
mtDNA Variation in the Altai-Kizhi Population of Southern Siberia: A Synthesis of Genetic Variation

CHRISTINE PHILLIPS-KRAWCZAK,^{1,2} ERIC DEVOR,³ MARK ZLOJUTRO,² KRISTIN MOFFAT-WILSON,³ AND MICHAEL H. CRAWFORD^{1,2}

Abstract The native peoples of Gorno Altai in southern Siberia represent a genetically diverse population and have been of great interest to anthropological genetics. In particular, the southern Altaian population is argued to be the best candidate for the New World ancestral population. In this study we sampled Altai-Kizhi from the southern Altaian village of Mendur-Sokkon, analyzed mtDNA RFLP markers and HVS-I sequences, and compared the results to other published mtDNA data from Derenko et al. (2003) and Shields et al. (1993) encompassing the same region. Because each independent study uses different sampling techniques in characterizing gene pools, in this paper we explore the accuracy and reliability of evolutionary studies on human populations. All the major Native American haplogroups (A, B, C, and D) were identified in the Mendur-Sokkon sample, including a single individual belonging to haplogroup X. The most common mtDNA lineages are C (35.7%) and D (13.3%), which is consistent with the haplogroup profiles of neighboring Siberian groups. The Mendur-Sokkon sample exhibits depressed HVS-I diversity values and neutrality test scores, which starkly differs from the Derenko et al. (2003) data set and more closely resembles the results for neighboring south Siberian groups. Furthermore, the multidimensional scaling plot of D_A genetic distances does not cluster the Altai samples, showing different genetic affinities with various Asian groups. The findings underscore the importance of sampling strategy in the reconstruction of evolutionary history at the population level.

To identify differences among various samples from the same geographic area and to investigate the effect of independent sampling on the reconstruction of human evolutionary history, we examined multiple data sets from a key region of southern Siberia, the Gorno Altai of the Altai Republic. This semi-autonomous republic of Russia is situated in Eurasia, adjacent to China, Mongolia, Kazakhstan, and Russian Siberia. In language, culture, geographic distribution, and genetics, the Altaian tribes can be divided into two distinct groups: the Northern

¹Genetics Program, University of Kansas, Lawrence, KS.

²Laboratory of Biological Anthropology, University of Kansas, Lawrence, KS 66045.

³Integrated DNA Technologies, Coralville, Iowa.

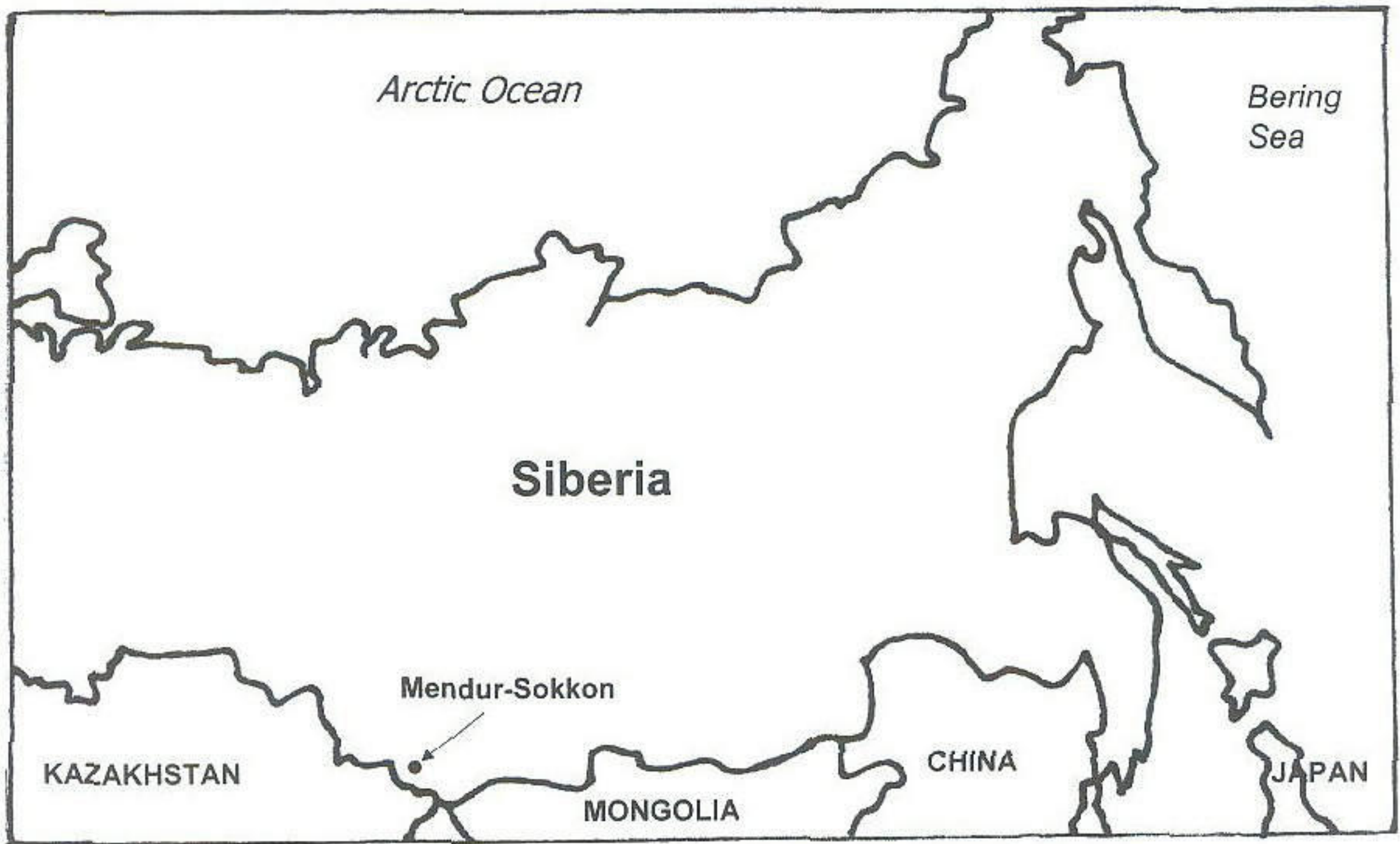


Figure 1. Map of Siberia and location of Mendur-Sokkon village.

Altaians and the Southern Altaians (Levin and Potapov 1964). The Chelkans, Kumandins, Tubalars, and Maimalars represent the Northern Altaian ethnic groups. The Southern Altaian groups consist of the Altai-Kizhi, also referred to as “proper Altai,” the Teleuts, and the Telenghits (Derenko et al. 2003). The focus of the present study, the village of Mendur-Sokkon, lies in the Southern Altaian region (Figure 1).

The rich history of the Gorno Altai contributes to its significant anthropological interest, especially regarding the peopling of the Americas across the Bering land bridge. For more than 50,000 years the region has experienced continuous *Homo* habitation, judging from several Middle Paleolithic Mousterian sites (Goebel 1999). The Altai was at the crossroads between Europe and Asia, with population movements and invasions from Mongolia, China, and Russia. Archeological evidence reveals that since the 5th and 6th centuries B.C., the Altai have had cultural contact with Mongolians in the form of both trade and invasion (Levin and Potapov 1964). In addition, historical texts document that the Altai have maintained trade connections with the Chinese. In 1756 the Altai established cultural and political contacts with the Russians, who were expanding into Siberia (Levin and Potapov 1964).

