

Diversity of Mitochondrial DNA Lineages in South Siberia

M. V. Derenko^{1,*}, T. Grzybowski², B. A. Malyarchuk¹, I. K. Dambueva³, G. A. Denisova¹, J. Czarny², C. M. Dorzhu⁴, V. T. Kakpakov⁵, D. Miścicka-Śliwka², M. Woźniak² and I. A. Zakharov⁵

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

²The Ludwik Rydygier Medical University in Bydgoszcz, Forensic Medicine Institute, ul. Skłodowskiej-Curie 9, 85-094 Bydgoszcz, Poland

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

⁴Tuva State University, 667035 Kyzyl, Russia

⁵Animal Comparative Genetics Laboratory, Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin Str. 3, 119991 Moscow, Russia

Summary

To investigate the origin and evolution of aboriginal populations of South Siberia, a comprehensive mitochondrial DNA (mtDNA) analysis (HVR1 sequencing combined with RFLP typing) of 480 individuals, representing seven Altaic-speaking populations (Altaians, Khakassians, Buryats, Sojots, Tuvinians, Tadjins and Tofalars), was performed. Additionally, HVR2 sequence information was obtained for 110 Altaians, providing, in particular, some novel details of the East Asian mtDNA phylogeny. The total sample revealed 81% East Asian (M*, M7, M8, M9, M10, C, D, G, Z, A, B, F, N9a, Y) and 17% West Eurasian (H, U, J, T, I, N1a, X) matrilineal genetic contribution, but with regional differences within South Siberia. The highest influx of West Eurasian mtDNAs was observed in populations from the East Sayan and Altai regions (from 12.5% to 34.5%), whereas in populations from the Baikal region this contribution was markedly lower (less than 10%). The considerable substructure within South Siberian haplogroups B, F, and G, together with the high degree of haplogroup C and D diversity revealed there, allows us to conclude that South Siberians carry the genetic imprint of early-colonization phase of Eurasia. Statistical analyses revealed that South Siberian populations contain high levels of mtDNA diversity and high heterogeneity of mtDNA sequences among populations ($F_{st} = 5.05\%$) that might be due to geography but not due to language and anthropological features.

Introduction

Analysis of mitochondrial DNA (mtDNA) polymorphism has become a useful tool for human population and molecular evolution studies, allowing researchers to infer the pattern of female migrations and peopling of different regions of the world (Wallace, 1995). The use

of the phylogeographic approach has allowed refinement of the analysis of maternal mtDNA lineages, suggesting the current model of complex demographic scenarios for the peopling of Eurasia (Richards *et al.* 2000). It has been shown that in the present-day Eurasian populations the mtDNA variation can be classified into two macrohaplogroups, M and N, both of them coalescing to the African macro-cluster L3, which can be considered as the most recent common ancestor of all non-Africans (Macaulay *et al.* 1999; Quintana-Murci *et al.* 1999). Macrohaplogroup M encompasses a number of East Asian-specific haplogroups, including C, D, G, E,

*Corresponding author: Dr. Miroslava V. Derenko, Genetics Laboratory, Institute of Biological Problems of the North, Portovaya str., 18, 685000 Magadan, Russia Fax/Phone: 7 41322 34463. E-mail: mderenko@mail.ru

Z and other recently described haplogroups M7, M8, M9, M10, designated previously as M* (Yao *et al.* 2002). Macrohaplogroup N encompasses multiple East Asian-specific lineages, including A, B, F, R9a, Y, and N9a, as well as the West Eurasian haplogroups HV, JT, UK, I, W, and X (Macaulay *et al.* 1999; Yao *et al.* 2002). While West Eurasian mtDNA variation is now quite well understood, there are only a few studies dealing with combined RFLP and HVR1 data for East Asian and Siberian populations (Torroni *et al.* 1993; Kolman *et al.* 1996; Starikovskaya *et al.* 1998; Schurr *et al.* 1999; Derenko *et al.* 2000; Forster *et al.* 2001; Derbeneva *et al.* 2002; Yao *et al.* 2002). Moreover, most of these studies were in fact generally motivated by research into the origin of Native Americans, and thereby the phylogeographic pattern of mtDNA differentiation in East Asian and Siberian populations is still poorly understood.

A particular significance of the genetic research of Altai and neighboring regions lies in the geographic location of the Altai area on the margin of southwestern part of Siberia, believed to be the main gateway for the initial peopling of the remainder of Siberia from the southern regions of Central Asia. Archaeological data suggest that South Siberia could have already been inhabited by modern humans in the late Pleistocene. The earliest Upper Paleolithic industries occurred in the Altai region and dated as 43300 ± 1600 years B.P. is thought to be linked with a gradual transformation of a Mousterian tradition by introduction of the more progressive elements during the early stage of the midlast glacial interstadial interval (45000 – 30000 years B.P.) (Derevianko, 1998; Goebel, 1999). A similar “mixed” technological character observed in the Altai can be found in most of the other Late Paleolithic stone industries in Siberia dating to second half of the Karginsk interstadial interval. Analogous combinations of both the Mousterian and Late Paleolithic elements are documented in the Transbaikal and Angara River basin regions, as well as in the upper Yenisei area (Chlachula, 2001). They also have been recorded in synchronous industries in Mongolia and North China thus testifying the similar processes of the cultural development as in Siberia (Okladnikov, 1981; Derevianko, 1998). During the following interval, i.e., at the end of the Karginsk interstadial and in the early last glacial (Sartan) stage, more progressive Late Paleolithic cultures with the advanced technique

of the prismatic core flaking reminiscent of the European traditions emerged in vast areas of Siberia: on the eastern margin of the West Siberian Lowland, in the upper Yenisei River basin, as well as in the Angara River basin (Derevianko & Zenin, 1996; Vasiliev *et al.* 1999). Such industries, however, are absent in the Altai area, indicative of complex and regionally divergent cultural evolution in different parts of Siberia (Chlachula, 2001).

Unfortunately, the archaeological records alone with the lack of human skeletal remains are inconclusive about the anthropological traits which were characteristic for the Upper Paleolithic Siberian population. Yet, the Upper Paleolithic artifacts from 18 000 B.P. already have been found in association with skeletal remains that bear similar morphology with contemporary anatomically modern humans teeth from Europe (Turner II, 1987). However, in their cranial traits, Siberians from the Neolithic period are seen as more closely related to northern Chinese and Mongolians than to Europeans (Alexeev & Gohman, 1984; Alexeev, 1989). Similarly, sinodontic teeth appear as characteristic for northern Chinese, Siberians, and Native Americans from 17000 B.P. onwards (Turner II, 1987).

Meanwhile, according to paleoanthropological data, the Europeoid traits became prevalent among steppe zone inhabitants of Tuva, Altai, Khakassia and West Mongolia since the Bronze Age or even earlier (Alexeev & Gohman, 1984; Alexeev, 1989). Subsequent migrations from Central and Eastern Asia led to the formation of anthropological traits seen in the contemporary population of South Siberia. Thus, the considerable ethnic diversity in Southern Siberia was largely shaped by migration processes that had occurred since the initial colonization of the region in Upper Paleolithic. Historically, a complex network of migrations can be traced from Central, Eastern Asia and Western Eurasia (Alexeev, 1989; Vasiliev, 1993; Derevianko, 1998).

Although classical genetic data (Cavalli-Sforza *et al.* 1994; Rychkov *et al.* 2000) cover most of the present-day populations living in South Siberia, the high-resolution mtDNA and Y-chromosomal data sets for the populations living in this region are either incomplete or virtually absent. To investigate the origin and evolution of aboriginal populations of South Siberia, we performed a comprehensive mtDNA analysis (HVR1 sequencing combined with RFLP typing) of seven

Altaic-speaking populations, occupying the broader area of Altai and Baikal regions, and compared them with the populations of Eastern and Central Asia that have had a great historical influence on Southern Siberians.

Material and Methods

Population Samples

A total of 480 hair root samples from unrelated individuals were collected from the following seven South Siberian autochthonous groups: Turkic-speaking Altaians, Tuvinians, Eastern Tuvinians (Todjins), Tofalars, Sojots, Khakassians and Mongolic-speaking Buryats (Figure 1). Information about birthplace, parents and grandparents was obtained from all donors.

The Altaian individuals ($n = 110$) came from five different districts of the Altai Republic. In detail, there were Telenghits from the Kosh-Agach and Ulagan districts, Altai-Kizhi from the Onguday, Ust-Koksa, Ust-Kan and Shebalinsk districts, Maimalars from the Shebalinsk and Maiminsk districts, and Tubalars and Chelkans from the Turochak district. The Altaians, the native people of the Altai Republic (South Siberia), number up to 70000 persons. 'Altaians' is the common denomination for seven formerly distinct Turkic-speaking groups - Altai-Kizhi, Teleuts and Telenghits, who represent Southern Altaians, and

Chelkans, Kumandins, Tubalars and Maimalars, who represent Northern Altaians. The differences between southern and northern Altaians are well established, on the basis of anthropological, linguistic and classical genetic-marker studies (Potapov, 1969; Alexeev & Gohman, 1984; Luzina, 1987). The analysis of the tribal structure of Altaians has shown that the present-day Altaians have retained their native language and ethnic identity. They have begun to mix with other ethnic groups (mostly Russians and Kazakhs) only recently, so the interethnic admixture is estimated to be less than 5% (Luzina, 1987; Osipova *et al.* 1997).

