Genetic Variation in the Enigmatic Altaian Kazakhs of South-Central Russia: Insights into Turkic Population History

Omer Gokcumen,¹ Matthew C. Dulik,¹ Athma A. Pai,¹ Sergey I. Zhadanov,^{1,2} Samara Rubinstein,¹ Ludmila P. Osipova,² Oleg V. Andreenkov,² Ludmila E. Tabikhanova,² Marina A. Gubina,² Damian Labuda,³ and Theodore G. Schurr^{1*}

¹Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104-6398 ²Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia 630090 ³Centre de Recherche de l'Hôpital Sainte-Justine, Université de Montréal, Montréal, Canada

KEY WORDS mtDNA; haplogroup; haplotype Central Asia; Mongolia

ABSTRACT The Altaian Kazakhs, a Turkic speaking group, now reside in the southern part of the Altai Republic in south-central Russia. According to historical accounts, they are one of several ethnic and geographical subdivisions of the Kazakh nomadic group that migrated from China and Western Mongolia into the Altai region during the 19th Century. However, their population history of the Altaian Kazakhs and the genetic relationships with other Kazakh groups and neighboring Turkicspeaking populations is not well understood. To begin elucidating their genetic history, we analyzed the mtDNAs from 237 Altaian Kazakhs through a combination of SNP analysis and HVS1 sequencing. This analy-

Kazakhs emerged as a unique ethno-linguistic group in the 13th century AD (Golden, 1992) after invading the region that now encompass modern Kazakhstan. At that time, many tribes of Turkic and Mongolian origin inhabited this region (Golden, 1992). The Kazakh ethnic identity, which was distinguishable by a distinctive Kipchak Turkic idiom, emerged in this complex sociocultural environment (Golden, 1992). However, the Kazakhs did not form a unified Khanate like other Turkic speaking groups. Instead, during the 15th and 16th centuries, they were subdivided into three major tribal confederations, including the Greater (Ulu), Middle (Orta) and Lesser (Kishi) Jüzs, or Hordes (Golden, 1992; Togan, 1994; Soucek, 2000) (see Fig. 1). This division likely occurred through already existing tribal lines, and marked a political, as well as a geographical, split.

The predecessors of the Altaian Kazakhs apparently emerged when groups from the Kerei tribe of the Middle Horde migrated into northwestern China and western Mongolia from central Kazakhstan in the 18th century (Golden, 1992). These groups subsequently established a seasonal nomadic life-style in the Altai-Sayan region, Xinjiang, and western Mongolia (Gladney, 1996). It is possible that, under the pressure of Russian migrations into traditionally Kazakh lands in the 19th century, other Kazakh groups may have been forced to follow the initial example of the Kerei tribe. Later in the 19th century, some segments of Chinese and Mongolian Kazakhs moved into the region encompassing the northern slopes of the Altai-Sayan Mountains (Gladney, 2003).

Until the late 19th century, Altaian Kazakhs maintained a traditional, seminomadic, way of life, and sis revealed that their mtDNA gene pool was comprised of roughly equal proportions of East (A-G, M7, M13, Y and Z) and West (H, HV, pre-HV, R, IK, JT, X, U) Eurasian haplogroups, with the haplotypic diversity within haplogroups C, D, H, and U being particularly high. This pattern of diversity likely reflects the complex interactions of the Kazakhs with other Turkic groups, Mongolians, and indigenous Altaians. Overall, these data have important implications for Kazakh population history, the genetic prehistory of the Altai-Sayan region, and the phylogeography of major mitochondrial lineages in Eurasia. Am J Phys Anthropol 136:278–293, 2008. © 2008 Wiley-Liss, Inc.

remained relatively isolated from neighboring indigenous groups within south Siberia (Konovalov, 1986; Gladney, 1996; Oktiabrskaya, 1997). After a frontier treaty between China and Russia was established in 1864, many tribes became sedentary and abandoned their traditional nomadic way of life (Samaev, 1991). This political division also divided the Altaian Kazakhs into separate clans, which now reside in remote mountainous parts of northern China, Mongolia, and the Altai Republic (Konovalov, 1986; Derevianko et al., 1994). Nevertheless, most Kazakh families in the region continue to live in yurts and move their herds seasonally, and appear to maintain some contact with other Kazakhs groups in south Siberia (Derevianko et al., 1994; Gladney, 1996; Oktiabrskaya, 1997).

*Correspondence to: Theodore G. Schurr, Ph.D., Department of Anthropology, University of Pennsylvania, 325 University Museum, 3260 South Street, Philadelphia, PA, 19104-6398, USA. E-mail: tgschurr@sas.upenn.edu

Received 8 November 2006; accepted 19 December 2007

DOI 10.1002/ajpa.20802

Published online 5 March 2008 in Wiley InterScience (www.interscience.wiley.com).

Grant sponsors: University of Pennsylvania; the National Science Foundation; Center for Undergraduate Research, Univ. of Penn. and the Social Sciences and Humanities Research Council of Canada (T.G.S.), the Russian Basic Fund for Research (L.P.O.).





Fig. 1. Map of Kazakhstan showing the locations of the three hordes, or tribes, of Kazakhs. Modified after the "Kazakh Hordes" map in Bregel (2000).

In light of its complex population history, and because the Altai-Sayan region is considered to be the ancestral homeland of Turkic speaking populations (Golden, 1992), we conducted an analysis of mitochondrial DNA (mtDNA) variation in Altaian Kazakhs. Through this analysis, we investigated several questions about their history. The first is their genetic relationships with Kazakh groups living outside of the Altai region, and what they indicate about the origin of this population. The second is the influence of the Mongolian expansion on the genetic diversity of populations from the Altai-Sayan region. A third concerns the genetic affinities of Altaian Kazakhs with other Turkic-speaking groups, such as Uighurs, Kirghiz and indigenous Altaians, and what they indicate about Turkic expansions. At a broader level, we were interested in providing a fuller description of genetic variation in Central Asian populations, as previous studies describing mtDNA diversity generally used limited data sets, leaving an incomplete picture of human phylogeography in the region.

Our results indicate that the Altaian Kazakhs have an extraordinarily diverse mtDNA gene pool compared to other Central Asian populations, with most of the known East and West Eurasian lineages being present in them. This level of diversity likely reflects the general antiquity of the haplotypes present in this region, and the complex biological and cultural origin of the Kazakhs, as well as gene flow from different populations over the past millennium.

MATERIALS AND METHODS **Populations and samples**

In 1999–2002, ethnographic fieldwork and sample collection were carried out in several Altaian Kazakh settlements within the southern part of the Altai Republic (see Fig. 2). A total of 227 Kazakh individuals living in Kosh-Agach, Cherny Anuy, Turata, and Zhan Aul villages donated samples at that time. Because of their geographic proximity, we grouped individuals from Cherny

Anuy and Turata into a single unit (CHAN) for comparison with those from Kosh-Agach (KOSH) and Zhan Aul (ZHAN). In addition, 10 samples collected from Kazakh individuals during fieldwork in the northern Altai Republic in 2003 (NKAZ) were included in the analysis, increasing the sample set to a total of 237 individuals.

Blood samples were drawn from all participants with informed consent under the human subjects guidelines of the Institute of Cytology and Genetics in Novosibirsk, Russia. Additional human approval was obtained at the University of Pennsylvania to conduct the genetic analyses of these samples. Genealogical data were also obtained from each person at the time of sample collection to ensure that the individuals were unrelated through at least three generations, and assess the level of admixture in these communities.

For comparative purposes, we examined haplogroup and HVS1 sequence information from Kazakhs, Kirghiz and Uighurs (Comas et al., 1998; Yao et al., 2004), Mon-golians (Kolman et al., 1996; Yao et al., 2004), and Turkic-speaking indigenous Altaians (Zhadanov et al., 2007) (see Fig. 2). In addition, we analyzed mtDNA variation in 89 Mongolian samples obtained from individuals living in western Mongolia. We included these data because previous studies of Mongolians either did not include much coding region SNP data (Kolman et al., 1996) or included samples obtained in Xinjiang (Yao et al., 2004). The latter point is important because of the known diversity of ethnic groups living in Mongolia, including the Khalkhs (e.g., Nyambuu, 1992; Katoh et al., 2005).

Molecular methods

Total DNA was isolated from whole blood using a standard phenol-chloroform extraction protocol. The mtDNAs from the 237 Altaian Kazakhs and 89 Mongolians were then surveyed for sequence variation through HVS1 sequencing, with variable positions between bp 16,000-16,500 being determined relative to the revised Cambridge Reference Sequence (rCRS) (Anderson et al.,



Fig. 2. Map of Central-East Asia showing the locations of the Altaian Kazakhs and other regional populations analyzed in this study. The population abbreviations are as follows: Altaian Kazakhs (AKAZ) and Mongolians (SMNG) from this study; Mongolians from Kolman et al. (1996) (KMNG); Southern and Northern Altaians (SALT and NALT) from Zhadanov et al. (2007); Kazakhs (CKAZ), highland and lowland Kirghiz (CKIR) and Uighurs (CUIG) from Comas et al. (1998); and Kazakhs (YKAZ), Mongolians (YMNG) and Uighurs (YUIG) from Yao et al. (2004). The map is modified after Figure 1 in Yao et al. (2004).