As the diverse ethnic history of the region would suggest, the Altaian mtDNA gene pool displays a complex mixture of East Asian and West Eurasian lineage (Derenko et al. 2001a, 2003). This region is of particular importance with regard to the peopling of the New World because it is the only region outside the Americas where all five founding New World mtDNA haplogroups—A, B,

C, D, and X—have been identified (Derenko et al. 2001b). For this reason the Altai region has been an area of intense interest to anthropological geneticists, and several field investigations have been conducted by various research teams.

Our goals in this paper are to genetically characterize the Southern Altaian population of Mendur-Sokkon village by using both mtDNA RFLP and HVS-I data and to compare these results with other published studies of the Altai (Shields et al. 1993; Derenko et al. 2003) and surrounding populations. In this paper we also address the reliability of sampling and genetic analysis of populations from the same geographic region by determining whether the multiple mtDNA data sets from the Altaian population yield similar diversity scores and phylogenetic results.

Materials and Methods

Population: The Altai-Kizhi of Gorno Altai. Approximately 60,000 native people live in the Gorno Altai region. The indigenous Altaian population is composed of various distinct Turkic-speaking groups to which the Altai-Kizhi belong (Levin and Potapov 1964). The Altaian valleys between the surrounding mountains of the Altai-Sayan range provide exceptional grazing grounds, and these pastures are used today by the Altai-Kizhi, a pastoral, seminomadic people who traditionally subsist by herding cattle (Crawford et al. 2002).

The sample consists of 98 individuals from the village of Mendur-Sokkon. Informed consent was obtained by the Russian members of the research team when the samples were collected in 1994. The village population numbers approximately 1,000 individuals subdivided into three major patrilineal clans (Irkut, Kipchak, and Todosh) plus a number of smaller clans.

The three Altaian samples considered here are referred to as Altai 1 (present study of Mendur-Sokkon), Altai 2 (Shields et al. 1993), and Altai 3 (Derenko et al. 2003). The surrounding populations used for comparison with the Altai samples consist of Kirghiz, Kazakh, Uighur, Khakassians, Tuva, Sojots, Buryats, Mongol, Han Chinese, and Russians (Figure 2).

mtDNA Analysis. DNA was extracted from the blood samples using a standard phenol-chloroform technique (Birren et al. 1997). The DNA samples were analyzed for mtDNA haplogroups corresponding to Asian- and European-specific lineages. Using standard PCR methods, we analyzed amplified mtDNA fragments for diagnostic variant sites using restriction fragment length polymorphisms (RFLPs). Table 1 identifies the polymorphic sites and the restriction enzymes used for the mtDNA haplogroup designations.

Restricted fragments were separated and typed by electrophoresis on 3% NuSieve agarose gels. Gels were stained with ethidium bromide, and DNA fragments were visualized using an ultraviolet transilluminator. Sequencing of the mtDNA HVS-I from nucleotide position 16151 to 16383 was performed at Integrated DNA Technologies on an ABI 310 capillary sequencer to further delineate

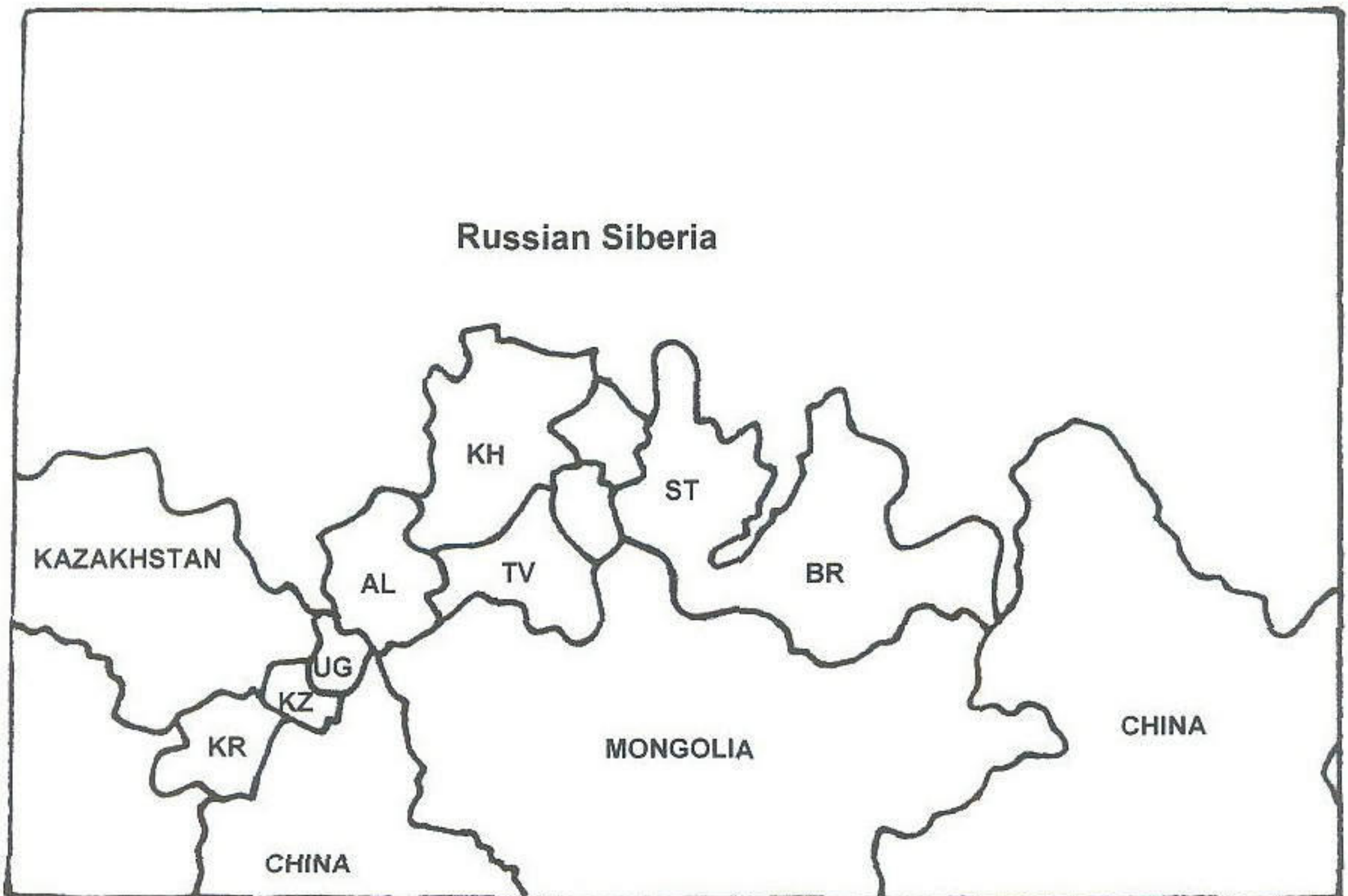


Figure 2. Geographic locations of southern Siberian populations [adapted from Derenko et al. (2003) and Comas et al. (1998)]. Populations: KR, Kirghiz; KZ, Kazakh; UG, Uighur; AL, Altaians; KH, Khakassians; TV, Tuva; ST, Sojots; BR, Buryats.

haplogroup membership into individual haplotypes. Region 16151–16383 was the longest consecutive sequence that encompassed the largest number of successful runs. Failed runs were also attempted using the reverse primer in addition to the forward primer, without success. The failed runs were likely caused by sample degradation, which prevented successful sequencing; thus 61 samples were used in the HVS-I analyses.