The remnant of Tofalars ($n = 58$), a small geographically isolated tribe of nomadic hunters and reindeer breeders, occupying the Taiga area on the northern slopes of the east Sayan Mountains and numbering approximately 600 individuals, was collected in the village of Alygdzher in the Nizhneudinsk administrative district of the Irkutsk region. The Tofalars originally spoke a Samoyed language, but later changed to a Turkic-group language. They appear to be of mixed origin, but related to Eastern Tuvinians (Todjins) (Levin & Potapov, 1964).

The Tuvinian samples ($n = 90$) were collected in Dzun-Khemchiksk, Mongun-Taiga, Bai-Taiga, Ovyursk, Tes-Khemsk, Erzinsk and Tandinsk districts of Tuva Republic. Taking into account the ethnoterritorial differences existing among Tuvinians as well as peculiar anthropological features observed in Eastern

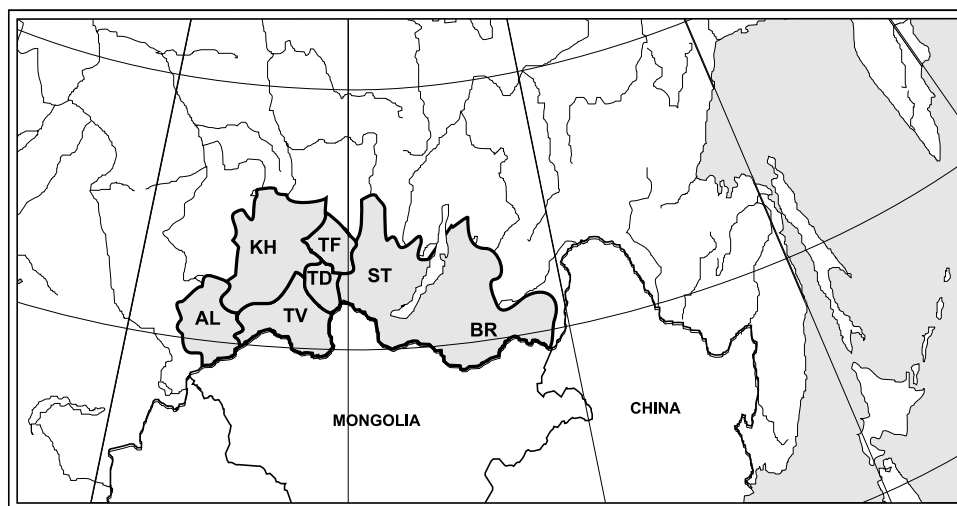


Figure 1 Geographic locations of the South Siberian populations studied. Populations are coded as: AL – Altaians, KH – Khakassians, BR – Buryats, ST – Sojots, TD – Todjins, TV – Tuvinians, TF – Tofalars.

Tuvinians – Todjins (Alexeev & Gohman, 1984), the Todjin samples ($n = 48$) were collected in the villages of Toora-Khem, Iij-Khem, Ulug-oo and Adyr-Kezhig in the Todja district of Tuva Republic. Most of the territory of the Tuva Republic is situated in the steppe zone in the centre of the Asian continent, bounded by the Sayan Mountains to the North and the Mongolian steppes to the South. Tuvinians number approximately 200 thousands individuals, and they still pursue traditional subsistence activities similar to those of Buryats. Buryats, who are also nomadic cattle breeders, live in the central southern part of the Siberia border to Mongolia and China (Levin & Potapov, 1964; Gurvich, 1980). They number up to 400,000 and represent the only Mongolic-speaking group in Siberia. Buryat samples ($n = 91$) were collected in the villages of Kizhinga, Khorinsk, Zakamensk, Eravna, Selenga, Barguzin and Kabansk districts of Buryat Republic, thus encompassing all territories inhabited by the modern Buryats.

Khakassians, the northern neighbours of Tuvinians, live in the middle reaches of the River Yenisey and in the upper reaches of its tributaries, the Abakan and the Chulym. On an administrative level, they belong to the Khakass Republic in the Krasnoyarsk Region of the Russian Federation – an area of some 61,900 square kilometres. Northern and eastern parts of the region are flat steppelands (the Abakan-Minusinsk Basin), whereas southern and western regions are mountainous. The Khakassian samples ($n = 53$) were collected in the settlements of Askiz, Shirinsk, Beisk and Ordzhonikidzevsk districts of Khakass Republic.

Additionally, the samples from the Sojots ($n = 30$) were collected in the Tunka and Okinsk districts of Buryat Republic. Historically, these samples represent well-defined Turkic-speaking tribes of cattle breeders, currently numbering approximately 2000 individuals.

DNA Extraction and Sequencing

DNA was extracted from the hair roots as described elsewhere (Walsh *et al.* 1991). PCR amplification of the entire noncoding region was performed using the primers L15926 and H00580. The temperature profile (for 30 cycles of amplification) was 94°C for 20 sec., 50°C for 30 sec., and 72°C for 2.5 min., (Thermal Cycler 9700; Perkin Elmer, USA). The resulting amplification prod-

uct was diluted 1000-fold and 4 mkl aliquots were added to an array of second-round, nested PCR reactions (32 cycles) to generate DNA templates for sequencing. The primer sets L15997/M13(-21)H16401 and M13(-21)L15997/H16401 were used to generate both strands of the HVR1. Similarly, the primer sets L00029/M13(-21)H00408 and H00408/M13(-21)L00029 were used for HVR2. The nucleotide sequences of HVR1 from position 15999 to 16400 and HVR2 from position 30 to 407 have been determined. Both primer sequences and nomenclature were used according to Sullivan *et al.* (1992). Negative controls were prepared for both the DNA extraction and the amplification process. PCR products were purified by ultrafiltration (Microcon 100; Amicon) and sequenced directly from both strands with a (-21)M13 primer using the BigDye Primer Cycle Sequencing Kit (Perkin Elmer) according to the manufacturer's protocol. Sequencing products were separated in a 4% PAGE gel on the ABI Prism™ 377 DNA Sequencer. Data were analyzed using the DNA Sequencing Analysis and Sequence Navigator programs (Perkin Elmer). The length polymorphisms located between 16180–16193 were disregarded from the analyses.

RFLP Analysis

Several amplified segments, mainly in the mtDNA coding regions, were analyzed by RFLP testing according to the method described by Torroni *et al.* (1993, 1996) and Macaulay *et al.* (1999), to screen haplogroup-specific sites (Table 1). The restriction fragments were resolved by electrophoresis in 8% PAGE gels and were visualized after ethidium bromide staining under UV.

Phylogenetic and Statistical Analyses

The phylogenetic relationships between mitochondrial haplotypes comprising various combinations of the HVR1 sequences and RFLPs were analyzed by the median-network method (Bandelt *et al.* 1995) and checked by the Network 3.1 program, from the Fluxus Engineering Web site. Coalescence times for mtDNA haplogroups were estimated according to the methods of Forster *et al.* (1996). Sequence classification into haplogroups was based on HVR1 and RFLP data.

Table 1 RFLP polymorphisms used to identify mtDNA haplogroups

Haplogroup	Characteristic restriction site(s)
M	+10394 <i>DdeI</i> , +10397 <i>AluI</i>
C	+10394 <i>DdeI</i> , +10397 <i>AluI</i> , -13259 <i>HincII</i> / +13262 <i>AluI</i>
D	+10394 <i>DdeI</i> , +10397 <i>AluI</i> , -5176 <i>AluI</i>
E	+10394 <i>DdeI</i> , +10397 <i>AluI</i> , -7598 <i>HhaI</i>
G	+10394 <i>DdeI</i> , +10397 <i>AluI</i> , +4830 <i>HaeII</i> / +4831 <i>HhaI</i>
A	+663 <i>HaeIII</i>
B	9-bp deletion
F	-12406 <i>HpaI</i> / <i>HincII</i>
HV	-14766 <i>MseI</i>
H	-14766 <i>MseI</i> , -7025 <i>AluI</i>
U	+12308 <i>HinfI</i>
U2	+15907 <i>RsaI</i>
K	+10394 <i>DdeI</i> , -9052 <i>HaeII</i> , +12308 <i>HinfI</i>
J	+10394 <i>DdeI</i> , -13704 <i>BstOI</i>
T*	+13366 <i>BamHI</i>
T1	-12629 <i>AvaII</i> , +13366 <i>BamHI</i>
I	-4529 <i>HaeII</i> , +8249 <i>AvaII</i> , +10032 <i>AluI</i> , +10394 <i>DdeI</i>
W	+8249 <i>AvaII</i> , -8994 <i>HaeIII</i>
X	-1715 <i>DdeI</i> , +14465 <i>AccI</i>

The nomenclatures of Macaulay *et al.* (1999), Richards *et al.* (2000) and Yao *et al.* (2002) were followed for the West Eurasian and East Asian mtDNA clusters.

The basic parameters of molecular diversity and population genetic structure (including analyses of molecular variance, AMOVA) were calculated using the computer program Arlequin 2.0 (Schneider *et al.* 2000). The statistical significance of *F*_{st}-values was estimated by permutation analysis using 10000 permutations.

Mitochondrial DNA HVR1 sequences from 103 Mongolians (Kolman *et al.* 1996), 66 Han Chinese and 64 South Koreans (Horai *et al.* 1996), 55 Kazakhs, 94 Kirghiz and 54 Uighurs (Comas *et al.* 1998) and 263 Han Chinese (Yao *et al.* 2002) were used for comparative analyses.