1981; Andrews et al., 1999). Amplified sequencing templates were purified through SAP/ExoI digestion, and then cycle sequenced using BigDyeTM Terminator Pre-Mix, v. 3.1 (Applied Biosystems). After purification with Centri-Sep columns (Princeton Separations), the DNA sequences were read on ABI 3100 Gene Analyzers at the University of Pennsylvania Core Sequencing facility. All sequence data were aligned and analyzed using Sequencher 4.0TM software (Gene Codes). In addition, all Altaian Kazakh samples were screened for a set of SNPs located in the coding region of the mtDNA genome that define the major West and East Eurasian haplogroups (Richards et al., 1998; Schurr et al., 1999; Herrnstadt et al., 2002; Quintana-Murci et al. 2004; Tanaka et al., 2004) through a PCR-RFLP approach (Table 1).

Haplogroup assignments

The HVS1 sequence data were initially used to define the sub-branches of these haplogroups, using reference information from previously published Kazakh, Kirghiz, Mongolian, and Uighur data (Kolman et al., 1996; Comas et al., 1998; Yao et al., 2004). Although there is still no consensus for naming some minor subhaplogroups, we followed the nomenclature of Tanaka et al. (2004) for East Eurasian haplogroups, and that of Derenko et al. (2003) for West Eurasian lineages that were seen in Central Asian and Siberian populations but not described in Tanaka et al. (2004). For the Eurasian haplotypes that were not identified through this approach, we used published information in Achilli et al. (2004) and Palanichamy et al. (2004) to make haplogroup assignments.

Statistical analysis

Haplotype diversity (h), nucleotide diversity (π) and the analysis of molecular variance (AMOVA) (Excoffier et al., 1992) were estimated using Arlequin 3.00

(Excoffier et al., 2005). In addition, we estimated neutrality indexes, such as Tajima's D (Tajima, 1989a) and Fu's $F_{\rm s}$ (Fu, 1997), with the same software. While initially designed to detect signs of positive selection, under the assumption of neutral evolution, these tests can be used to estimate the neutrality of genetic variation (i.e., to detect population expansions (or reductions) (Tajima, 1989b).

To estimate the genetic distances between Altaian Kazakhs and other Central Asian populations, we generated pairwise $F_{\rm ST}$ values from HVS1 sequences using the population pairwise distance method of Reynolds et al. (1983) and the Nei and Li (1979) estimate of the average number of differences between populations (ARLEQUIN; Excoffier et al., 2005). For these calculations, we assumed Tajima and Nei's (1984) model for nucleotide substitution. To visually represent the genetic relationships between populations, we used multidimensional scaling (MDS) to plot the pairwise $F_{\rm ST}$ values in two-dimensional space (SPSS 11.0 software). Likewise, we utilized principal component analysis (PCA) to visualize the differences in haplogroup frequencies between populations (SPSS 11.0 software).

We also used an analysis of molecular variance (AMOVA) with our data to determine the nongenetic influences on genetic variation in these populations. For these calculations, the Altaian Kazakhs and other populations were grouped based on their linguistic, ethnic, geographical or religious affiliations. Three geographic groupings were considered: (1) West of the Altai Mountains (Uighurs, Kazakhs and Kirghiz); (2) Proximate to the Altai Mountains (Altaians and Altaian Kazakhs); and (3) East of the Altai Mountains (Mongolians). We also compared the diversity in Turkic speakers (Kazakhs, indigenous Altaians, Kirghiz, Uighurs) relative to Mongolic speakers (Mongolians) (language), as well as in different ethnic groups (Kazakhs, Mongolians, Uighurs, indigenous Altaians). To better understand the influence of religious affiliation on the genetic diversity

GENETIC VARIATION IN ALTAIAN KAZAKHS

Hg	SNP/Polymorphism	Primers (5'-3')	$T_{\rm m}~({ m C^o})$	Size
A	663/+HaeIII 663	534-553/725-706	59	191
В	9-bp Del*	8188-8207/8366-8345	49	116
С	13263/-HincII 13259	13001-13020/13403-13384	45	402
D	5178A/-AluI 5176	5151-5170/5481-5464	55	330
E	7598/—HhaI 7598	7367-7384/7628-7610	49	261
F	12406/-HincII 12406	12385-12405/12576-12595	55	191
G	4833/+HaeII 4830	4651-4670/4952-4934	49	301
Н	7028/-AluI 7025	6890-6909/7131-7115	47	241
HV	14766/-MseI 14766	$14407 \cdot 14424/14810 \cdot 14791$	51	403
I, X	1719/-DdeI 1715	1615-1643/1899-1879	54	284
Ι	4577/–NlaI 4577	4500-4519/4678-4659	51	178
J	13708/-BstOI 13704	13537-13556/13851-13832	51	314
K	9055/-HaeII 9052	8925-8953/9100-9081	53	175
Μ	10400/+AluI 10398	10279-10296/10569-10550	51	290
Ν	10398/+DdeI 10304	10279-10296/10569-10550	51	290
Ν	10873/-MnlI 10871	10718-10738/10930-10911	55	212
R	12705/+MboII 12705	12599-12618/12785-12766	51	186
Т	15928/-MspI 15928	$15838 \cdot 15857 / 16159 \cdot 16141$	51	321
Т	15607/+AluI 15607	15409-15428/15728-15709	51	319
UK	12308/+HinfI 12308	12104-12124/12338-12309	53	234
V	4529/-HaeII 4529	4500-4519/4678-4659	51	178
Y	7933/+MboI 7933	7871-7890/8020-8001	51	149

TABLE 1. Single nucleotide polymorphisms (SNPs) analyzed in Altaian Kazakhs

* This mutation is detected by directly sizing the PCR amplicons, not through RFLP analysis.

of these populations, we also divided them into Muslim (Uighurs, Kirghiz, and Kazakhs) and non-Muslim and Burkhanist/Buddhist (Mongolians, indigenous Altaians) groups.

Phylogenetic analysis

The combination of SNP and HVSI sequence data for each individual was used to construct a network of haplotypes using Network 4.0 (Bandelt et al., 1995, 1999). The weighting scheme suggested in Bandelt et al. (2002) was used with slight modifications. According to this schema, hypervariable sites such as 16093, 16129, 16189, 16311 and others, were given lower weights relative to other less mutable sites. Networks were created using both median joining and reduced median joining approaches to test the effect of removing terminal reticulations from the phylogenetic branches. To highlight key mutations occurring along these branches, we redrew the median joining networks by hand. The framework for haplogroup assignment in this network was based on networks compiled by Tanaka et al. (2004), Starikovskaya et al. (2005), and Palanichamy et al. (2004).

RESULTS

mtDNA diversity in Altaian Kazakhs

Haplogroup diversity. Virtually every known West and East Eurasian haplogroup was present in the Altaian Kazakhs (Table 2). East Eurasian lineages made up $\sim 66\%$ of the total, whereas West Eurasian lineages constituted 34% of the Altaian Kazakh genetic diversity. The East Eurasian haplogroups included those deriving from both macrohaplogroups M (C, D, G, M7, M8, and Z) and N (A, B, F, N9, and Y) (Macaulay et al., 1999; Schurr et al., 1999; Torroni et al., 2001; Yao et al., 2002). The West Eurasian haplogroups derived from both macrohaplogroup N (I, J, N1, T, W, and X) and R (HV, H, R1, and UK) (Richards et al., 2000; Quintana-Murci et al., 2004).

Interestingly, certain haplogroups were confined to particular settlements (Table 2). For instance, R1 was present in only CHAN and Z appeared in only KOSH, while both N1a and N9a were nearly exclusive to ZHAN. In addition, the more frequent haplogroups, such as B, C, and D, also differed in their distribution in the Altaian Kazakh settlements. Haplogroup B represented 14% of the mtDNAs in ZHAN, whereas this lineage appeared at much lower frequencies in the other two settlements. Similarly, D comprised nearly 25% of the mtDNAs in ZHAN, but appeared at a much lower frequency in CHAN (Table 3).

These data, and particularly the somewhat higher frequencies of West Eurasian H and R, may reflect the distinct demographic history of the Cherny Anuy settlement. This location was established by Christianized Kazakhs in the 19th Century, and has experienced considerable admixture with ethnic Russians since that time (Karih, 2000). The higher frequencies of G2a and R may also hint at founder effects or the relative isolation of this population compared to other indigenous groups within the Altai region. Overall, the geographic localization of mtDNA haplogroups suggests differences in the events that shaped the genetic composition of these settlements.