Analytical Methods

Diversity Measures. Gene diversity,

$$H = \left(\frac{n}{n-1} \right) \left(1 - \sum_{i=1}^k p_i^2 \right), \quad (1)$$

θ_S , and θ_π were calculated for the sequence data in the Altai and the comparative population data set using Arlequin, version 2.0 (Schneider et al. 2000). The parameter θ equals $2N_e\mu$ for haploid genetic systems, where N_e is effective population size and μ is the mutation rate. The two θ estimators are derived from different aspects of the mtDNA sequence data, each providing an alternative

Table 1. Polymorphic Sites Determining mtDNA Haplogroups

<i>mtDNA Group</i>	<i>Restriction Sites</i>
A	+633 <i>HaeIII</i>
B	-8250 <i>HaeIII</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>
F	-12406 <i>HpaI/HincII</i>
G	+10394 <i>DdeI</i> , +10397 <i>AluI</i> , +4830 <i>HaeII</i> , +4831 <i>HhaI</i>
H	-7025 <i>AluI</i>
I	-4529 <i>HaeII</i> , +8249 <i>AvaII</i> , +16389 <i>BamHI</i> , +10032 <i>AluI</i>
J	-13704 <i>BstNI</i>
K	-9052 <i>HaeII</i>
M	+10394 <i>DdeI</i> , +10397 <i>AluI</i>
T	+13366 <i>BamHI</i>
U	+12308 <i>HinfI</i>
X	-1715 <i>DdeI</i> , +14465 <i>AccI</i>

picture of past evolutionary processes. θ_π , the mean number of pairwise differences between sequences, and θ_S both reflect recent and ancient demographic events, but θ_S has been more responsive to changes in effective population size during recent history (Rogers and Harpending 1992; Helgason et al. 2000).

Median-Joining Network Analysis. Median-joining network analysis was carried out, and phylogenetic networks were constructed by hand with the aid of Network, version 4.0 (Fluxus Engineering; available at <http://www.fluxus-engineering.com>) for 61 sequences from Mendur-Sokkon village. The median-joining algorithm used generates the most parsimonious tree composed of proportionately sized circles that represent observed haplotype frequencies (Bandelt et al. 1995, 1999). It resolves parallelisms and reversals with a compatibility argument justified by character-state frequency criteria.

Mismatch Distributions. Mismatch distributions were constructed using the program IWAVE (Sherry et al. 1994). These distributions were used to identify characteristics in the genetic structure of natural populations that can be attributed to the action of demographic processes (Kingman 1982). Two common signatures of population growth are starlike genealogies (Slatkin and Hudson 1991) and unimodal distributions of pairwise differences (mismatch distributions) (Harpending et al. 1993). Mismatch distributions for the three Altai population samples, as well as the distributions for the comparative population data sets, were constructed.

Coalescent dates were computed using the Tau (τ) parameter from Rogers and Harpending's mismatch model using the HVS-I mutation rate (μ) of 16.5%

per nucleotide site per million years (Ward et al. 1991). Because τ was estimated from the mismatch distribution for the total sample, the corresponding coalescent date is not population specific (i.e., relating to ethnogenesis) but is more a reflection of the demographic processes that molded the Asian mtDNA landscape due to the disproportionate influence of intermediate- and high-frequency variants of the ancestral branches between the major haplogroups on the modal value of a mismatch distribution.

Neutrality Test Statistics. The statistics Tajima's D (Tajima 1989a, 1989b) and Fu's F_s (1997) were used to test for departures from the null hypothesis of constant population size. Tajima's D is defined as

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_s}{\left[\hat{V}(\hat{\theta}_\pi - \hat{\theta}_s) \right]^{1/2}}, \quad (2)$$

where $\theta_\pi = \pi$,

$$\theta_s = S / \sum_{i=1}^{n-1} \frac{1}{i}, \quad (3)$$

and S is the number of segregating sites in the sample.

The F_s statistic is based on the probability of observing a random neutral sample with a number of alleles (k) similar to or smaller than the observed value, given the observed number of pairwise differences, taken as an estimator of θ . If there is an excess of rare alleles, a product of a starlike genealogy, then F_s will have large negative values. Both measures were calculated using Arlequin, version 2.0. The statistical significance was tested by generating random samples under selective neutrality and population equilibrium using a coalescent simulation algorithm, with p values representing the proportion of random statistics less than or equal to the observation (Hudson 1990).

Multidimensional Scaling. Multidimensional scaling (MDS) was performed using the software package NTSYS (Applied Biostatistics, Port Jefferson, New York) on Nei's intermatch-mismatch distance (D_A) matrices (Nei and Li 1979) to provide a two-dimensional representation of the genetic relationships between the various populations considered in this study.

Results

mtDNA Haplogroups. European populations are mainly composed of 10 mtDNA haplogroups: HV, H, V, J, T, U, K, I, W, and X. The common Asian

Table 2. Mitochondrial DNA Haplogroup Frequencies in Mendur-Sokkon ($n = 98$)

<i>mtDNA Haplogroup</i>	<i>Haplogroup Frequency (%)</i>
A	5.1
B	2.0
C	35.7
D	13.3
F	3.1
G	5.1
M (not C, D, E, or G)	10.2
H	7.1
J	2.0
K	4.1
U	6.1
X	1.0
Other	5.1

haplogroups are A, B, C, D, E, F, G, Y, Z, and M (Derenko et al. 2001a). As the Altaians' diverse ethnic history would suggest, the Altaian mtDNA gene pool displays both European and Asian mtDNA components. The mtDNA haplogroups present in Mendur-Sokkon are summarized in Table 2 and were identified through RFLP analyses.

Twelve identified haplogroups are present in the Mendur-Sokkon population, with 5.1% of the sample set unclassified. The most common haplogroup is C (35.7%), followed by D (13.3%) and "other" M haplogroups (10.2%), that is, samples possessing the +10394 *DdeI* and +10397 *AluI* sites that define macrohaplogroup M but not belonging to haplogroups C, D, E, or G. One individual in the sample of the community exhibited haplogroup X.

mtDNA Sequences. The mtDNA HVS-I sequences found in the Mendur-Sokkon sample are listed in Table 3, based on the nucleotide range 16151–16383, and are referenced against the Cambridge Reference Sequence (Anderson et al. 1981).

Haplogroups are listed in the far left-hand column, with the frequency (number of sequences shared by individual samples) of each haplotype indicated in the adjacent column. Thirty different haplotypes were identified, representing 32 transitions and 2 transversions. Figure 3 displays the median-joining network of the Mendur-Sokkon mtDNA sequence data.

The network displays proportionately sized circles, which represent observed haplotype frequencies. The reticulation observed within haplogroup C indicates an unresolved evolutionary pathway, with the parallel lines representing the same mutation. The network also displays starlike components within haplogroups C and D, which are features typically associated with an expanding population. Outside haplogroups C and D, a large degree of fragmentation with long internal branches is observed, which may be suggestive of gene flow.

