Spitsyn VA 2002. Distribution of the ABO blood groups and the HP, TF, GC, PI, and C3 serum proteins in Yakuts. *Russian Journal of Genetics*, **38**: 548-553.
- Templeton AR, Sing CF 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**: 659-669.
- Templeton AR, Routman E, Phillips CA 1995. Separating population structure from population history: a cladistic analysis of the geographic distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*. *Genetics*, **140**: 767-782.
- Tokarev SA, Gurvich IS 1956. The Yakuts. In: Levin MG, Potapov LP (Eds.): *Narody Sibiri*. Moscow: Russian Academy of Science pp. 267-328.
- Torroni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC,

- Lawrence DN, Weiss KM, Wallace DC 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics*, **130**: 153-162.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *American Journal of Human Genetics*, **53**: 563-590.
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proceedings of the National Academy of Sciences, USA*, **86**: 9350-9354.
- Watterson GA 1978. The homozygosity test of neutrality. *Genetics*, **88**: 405-417.
- Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Research*, **15**: 529-542.
- Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, **70**: 635-651.