Looking more broadly at patterns of haplogroup variation, a comparison of the three Kazakh populations (Altaian, Kazakhstan, Xinjiang) showed them to have a similar mtDNA composition consisting of mainly East Eurasian haplogroups. In fact, haplogroups A, B, C, D, F1, G2a, H, and M were present in all of them, suggesting that these lineages represent the common maternal gene pool from which these different Kazakh populations emerged (Table 3). Even so, there were also some notable differences between them. Altaian Kazakhs had higher frequencies of HV, T, U, and Z mtDNAs than the other two Kazakh groups, who had generally higher frequencies of West Eurasian lineages, not to mention that only Altaian Kazakhs had N1a haplotypes. Furthermore, as indicated above, Altaian Kazakhs simply had more haplogroups present in their mtDNA gene pool compared to the groups from Kazakhstan and Xinjiang. These differences again implied that the population history of the Altaian Kazakhs was somewhat distinctive from those of the other two Kazakh groups.

O. GOKCUMEN ET AL.

TABLE 2. mtDNA haplotypes in Altaian Kazakhs

#	Hg	SNPs	HVS1 Sequence	CHAN	KOSH	ZHAN	NKAZ	Total
1	А	663-10873	147A-223-242-290-319	1				1
2	Α	663-10873	086-187-223-231-290-319			1		1
3	A2	663-10873	124-290-319-362	1				1
4	A2	663-10873	189 - 223 - 256 - 290 - 319 - 356 - 362		1			1
5	A2	663-10873	223-290-319-325-362		1	1		2
6	A2	663-10873	223-290-319-362		1	3		4
7	B4	10873-12705-9bpDel	189-217-261		1	2		3
8	B4a	10873-12705-9bpDel	129-189-261			2		2
9	B4a	10873-12705-9bpDel	189-194C-195-217-261-266G			2		2
10	B40 D4b	10873-12705-90pDel	080-130-154-189-217-218	1		び 1		4
11	D40 D5	10873-12705-90pDei 10873-12705-90pDei	080-104-189-217-218			1		1
12	Б0 В5	10873 10398 12705 9bpDel	111-140-109-201 190 180 934 943 978			2		1
14	B5	10873 - 10398 - 12705 - 900 Del	1/0-189-2/3	1		2		1
15	B5	$10873 \cdot 10398 \cdot 12705 \cdot 9 \text{bpDef}$	140-189-234-243-278	1		1		1
16	B5	10873-10398-12705-9bpDel	111-140-189			1		1
17	B5	10873-10398-12705-9bpDel	111-140-189-234-243		1	2		3
18	B5	10873-10398-12705-9bpDel	111-140-189-234-243-278		-	$\overline{1}$		1
19	С	10398-10400-13263	093-129-223-298-327				1	1
20	С	10398-10400-13263	093-129-223-289-298-327		1	1		2
21	С	10398-10400-13263	129-150-223-298-327		1			1
22	С	10398-10400-13263	129-223-298-327			1		1
23	С	10398-10400-13263	223-298-311-327			1		1
24	C^*	10398-10400-13263	223-298-327	3	1	2	1	7
25	C4a	10398-10400-13263	167-171-223-298-327-344-357	2	1			3
26	C4b	10398-10400-13263	223-291-298-327	2		_	2	4
27	C5	10398-10400-13263	093-223-261-288-298			2		2
28	C5	10398-10400-13263	093-223-288-298-327-390	2			1	2
29	C5	10398-10400-13263	148-223-288-298-327	1			1	1
30	D	5178A-10398-10400	051-213-223-362	1		1		1
31	D	5178A-10398-10400 5178A 10908 10400	111-107-223-302			1		1
ാ∠ ୨୨	D D	5178A 10396-10400 5178A 10208 10400	172-223-311-302		1	1		1
34	D D	51784-10398-10400	222-222-290-262		1	1		1
35	D	5178A-10398-10400	223-232-362			1		1
36	D	5178A-10398-10400	223-278-295-362			1		1
37	D	5178A-10398-10400	223-311-362			2		2
38	D*	5178A-10398-10400	223-362			2		2
39	D4	5178A-10398-10400	176-223-343-362			1		1
40	D4a	5178A-10398-10400	129-158-223-234-260-292-311-362		1	2		3
41	D4a	5178A-10398-10400	129-223-256-362			2		2
42	D4b	5178A-10398-10400	093-172-173-215-223-319-362		1			1
43	D4b	5178A-10398-10400	223-284-319-362		1	3		4
44	D4	5178A-10398-10400	223-245-362			1		1
45	D4 D4	5178A-10398-10400	223-245-362-368	1		_		1
46	D4 D4	5178A-10398-10400	185-189-224-245-292-362			1		1
47	D4 D4	5178A-10398-10400	224-245-292-362		1	Z		2
48	D4 D4	5178A-10398-10400 5178A 10208 10400	107-172-174-223-287-302		1	2		1
49 50	D4 D4	5178A 10398 10400	174-220-002 170 993 977 911 969		1	Э	1	4
51	D4 D4	5178A-10398-10400	223-294-362			1	T	1
52	D4	5178A-10398-10400	223-274-311-362			1		1
53	D4	5178A-10398-10400	218-223-362	1		1		2
54	D5	5178A-10397-10398-10400	148-189-223-362		1			1
55	D5a	5178A-10397-10398-10400	051-164-172-189-223-266-362			2		2
56	D5c	5178A-10397-10398-10400	188Ci-193Ci-223-311-316-362			2		2
57	F1a	10873-12406-12705	129-172-304			4		4
58	F1a	10873-12406-12705	129-172-304-311	3		1		4
59	F1b	10873-12406-12705	129-189-194C-195-304			1		1
60	F1b	10873-12406-12705	189-304	2	1			3
61	F1b	10873-12406-12705	172-179-189-232A-249-304-311		1	1		2
62	F1b	10873-12406-12705	172-179-189-232A-249-304-311-354			1		1
63	F1b	10873-12406-12705	189-232A-249-304-311			1	_	1
64 67	F2a	7828-10873-12705	092A-291-304			0	1	1
65 66	G	4833-10398-10400	223-294-362			2		2
00 67	G2 C2	4000-10000 10400	UJJ-22J-214-210-201-302 109 999 978 969	9		T		1
68	G2 C2	4000-10000-10400 4833 10308 10400	174-440-410-004 993 934 978 300 969	J	1			ა 1
60	G20	4000-10000-10400 /833-10308-10/00	220-204-210-007-002 993-997-978-369	5	T	2		L Q
70	G2a	4833-10398-10400	169-223-227-265C-278-362	U	1	U		1
$\overline{71}$	G2a	4833-10398-10400	093-223-227-234-278-309-362	1	Ŧ	1		2
	2.24			-		-	(con	tinued)

GENETIC VARIATION IN ALTAIAN KAZAKHS

			TABLE 2. (Continued)					
#	Hg	SNPs	HVS1 Sequence	CHAN	KOSH	ZHAN	NKAZ	Total
72	G4a	4833-10398-10400	189-223-265C-274-362		2	2		4
73	Η	7028-10873-12705-14766	184	1				1
74	H	7028-10873-12705-14766	189	1		1		2
75 70	H	7028-10873-12705-14766	209	1		1		1
76	н ц	7028-10873-12705-14766	261 351		1	1		1
78	H	7028-10873-12705-14706	129-260-261		1	1		1
79	H	7028-10873-12705-14766	129-260-261-309			2		2
80	Н	7028-10873-12705-14766	192-261-289		1			1
81	Η	7028-10873-12705-14766	278-292	1				1
82	H	7028-10873-12705-14766	311	0	2	1	1	4
83	H	7028-10873-12705-14766	CRS	2	2	2		6
04 85	п Н	7028-10873-12705-14706	554 176	1		1		1
86	H	7028-10873-12705-14766	093	1	1	1		1
87	HV	10873-12705-14766	129-249-311		1	1		1
88	Ι	1715-10398-10873	129-223	1				1
89	Ι	1715-10398-10873	129-172-223-311-391			2		2
90	I	1715-10398-10873	129-223-391		2			2
91	J	10398-10873-12705-13708	069-092-126-193-261	1	1			1
92	J T*	10398-10873-12705-13708	069-126-260A1-299 060-126	1				1
93 94	J 12a	10398-10873-12705-13708	069-126	4			1	1
95	J2a	10398-10873-12705-13708	069-126-145-231-239-260-261-311		1		1	1
96	J2b	10398-10873-12705-13708	069-126-193-300-309		-	1		1
97	Μ	10398-10400	126-185-223-325		1			1
98	Μ	10398-10400	126-223-325			1		1
99	M	10398-10400	129-152-153-179-192-223-362	1				1
100	M	10398-10400	129-152-179-192-223-362	1		2		3
101	M M13	10398-10400	221-223-200 145 168 188 993 957 311		1	1		1
102	M7h	10398-10400	129-192-223-297		1	1		1
104	M7c	10398-10400	223-278-295-362			1		1
105	M8a	10398-10400	184-223-298-319		1			1
106	M9a	10398-10400	223-234-291-316-362			1		1
107	N1a	1715-10398-10873	147A-172-181-189-223-248-320-355			1		1
108	NIa N1-	1715-10398-10873	147A-172-189223-248-320-355			1		1
109	N1a N1a	1715-10398-10873	147A-172-189-223-248-320-355		1	4		4
111	N1b	1715-10398-10873	145-176G-223-390		T	1		1
112	N9a	10873	131G-223-257A-261			1		1
113	N9a	10873	111-129-223-257A-261			1		1
114	N9a	10873	172-223-257A-261			2		2
115	R	10873-12705	304	1				1
116	K D1	10873-12705	192-304-309-390	I F				1
117	л1 T1	10873-12705	270-311 093-126-163-186-189-294-296	9	1			5 1
119	T2	10873-12705-15607-15928	126-189-292-294-296		1	1		1
120	T2	10873-12705-15607-15928	126-294-296-304-362		1	1		2
121	U	10873-12308-12705	93		2	1		3
122	U	10873-12308-12705	126-304			2		2
123	U	10873-12308-12705	129-189-234		1	1		2
124	U	10873-12308-12705	129-189-234-292-311-327		1	1		1
120	U	10873 19308 19705	174-109-192-209-311 180 994 909 911 997		1	1		1
$120 \\ 127$	U1	10873-12308-12705	189-249	2		1		2
128	Ū3	10873-12308-12705	148-311-343-390	-		1		1
129	U4	10873-12308-12705	356	1				1
130	U4	10873-12308-12705	356-362			2		2
131	U5a	10873-12308-12705	192-256-270-304-399			1		1
132	Uba Ufa	10873-12308-12705	192-296-270-399 109-941-956-970-997-904-995-900			1	1	1
133 134	Upa U5a	10873-12308-12705	192-241-200-270-287-304-320-399 256-270-399			9	T	1
135	W	10873	223-292			2 1		2 1
136	Y	1719-7933-10873	126-231-266		1	1		1
137	Y	1719-7933-10873	126-231-266-278	1				1
138	Z	10398-10400	185-223-260-298		1			1
	Total			54	45	128	10	237