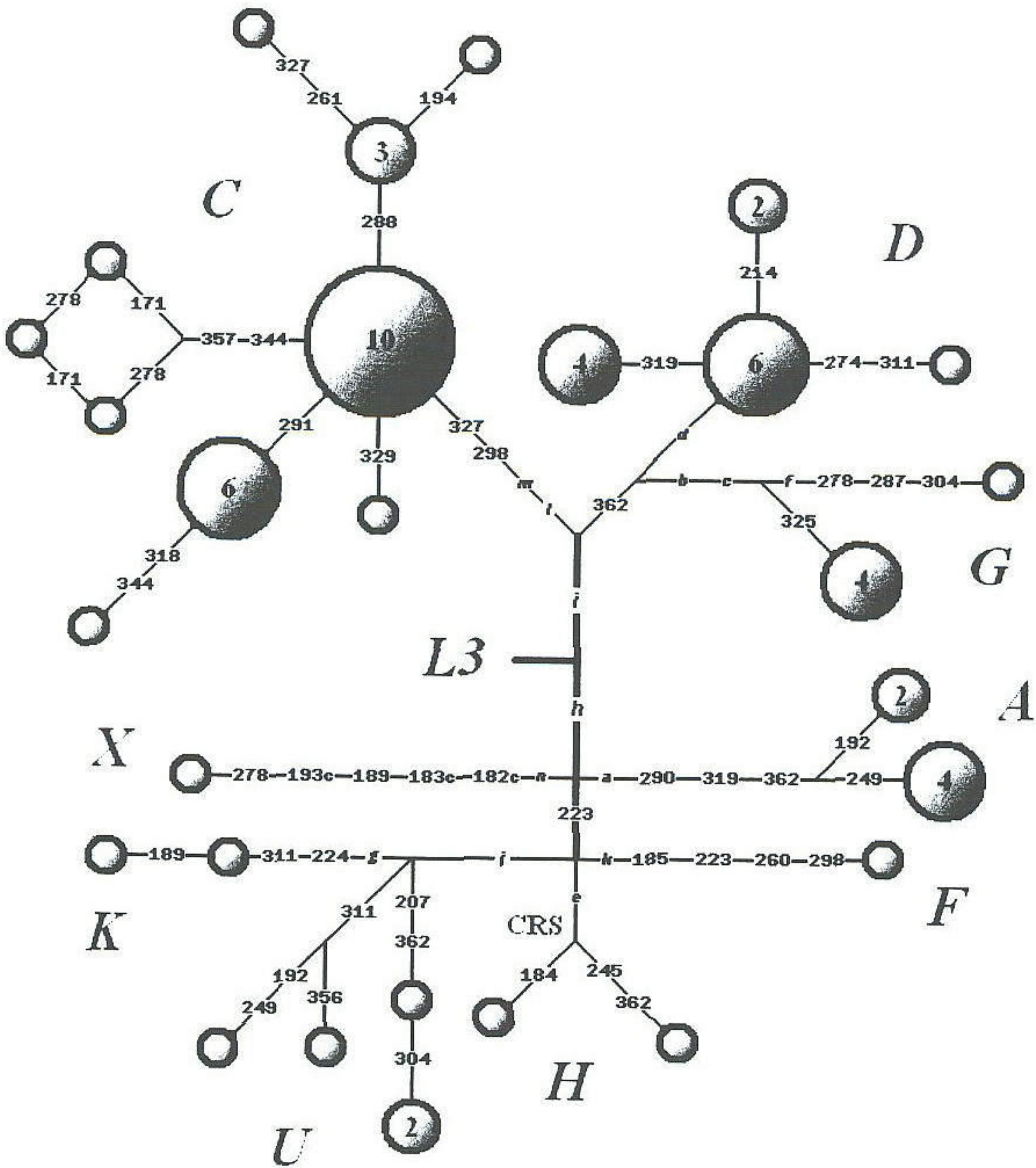


Figure 3. Mendur-Sokkon median-joining network. a = +*Hae*III 663; b = +*Hae*II 4830; c = +*Hha*I 4831; d = -*Alu*I 5176; e = -*Hha*I 7598; f = +*Dde*I 10394; g = +*Alu*I 10397; h = +*Hin*fI 12308; i = -*Hpa*I/*Hinc*II 12406; j = -*Hinc*I 13259; k = +*Alu*I 13262; l = +*Acc*I 14465.

Table 4 contains a comparison of the numbers of shared haplotypes (each haplotype is counted only once) among the three distinct samples from the Altai and various Eurasian populations.

Altai 1 and Altai 3 share the greatest number of common haplotypes (9), whereas Altai 2 shares no haplotypes with Altai 3 and has only one haplotype in common with Altai 1. The Altai 2 sample is based on only 17 individuals,

Table 4. Shared Haplotypes Among Altai and Eurasian Populations (np 16151–16383)

	<i>Altai 1</i>	<i>Altai 2</i>	<i>Altai 3</i>
Altai 1	–	1	9
Altai 2	1	–	0
Altai 3	9	0	–
Buryat	5	3	5
Tuva	7	1	7
Mongol	4	2	11
Kazakh	5	2	8
Kirghiz	4	1	5
Uighur	2	1	5
Han	3	0	15
Russian	2	1	5

whereas the Altai 1 and Altai 3 samples are based on 61 and 110 individuals, respectively; thus individuals from Altai 1 and 3 are more likely to share common haplotypes. Mongol and Han Chinese sequences reflect their historical connections with the Altai population by sharing 17 and 18 haplotypes with the three combined Altai samples.

Measures of HVS-I sequence diversity are summarized in Table 5 for the three Altai samples, along with the south Siberian, central Asian, Chinese, and Russian populations.

The two θ indexes are derived from different aspects of the mtDNA sequence data, each providing an alternative picture of past evolutionary processes. All the populations in Table 5 display a negative Tajima's D , indicating population expansion. However, two of the three Altai samples fail to exhibit statistically significant deviation from 0 ($p < 0.05$). Fu's F_s values for all three Altai samples deviate significantly from the model and provide additional evidence of population expansion. Gene diversity, a measure equivalent to expected heterozygosity, for the three Altai samples ranges from 0.946 to 0.993. The gene diversities are graphically represented in Figure 4.

It appears that the Altai 1 sample more closely resembles Tuva and Khakasian populations than the other two Altai samples. Altai 2 and Altai 3 more closely resemble Mongol, Kazakh, Kirghiz, Uighur, and Han Chinese populations.

Mismatch distributions are used to identify characteristics of genetic structure in natural populations that can be attributed to the action of demographic processes (Kingman 1982). Mismatch analyses of the three Altai samples show unimodal peaks, which are predicted for expanding populations (Figure 5).