Results and Discussion

MtDNA Composition of the South Siberian Populations

Four hundred and eighty South Siberian samples were analyzed by HVR1 sequencing and assaying additional RFLP markers (Table 1). One hundred and eighty four

different HVR1 haplotypes were observed. Altogether, 150 haplotypes belong to Asian-specific mtDNA haplogroups, M*, M7, M8, M9, M10, A, B, C, D, F, G, Y, Z, N9a, R*, whereas 34 belong to West Eurasian-specific haplogroups, H, J, T, U, X, I, N1a (Table 2). The haplogroup frequencies observed in South Siberia are summarized in Table 3. The populations studied exhibit a high percentage of haplogroup M mtDNA lineages (M*, M7, M8, M9, M10, C, D, G, Z), ranging from 48.2% in Altaians to 80.2% in Buryats. Haplogroup C is the most frequent haplogroup within the Asian-specific fraction, closely followed by haplogroup D. Haplogroups C and D together account from about one third to over half of the Asian-specific fraction in all populations studied. The spread of haplogroups C and D reveals some characteristic differences among populations of South Siberia. For example, haplogroup C variants are predominantly found among populations living in the East Sayan region, Tofalars, Tuvinians and Todjins. On the other hand, the highest frequency of haplogroup D was observed among Buryats and Sojots (33% and 46.7%, respectively) living in the Baikal region.

The geographic distribution of haplogroup Z contrasts with that of its sister haplogroup C (Table 3). Among Khakassians, Sojots and Todjins haplogroup Z has not been found, as opposed to the high frequency of haplogroup C among them. Haplogroup Z is found in Altaians, Buryats, Tuvinians and Tofalars with an overall frequency of 2.1%.

Haplogroup G mtDNAs, which are widely distributed in the northeast Siberian populations of Koryaks, Evens, Chukchi and Itelmens (Derenko & Shields, 1997; Starikovskaya *et al.* 1998; Schurr *et al.* 1999), were practically absent in the majority of South Siberian populations, with the exception of Todjins where the frequency is rather high - 18.8%. In contrast, mtDNAs harbouring both G and E specific RFLPs (+4830 *HaeII*/+4831 *HhaI* for G and -7598 *HhaI* for E) were identified in five out of seven populations studied - Altaians, Buryats, Sojots, Tuvinians and Tofalars, with frequencies varying from 0.9% in Altaians to 14.3% in Buryats. Such 'E/G' mtDNA variants originated on the background of haplogroup G due to mutation at np 7600, which gives a similar E-specific RFLP pattern, and therefore should be considered as subgroup G2 within haplogroup G (Yao *et al.* 2002). It is noteworthy

Table 2 HVR1 sequence variation and mtDNA haplogroup (HG) status of 480 South Siberian samples

HG	HVR1 sequence	110 AL	53 KH	91 BR	30 ST	48 TD	90 TV	58 TF
A	086 223 290 319 362					1		
A	183 223 274 290 319 362						1	
A	189 223 290 319 362			1				
A	223 242 290 293C 319							3
A	223 290 292A 319 362			1				
A	223 290 297 311 319 362				3			
A	223 290 319 362		2			1		
B4	189 217			2	1			
B4	189 217 240			1				
B4	093 145 189 217 266 362					1	4	
B4	129 153 189 217 223 247 320			1				
B4a	129 189 217 261 356						1	
B4a	167 189 217 261 317T					1		2
B4a	189 217 261 299						1	
B4b	086 136 189 217	2	2				1	
B4b	086 136 189 217 293A/G	1						
B5b	111 140 189 234 243 304			1				
B5b	140 189 243 274	1		1				
C	025 093 129 223 235 298 327 390		1					
C	093 129 223 235 298 327 390							19
C	129 223 235 298 327 390		1					
C	223 235 298 327 390							1
C	093 129 150 189 223 298 327						1	
C	093 129 223 298 327	6		3			6	
C	093 129 223 298 327 381						1	
C	093 129 223 327			1				
C	129 150 178 223 298 327		2					
C	129 150 223 298 327			1				
C	129 140 171 223 291 298 327 344 357		1					
C	129 140 171 223 298 327 344 357		2					
C	167 171 223 298 327 344 357	1						
C	171 223 224 298 327 344 357							1
C	171 223 298 327 344						1	
C	171 223 298 327 344 357	3	2			3	2	3
C	223 298 327 344 357							1
C	093 223 288 298 327 390					2		
C	148 164 223 288 298 327		1					
C	148 223 288 298 301 327						1	
C	148 223 288 298 327			1	1	4	5	5
C	148 223 327					1		
C	093 223 288 291 298 327		1					
C	093 223 261 288 298			2				
C	223 261 288 298			2		1	1	
C	223 298 327	5	4	2	1	4	10	5
C	223 298 311 327			2		2	5	
C	189 223 294 298 311 327					1		
C	223 294 298 311 327					2	6	
C	223 259insA 294 298 327			1				
C	223 259insA 298 327			3	4			
C	223 270 298 327			3				
C	223 291 298 327	4				2	2	
C	223 293 298 327			1		1		1
C	129 223 298 327		3	3			2	

Table 2 (continued)

HG	HVR1 sequence	110 AL	53 KH	91 BR	30 ST	48 TD	90 TV	58 TF
C	025 223 298 327		1					
C	175 223 298 327			1				
C	223 298 327 343	1						
C	223 242 298 327	1						
D	042 172 223 362			1				
D5a	051 172 189 223 266 362			1				
D5a	092 126 164 189 223 266 362	1	1					
D5a	092 164 172 189 223 266 362						2	
D5b	126 136 189 223 360	1						
D	082 147A 189C/T 223 362	1						
D	082 147A 223 362	1						
D	082 223 362		3	1				
D	092 129 148 223 271 362			1				
D	092 223 316 362	1						
D	093 164 189 223 228 362			1				
D	093 164 223 245 362			1				
D	093 172 173 215 223 319 362	2						
D	093 223 239 243 319 362						1	
D	129 145 223 311 319 362			2				
D	129 173 223 319 362			1				
D	093 223 232 290 362			1				
D	093 223 362						1	
D	129 152 179 192 223 362						2	
D	147 223 362			1				
D	140 223 274 311 362	2						
D	171 223 311 362			1				
D	171T 223 355A 362			1	4			
D	174 223 262		1	1	4			
D	184C/T 223 311 362			1				
D	192 223 362			1				
D	218 223 362			1				
D	221 223 245 362			1				
D	223 232 290 362			1				
D	223 245 362			1				
D	223 291 362	1						
D	223 294 362			1				
D	223 311 362			3				
D	223 319 362	5		1	1		8	
D	182 223 362			1				
D	223 274 362			2				
D	223 362	2	2	2	5	2	2	
G2	223 278 362			5				
G2	051 150 223 278 362			1				
G2	129 223 274 278 362				1			
G2	145 223 278 362			1				
G2	223 278 287 304 362	1						
G2a	003 105 107G 223 227 278 362							1
G2a	189C/T 223 227 278 362			1				
G2a	223 227 234 278 362			1				
G2a	223 227 274 278 362			2	1		2	
G2a	223 227 278 362			2				
G3	093 223 274 362 390						1	
G3	156 223 274 362 390						1	

Table 2 (continued)

HG	HVR1 sequence	110 AL	53 KH	91 BR	30 ST	48 TD	90 TV	58 TF
G4	223 325 362					3		
G4	129 223 325 362 365	1						
G4	218 223 260 325 362					4	1	
G4	223 260 325 362					2		
G4	223 260C/T 325 362						1	
F1	093 207 304 362 399	1						
F1a	129 162 172 304 399		1					
F1a	162 172 304		2					
F1b	189 304	5	4					
F1b	114A 189 232A 249 304 311			1				
F1b	129 189 232A 249 304 311 344		1					
F1b	172 179 189 232A 249 304 311	3				1	2	
F1b	179 189 232A 249 304 311		1					
F1b	189 232A 249 304 311		3					
F2a	092A 291 304	1						
R*	CRS							1
R*	124 148 290 304 309 390						1	
R*	145 192 243 304 309 362 390				1			
R*	145 192 243 304 309 390					1	1	
R*	051 168 172 311					1		
H	304					1		
H	092 245 362	1						
H	169 184	2						
H	288 362	1	2					
H	311	1						
H	354	2						
H	093 129 168 291							1
H	CRS			2				3
H	220C 235 291						1	
I	129 223 391	2		1				
J*	069 126			2				5
J1	069 126 145 172 222 261		1				4	
J1	069 126 145 261 290	4						
J1	069 126 145 172 261 278						1	
M*	145 148 188 189 223 381			1				
M7b1	129 192 223 297					1		
M7c	223 248 295 319			1				
M7c	145 223 295 304					1		
M8	184 189 223 298 355 362	2						
M8a	134 184 223 287 298 319	1						
M8a	148 223 298 319	1						
M8a	184 223 298 319	1						
M9	223 234 291 316 362	1						
M9	223 234 316 362	1						
M10	129 193 223 311 357			1				
M10	093 193C/T 223 311 357 381	1						
N1a	147A 172 189 223 248 320 355	2						
N1a	147G 172 189 223 248 320 355	1						
N9a	111 129 223 257A 261						1	
N9a	172 223 257A 261		1					
N9a	189 223 257A 261	2						
N9a	223 248 257A 261 311	3						
T*	051 126 189 294 296						1	