CHAN, KOSH and ZHAN are southern Altai Republic villages; NKAZ, Kazakhs from the northern Altai Republic.

237

10

O. GOKCUMEN ET AL.

TABLE 3. Haplogroup frequencies in selected Turkic- and Mongolic-speaking populations of Eurasia

			1 0	1, 1				0			,		
Hg	n = 54 CHAN	n = 45 KOSH	n = 128 ZHAN	n = 10 NKAZ	n = 237AKAZ	n = 55 CKAZ	n = 53 YKAZ	n = 89 SMNG	n = 103 KMNG	n = 49 YMNG	n = 95 CKIR	n = 55 YUIG	n = 47 CUIG
С	16.7	8.9	5.5	50.0	10.5	7.3	13.2	18.0	14.6	10.3	12.6	1.8	6.4
Z		2.2			0.4	1.8	11.3	3.4	4.9		1.1		
M8a		2.2			0.4		1.9		1.0				6.4
D	5.6	11.1	21.1	10.0	15.6	16.4	13.2	29.2	19.4	18.4	20.0	16.4	6.4
D5		2.2	3.1		2.0	1.8		2.2	4.9	2.0			4.3
G			1.6		0.8		1.9	2.2	1.0	4.1			
G2	5.6	2.2	0.8		2.1				1.0	6.1		1.8	
G2a	11.1	2.2	3.1		4.6	5.5	1.9	4.5	5.8	6.1	8.4		10.6
G4		4.4	1.6		1.7								2.1
Μ	3.7	2.2	3.1		3.0	5.5	3.8	7.9	3.8		4.2	3.6	2.1
M7			1.6		0.8		1.9			2.0		5.5	6.4
M9a			0.8		0.4	1.8	1.9		1.0	2.0	2.1	1.8	
M13		2.2	0.8		0.8					2.0			
N9a			3.1		1.7		1.9	1.1	1.0	2.0	3.2	1.8	
Y1	1.9	2.2			0.8	1.8			1.9				
Α	3.7	6.7	3.9		4.2	9.1	3.8	3.4	4.9	8.2	4.2	7.3	4.3
R	13.0				2.5						3.2	12.7	
B4	1.9	2.2	7.8		5.1	3.6	3.8		6.8	2.0	3.2	1.8	
B5	1.9	2.2	6.3		4.2	1.8		2.2	2.9		3.2	5.5	
F1	9.3	4.4	7.0		6.8	3.6	5.6	4.5	4.9	10.3	2.1	7.3	6.4
F2				10.0	0.4		1.9	2.2	1.0	8.2			
N1a		2.2	4.7		3.0								
N1b			0.8		0.4								
Ι	1.9	4.4	1.6		2.1		1.9						
W			0.8		0.4	3.6		1.1					6.4
Т		4.4	1.6		1.7	7.3	7.5	2.2	1.0		1.1	1.8	2.1
J	5.6	4.4	0.8	10.0	3.0		1.9	6.7			5.3		4.3
U		8.9	4.7		3.4	1.8			2.9			9.1	8.4
U1	3.7				0.8				1.0		3.2	1.8	
U3			0.8		0.4								
U4	1.9		1.6		1.3	1.8			3.9	2.0		5.5	4.3
U5			3.1	10.0	2.1	1.8	3.8	1.1		6.2	1.1		
Κ			1.6		0.8			3.4	1.0			3.6	2.1
HV			0.8		0.4	5.5	5.6			2.0	6.3		4.3
Η	13.0	15.6	7.8	10.0	10.5	16.4	7.5	5.6	1.9	4.1	14.7	10.9	10.6
Other						1.8	3.8			2.0	1.1		2.1

CKIR = highland and lowland Kirghiz populations from Comas et al (1998). "Other" haplotypes belong to haplogroups or their branches not listed in the table.

Haplotype diversity. Additional insights into the pattern of genetic diversity in Altaian Kazakhs were obtained through HVS1 sequencing. The sequencing of 237 Altaian Kazakh mtDNAs revealed a total of 138 distinct haplotypes, with six of them appearing in individuals from the northern Altai Republic (Table 3). While the three settlements in the southern Altai Republic had many haplogroups in common, the frequencies of individual haplotypes differed considerably between them. In fact, only two haplotypes (#25 and #83) were present in all three settlements. Otherwise, CHAN and KOSH shared four haplotypes, CHAN and ZHAN shared eight haplotypes, and KOSH and ZHAN shared 15 haplotypes.

These patterns were supported by the summary statistics. Nei's D estimates showed that CHAN ($D = 0.975 \pm 0.010$) was less diverse than KOSH ($D = 0.993 \pm 0.007$) or ZHAN (0.994 ± 0.002) (Appendix A). Similarly, the $F_{\rm ST}$ distances between CHAN and ZHAN, and between CHAN and KOSH, were small but significant (0.022 and 0.012, respectively). Thus, despite their general similarity in terms of haplogroup frequencies, these settlements showed genetic differences between them at the haplotype level.

We also calculated the neutrality indexes, Tajima's D (Tajima, 1989a) and Fu's $F_{\rm s}$ (Fu, 1997), to look for evidence of population expansions. For CHAN, Tajima's D

and Fu's $F_{\rm s}$ values were significantly negative (P < 0.05), whereas KOSH and ZHAN had significantly negative Fu's $F_{\rm s}$ values but insignificant Tajima's D values (P > 0.05). Such results suggested that the Altaian Kazakhs had undergone a recent population expansion, particularly those living in CHAN. However, similar contradictions between Tajima's D and Fu's $F_{\rm s}$ values were observed in previous studies of Central Asian populations (Quintana-Murci et al., 2004), and explained as reflecting the influence of the heterogeneous mutation rate along the HVS1 on the Tajima's D estimates (Aris-Brosou and Excoffier, 1996).

In spite of the evidence for their heterogeneity, the $F_{\rm ST}$ values did not allow us to statistically distinguish the three settlements from each other. Thus, for all subsequent analyses, the data from these locations were combined and treated as coming from a single Kazakh population.

Phylogeography of Altaian Kazakh mtDNAs

The network of Altaian Kazakh mtDNAs confirmed the considerable diversity of both East and West Eurasian haplogroups in this population (see Fig. 3). Haplogroups D and H each exhibited a wide array of haplotypes, suggesting the recent expansion of these



O. GOKCUMEN ET AL.

TIDDD If Summary statistics for Some at and Dast Tistan populations	TABLE 4.	Summary statistics	for	Central d	and	East Asian	populations
--	----------	--------------------	-----	-----------	-----	------------	-------------

				v	,						
	AKAZ	CKAZ	YKAZ	CUIG	YUIG	CKIR	SMNG	KMNG	YMNG	SALT	NALT
Sample size	237	55	53	56	47	93	89	103	49	264	222
# Haplotypes	139	45	47	46	43	70	61	83	38	75	61
Gene	0.997	0.990	0.996	0.992	0.995	0.990	0.990	0.995	0.988	0.975	0.964
diversity (S.E.)	(± 0.001)	(± 0.006)	(± 0.005)	(± 0.005)	(±0.006)	(± 0.004)	(±0.003)	(±0.002)	(±0.007)	(± 0.003)	(±0.004)
Mean # pairwise differences	6.725	6.163	6.425	5.695	6.193	6.107	6.249	6.624	6.129	6.179	7.031
Tajima's D	-1.899	-1.853	-1.835	-2.013	-1.883	-1.957	-1.904	-1.829	-1.694	-1.573	-1.150
(P-value)	(0.002)	(0.008)	(0.008)	(0.007)	(0.009)	(0.002)	(0.010)	(0.014)	(0.024)	(0.027)	(0.115)
Fu's $F_{\rm s}$	-24.625	-25.234	-25.174	-25.349	-25.230	-25.166	-25.143	-25.011	-25.244	-24.711	-24.584
(P-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)

Population abbreviations are as follows: Altaian Kazakhs (AKAZ) and Mongolians SMNG) from this report; Mongolians from Kolman et al. (1996) (KMNG); Northern Altaians (NALT) and Altaian-Kizhi (AKIZ) from Zhadanov et al. (2007); Kazakhs (CKAZ), Kirghiz (CKIR) and Uighurs (CUIG) from (Comas et al. 1998); and Uighurs (YUIG) and Kazakhs (YKAZ) from Yao et al. (2004).

lineages, whereas other haplogroups, such as A, B, F, and G, had a different pattern in which deep, differentiated branches appeared, implying a long-term presence for them in the region.