The distribution of the Altai 1 sample has a mode of 4 and has an additional minor peak at $i = 1$; this differs from the other two Altai samples, which exhibit modes of 5 and have no secondary minor peaks. Although the Altai 1 distribution

Table 5. HVS-I Sequence Diversity for Altai and Eurasian Populations (16151–16383)

Population	<i>n</i>	θ_π (SD) ^a	θ (SD) ^a	Gene Diversity (SD) ^b	Tajima's <i>D</i> ^c	<i>F_s</i> ^d	Reference
Altai 1	61	4.868 (2.668)	8.517 (2.584)	0.971 (0.011)	-1.417	-25.590 ^f	Present study
Altai 2	17	4.978 (2.850)	7.099 (2.825)	0.993 (0.023)	-1.199	-11.539 ^f	Shields et al. (1993)
Altai 3	110	5.440 (2.923)	11.758 (3.128)	0.976 (0.005)	-1.714 ^e	-25.305 ^f	Derenko et al. (2003)
Buryat	91	4.305 (2.383)	10.821 (2.995)	0.971 (0.009)	-1.945 ^f	-5.763 ^f	Derenko et al. (2003)
Tuva	90	5.013 (2.724)	9.859 (2.763)	0.941 (0.015)	-1.580 ^e	-23.406 ^f	Derenko et al. (2003)
Mongol	103	5.368 (2.891)	11.715 (3.151)	0.991 (0.003)	-1.738 ^e	-25.352 ^f	Kolman et al. (1996)
Kazakh	55	5.570 (3.014)	10.709 (3.231)	0.987 (0.008)	-1.634 ^e	-25.384 ^f	Comas et al. (1998)
Kirghiz	94	5.145 (2.786)	12.707 (3.440)	0.985 (0.005)	-1.934 ^f	-25.449 ^f	Comas et al. (1998)
Uighur	55	4.866 (2.673)	11.584 (3.467)	0.989 (0.006)	-1.983 ^f	-25.592 ^f	Comas et al. (1998)
Han	250	6.474 (3.401)	17.551 (3.972)	0.996 (0.001)	-1.926 ^f	-24.654 ^f	Yao et al. (2002)
Russian	103	3.339 (1.914)	8.642 (2.411)	0.930 (0.020)	-1.929 ^f	-26.237 ^f	Orekhov et al. (1999)

- a. Two different estimators of $\theta = 2N_e\mu$: θ_π is equivalent to the mean number of pairwise differences between sequences (π), and θ_S is based on the number of variant sites.
- b. Gene diversity is equivalent to the expected heterozygosity: $(n/n - 1)(1 - \sum p_i)$.
- c. Tajima's *D* compares the two θ estimators, with negative scores indicating larger θ_S values relative to θ_π and suggesting the possible effects of a population expansion.
- d. *F_s* is based on the probability of observing a random neutral sample with a number of alleles (*k*) similar to or smaller than the observed value given θ_π . Negative scores indicate an excess of alleles, a signature of population expansion.
- e. $P < 0.05$.
- f. $P < 0.01$.

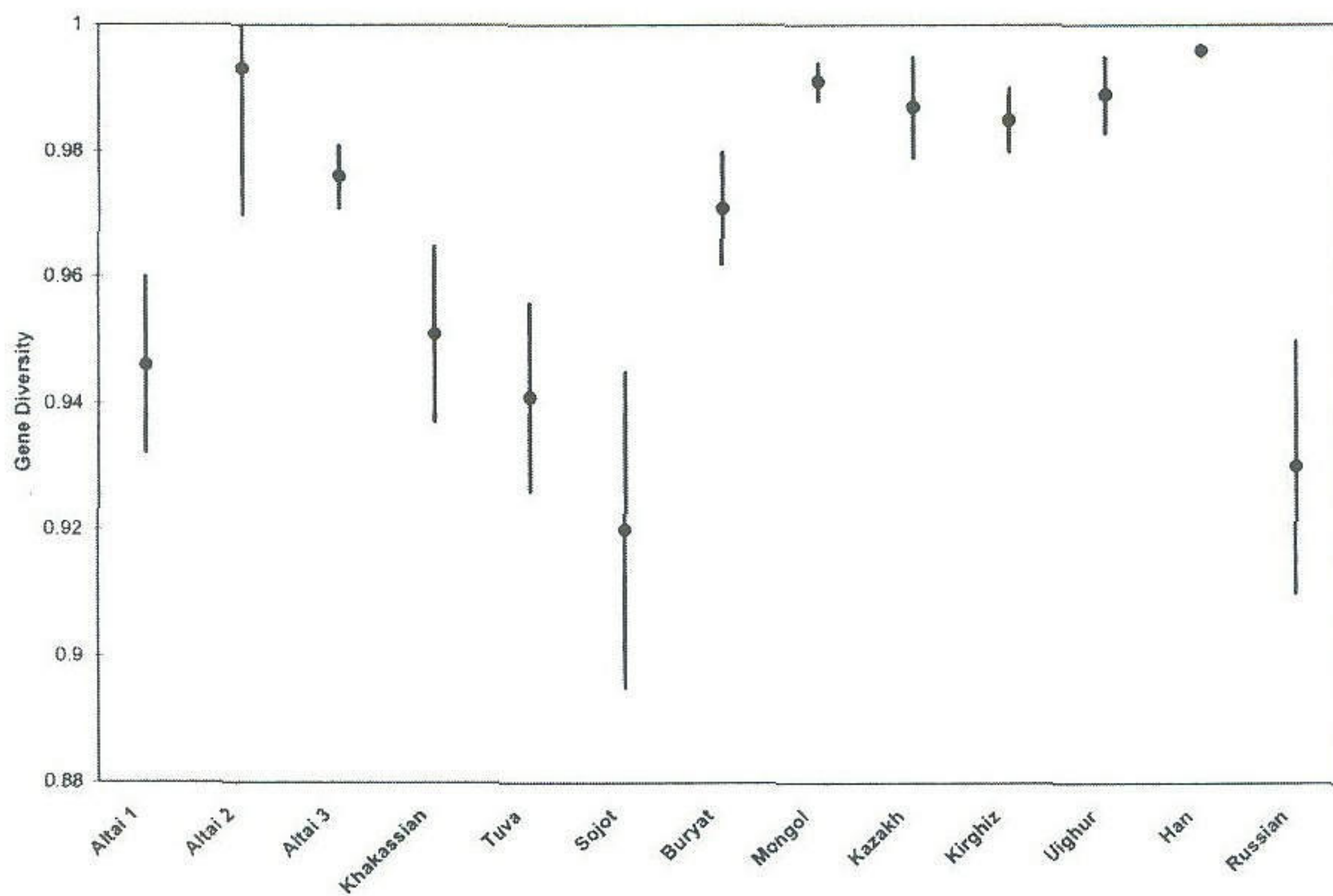


Figure 4. Gene diversity ranges.