Table 2 (continued)

HG	HVR1 sequence	110 AL	53 KH	91 BR	30 ST	48 TD	90 TV	58 TF
T*	126 168 294 296 324		1					
T*	126 294 296			1				
T*	126 294 304							3
T1	126 163 186 189 294	1						
U2	051 129C 189 214 258 362	5						
U2	051 129C 189 294 362		1					
U2	189 214 362	1						
U3	343	2				3	1	
U4	311 356	2	4					
U4	356	4	1					
U5a	172 192 256 270 291 311 399				1			
U5a	192 241 256 270 287 304 325 399	1						
U5a	192 256 270			1				
U5b	189 261 270						1	
U5b	189 270						1	
U5b	192 249 311	3						
K	224 311				1			
X	189 223 278 ^a	3						
Y	126 189 231 266 311					1		
Y	126 193 231 266						1	
Y	126 231 266			1				
Y	126 231 266 319 399			1	1			
Z	129 185 223 224 260 298			1				
Z	185 223 260 298	5						
Z	185 223 260 298 360							3
Z	185 223 260 298 399						1	

Note: Variant positions from the Cambridge Reference Sequence (CRS) of Anderson *et al.* (1981) are shown minus 16000. Transversions are further specified by the appropriate base change. Ins indicates an insertion. Heteroplasmic variants are shown by a/. Populations are coded as: AL – Altaians, KH – Khakassians, BR – Buryats, ST – Sojots, TD – Tadjins, TV – Tuvinians, TF – Tofalars. ^aThis haplotype has already been published (Derenko *et al.* 2001).

that four Buryat and three Tuviniian mtDNAs which we classified previously as E haplotypes (Derenko *et al.* 2000), also harbour +4830 *Hae*II/+4831 *Hha*I site gains characteristic for haplogroup G. Thus, it seems that haplogroup E has a very restricted distribution: it is virtually absent in South Siberia and occurs as rarely as 1.5–5% in Tibet and southern China (Torrioni *et al.* 1994; Kivisild *et al.* 2001).

Haplogroup M sub-lineages, M7, M8, M9, M10 and M* were detected in Tadjins, Buryats and Altaians with frequencies of 4.2%, 3.3% and 7.3%, respectively. Haplogroups B and F encompassing almost all East Asian R lineages (Richards & Macaulay, 2000; Yao *et al.* 2002) are found in South Siberian populations with considerable differences in geographic distribution. Haplogroup B mtDNAs are widely spread, although at low frequencies (ranging from 3.3% to 7.8%), among all South Siberian populations studied here. Haplogroup F has not been

found in Sojots and Tofalars; it is also very rare in Buryats, Tuvinians and Tadjins, whereas in Khakassians and Altaians it is found with frequencies of 22.6% and 9.1%, respectively.

Haplogroup Y, widely spread in Northeastern Asia where it is found with considerable frequencies in Evens, Koryaks, Itelmens, Nivkhs and Ainu (Horai *et al.* 1996; Derenko & Shields, 1997; Schurr *et al.* 1999), is much less frequent in South Siberia, being found in Buryats, Sojots, Tuvinians and Tadjins. Haplogroup N9a, which has a predominantly East Eurasian distribution, was also found at very low frequencies among Altaians, Khakassians and Tuvinians.

Despite the fact that the majority of maternal lineages of South Siberian populations belong to East Asian specific mtDNA haplogroups, a substantial West Eurasian fraction was revealed in gene pools of the populations studied (Table 3). Lineages characteristic of West

Table 3 MtDNA haplogroup distribution (no. of individuals and % values in parenthesis) in South Siberian populations

	Altaians (110)	Khakassians (53)	Buryats (91)	Sojots (30)	Todjins (48)	Tuvinians (90)	Tofalars (58)	In total (480)
A	0	2 (3.8)	2 (2.2)	3 (10.0)	2 (4.2)	1 (1.1)	3 (5.2)	13 (2.7)
B	4 (3.6)	2 (3.8)	6 (6.6)	1 (3.3)	2 (4.2)	7 (7.8)	2 (3.5)	24 (5.0)
M*	8 (7.3)	0	3 (3.3)	0	2 (4.2)	0	0	13 (2.7)
C	21 (19.1)	19 (35.9)	26 (28.6)	6 (20.0)	23 (47.9)	43 (47.8)	36 (62.1)	174 (36.3)
D	17 (15.5)	7 (13.2)	30 (33.0)	14 (46.7)	2 (4.2)	16 (17.8)	0	86 (17.9)
G2	1 (0.9)	0	13 (14.3)	2 (6.7)	0	2 (2.2)	1 (1.7)	19 (4.0)
G	1 (0.9)	0	0	0	9 (18.8)	4 (4.4)	0	14 (2.9)
Z	5 (4.6)	0	1 (1.1)	0	0	1 (1.1)	3 (5.2)	10 (2.1)
F	10 (9.1)	12 (22.6)	1 (1.1)	0	1 (2.1)	2 (2.2)	0	26 (5.4)
N9a	5 (4.5)	1 (1.9)	0	0	0	1 (1.1)	0	7 (1.5)
Y	0	0	2 (2.2)	1 (3.3)	1 (2.1)	1 (1.1)	0	5 (1.0)
H	7 (6.4)	2 (3.8)	2 (2.2)	0	1 (2.1)	1 (1.1)	4 (6.9)	17 (3.5)
U	18 (16.4)	6 (11.3)	1 (1.1)	1 (3.3)	3 (6.3)	3 (3.3)	0	32 (6.7)
K	0	0	0	1 (3.3)	0	0	0	1 (0.2)
T	1 (0.9)	1 (1.9)	1 (1.1)	0	0	1 (1.1)	3 (5.2)	7 (1.5)
J	4 (3.6)	1 (1.9)	2 (2.2)	0	0	5 (5.6)	5 (8.6)	17 (3.5)
R*	0	0	0	1 (3.3)	2 (4.2)	2 (2.2)	1 (1.7)	6 (1.3)
X	3 (2.7)	0	0	0	0	0	0	3 (0.6)
N1a	3 (2.7)	0	0	0	0	0	0	3 (0.6)
I	2 (1.8)	0	1 (1.1)	0	0	0	0	3 (0.6)

Eurasian populations were found with the highest frequency among Altaians (34.5%), Khakassians (18.9%) and Tofalars (20.7%), but are less frequent among Tuvinians, Todjins, Sojots and Buryats. Haplogroup U is the most frequent haplogroup within the West Eurasian fraction, closely followed by haplogroups H and J. A high percentage of haplogroups U and H was observed in the mtDNA pool of Altaians (16.4% and 6.4%, respectively) and Khakassians (11.3% and 3.8%, respectively), whereas the highest frequencies of haplogroups J and T (8.6% and 5.2%, respectively) were detected in Tofalars. Altaians also possess mtDNA sequences belonging to some rare West Eurasian haplogroups such as N1a, X, and I.

Thus, the mtDNA haplogroup distribution data indicate that contemporary South Siberian maternal lineages evolved largely on the basis of Asian-specific substratum, with the West Eurasian component accounting for 7%–35% of mtDNA haplotypes.

Phylogenetic Analysis of South Siberian mtDNA Lineages

To provide further insight into the variation of mtDNA haplogroups revealed in South Siberian populations, a detailed phylogenetic analysis of HVR1 sequences was

performed. Moreover, in order to obtain some additional information on mtDNA classification, we determined HVR2 sequences in Altaians ($n = 110$) characterized by the highest level of mtDNA variability among populations studied here.

Figure 2 shows the median network of haplogroups C, Z and M8, which are defined by a transition at nucleotide position (np) 16298. According to the phylogenetic data based on HVR1 and HVR2 variation in Altaians, both haplogroup C and Z sequences are characterized by the deletion of an adenine residue at np 249 (Figure 3). In addition, based on whole mitochondrial genome sequencing data, these haplogroups share polymorphisms at nps 4715, 7196CA, 8584 (Finnilä *et al.* 2001; Maca-Meyer *et al.* 2001) and therefore should be considered as sister haplogroups (Yao *et al.* 2002).