Looking closer at haplotype differentiation within haplogroup C, we observed a number of distinct subbranches. These included previously identified clusters of C4a, C4b, and C5, as defined by the 16357, 16291, and 16288 mutations, respectively (Tanaka et al., 2004) (Table 2; Fig. 3). C4a is particularly widespread in northern Asia, being seen in a number of indigenous Siberian populations (Schurr et al., 1999; Derenko et al., 2003; Pakendorf et al., 2003; Schurr and Wallace, 2003; Starikovskaya et al., 2005). A few of the haplotypes also appeared to be unique to the Altaian Kazakhs.

A similar pattern was observed in haplogroup D, where the previously defined branches of D4a, D4b, D4d and D5 were present (Tanaka et al., 2004) (Table 2; Fig. 3). Overall, haplogroup D was the most diverse mtDNA lineage in the Altaian Kazakhs, and exhibited a number of unique haplotypes. In addition, the presence of the distinctive D5c haplotypes in the Altaian Kazakhs was very interesting, as our work suggested that D5c represents a genetic connection between indigenous Altaian and Japanese populations (Dulik et al., 2007), which, in the case of the Altaian Kazakhs, may reflect gene flow with these Altaian populations.

Haplogroup H was also quite diverse, with its haplotypes clustering around a central node. In fact, the nodal haplotype, defined by the rCRS, was the most frequent of these in the Altaian Kazakhs. The one exception to this pattern was the branch defined by the 16261 transition, which showed some degree of reticulation. Such a degree of diversity was consistent with the introduction of this lineage into Central Asia during the Neolithic, as observed with ancient DNA data (e.g., Lalueza-Fox et al., 2004), and the general antiquity and diversity of this haplogroup in Eurasia as a whole (Richards et al., 2000; Achilli et al., 2004).

Haplogroup B also showed considerable haplotypic diversity, despite being less frequent than some of the other haplogroups in Altaian Kazakhs. Subhaplogroups B4 and B5 encompassed all of these mtDNAs, with the most common haplotype in the Altaian Kazakhs (#10; Table 3) belonging to B4b. The frequency of this particular haplotype was intriguing because it had not previously been observed in other Kazakh populations surveyed for mtDNA variation (Comas et al., 1998; Yao

et al., 2004), although the frequency of haplogroup B was fairly similar amongst them, and was not observed in indigenous Altaian or Mongolian groups. At the same time, the overall frequencies of B4 and B5 in the Altaian Kazakhs were consistent with those observed in Mongolians, Kirghiz and Uighur populations from the surrounding regions (Table 3).

Haplogroup N1a was also present in the Altaian Kazakhs. Seeing as how there were no occurrences of this lineage in other Kazakh populations or neighboring populations (Kolman et al., 1996; Comas et al., 1998; Yao et al., 2004), this finding was intriguing (Table 3). The haplotypic variation within the seven N1a samples was relatively high (Table 2), with these haplotypes belonging to both the European and Central Asian branches of this haplogroup, as recently defined by Haak et al. (2005). Thus, the source of N1a haplotypes in Altaian Kazakhs was unclear, although they seemed to have originated west of this part of Central Asia (Gokcumen et al., 2007).

mtDNA diversity in Central Asia

The observed gene diversity in Altaian Kazakhs was very similar to that observed in other Kazakhs, indigenous Altaians, Uighurs and Kirghiz, and Mongolians (Table 4; Appendix B). For these comparisons, we used haplotype information for Mongolian populations from published sources (Comas et al., 2004; Yao et al., 2004) and our analysis of 89 Mongolian samples, in which 60 different haplotypes were identified (Appendix C). Overall, these data were consistent with the general trend showing that Central Asian populations are genetically more diverse than the surrounding populations (e.g., Wells et al., 2001; Quintana-Murci et al., 2004). The slightly higher diversity in the Altaian Kazakhs compared to western Kazakh populations could possibly be attributable to recent admixture with Mongolians or other Turkic speaking groups, who would have contributed new genotypes to this population.

To visualize the extent of genetic diversity among these populations, we plotted pairwise $F_{\rm ST}$ values estimated from HVS1 sequence data through MDS analysis (see Fig. 4). Not surprisingly, Altaian Kazakhs clustered together with other Kazakh populations, with which they showed a similar haplogroup composition. The Uighurs and Kirghiz living in eastern Kazakhstan were also genetically proximate to the Altaian Kazakhs. By contrast, the Altaian Kazakhs were quite genetically dis-



Fig. 4. A multidimensional scaling plot (MDS) of the genetic distances estimates for Central Asian populations and Altaian Kazakhs. The populations represented on the plot are Altaian Kazakhs and Mongolians from this study; Mongolians from Kolman et al. (1996); Southern and Northern Altaians from Zhadanov et al. (2007); and Kazakh, Kirghiz and Uighur groups from Comas et al. (1998). All population designations are the same as those indicated in Figure 2. The stress value for this plot is 0.0583. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tinct from indigenous Altaian populations. The latter groups were also positioned at some distance from other Central Asian populations, although the Southern Altaians, which occupy the same region as the Altaian Kazakhs, clustered closest to Mongolians. These observations suggested that there has been relatively little admixture between Altaian Kazakhs and indigenous Altaians after the former group immigrated into the Altai region. On the other hand, the relatively small genetic distance between Mongolian groups and Altaian Kazakhs suggested a more direct genetic interaction between them.

The PCA plot of haplogroup frequencies showed a slightly different picture than that produced by the MDS approach (see Fig. 5). It revealed a strong clustering of Kazakh groups, as previously observed, and the close positioning of Kirghiz to the Kazakhs. However, Uighur groups, as well as Mongolians and Altaians, were separated from them. This difference in clustering in the MDS and PCA plots was not entirely unexpected, as the latter takes into account only the haplogroup frequencies and ignores the haplotypic differences within them.

In an effort to further delineate possible population interactions and genetic affinities, we determined the extent of haplotype sharing among Central Asian populations (Table 5). The Altaian Kazakhs were found to share 19 haplotypes with Kazakhstan Kazakhs and 11 with Xinjiang Kazakhs, but only six appeared in all three of them. These six haplotypes included the founder haplotypes for C (#24) and D (#38), as well as one each from F1b, G2a, T2b, and Z. Thus, despite overall similarities at the haplogroup level, each of the Kazakh groups was somewhat differentiated at the haplotype level.

Similar patterns were seen in comparisons with other Central and East Asian populations. The Altaian Kazakhs shared an average of 11 haplotypes with each Mongolian group, 10 haplotypes with the Kirghiz, and 10 with each Uighur group. In all cases, however, only a subset (3–4) was shared by all of the populations being compared. The rCRS and the haplogroup D founder type (#38) were



Fig. 5. Principal components analysis (PCA) plot of based on mtDNA haplogroup frequencies in Central Asian populations and Altaian Kazakhs (see Table 3). All population designations are the same as those indicated in Figure 2. [Color figure can be viewed in the online issue, which is available at www.interscience. wilev.com.]

the most consistently shared haplotypes between the Altaian Kazakhs and other Central and East Asian populations.

The AMOVA results also shed light on patterns of genetic variation in this region (Table 6). To begin with, the linguistic affiliations of these populations did not produce any significant variances between these groups, with most ($\sim 98\%$) of the observed variation occurring within them. In contrast, religious affiliation provided a small but significant difference between groups (1.41%), as did ethnicity (1.39%). These two values were intriguing, and could indicate the effect of kinship structures in shaping of the overall genetic diversity in the region. Similarly, geography accounted for small but significant of within group variation (1.58%), which increased (1.69%) when the Altaian Kazakhs were added to their hypothetical region of origin, southwest of the Altai Mountains. Conversely, when Altaian Kazakhs were analyzed according to their current geographic location, the within group variation became nonsignificant. Regardless of the category being analyzed, however, almost all of the genetic variation was shared across populations, suggesting that either gene flow between the populations was considerable, or they originated from a common ancestral gene pool.