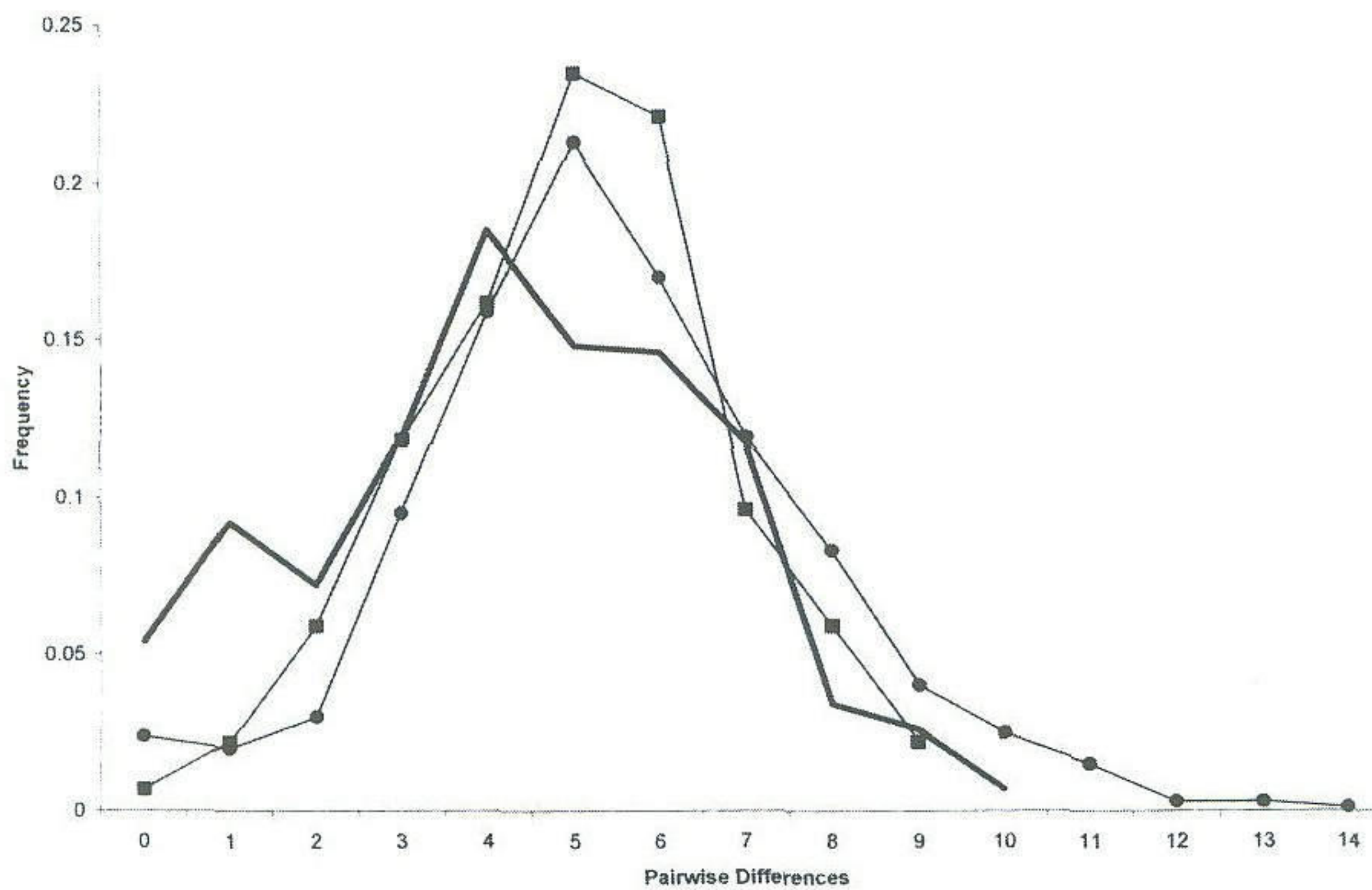


Figure 5. Mismatch distribution of Altai population samples. Bold solid line, Altai 1; squares, Altai 2; circles, Altai 3.

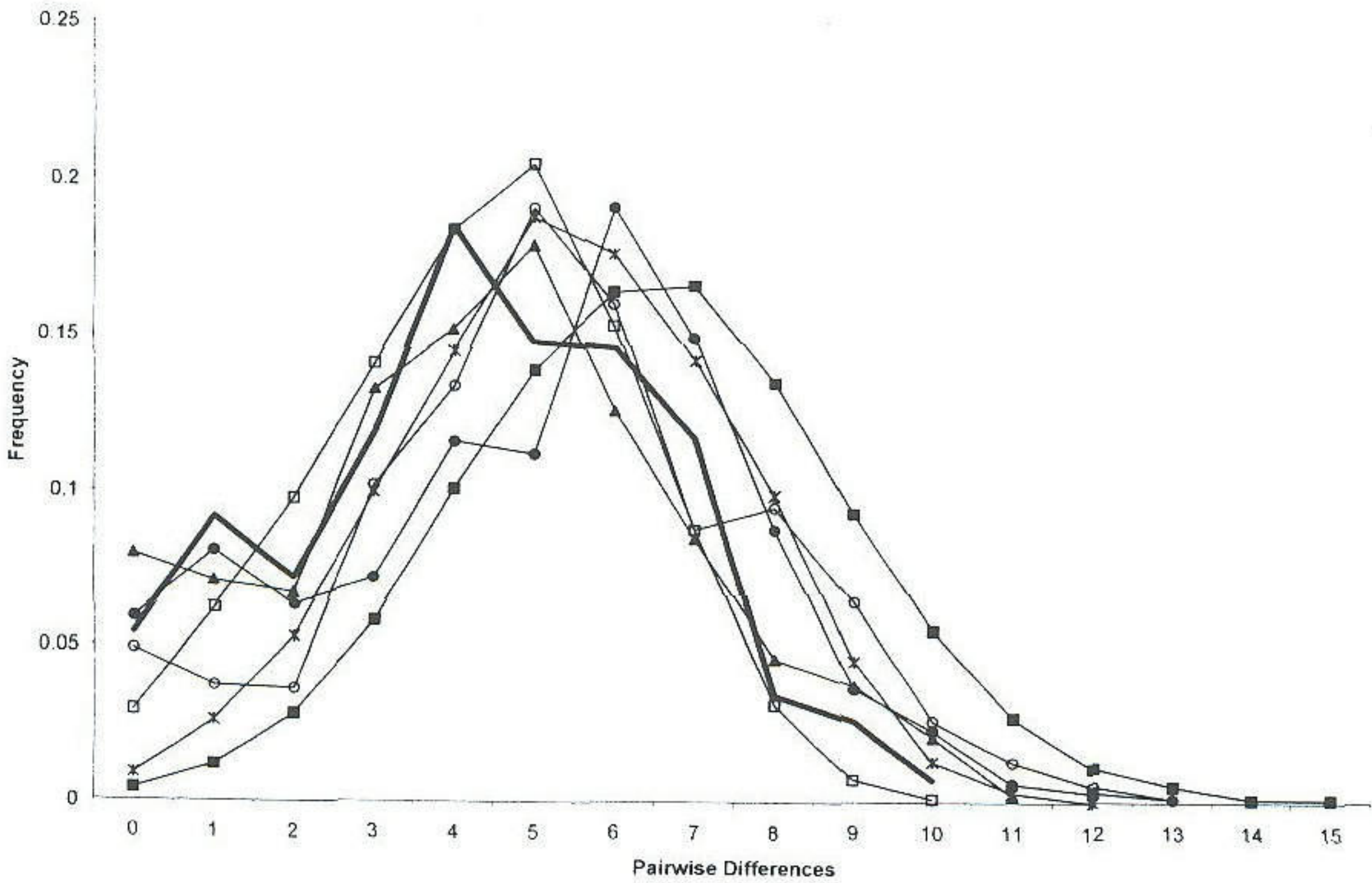


Figure 6. Mismatch distribution of Mendur-Sokkon (Altai 1) vs. other Eurasian populations. Bold solid line, Altai 1; filled squares, Han; open squares, Buryat; filled circles, Tuva; open circles, Khakassian; X's, Mongols; triangles, Sojot.

appears more ragged than those of the other Altai samples, Harpending's raggedness index (Harpending 1994) is not significant ($p > 0.05$) for this distribution. When the Altai 1 distribution is compared to the distributions of other Eurasian populations (Figure 6), it appears to be qualitatively more similar to the distributions of the Tuva, which also display a minor peak at $i = 1$, and the Sojots, with a small peak at $i = 0$. Similar to Altai 1, neither the Tuva nor the Sojots were significant for the raggedness index.