In haplogroup C, the most frequent haplotype, represented by HVR1 motif 16223–16298–16327, is observed in all populations studied. This haplotype happens to be the ancestral type from which several one-step-related sequences derived. Besides that, at least three distinct clusters were observed, the first being determined by a transition at np 16129, the second by a transition at np 16288, and the third by the HVR1 motif 16171–16344–16357. The first cluster appears to be

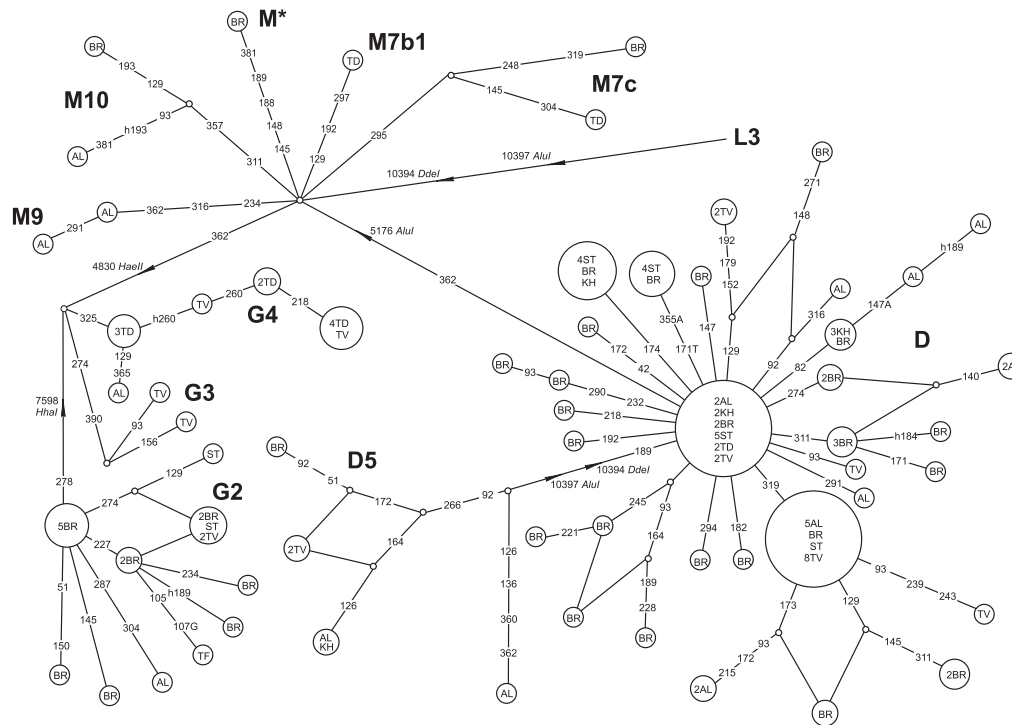


Figure 4 Phylogenetic network of South Siberian D, G and M7, M9, M10 and M* HVR1 sequences. Designations as shown in Figure 2.

contrast, the largest cluster G2 within haplogroup G, determined by transition at np 16278 and *HhaI* site loss at np 7598 (characteristic for haplogroup E), was present in most South Siberian populations studied. Cluster G2 may be fairly ancient since its coalescence age is estimated as 27600 ± 12400 years B.P. In addition to its considerable age, cluster G2 has a striking phylogeographic distribution, restricted to Central and East Asia, being found in Central Asians, Han Chinese, Tibetans and Ainu (Torroni *et al.* 1994; Horai *et al.* 1996; Comas *et al.* 1998; Yao *et al.* 2002).

It should be noted that the additional haplogroup G cluster G1 was previously described in Northeast Siberian populations of Koryaks, Evens, Chukchi and Itelmens (Derenko & Shields, 1997; Starikovskaya *et al.* 1998; Schurr *et al.* 1999). Cluster G1 is characterized by a transition at np 16017, which clearly distinguishes its haplotypes from other haplogroup G mtDNA sequences described here. The presence of at least four distinct clusters within haplogroup G, as well as its obvious geographical substructuring, implies a considerable degree of divergence of these mtDNAs in Siberia.

The remaining M mtDNAs are represented in South Siberia by several minor branches occurring at very low frequencies among Altaians, Buryats and Todjins (Figure 4). The only Todjin HVR1 sequence 16129–16192–16223–16297 probably belongs to haplogroup M7b1, whereas the mtDNA haplotypes defined by 16234–16316 and 16311–16357 HVR1 motifs could be assigned to haplogroups M9 and M10, respectively (Yao *et al.* 2002). Two other HVR1 sequences with the 16295 motif were classified as M7c mtDNAs according to the East Asian mtDNA classification (Yao *et al.* 2002). The Buryat mtDNA with the 16145–16148–16188–16189–16223–16381 HVR1 sequence belongs to the still unassigned M* haplotypes.

Figure 5 shows mtDNA haplotypes classified into various haplogroups of macrohaplogroups N and R. The latter is considered as a subhaplogroup of N shared between Eastern Asians and Western Eurasians. Contrary to the widely spread M haplogroups, East Asian-specific N haplogroups A, N9a, and Y were found in South Siberian populations with frequencies less than 3%. Haplogroup A mtDNAs were found to harbour seven distinct HVR1 sequences. Most of them were

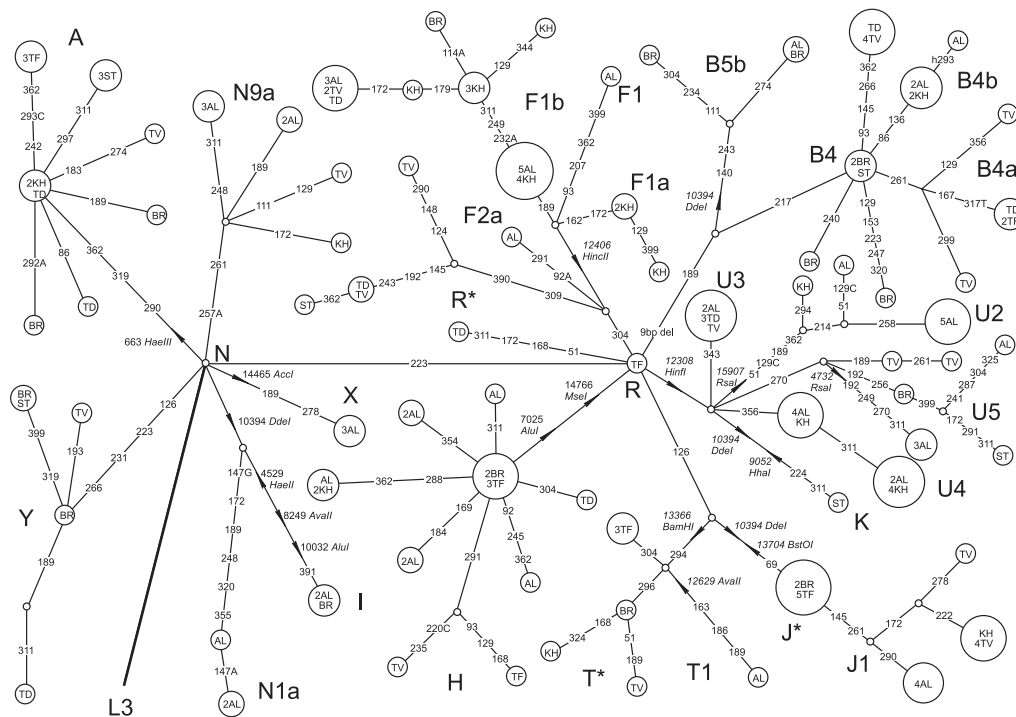


Figure 5 Phylogenetic network of R and N macrohaplogroup HVR1 mtDNA lineages observed in South Siberian populations. Designations as shown in Figure 2.

identified previously in Central and East Asian populations (Kolman *et al.* 1996; Comas *et al.* 1998; Yao *et al.* 2002). It is noteworthy that South Siberian haplogroup A mtDNAs lack polymorphisms at nps 16111 and 16192 which are typical for Northeast Asian and New World populations.

Four different HVR1 sequences were observed within haplogroup Y. Similar to haplogroup A mtDNAs, most of them were described earlier in Central and Eastern Asia, whereas only one HVR1 sequence found in Todjins belongs to the Northeast Asian Y-subcluster defined by a transition at np 16189.

Haplogroup N9a, which is a sister haplogroup to Y according to the East Asian phylogenetic tree (Yao *et al.* 2002), was found in Altaians, Khakassians and Tuvinians. All South Siberian N9a haplotypes were similar or identical to those described previously in Han Chinese (Yao *et al.* 2002) and Central Asians (Comas *et al.* 1998).

The remaining East Asian-specific sequences of South Siberian gene pools belong to the two major haplogroups within macrohaplogroup R, namely B and F. Haplogroup B, comprising 5% of the total data set, is

represented in South Siberia by two major clusters, B4 and B5. Subcluster B5 as a whole has the 16189–16140 motif in association with the presence of a DdeI site at np 10394, whereas B4 is determined by the 16189–16217 motif. In the populations studied, cluster B5 encompassed only two different HVR1 sequences, having an additional transition at np 16243 which is characteristic for the B5b mtDNA haplotypes (Yao *et al.* 2002). In contrast, cluster B4 seems to be the most frequent, covering almost all haplogroup B mtDNAs found in South Siberia. Cluster B4 includes at least two well-represented groups of HVR1 sequences – B4a, defined by a transition at np 16261, and B4b, determined by a transition at np 16136. One should note that the South Siberian B4b mtDNAs had the np 16086C mutation. This polymorphism has been observed in all subcluster B4b mtDNAs of the Altaians, Khakassians and Tuvinians presented here, as well as in Mongolians (Kolman *et al.* 1996) but is absent in those from East Asian populations (Horai *et al.* 1996; Qian *et al.* 2001; Yao *et al.* 2002). The majority of B4a mtDNA haplotypes found in Tuvinians, Todjins and Tofalars were not found in the published data sets and appear to be unique

to South Siberian populations. One exception is the Tuvininan HVR1 sequence with an additional transition at np 16299, which was found previously in a Mongol sample described by Kolman *et al.* (1996).