DISCUSSION

Maternal genetic diversity in Altaian Kazakhs

The population history of the Altaian Kazakhs is complex, as seen in their diverse mtDNA composition. The specific combination of haplogroups and haplotypes within the Altaian Kazakh population distinguishes it from that seen in the linguistically related Kazakhs to the west and south (Xinjiang), and their geographical neighbors, indigenous Altaians. Yet, they share a set of eight different haplogroups, and a set of six haplotypes, amongst them, implying they have some degree of common genetic ancestry.

At the same time, the Altaian Kazakhs show evidence of population subdivision. The CHAN settlement is 288

O. GOKCUMEN ET AL.

TABLE 5.	Haplogroup	sharing in	selected	Central Asian	populations	
----------	------------	------------	----------	---------------	-------------	--

HVS1-Seq	Hg	AKAZ	CKAZ	YKAZ	SMNG	KMNG	YMNG	CKIR	CUIG	YUIG
223-290-319-362	A2	4	1			1		1		1
189-194C-195-217-261-266G	B4a	2		1						
140-189-243	B5	1				2				
111-14-189-234-243-278	B5b	1	1							
223-298-327	С	7	6	2	5	1		1		
223-298-311-327	С	1	1			1	1	1		
129-223-298-327	С	1		1	1	1		1		1
093-129-223-298-327	С	1	1	1	2					1
167-171-223-298-327-344-357	C4a	3	1		1	1	1			
093-223-288-298-327-390	C5	2			1					
148-223-288-298-327	C5	1						2		
223-362	D	2	3	1	3	2	3	6	3	2
223-232-29-362	D	1			1					
223-294-362	D	1	3		1		1	1		
176-223-343-362	 D4	1			$\overline{2}$					
223-245-362	D4d	1							1	
223-245-362-368	D4d	1					2	1	_	
224-245-292-362	D4d	$\overline{2}$	1				_	_		
172-179-189-232A-249-304-311	F1b	2		1			1		2	
189-232A-249-304-311	F1b	1	2	1			$\overline{2}$			1
189-304	F1b	3				1				
093-223-227-234-278-309-362	G2a	2	1	1				1		1
223-227-278-362	G2a	8			3	1		$\overline{2}$		1
184	Н	1						$\frac{-}{2}$		
189	Н	$\overline{2}$						1	1	
CRS	H2	6	4		2	1	2	5	$\overline{2}$	1
311	H11	4	1		1	_	_	-	1	1
354	H2a	1	-		-	1			2	-
129-223	I	ī				-		1	-	
129-172-223-311-391	Ī1	2		1				_		
129-152-153-179-192-223-362	M	1		-				1		
111-129-223-257A-261	N9a	ī						1	1	
093-126-163-186-189-294-296	T1a	ī	1					-	-	
126-294-296-304-362	T2h	2	1	1						
356	Ū4	1	1	-			1		3	2
223-292	Ŵ	1	1		1		*		0	3
126-231-266	Y1b	1	1		*					0
185-223-26-298	\mathbf{Z}^*	1	1	1		1				

Population abbreviations are defined in the legend of Table 4.

genetically distinctive from the other two, KOSH and ZHAN, in terms of its overall genetic diversity, haplogroup composition and haplotype sharing. These findings could suggest that (a) multiple groups of Kazakhs migrated into the Altai region, (b) there is some clan or population structure in the Altaian Kazakhs, or (c) Kazakh groups have recently become differentiated due to population admixture and/or founder events. In light of their limited period of residence in the Altai-Sayan region, the intra-population differentiation seen in the Altaian Kazakhs probably developed before their arrival in this region, since they are not genetically close to indigenous Altaian groups.

Thus, it seems likely that such diversity reflects their having a heterogeneous ancestry dating back to pre-Turkic Central Asia, and perhaps to recent population interactions in China, Mongolia and the Altai region over the past three centuries. Evidence in support of the latter interpretation is the presence of haplogroups M7, M8, M13, N1a, and R1 in one of the three settlements to the exclusion of the others. On the other hand, ethnographic and demographic information also suggests that differences between CHAN and KOSH/ZHAN could be partly attributable to historical conversion and admixture with Russian populations by members of CHAN (Konovalov, 1986; Oktiabrskaya, 1997; Karih, 2000). Although the diversity indices and $F_{\rm ST}$ values suggest that each Altaian Kazakh settlement has experienced a distinct population history that defines its contemporary genetic make-up, they could not be statistically distinguished from each other, i.e., as subgroups separate from the entire Kazakh population. One explanation for this finding is that Kazakhs share a common biological ancestry, as seen in the sharing of haplogroups and haplotypes, but one that was later structured through clan divisions in the subsequent centuries, as suggested for the Y-chromosome variation in Central Asian populations (Chaix et al., 2004). However, this interpretation must be tested with additional genetic data (e.g., Y-chromosome and autosomal) and the further review of demographic information from Altaian Kazakh populations.

Common maternal ancestry of Turkic populations

Current data suggests that the spread of Turkic languages occurred through cultural diffusion, rather than a population expansion resulting in the replacement of autochthonous genetic lineages, at least in the maternal context (Comas et al., 1998; Quintana-Murci et al., 2004). This interpretation receives support from the observation that Altaic speaking populations share a common set of paternal haplotypes (Wells et al., 2001).

GENETIC VARIATION IN ALTAIAN KAZAKHS

Source of variation	Degrees of freedom	Percentage of variation	<i>P</i> -value
Ethnicity			
Among group	4	1.39	0.019 ± 0.004
Among population within group	6	0.73	0
Within populations	1257	97.88	0
Language			
Among group	1	-0.10	0.641 ± 0.013
Among population within group	9	1.95	0
Within populations	1247	98.15	0
Religion			
Among group	1	1.41	0.016 ± 0.004
Among population within group	9	1.10	0
Within populations	1247	97.49	0
Geography (Altaian Kazakhs in their h	ypothetical origin west of the Alta	i)	
Among group	3	1.69	0
Among population within group	7	0.55	0
Within populations	1257	97.76	0

TABLE 6. AMOVA of mtDNA haplotypes in Central and East Asian populations

Ethnicity: Group 1 = Kazakhs (AKAZ, CKAZ, YKAZ); Group 2 = Uighurs (CUIG, YUIG); Group 3 = Kirghiz (CKIR); Group 4 = Mongolians (YMNG, KMNG, SMNG); Group 5 = Altaians (NALT, SALT); Religion: Group 1 = Muslim (AKAZ, CKAZ, YKAZ, CUIG, YUIG, CKIR); Group 2 = Non-Muslim (YMNG, KMNG, SMNG, NALT, SALT); Language: Group 1 = Turkic (AKAZ, CKAZ, YKAZ, CKAZ, CUIG, YUIG, CKIR, NALT, SALT); Group 2 = Mongolic (YMNG, KMNG, SMNG); Geography: Group 1 = Southwest of Altai (AKAZ, CKAZ, CKAZ, CKAZ, CUIG, CKIR); Group 2 = Altai (NALT, SALT); Group 3 = South of Altai (YKAZ, CVIG, CKIR); Group 4 = Altai (NALT, SALT); Group 3 = South of Altai (YKAZ, YUIG, YMNG); Group 4 = Mongolia (KMNG, SMNG); When the Altaian Kazakhs were removed from the Geography analysis, the corresponding Among Group value was 0.33, p = 0.421.

Under such a scenario, while dispersing throughout Central Asia, Turkic groups (men) married women from local populations (Quintana-Murci et al., 2004), with religious, economic and political traditions developing differently among Turkic groups, especially after their westward expansion in the 10–11th centuries (Golden, 1992).

In this study, we also find that all Turkic and Mongolic groups possess a common set of maternal haplogroups (C, D, G2a, H), and a minimal number of haplotypes from these lineages at appreciable frequencies. However, the overall patterns of haplotype sharing amongst these groups vary considerably. This finding is not necessarily incompatible with the cultural diffusion model per se, but implies that present day Turkic-Mongolic ethnic groups emerged from a common mtDNA pool that was widely distributed in Central and East Asia.

These data more generally support previous interpretations of the genetic diversity in Central Asian populations. For instance, Quintana-Murci et al. (2004) documented that linguistic traditions do not correlate with mtDNA diversity among the present day populations living in the "Southwest Corridor" and "Central Asia", and our data indicate the same trend. On the other hand, religious tradition and ethnicity showed small but significant withingroup variance. These findings suggest some influence of current cultural practices on genetic variation in Turkic groups, and perhaps also the religious and socio-economic differences between indigenous Altaians and Altaian Kazakhs (Badenkov, 2002; Halemba, 2003). In addition, tribal and clan connections, rather than strictly ethnic ones, may have played an important role in shaping the genetic structure within Central Asia.

Genetic prehistory of Central Asia

Our study reveals several genetic patterns on different temporal levels in Central Asia. The first is the decrease in East Eurasian haplogroup frequencies in a westward direction, and the opposite trend for West Eurasian haplogroups. From a more local perspective, however, the spatial distribution of individual haplogroups and their haplotypes does not show the same pattern. Thus, it would be inaccurate to attribute the pattern of Central Asian genetic diversity to a single sweeping population event (but see Comas et al., 1998).