Coalescent dates in Table 6 were calculated from the τ parameter estimates based on mismatch analyses for the three Altai samples and the other comparative populations.

Overall, the ages range from 67,300 to 95,500 B.P., with the exception of the Russians, which have a later coalescent date (47,300 B.P.), consistent with the findings for other European populations (Richards et al. 2000). Interestingly, the dates for the three Altai samples are remarkably close, ranging from 74,100 to 79,700 B.P., and thus appear to be less affected by sampling than other population genetic statistics.

Multidimensional Scaling. In the MDS plot of genetic distances (Figure 7), Altai 2 and Altai 3 cluster separately from Altai 1 along the first dimension and show closer genetic ties to the Central Asian cluster that includes the Mongols.

Table 6. Coalescent Dates for Altai and Eurasian Populations

<i>Population</i>	τ (95% CI) ^a	<i>Age – 15% Rate (95% CI)</i>
Altai 1	5.182 (2.105–6.938)	74,100 (30,100–99,300)
Altai 2	5.283 (3.025–6.874)	75,600 (43,300–98,300)
Altai 3	5.574 (4.127–6.104)	79,700 (59,000–87,300)
Khakassian	5.778 (2.691–7.484)	82,700 (38,500–107,100)
Tuva	6.268 (2.814–8.826)	89,700 (40,300–126,300)
Sojot	5.267 (2.187–7.969)	75,400 (31,300–114,000)
Buryat	4.718 (2.723–5.823)	67,500 (39,000–83,300)
Mongol	5.662 (3.979–6.377)	81,000 (56,900–91,200)
Kazakh	5.880 (3.213–7.954)	84,100 (46,000–113,800)
Kirghiz	5.432 (3.545–6.319)	77,700 (50,700–90,400)
Uighur	4.707 (3.058–6.131)	67,300 (43,700–87,700)
Han	6.677 (5.347–7.150)	95,500 (76,500–102,300)
Russian	3.309 (1.587–5.598)	47,300 (22,700–80,100)

a. Estimation of coalescence time in units of mutation ($1/2\mu$) was calculated using the Rogers and Harpending (1992) mismatch model.

The first dimension also separates the Russian sample as an outlier with respect to the other populations. Along the second dimension, the three Altai samples display a pattern of relatedness by clustering together. Overall, the Altai 1 sample appears to cluster most closely with the Tuva. With the stress value of 0.0715, according to Kruskal (1964), this MDS plot fits well with the original D_A distance matrix.

Discussion

Our primary goals in this paper were to examine the maternal ancestry of the Southern Altai population of Mendur-Sokkon using both mtDNA RFLP and HVS-I data and to compare the results of the current study with those of other studies of the Altai and surrounding populations. The results show that Mendur-Sokkon does indeed resemble the other Altai samples, but it also is similar genetically to the neighboring Tuva population.

Sampling methodology and the criteria used to define a population are critical to the successful and reproducible characterization of distinct groups. The sequence data of Derenko et al. (2003) are a compilation of both Southern Altaian and Northern Altaian samples, even though they have been identified as two distinct groups based on language, culture, geographic distribution, and genetics. The combining of data from both Southern and Northern Altaian populations may have obscured some population structure and may have contributed to

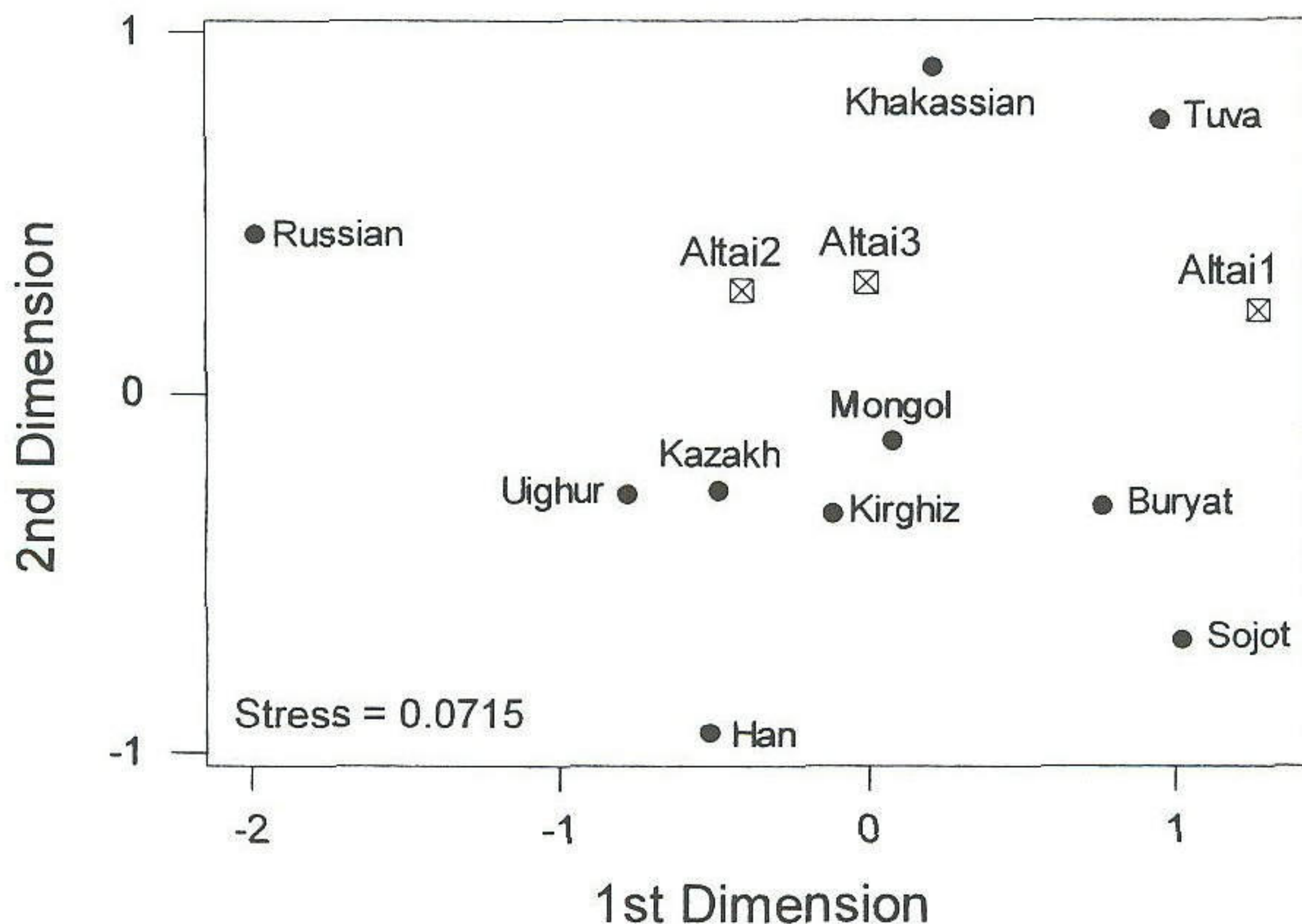


Figure 7. Multidimensional scaling plot of D_A distances.

the differences seen between Derenko's data and the Mendur-Sokkon data. HVS-I θ values show depressed levels for Mendur-Sokkon and the Shields et al. (1993) Altai data set relative to the Derenko et al. (2003) Altai data set, which exhibits values closer to the Mongols and Central Asian Turkic groups. Neutrality test statistics show a similar trend to the θ values, as the Derenko et al. (2003) sample has larger negative scores for Tajima's D and Fu's F_s , which are both significant at the $P = 0.05$ level. This finding, along with the θ values, is probably the consequence of combining the Northern and Southern Altaian genetic elements in a single sample, thus inflating its diversity as well as its neutrality test scores because of the inclusion of more low-frequency haplotypes and their corresponding singleton variants.