According to the East Asian mtDNA phylogeny (Yao *et al.* 2002) haplogroup F, which was defined originally by the loss of the *HincII* site at np 12406, represents only one particular branch, named as F1, within haplogroup R9. Haplogroup R9 includes both F and R9a and is identified by the deletion of one A at np 249, as well as a transition at np 10310 (Yao *et al.* 2002). Haplogroup F is therefore defined more widely and may be distinguished by a transition at np 16304. Haplogroup F, comprising 5% of the South Siberian data includes two clusters. Only one mtDNA haplotype from the F2a subcluster was observed in Altaians. This mtDNA, characterized by the 16092A–16291 HVR1 motif and the deletion of A at np 249 in HVR2 (Figures 3 and 5) is likely identical to that described in Han Chinese (Yao *et al.* 2002). The remaining F-sequences are defined by the *HincII* site at np 12406 and belong to the F1 cluster, which may be further subdivided into two additional subclusters. The first subcluster, F1a, defined by transitions at nps 16162 and 16172, encompassed two different HVR1 sequences found in Khakassians. The second subcluster, F1b, determined by a transition at np 16189, is linked with six different HVR1 sequences observed in the majority of populations studied here. It should be noted that only one F1b mtDNA with a 16114A variant seems to be unique for South Siberia, whereas the remaining sequences were described previously in Central and Eastern Asians (Kolman *et al.* 1996; Horai *et al.* 1996; Comas *et al.* 1998; Yao *et al.* 2002). One Altaiian F1 mtDNA, having a 16093–16207–16304–16362–16399 HVR1 sequence, could not be classified as F1a or F1b. Similarly, five R* haplotypes found in Tuvinians, Tadjins, Tofalars and Sojots could not be further specified (Figure 5). Evidently, some of these R* mtDNAs belong to specific subgroups (one with HVR1 motif 16304–16309–16390 and another with 16051–16168–16172–16311), the phylogenetic position of which is not clear.

The West Eurasian fraction of the mtDNA pool of South Siberians is represented by haplogroups U, J, T, and H, belonging to macrohaplogroup R and by haplogroups I, N1a, and X from macrohaplogroup N

(Figure 5). Haplogroup U, which accounts for nearly 7% of the total data set, contains five clusters U2, U3, U4, U5, and K. Although cluster K seems to be non-typical for South Siberian populations, being found in only one Sojot individual, other U-clusters are widely distributed in the populations studied. In contrast to U3 and U5, which are found in the majority of South Siberian populations, clusters U2 and U4 appear to be specific to Altaians and Khakassians. Moreover, U2 haplotypes defined by a transition at np 16214 seem to be unique for Altaians as apart from one occurrence in Northern Caucasians, similar or identical haplotypes have not been found in Eurasian populations (according to the database of Richards *et al.* 2000). Cluster U5 showed the greatest diversity of HVR1 sequences relative to other clusters of haplogroup U. South Siberian cluster U5 encompassed six haplotypes, which belong to subclusters U5a (with 16192–16256–16270 HVR1 motif) and U5b (with 16189–16270 motif). It should be noted that the Altaiian haplotype 16192–16249–16311, similar to U1 sequences, belongs nevertheless to subcluster U5b, since it is characterized by the presence of an *RsaI* site at np 4732 and transition at np 150 in HVR2 (Figures 3 and 5).

HVR1 sequences from haplogroup H, which is the largest in Europe, were detected in mtDNA pools of South Siberians with an overall frequency of 3.5%. Two Buryat and three Tofalar HVR1 sequences were found to be identical to CRS, whereas the remaining H-haplotypes differed from it by one to four nucleotide substitutions. As expected all Altaiian H-sequences have the 73A variant in HVR2 with the exception of haplotype 16169–16184, characterized by 73G (Figure 3). Interestingly, that the same H-haplotype was found recently in the Northwest Siberian Mansi population (Derbeneva *et al.* 2002).

Haplogroup J, which was revealed at low frequencies in the majority of South Siberian populations, is represented by two clusters, J* and J1. The central haplotype of J* with the 16069–16126 HVR1 sequence, was shared by two Buryats and five Tofalars and also spread in Yakuts and Evens (Derenko & Shields, 1997); it may have a wide geographic distribution, whereas J1 mtDNA haplotypes are restricted to South Siberia being found in Altaians, Tuvinians and Khakassians. Haplogroup T, which has a common origin with haplogroup

J, encompassed four T* haplotypes found in Tofalars, Buryats, Khakassians and Tuvinians, and only one T1 haplotype revealed in Altaians. We have not found any identical T-sequences in the neighbouring populations of Central and East Asia, although several T* and T1 mtDNAs were observed previously in Mongolians (Kolman *et al.* 1996), Kazakhs, Kirghizs and Uighurs (Comas *et al.* 1998).

Altaians also possess mtDNAs from haplogroups X, N1a and I. Haplogroups N1a and I may have a common origin since they share transitions at nps 199 and 204 in the HVR2 (Figure 3), as well as several coding-region variants (Kivisild *et al.* 1999). Haplogroup N1a seems to be very rare, occurring as only 2.7% of the present data set, and being found only in Altaians. N1a-sequences have been detected so far with low frequencies in different populations of Europe and West Asia (Richards *et al.* 2000), but were not found in the Central and East Asian populations (Kolman *et al.* 1996; Comas *et al.* 1998; Yao *et al.* 2002). The same remark is true for the haplogroup I sequence, revealed in Altaian and Buryat individuals.

Haplogroup X mtDNAs found in Altaians are represented by HVR1 motif 16189–16223–16278 which was proposed to be a root sequence of the X phylogeny (Brown *et al.* 1998). Haplogroup X has a remarkable geographic distribution – it occurs with low frequencies in Western Eurasian populations and amongst Native Americans, but has not been found in Asians, including Siberians, suggesting that it may have come to the Americas via a Eurasian migration (Brown *et al.* 1998). The only exception are Altaian X mtDNA variants occupying intermediate phylogenetic positions between

European and Native American X haplotypes, as it has been shown earlier (Derenko *et al.* 2001).

Thus, the data presented in this study demonstrate that South Siberian populations represent a complex pattern of the mtDNA structure, reflecting diverse interactions that occurred at different times between eastern and western Eurasian populations. Moreover, the South Siberian gene pool contains traces of the source of different expansions from the Central Asia/South Siberia region into the Americas and North Eurasia, reaching the northern European territories.

Sequence Diversity and Genetic Structure of South Siberian Populations

Table 4 lists some diversity parameters estimated for South Siberian HVR1 data. The nucleotide diversity ranged from 0.012 in Sojots to 0.017 in Altaians and Khakassians, while the haplotype diversity ranged from 0.867 in Tofalars to 0.991 in Buryats. When these values are compared with those of populations from Central and East Asia, it is apparent that the nucleotide diversity values of South Siberian populations were similar to those found in Central and Eastern Asian populations (0.014–0.019), while the haplotype diversity was higher in Central and Eastern Asia (0.980 – 0.999). The mean number of pairwise nucleotide differences was fairly uniform across different South Siberian groups, ranging from 4.9 in Sojots to 6.7 in Altaians (Table 4). These estimates are within the range of mean pairwise differences found in Central and Eastern Asian populations (5.5–7.6). All the mismatch distributions for the South Siberian groups were approximately

Table 4 MtDNA diversity parameters in South Siberian populations

Population	Sample size	No. of lineages	No. of polymorphic sites	Haplotype diversity	Nucleotide diversity	Mean number of pairwise differences	Mismatch observed variance	Tau	Raggedness index
Altaians	110	52	76	0.981	0.017	6.70	6.74	6.51	0.010
Buryats	91	65	77	0.991	0.014	5.50	5.93	4.85	0.015
Sojots	30	14	34	0.920	0.012	4.86	8.63	2.91	0.013
Todjins	48	27	44	0.971	0.015	5.84	8.05	4.35	0.006
Tuvinians	90	42	65	0.964	0.016	6.50	10.01	4.63	0.007
Khakassians	53	31	48	0.976	0.017	6.64	8.60	5.23	0.013
Tofalars	58	16	37	0.867	0.015	5.91	10.95	3.67	0.036

bell-shaped, suggesting prehistoric population expansions. The raggedness statistic for the Siberian mismatch distributions varied from 0.006 to 0.036 (Table 4); values of raggedness index less than 0.05 are also indicative of prehistoric population expansions (Harpending *et al.* 1993). Assuming that the mismatch distributions did therefore reflect past population expansions, we estimated tau, the time of population expansion in units of mutational time (Table 4). The estimated tau-values varied from 2.9 to 6.5, which corresponds to estimated expansion times of 22000–49000 years ago, assuming a rate of human mtDNA divergence of 33% per million years (Ward *et al.* 1991).

The population structure of South Siberian mtDNA sequences was investigated by the AMOVA procedure (Excoffier *et al.* 1992). AMOVA showed that, when the seven populations were treated as a single group, 94.95% of the total variance was within populations and 5.05% (which was statistically significant at $p < 0.001$) was between populations. Eliminating Tofalars, which show reduced haplotype diversity due to possible founder effect, results in an average F_{st} value 2.86% for the remaining populations. Thus, overall, high level of between-population differentiation is observed in South Siberian populations.