Instead, this pattern likely emerged through the combined effects of multiple population interactions that occurred at different periods of time. Such interactions would include the movement of Samoyedic speakers to northwest Siberia, the entry of West Eurasian steppe populations several millennia ago, the westward expansion of the Huns (Xiongnu), and later expansions of Turkic and Mongolic groups between the 9th and 13th centuries (Golden, 1992; Hiebert, 1994; Soucek, 2000; Frachetti, 2002; Renfrew, 2002).

The second is the occurrence of admixture between neighboring populations throughout the history of the Central Asian steppe. During and after their expansions, Turkic and Mongolic populations may have contributed to the gene pool of many other Central Asian populations, thereby creating genetic links between previously isolated groups. In addition, under the domination of these groups, the political and cultural landscape of Central Asia was remade. In fact, the emergence of the Kazakhs as a distinct group may partly attributable to the political and cultural landscape of 13th Century Kazakhstan, which was under the rule of the Golden Horde at that time (Golden, 1992).

These ever-changing economic and cultural interactions, along with the emergence of new subsistence patterns and technological developments (e.g., pastoral nomadism, oasis agriculture) (Bulliet, 1975; Hiebert, 1994) helped to create new social systems and connections between groups in this region (e.g., Frachetti, 2002). For instance, war-like nomadism, which gave rise to massive and ethnically diverse armies, greatly influenced the Central Asian genetic landscape (e.g., Renfrew, 2002). Similarly, during the Bronze Age, the Cen-

APPENDIX A

		0	
	CHAN	KOSH	ZHAN
Gene diversity	0.975 ± 0.010	0.993 ± 0.007	0.994 ± 0.002
Sample size	54	45	128
# haplotypes	34	39	88
Mean pairwise differences	6.325 ± 3.046	8.895 ± 4.177	10.565 ± 4.845
Tajima's D	-1.673	-1.551	-1.156
Fu's Fs	-21.919	-24.765	-24.320

TABLE A1. Genetic diversity in Altaian Kazakh Villages

TABLE A2. F_{ST} estimates for Altaian Kazakh settlements

	CHAN	KOSH	ZHAN
CHAN	*	0.046	0.001
KOSH	0.012	*	0.487
ZHAN	0.022	-0.001	*

Numbers in bold face are statistically significant (P < 0.05).

CHAN, Cherny Anuy; KOSH, Kosh Agach; ZHAN, Zhan Aul.

APPENDIX B

TABLE B1. F _{ST} Estimates for Central Asian population	TABLE B1.	$F_{\rm ST} E$	Estimates	for	Central	Asian	populations
--	-----------	----------------	-----------	-----	---------	-------	-------------

	AKaz	CKaz	YKaz	CUig	YUig	CKir	SMon	KMon	YMon	AKiz	NAlt
AKaz	*	0.425	0.051	0.099	0.613	0.593	0.000	0.011	0.446	0.000	0.000
CKaz	0.000	*	0.124	0.488	0.917	0.540	0.007	0.132	0.629	0.000	0.000
YKaz	0.006	0.006	*	0.014	0.373	0.124	0.097	0.188	0.103	0.011	0.002
CUig	0.004	-0.001	0.016	*	0.432	0.031	0.000	0.007	0.089	0.000	0.000
YUig	-0.002	-0.007	0.001	-0.000	*	0.746	0.026	0.205	0.605	0.000	0.000
CKir	-0.001	-0.001	0.005	0.010	-0.003	*	0.012	0.084	0.681	0.000	0.000
SMon	0.017	0.016	0.006	0.030	0.012	0.010	*	0.090	0.012	0.001	0.000
KMon	0.006	0.004	0.003	0.014	0.003	0.005	0.004	*	0.343	0.000	0.000
YMon	-0.000	-0.003	0.008	0.008	-0.003	-0.003	0.016	0.001	*	0.000	0.000
AKiz	0.033	0.036	0.015	0.053	0.029	0.027	0.014	0.019	0.032	*	0.000
NAlt	0.031	0.035	0.021	0.044	0.029	0.027	0.024	0.024	0.029	0.012	*

The numbers in boldface are statistically significant (P < 0.05). The $F_{\rm ST}$ values were estimated using the Tajima and Nei (1984) correction for nucleotide substitution. They appear above the diagonal, and the associated P values appear below it.

APPENDIX C

TABLE C1.	mtDNA	haplotypes	in	Mongolians

					1 0	1	0		
#	Hg	SNPs	HVS1 sequences	п	#	Hg	SNPs	HVS1 sequences	n
1	А	10873	093-223-290-293G-319	1	31	D4b	10398-10400	093-171-223-319-362	1
2	Α	10873	223-242-290-319	1	32	D4b	10398-10400	192-223-290-319-362	1
3	A4	10873	223-290-292A-319-362	1	33	D4b	10398-10400	223-319-362	3
4	B5	10873-10398-12705	126-140-189-311	1	34	D5a	10397-10398-10400	189-223-266-362	2
5	B5b	10873-10398-12705	140-189-243-260-278	1	35	F1	10873-12705	108-129-162-172-304	1
6	С	10398-10400	069-093-129-150-223-298-327	1	36	F1	10873-12705	124-148-290-304-309-390	1
7	С	10398-10400	093-129-223-298-327	2	37	F2b	10873-12705	203-304	2
8	C5	10398-10400	093-223-288-298-327	1	38	F1a	10873-12705	129-162-172G-304-311	1
9	C5	10398-10400	093-223-288-298-327-390	1	39	F1b	10873-12705	189-232A-249-304-311	1
10	C2	10398-10400	093-223-298-325-327-356	1	40	G2a	10398-10400	223-227-274-278-293-362	1
11	С	10398-10400	129-184-223-239-298-327-355	1	41	G2a	10398-10400	223-227-278-362	3
12	С	10398-10400	129-223-239-298-327	1	42	G3	10398-10400	223-265-274-362	2
13	С	10398-10400	129-223-298-327-434	1	43	Η	10873-12705	311	1
14	C5	10398-10400	164-223-261-288-298	1	44	Η	10873-12705	093-311	2
15	C4a	10398-10400	167-171-223-298-327-344-357	1	45	Η	10873-12705	m rCRS	2
16	C4a	10398-10400	171-223-298-327-344-357	1	46	\mathbf{J}	10873-10398-12705	069-126-193	1
17	C^*	10398-10400	223-298-327	4	47	\mathbf{J}	10873-10398-12705	069-126-362	2
18	D	10398-10400	042-214-223-362	2	48	\mathbf{J}	10873-10398-12705	069-126-366	3
19	D	10398-10400	093-223-232-290-362	3	49	Κ	10873-10398-12705	093-224-311	1
20	D	10398-10400	093-232-290-362	1	50	Κ	10873-10398-12705	224-311-368	1
21	D	10398-10400	171-223-311-362	1	51	Κ	10873-10398-12705	224-311	1
22	D	10398-10400	174-223-241-362	1	52	\mathbf{M}	10398-10400	215-223-274	1
23	D	10398-10400	223-224-245-292-362	2	53	\mathbf{M}	10398-10400	203-223-304	2
24	D	10398-10400	223-256-311-316-362	1	54	M8	10398-10400	189-223-298-355-362	1
25	D	10398-10400	223-286-362	1	55	M8c	10398-10400	093-223-261-288-298	3
26	D	10398-10400	223-287-362	2	56	N9a	10873	111-129-223-257A-261-273T	1
27	D	10398-10400	223-294-362	1	57	Т	10873-12705	078-126-294-296	2
28	D	10398-10400	223-262	3	58	U5	10873-12705	256-270-320-299	1
29	D4	10398-10400	245-362-368	1	59	W	10873	223-292	1
30	D4a	10398-10400	129-152-179-192-223-362	2	60	Z	10398-10400	129-185-223-224-260-298	3
								Total	89

tral Asian steppe, which was a barrier for population movement, became an important landscape for economic, political and, consequently, genetic interaction between Eurasian groups (Hiebert, 1994; Comas et al., 1998; Wells et al., 2001). In the later periods, the establishment of the Silk Road also facilitated links between the East and West (Comas et al., 1998, 2004; Whitefield, 1999; Yao et al., 2004).

Within Central Asia, the spatial distribution of specific maternal sublineages is complex. For instance, the presence of D5c in the Altaian Kazakhs may reflect admixture with indigenous Altaians or Mongolians, and certainly indicates ancient genetic connections between indigenous Altaians and Japanese, probably as a consequence of Altaic expansions in North Asia (Dulik et al., 2007). Similarly, subhaplogroup C4b, which is widespread in Siberia, does not appear in Kazakhs from Kazakhstan. Thus, its presence in the Altaian Kazakhs probably reflects recent admixture with populations from Xinjiang or Mongolia.

Another interesting finding is the considerable frequency and diversity of N1a haplotypes in Altaian Kazakhs. This haplogroup appears in geographically widespread populations, but is generally rare in Eurasia and Africa (Haak et al., 2005). Interestingly, Ricaut et al. (2004) reported a HVS1 sequence from a 2,500-year-old Altaian individual that also appears in four Altaian Kazakh individuals (#110; Table 2). In addition, we identified an N1a haplotype in a single Altai-kizhi individual (Zhadanov et al., 2007). Thus, the presence of this haplogroup in Central Asia may date back to the initial expansion of modern humans into eastern Eurasia.