Effects of different sampling methodology are also apparent when calculating gene diversity. The lower gene diversity observed in Mendur-Sokkon can be attributed, at least in part, to the fact that the sample was obtained from a single village, whereas the other two Altai samples are composed of more than one village. With regard to the shared sequences, the small sample size of the Shields et al. (1993) study fails to identify the relationship between the Altaian population and the Han Chinese. Without this information a portion of the ancestral picture of the Altaian population is missing.

Despite the differences between the Altai samples, the coalescent dates are similar. Even though these coalescent dates are not specific to the Altai region but

are more a reflection of the initial colonization and expansion into the Eurasian landmass, their similarity is reassuring because it is unlikely that inadequate sampling will disrupt the deep interior branches of the Altai's actual mtDNA structure.

The importance of the Altaian region is evident when addressing the peopling of the New World. The five founding haplogroups, A, B, C, D, and X, found in Native American populations are also present in the Altai (Starikovskaya et al. 2005), suggesting that the Altaian population is a likely ancestral population to Native Americans. It should be noted that the Native American haplogroup X is a specific subhaplotype (X2a) not presently found outside the New World (Reidla et al. 2003). The Altaian haplogroup X has been postulated to hold an intermediate position between European and Native American haplogroup X lineages (Derenko et al. 2001b). Alternative theories suggest that the haplogroup X mtDNAs found in the Altai are not related to the Native American lineages and are explained by recent gene flow from Europe or West Asia into the Altai region (Reidla et al. 2003). Because Native American haplogroup X sequences have been designated as a subgroup that is phylogenetically distinctive from Eurasian haplogroup X, the ancestral origin of this Native American lineage remains unresolved.

In summary, Mendur-Sokkon exhibits depressed θ estimates and gene diversity scores relative to other Altai samples, which demonstrates Mendur-Sokkon's similarity to some neighboring south Siberian populations and may suggest the effects of a population bottleneck, perhaps a genetic signature of Russian contact and the corresponding decline in native Siberian population numbers (Forsyth 1992). The Derenko et al. (2003) Altai mtDNA data, on the other hand, are more consistent with the statistics of groups with known histories of expansion, such as the Mongols, and likely represent an artifact of the sampling strategy, which involved greater geographic coverage of the Altai region. Thus independent sampling of a single population can result in a different picture of a population's evolutionary history.

Acknowledgments We thank the members of the research team from the Siberian Branch of the Academy of Sciences who were instrumental in the collection of the blood samples from the study participants. In particular, we thank L. Osipova, T. Karafet, and O. Posukh; without their help this study would not have been possible. We are most thankful to the villagers of Mendur-Sokkon for their assistance and hospitality. We also extend our gratitude to Theodore Schurr for his expertise and use of his laboratory during the commencement of these mtDNA analyses. This project was supported by a General Research Grant from the University of Kansas and by the Siberian Branch of the Russian Academy of Sciences.

Received 13 December 2005; revision received 16 June 2006.

Literature Cited

- Anderson, S., A. T. Bankier, B. G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290(5806):457–465.
- Bandelt, H.-J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.
- Bandelt, H.-J., P. Forster, B. C. Sykes et al. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.
- Birren, B., E. D. Green, S. Klapholz et al. 1997. *Genome Analysis: A Laboratory Manual*, v. 1. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Comas, D., F. Calafell, E. Mateu 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.
- Crawford, M. H., J. McComb, M. S. Schanfield et al. 2002. Genetic structure of pastoral populations of Siberia: The Evenki of central Siberia and the Kizhi of Gorno Altai. In *Human Biology of Pastoral Populations*, W. R. Leonard and M. H. Crawford, eds. Cambridge, England: Cambridge University Press, 11–49.
- Derenko, M. V., G. A. Denisova, B. A. Malyarchuk et al. 2001a. The structure of the gene pools of the ethnic populations of Altai-Sayan region based on mitochondrial DNA polymorphism data. *Russ. J. Hum. Genet.* 37(10):1177–1184.
- Derenko, M. V., T. Grzybowski, B. A. Malyarchuk et al. 2001b. The presence of mitochondrial haplogroup X in Altaians from south Siberia. *Am. J. Hum. Genet.* 69:237–241.
- Derenko, M. V., T. Grzybowski, B. A. Malyarchuk et al. 2003. Diversity of mitochondrial DNA lineages in south Siberia. *Ann. Hum. Genet.* 67:391–411.
- Forsyth, J. 1992. *A History of the Peoples of Siberia: Russia's North Asian Colony, 1581–1990*. Cambridge, England: Cambridge University Press.
- Fu, Y.-X. 1997. Statistical tests of neutrality of migrations 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. *Evol. Anthropol.* 8:208–227.
- Harpending, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* 66(4):591–600.
- Harpending, H. C., S. T. Sherry, A. R. Rogers et al. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–496.
- Helgason, A., S. Sigurethardottir, J. R. Gulcher et al. 2000. mtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *Am. J. Hum. Genet.* 66(3):999–1016.
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*, v. 7, D. Futuyma and J. Antonovics, eds. Oxford, England: Oxford University Press, 1–44.
- Kingman, J. F. C. 1982. The coalescent. *Stochastic Proc. Appl.* 13:235–248.
- Kolman, C. J., A. M. Sambuughin, and E. Bermingham. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321–1334.
- Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.
- Levin, M. G., and L. P. Potapov. 1964. *The Peoples of Siberia*. Chicago: University of Chicago Press.
- Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76:5269–5273.
- Orekhov, V., A. Poltoraus, L. A. Zhivotovsky et al. 1999. Mitochondrial DNA sequence diversity in Russians. *Fed. Eur. Biochem. Soc. Lett.* 445(1):197–201.
- Reidla, M., T. Kivisild, E. Metspalu et al. 2003. Origin and diffusion of mtDNA haplogroup X. *Am. J. Hum. Genet.* 73:1178–1190.

- Richards, M., V. Macaulay, E. Hickey et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1251–1276.
- Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9(3):552–569.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin, Version 2.000: A Software for Population Genetics Data Analysis*. Geneva: Genetics and Biometry Laboratory, University of Geneva.
- Sherry, S. T., A. R. Rogers, H. Harpending et al. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* 66(5):761–775.
- Shields, G. F., A. M. Schmiechen, B. L. Frazier et al. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* 53:549–562.
- Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129(2):555–562.
- Starikovskaya, E. B., R. I. Sukernik, O. A. Derbeneva et al. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.* 69:67–89.
- Tajima, F. 1989a. The effect of change in population size on DNA polymorphism. *Genetics* 123(3):597–601.
- Tajima, F. 1989b. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
- Ward, R. H., B. L. Frazier, K. Dew-Jager et al. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720–8724.
- Yao, Y.-G., Q.-P. Kong, and H.-J. Bandelt. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* 70:635–651.