The pairwise F_{st} values were statistically significant for all South Siberian population pairs, except for Altaians and Khakassians, Tuvinians and Todjinians, Buryats and Sojots, thus providing some evidence for genetic structuring. The populations were then subdivided

into Turkic-speaking (Altaians, Khakassians, Sojots, Tuvinians, Todjins and Tofalars) and Mongolic-speaking (Buryats). According to the AMOVA, the fraction of genetic variance that could be attributed to language was 1.09% (not statistically different from zero) thus indicating that language does not reflect any difference in the mtDNA pool of South Siberian populations. When populations were grouped on the basis of anthropological characteristics into Central Asian (Altaians, Khakassians, Sojots, Buryats and Tuvinians) and Baikalian (Tofalars and Todjins) groups, the proportion of the genetic variance that was due to the differences between groups was again less than between populations within groups (3.32% and 3.69%, respectively). However, when the populations were classified according to geographic proximity, geography could account for 5.01% of the mtDNA genetic variance. In that case Altaians were grouped with Khakassians, Tuvinians with Todjins and Buryats with Sojots, whereas Tofalars, representing a small, geographically isolated tribe, were treated as a separate group.

In order to place the seven populations studied here in a broader geographical context, we expanded the analysis to include seven additional populations from Central and Eastern Asia: Mongolians (Kolman *et al.* 1996), Han Chinese and Koreans (Horai *et al.* 1996), Kazakh, Kirghiz and Uighur (Comas *et al.* 1998) and Han Chinese (Yao *et al.* 2002). The pairwise F_{st} values indicated that all seven South Siberian populations, as well as two Chinese populations, differ significantly from Eastern

Table 5 Pairwise F_{st} -values between South Siberian, East Asian and Central Asian populations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Altaians	–													
2 Buryats	0.025 ^a	–												
3 Sojots	0.046 ^a	0.009 ^d	–											
4 Todjins	0.035 ^b	0.038 ^a	0.062 ^b	–										
5 Khakassians	0.009 ^d	0.045 ^a	0.073 ^a	0.031 ^c	–									
6 Tofalars	0.118 ^a	0.139 ^a	0.191 ^a	0.085 ^a	0.091 ^a	–								
7 Tuvinians	0.025 ^a	0.025 ^a	0.055 ^b	0.005 ^d	0.020 ^c	0.066 ^a	–							
8 Chinese	0.026 ^a	0.047 ^a	0.056 ^a	0.091 ^a	0.035 ^b	0.170 ^a	0.071 ^a	–						
9 Mongolians	0.006 ^c	0.013 ^c	0.028 ^b	0.046 ^a	0.022 ^b	0.138 ^a	0.033 ^a	0.018 ^b	–					
10 Koreans	0.022 ^b	0.014 ^c	0.025 ^b	0.069 ^a	0.053 ^a	0.189 ^a	0.060 ^a	0.027 ^b	0.006 ^d	–				
11 Kazakh	0.012 ^c	0.034 ^a	0.051 ^a	0.075 ^a	0.030 ^b	0.166 ^a	0.059 ^a	0.015 ^c	0.004 ^d	0.023 ^b	–			
12 Kirghiz	0.008 ^c	0.017 ^b	0.036 ^a	0.055 ^a	0.028 ^a	0.150 ^a	0.041 ^a	0.020 ^b	0.005 ^d	0.015 ^b	0.000 ^d	–		
13 Uighur	0.022 ^a	0.051 ^a	0.068 ^a	0.102 ^a	0.044 ^a	0.193 ^a	0.082 ^a	0.011 ^c	0.013 ^c	0.030 ^a	–0.003 ^d	0.008 ^d	–	
14 Chinese Han	0.015 ^a	0.030 ^a	0.043 ^a	0.074 ^a	0.035 ^a	0.158 ^a	0.058 ^a	0.001 ^d	0.007 ^c	0.010 ^b	0.009 ^c	0.011 ^b	0.009 ^c	–

Note: ^a $P < 0.001$; ^b $0.005 > P > 0.001$; ^c $0.05 > P > 0.005$; ^d $P > 0.05$

and Central Asian populations (Table 5). On the contrary, Mongolians were indistinguishable from Koreans, Kazakh and Kirghiz, while the two latter populations were closely associated both with each other and with Uighurs (Table 5). When applied to Central Asian populations (Kazakh, Kirghiz and Uighur), the AMOVA revealed that only 0.22% ($p > 0.1$) of the total variance could be attributed to differences between populations. East Asian populations of Mongolians, Chinese and Koreans exhibited a slightly higher F_{st} -value of 0.85% ($p < 0.001$), while South Siberian mtDNA differentiation was six times higher. Summarizing, South Siberian populations contain high levels of mtDNA diversity and high heterogeneity of mtDNA sequences among populations that could be due to geography, but not due to language and anthropological features. It should be noted that the pattern of South Siberian mtDNA differentiation is consistent with analysis of classical genetic markers, which reveals a high degree of heterogeneity among Siberian populations (Rychkov & Sheremetyeva, 1977).

Conclusions

Four hundred and eighty individuals from seven different Altaic-speaking populations of South Siberia were studied in order to estimate the mtDNA variation, and determine the relative contribution of East Asian and West Eurasian lineages to the gene pool of the present-day South Siberians. In summary, our phylogenetic analysis shows that the majority of South Siberian mtDNA sequences can be perfectly classified into specific sub-haplogroups of the Eurasian founder macrohaplogroups M, N, and R. The total sample revealed as much as 81% East Asian and 17% West Eurasian contribution to the total mtDNA pool. The amount of West Eurasian ancestry varies widely and has distinct patterns in different regions of South Siberia. The highest influx of West Eurasian mtDNA lineages was observed in populations of the Altai region (18.9% in Khakassians and 34.5% in Altaians). Also, a high West Eurasian mtDNA contribution in the East Sayan populations of Tuvinians, Todjins and Tofalars was observed (ranging from 12.5% to 22.4%), whereas in Buryats and Sojots from the Baikal region it was markedly lower (less than 10%).

Such east-to-west cline in the frequencies of West Eurasian-specific mtDNA haplotypes observed in South Siberia is consistent with archaeological and paleo-anthropological views about the presence of Europeoid-specific traits in inhabitants of the Altai and Sayan region since the Bronze Age. However, beginning from the early Iron Age, the presence of Mongoloid component has been increasing, becoming prevalent in modern times. Thus, dynamics of the anthropological composition of the Altai and Sayan region populations can be characterized by definitely directed replacement of the Europeoid component by the Mongoloid one (Alexeev, 1989). It should be noted in this respect that this process was nonuniform. Anthropological data demonstrate the territorial differences in the chronology of ethnogenetic processes in at least three largest groups of the Altai and Sayan region: Khakassians, Altaians, and Tuvinians. Most intensive process of Khakassians formation dates back to the end of the first millennium A.D., while the admixture of Mongoloids and Europeoids in the Altai area was completed between the first and second millenniums A.D. In Tuva, the prevalence of Europeoids can be traced back up to the pre-Mongolian time. Moreover, modern Tuvinians display anthropological features specific for southern Europeoids, whereas Khakassians demonstrate the influence of Eastern European anthropological traits (Alexeev & Gohman, 1984). On the other hand, Baikalian populations ancestral to the present-day Buryats were characterized by Mongoloid-specific anthropological features since Neolithic (Alexeev & Gohman, 1984).

Similarly, Y chromosome data revealed dual affinities of the South Siberian male lineages: they are generally characterized by a subset of southern East Asian haplotypes, while Kets, Selkups, Altaians, Shors and Khakassians also demonstrate the presence of a paleo-Europeoid component ancestral to the Native American and European Y-chromosome lineages (Santos *et al.* 1999; Karafet *et al.* 1999; Derenko *et al.* 2002).

The majority of South Siberian mtDNA sequences described in this study belongs to East Asian-specific haplogroups and therefore may have Central and/or East Asian roots. Moreover, the results of the present study clearly demonstrate that a subset of Asian-specific mtDNA haplogroups (A, B, C, D, F, G, Y, and Z) described previously in Siberians (Torroni *et al.* 1993;

Starikovskaya *et al.* 1998; Schurr *et al.* 1999; Derenko *et al.* 2000) could be further extended by inclusion of additional M7, M8, M9, M10, N9a, F2 haplogroups that have been revealed for the first time in South Siberian populations. The considerable substructure within South Siberian haplogroups B, F, and G, together with a high degree of haplogroup C and D diversity revealed there, allows us to conclude that South Siberians carry the genetic imprint of an early-colonization phase of Eurasia. Moreover, the early presence of Europeoids in South Siberia region is confirmed by the occurrence of unique U2-16214 haplotypes in Altaians as well as by the relatively high frequency of subhaplogroup U4 revealed in Altaians and Khakassians. These U4 mtDNAs are identical to those described recently in the Northwest Siberian Mansi where they were considered as an indicator of Upper Paleolithic population of Europeans preserved in Siberia (Derbeneva *et al.* 2002).

Intriguingly, despite numerous historically recorded migrations and substantial gene flow across South Siberia from the Bronze Age to the present time, the high degree of between-population differentiation has been maintained, suggesting the influence of specific demographic factors. More extensive sampling of East and Central Asian populations should provide more precise and reliable information about the relationships between their mitochondrial gene pools, and might reveal continent-wide patterns in the distribution of particular haplotypes or haplogroups; this, in turn, will contribute to our understanding of the demographic history of modern Eurasian populations.

Acknowledgements

This research was supported by grants from the Russian Foundation for Basic Research (99-06-80430), the State Committee for Scientific Research (3 P04C 048 23) and Ludwik Rydygier Medical University in Bydgoszcz, Poland (BW66/02). We thank Ewa Lewandowska for her excellent technical assistance and three anonymous reviewers for helpful comments on the manuscript.