Recent aDNA studies have also shed light on Central Asian population history. For instance, the Altaian Kazakhs possessed almost all of the haplotypes appearing in seven individuals from Bronze/Iron Age Kazakhstan (Lalueza-Fox et al., 2004). However, linking these ancient people with Altaian Kazakhs in terms of ethnicity or language is difficult, as both Turkic languages and the Kazakh ethnic identity emerged considerably later than the Iron Age (Golden, 1992). In addition, there are several unique haplotypes among Altaian Kazakhs that differ slightly from more widely observed West and East Eurasian haplotypes to which they are closely related. Given the considerable time depth of Central Asian genetic diversity, these haplotypes are likely to be unique to this region. Overall, it appears that this region has been not only a recipient but also a contributor of East Eurasian genetic diversity.

CONCLUSIONS

Our results reveal that the Altaian Kazakhs are one of the most genetically diverse groups in the region with their mtDNA pool encompassing most of the major East and West Eurasian lineages. This level of diversity likely reflects the general antiquity of the haplotypes present in this region, as well as population interactions and gene flow over the past several millennia. In addition, our study suggests that the genetic composition of the Kazakhs preceded the development of their distinct ethnic identity, and perhaps even the emergence of Turkic languages. This interpretation supports the emerging hypothesis that Central Asian genetic structure has been shaped mainly by clan social organization (Chaix et al., 2004). While the great maternal genetic diversity within the Altaian Kazakhs demonstrates the richness of Central Asian population history, it provides a singular view of biological variation in this group. Our ongoing analyses of paternal and autosomal genetic variation in Altaian Kazakhs will surely reveal new insights into its complex history, and contribute to our understanding of the way in which human groups formed, moved and interacted within the Altai-Sayan region.

ACKNOWLEDGMENTS

The authors would like to thank all of the Altaian Kazakh participants for their involvement in this project.

LITERATURE CITED

- Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral M, Dugoujon J, Roostalu U, Loogvali E, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti S, Semino O, Torroni A. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75:910–918.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457-465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.
- Aris-Brosou S, Excoffier L. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494–504.
- Badenkov YP. 2002. Poverty and remoteness contribute to ethnic tensions in the sacred Altai Mountains: a interview with Vladimir Sabin, member of Parliament. Altai Republic. Mountain Res Dev 22:320–323.
- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human-populations using median networks. Genetics 141:743–753.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48 (www.fluxus-engineering.com).
- Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA Data. Am J Hum Genet 71:1150–1160.
- Bregel Y. 2000. Historical maps of Central Asia, 9th–19th centuries A.D. Bloomington: Indiana University Research Institute for Inner Asian Studies.
- Bulliet RW. 1975. The camel and the wheel. Cambridge: Harvard University Press.
- Chaix R, Austerlitz F, Khegay T, Jacquesson S, Hammer MF, Heyer E, Quintana-Murci L. 2004. The genetic or mythical ancestry of descent groups: lessons from the Y-chromosome. Am J Hum Genet 75:1113–1116.
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am J Hum Genet 63:1824–1838.
- Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495–504.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M. 2003. Diversity of mitochondrial DNA lineages in South Siberia. Ann Hum Genet 67:391–411.

- Derevianko AP, Molodin VI, Savinov DG. 1994. Ethno-political situation in Ukok plateau in 19–20th centuries. In: Ancient Cultures of Berketsk valley (Mountain Altay, Plateau Ukok). Novosibirsk. p 187–192.
- Dulik M, Gokcumen O, Zhadanov S, Osipova L, Schurr TG. 2007. The phylogeography of haplogroup D5 [abstract]. Am J Phys Anthropol Suppl 44:102.
- Excoffier L, Smouse P, Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplo-types: application to human mitochondrial DNA restriction data. Genetics 131:479–491.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.
- Frachetti M. 2002. Bronze Age exploitation and political dynamics of the eastern Eurasian steppe zone. In: Boyle K, Colin R, Levine M, editors. Ancient interactions: East and West in Eurasia. Cambridge: McDonald Institute Press. p 161–170.
- Fu YX. 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection Genetics 147: 915–925.
- Gladney DC. 1996. Relational alterity: constructing Dungan (Hui). Uygur, and Kazakh identities across China, Central Asia, and Turkey. Hist Anthropol 9:445–477.
- Gladney DC. 2003. Dislocating China: Muslims, Minorities, and other subaltern subjects. London: C. Hurst.
- Gokcumen O, Dulik M, Zhadanov S, Osipova L, Schurr TG. 2007. The phylogeography of haplogroup N1a [abstract]. Am J Phys Anthropol Suppl 44:117.
- Golden PB. 1992. An introduction to the history of the Turkic peoples: ethnogenesis and state-formation in Medieval and Early Modern Eurasia and the Middle East. Wiesbaden: O. Harrassowitz.
- Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, Burger J. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310:1016–1018.
- Halemba A. 2003. Contemporary religious life in the Republic of Altai: the interaction of Buddhism and Shamanism. Sibirica 3:165–182.
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N. 2002. Reduced median network analysis of complete mitochondrial DNA coding region sequences for the major African. Asian, and European haplogroups. Am J Hum Genet 70:1157–1171.
- Hiebert FT. 1994. Origins of the Bronze Age Oasis Civilization in Central Asia. Cambridge: Peabody Museum of Archaeology and Ethnology Press.
- Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungerel G, Chae G-T, Han H, Jia G-J, Tokunaga K, Munkhtuvshin N, Tamiya G, Inoko H. 2005. Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. Gene 346:63– 70.
- Karih EV. 2000. Interethnic relationships in Altay region household in 19–20th centuries. In: Actual questions of Siberian history. Barnaul. p 274–275.
- Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334.
- Konovalov AV. 1986. Kazakhs in Southern Altai: the problem of ethnic origin. Alma-Ata, Nauka Kazakh SSR (In Russian).
- Lalueza-Fox C, Sampietro ML, Gilbert MTP, Castri L, Facchini F, Pettener D, Bertranpetit J. 2004. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. Proc R Soc Lond B 217:942–947.
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozari R, Bonne-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273.

- Nyambuu K. 1992. Introduction to the ethnography of Mongolia. Ulaanbaatar: Mongolian State Press.
- Oktiabrskaya VI. 1997. Kazakhs of Altai. History and the modern period. Ètnograficeskoe obozrenie 6:92–102.
- Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. Am J Phys Anthropol 120:211–224.
- Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Khang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75:966–997.
- Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K. 2004. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74: 827–845.
- Renfrew C. 2002. Pastoralism and interaction: some introductory questions. In: Boyle K, Colin R, Levine M, editors. Ancient interactions: East and West in Eurasia. Cambridge: McDonald Institute Press. p 1–24.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779.
- Ricaut FX, Keyser-Tracqui C, Bourgeois J, Crubezy E, Ludes B. 2004. Genetic analysis of a Scytho-Siberian skeleton and its implications for ancient Central Asian migrations. Hum Biol 76:109–112.
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychov I, Rychov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu, Hatina J, Belledi M, Di Rienzo A, Novelleto A, Oppenheim A, Norby S, Al-Zaheri, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H-J. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67(5):1251–76.
- Samaev GP. 1991. Mountain Altay in 17–19th centuries. Problems of political history and annexation to Russia. Gorno-Altaysk, Altai Books (In Russian).
- Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC. 1999. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea Bering Sea region during the Neolithic. Am J Phy Anthropol 108:1–39.
- Schurr TG, Wallace DC. 2003. Genetic prehistory of Paleoasiatic-speaking peoples of northeastern Siberia and their links to Native American populations. In: Kendall L, Krupnik I, editors. Constructing cultures then and now: celebrating Franz Boas and the Jesup North Pacific Expedition. Baltimore: Smithsonian Institution Press. p 239–258.
- Soucek S. 2000. A history of inner Asia. Cambridge: Cambridge University Press.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet 69 (Part 1):67–89.
- Tajima F, Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269-285.
- Tajima F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
- Tajima F. 1989b. The effect of change in population size on DNA polymorphism. Genetics 123:597-601.
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, Hirose R, Fujita Y, Kurata M, Shinoda K,

Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H. 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14:1832–1850.

- Togan ZV. 1994. The origins of Kazaks and Ozbeks. In: Paksoy HB, editor. Central Asia reader, Armonk: M.E. Sharpe. p 25– 29.
- Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-abrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R. 2001. A signal from human mtDNA of postglacial recolonization in Europe. Am J Hum Genet 69:844–852.
- Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonné-Tamir B, Scozzari R. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137–1152.

- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249.
- Whitefield S. 1999. Life along the Silk Road. Berkeley: University of California Press.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Khang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651.
- Yao Y-G, Qing-Peng Kong Q-P, Wang C-Y, Zhu C-L, Zhang Y-P. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. Mol Biol Evol 21:2265–2280.
- Zhadanov SI, Dulik M, Osipova LP, Schurr TG. 2006. Mitochondrial DNA variation in northern Altaians: affinities with Siberian and Turkic populations. Am J Phys Anthropol Suppl 129 (S42): 192.