References

- Alexeev, V. P. (1989) *Historical anthropology and ethnogenesis*. Moscow: Nauka (in Russian).
 Alexeev, V. P. & Gohman, I. I. (1984) *Anthropology of Asiatic part of the USSR*. Moscow: Nauka (in Russian).

- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R. & Drouin, J. *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
 Bandelt, H.-J., Forster, P., Sykes, B. C. & Richards, M. B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.
 Bandelt, H.-J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48.
 Brown, M. D., Hosseini, S. H., Torroni, A., Bandelt, H.-J., Allen, J. C. & Schurr, T. G. *et al.* (1998) MtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* **63**, 1852–1861.
 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
 Chlachula, J. (2001) Pleistocene climate change, natural environments and palaeolithic occupation of the Altai area, west-central Siberia. *Quaternary International* **80–81**, 131–167.
 Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E. & Martinez-Arias, R. *et al.* (1998) Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* **63**, 1824–1838.
 Delghandi, M., Utsi, E. & Krauss, S. (1998) Saami mitochondrial DNA reveals deep maternal lineage clusters. *Hum Hered* **48**, 108–114.
 Derbeneva, O. A., Starikovskaya, E. B., Wallace, D. C. & Sukernik, R. I. (2002) Traces of early Eurasians in the Mansi of Northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* **70**, 1009–1014.
 Derenko, M. V., Grzybowski, T., Malyarchuk, B. A., Czarny, J., Miścicka-Śliwka, D. & Zakharov, I. A. (2001) The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am J Hum Genet* **69**, 237–241.
 Derenko, M. V., Malyarchuk, B. A., Dambueva, I. K., Shaikhaev, G. O., Dorzhu, C. M., Nimaev, D. D. & Zakharov, I. A. (2000) Mitochondrial DNA variation in two South Siberian aboriginal populations: Implications for the genetic history of North Asia. *Hum Biol* **72**, 945–973.
 Derenko, M. V., Malyarchuk, B. A., Denisova, G. A., Dorzhu, Ch. M., Karamchakova, O. N. & Luzina, F. A., *et al.* (2002) The Y-chromosome biallelic loci variability in populations of Altai and Sayan region. *Genetika* **38**, 393–399 (in Russian).
 Derenko, M. V. & Shields, G. F. (1997) Mitochondrial DNA sequence diversity in three North Asian aboriginal population groups. *Mol Biol (Moscow)* **31**, 665–669.
 Derevianko, A. P. (1998) Human occupation of nearby regions and the role of population movements in the Paleolithic of Siberia. In: *The Paleolithic of Siberia: New*

- Discoveries and Interpretations* (ed. A. P. Derevianko), pp. 336–351. Urbana, IL: University of Illinois Press.
- Derevianko, A. P. & Zenin, V. N. (1996) Preliminary results of field investigations of the Late Palaeolithic site Shestakhovo in (1996). In: *New archaeological and ethnographic discoveries in Siberia* (eds. A. P. Derevianko & V. I. Molodin), pp. 60–63. Novosibirsk: Institute of Archaeology and Ethnography Press (in Russian).
- Excoffier, L., Smouse, P. & Quattro, J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.
- Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**, 935–945.
- Forster, P., Torroni, A., Renfrew, C. & Röhl, A. (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* **18**, 1864–1881.
- Finnilä, S., Lehtonen, M. S. & Majamaa, K. (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* **68**, 1475–1484.
- Goebel, T. (1999) Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. *Evol Anthropol* **8**, 208–227.
- Gurvich, I. S. (1980) *Ethnogenesis of North Indigenous Peoples*. Moscow: Nauka (in Russian).
- Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) The genetic structure of ancient human populations. *Curr Anthropol* **34**, 483–496.
- Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y. & Fucharoen, G. *et al.* (1996) MtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am J Hum Genet* **59**, 579–590.
- Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A. & Long, J. *et al.* (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* **64**, 817–831.
- Kivisild, T., Bandelt, H.-J., Wang, J., Derenko, M., Mal'yarchuk, B. & Golubenko, M. *et al.* (2001) Mitochondrial DNA tree for Eastern Asian populations. In: *Abstracts of the First workshop on information technologies application to problems of biodiversity and dynamics of ecosystems in North Eurasia (WITA'2001)*, p. 302. Novosibirsk: Institute of Cytology and Genetics.
- Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S. & Villem, R. (1999) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. In: *Genomic diversity: applications in human population genetics* (eds. S. Papiha, R. Deka & R. Chakraborty), pp. 135–152. New York: Kluwer Academic/Plenum Publishers.
- Kolman, C. J., Sambuughin, N. & Bermingham, E. (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**, 1321–1334.
- Levin, M. G. & Potapov, L. P. (1964) *The peoples of Siberia*. Chicago: University of Chicago Press.
- Luzina, F. A. (1987) *Hereditary polymorphism and genetic processes in indigenous populations of Altai region*. PhD thesis, Moscow: Moscow State University (in Russian).
- Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* **2**, 13.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F. & Guida, V. *et al.* (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.
- Okladnikov, A. P. (1981) *The Palaeolithic of Central Asia. Moltyn-Am (Mongolia)*. Novosibirsk: Nauka (in Russian).
- Osipova, L. P., Kashinskaya, Yu. O., Posukh, O. L., Ivakin, E. A. & Kryukov, Yu. A. (1997) A genetic demographic study of the South Altaian Population of the Mendur-Sokkon Village, Altai Republic. *Genetika* **33**, 1559–1564 (in Russian).
- Potapov, L. P. (1969) *Ethnic structure and origin of Altaians*. Leningrad: Nauka (in Russian).
- Qian, Y. P., Chu, Z.-T., Dai, Q., Wei, C.-D., Chu, J. Y., Tajima, A. & Horai, S. (2001) Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet* **46**, 211–220.
- Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence for an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nature Genet* **23**, 437–441.
- Richards, M. & Macaulay, V. (2000) Genetic data and the colonization of Europe: genealogies and founders. In: *Archaeogenetics: DNA and the population prehistory of Europe* (eds. C. Renfrew & K. Boyle), pp. 139–151. Cambridge: McDonald Institute for Archaeological Research.
- Richards, M. B., Macaulay, V. A., Hickey, E., Vega, E., Sykes, B. & Guida, V. *et al.* (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.
- Rychkov, Yu. G. & Sheremetyeva, V. A. (1977) The genetic process in the system of ancient human isolates in North Asia. In: *Population structure and human variation* (ed. G. A. Harrison), pp. 47–108. Cambridge: Cambridge University Press.
- Rychkov, Yu. G., Zhukova, O. V., Sheremetyeva, V. A., Spitsyn, V. A., Brook, S. I. & Lebedeva, I. A. *et al.* (2000) Gene pool of population of Russia and contiguous

- countries. In: *Gene pool and genegeography of population* (ed. Yu. G. Rychkov), v. 1. Saint Peterburg: Nauka (in Russian).
- Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D. J., Schanfield, M. & Leonard, W. R. *et al.* (1999) The Central Siberian origin for Native American Y chromosomes. *Am J Hum Genet* **64**, 619–628.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin v. 2.000: a software for population genetics data analysis*. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.
- Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B. & Wallace, D. C. (1999) Mitochondrial DNA variation in Koryaks and Itel'men: Population replacement in the Okhotsk Sea – Bering Sea region during the Neolithic. *Am J Phys Anthropol* **108**, 1–39.
- Starikovskaya, Y. B., Sukernik, R. I., Schurr, T. G., Kogelnik, A. M. & Wallace, D. C. (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* **63**, 1473–1491.
- Sullivan, K. M., Hopgood, R. & Gill, P. (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int J Legal Med* **105**, 83–86.
- Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L. & Scozzari, R. *et al.* (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.
- Torroni, A., Miller, J. A., Moore, L. G., Zamudio, S., Zhuang, J., Droma, T. & Wallace, D. C. (1994) Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* **93**, 189–199.
- Torroni, A., Sukernik, R. I., Schurr, T. G., Starikovskaya, Y. B., Cabell, M. F. & Crawford, M. H. *et al.* (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* **53**, 591–608.
- Turner II, C. G. (1987) Late Pleistocene and Holocene population history of east Asia based on dental variation. *Am J Phys Anthropol* **73**, 305–321.
- Vasiliev, S. A. (1993) The Upper Paleolithic of northern Asia. *Curr Anthropol* **34**, 82–92.
- Vasiliev, S. A., Yamskikh, A. F., Yamskikh, G. I., Svezhentsev, Y. S. & Kasparov, A. K. (1999) Stratigraphy and palaeoecology of the Upper Palaeolithic sites near the Maima village (Upper Yenisei valley, Siberia). In: *Quaternary of Siberia. Quaternary Geology, Paleogeography and Palaeolithic Archaeology* (eds. J. Chlachula, R. A. Kemp & J. Tyráček), pp. 29–35. *Anthropozoikum* **23**.
- Wallace, D. C. (1995) Mitochondrial DNA variation in human evolution, degenerative disease and aging. *Am J Hum Genet* **57**, 201–223.
- Walsh, P. S., Metzger, D. A. & Higuchi, R. (1991) Research report: Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**, 506–513.
- Ward, R. H., Frazier, B. L., Dew, K. & Pääbo, S. (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* **88**, 8720–8724.
- Yao, Y.-G., Kong, Q.-P., Bandelt, H.-J., Kivisild, T. & Zhang, Y.-P. (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* **70**, 635–651.

Received: 14 May 2002

Accepted: 29 October 2002