

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Gene xx (2004) xxx–xxx

**GENE**  
 SECTION  
 EVOLUTIONARY GENOMICS
[www.elsevier.com/locate/gene](http://www.elsevier.com/locate/gene)

## Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis

Toru Katoh<sup>a</sup>, Batmunkh Munkhbat<sup>a,b</sup>, Kenichi Tounai<sup>a</sup>, Shuhei Mano<sup>a</sup>, Harue Ando<sup>a</sup>, Ganjuur Oyungerel<sup>b</sup>, Gue-Tae Chae<sup>c</sup>, Huun Han<sup>d</sup>, Guan-Jun Jia<sup>e</sup>, Katsushi Tokunaga<sup>f</sup>, Namid Munkhtuvshin<sup>b</sup>, Gen Tamiya<sup>a</sup>, Hidetoshi Inoko<sup>a,\*</sup>

<sup>a</sup>Molecular Life Science, School of Medicine, Tokai University, Bohseidai, Isehara, Kanagawa 259-1193, Japan

<sup>b</sup>Central Scientific Research Laboratory, National Institute of Medicine, Ulaanbaatar, Mongolia

<sup>c</sup>Institute of Hansen's Disease, The Catholic University of Korea, Seoul, Republic of Korea

<sup>d</sup>Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

<sup>e</sup>Harbin Red Cross Blood Center, Harbin, China

<sup>f</sup>Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

Received 23 June 2004; received in revised form 12 October 2004; accepted 21 October 2004

Received by T. Gojobori

### Abstract

About 20 ethnic groups reside in Mongolia. On the basis of genetic and anthropological studies, it is believed that Mongolians have played a pivotal role in the peopling of Central and East Asia. However, the genetic relationships among these ethnic groups have remained obscure, as have their detailed relationships with adjacent populations. We analyzed 16 binary and 17 STR polymorphisms of human Y chromosome in 669 individuals from nine populations, including four indigenous ethnic groups in Mongolia (Khalkh, Uriankhai, Zakhchin, and Khoton). Among these four Mongolian populations, the Khalkh, Uriankhai, and Zakhchin populations showed relatively close genetic affinities to each other and to Siberian populations, while the Khoton population showed a closer relationship to Central Asian populations than to even the other Mongolian populations. These findings suggest that the major Mongolian ethnic groups have a close genetic affinity to populations in northern East Asia, although the genetic link between Mongolia and Central Asia is not negligible.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Y chromosome; Mongolian; Binary haplogroup; STR haplotype

### 1. Introduction

At present, nearly 20 ethnic groups are distributed in the various regions of Mongolia. Anthropological studies have suggested that contemporary Mongolians are direct descendants of ancient Mongolian and Turkish nomads

whose remains, dating back to at least the end of the Neolithic period, have been discovered throughout the territory of Mongolia (Tumen, 1992). A population genetic study based on classical enzyme markers has placed Mongolians in the Northeast Asian cluster, including Tibetan, Korean, Japanese, and Ainu populations, implying a close relationship between Mongolia and East Asia (Cavalli-Sforza et al., 1994). Furthermore, a recent Y-chromosomal study suggested that one specific Y-chromosomal haplotype, carried by likely male-line descendants of Genghis Khan of the Mongol empire, recently and rapidly spread throughout the Eurasian continent (Zerjal et al., 2003). Therefore, it is believed

*Abbreviations:* mtDNA, mitochondrial DNA; STR, short-tandem-repeat; PCR, polymerase chain reaction; MJ, median-joining; TMRCA, time to the most common recent ancestor; MDS, multidimensional scaling.

\* Corresponding author. Tel.: +81 463 93 1121x2312; fax: +81 463 96 4137.

E-mail address: [hinoko@is.icc.u-tokai.ac.jp](mailto:hinoko@is.icc.u-tokai.ac.jp) (H. Inoko).

0378-1119/\$ - see front matter © 2004 Elsevier B.V. All rights reserved.  
doi:10.1016/j.gene.2004.10.023

that Mongolians have had a pivotal role in the peopling of Central and East Asia. However, the genetic relationships among the ethnic groups in Mongolia and their detailed relationships with adjacent populations have remained obscure.

It has been known that the nonrecombining region of the human Y chromosome is inherited paternally as a haploid genome, and that the polymorphisms in this region can be used as informative markers for tracing paternal genetic variations. Thus, there have been many reports on Y-chromosome variations in various populations, including those in Central and East Asia (Karafet et al., 2001; Qamar et al., 2002; Tajima et al., 2002; Zerjal et al., 2002, 2003). However, relatively little data on Mongolian populations has been available.

To decipher the genetic features of Mongolian populations, we investigated genetic variations based on binary and short-tandem-repeat (STR) markers on the Y chromosome. We examined four indigenous ethnic groups in Mongolia: the Khalkh, Uriankhai, Zakhchin, and Khoton. The Khalkh, who make up the largest ethnic group, are considered direct descendants of the core Mongolian tribes who inhabited the present territory of Mongolia from the eighth to the twelfth centuries (Nyambuu, 1992). The Uriankhai and Zakhchin are relatively small ethnic groups inhabiting primarily western Mongolia. It is generally accepted that the ancestors of the Uriankhai and Zakhchin belonged to the so-called Oirad Mongols, who came from several western Mongolian tribes of Mongolian and Turkish origin (Nyambuu, 1992). Although the Uriankhai are recognized as one of the ancient tribes inhabiting the mountainous areas of western and northern Mongolia, historical data indicate that this group was formed by the gradual intermixing of nomadic groups of Mongol and Turkish origin in the twelfth century (Nyambuu, 1992). Ethnological data indicate that the Zakhchin descended from a recent admixture of Oirad Mongolian groups such as the Torguud, Durvud, and Oold during the fifteenth to seventeenth centuries (Nyambuu, 1992). The Khoton population is one of the smallest groups, residing only in northwestern Mongolia. This population is considered to be a small group of Turkish origin that migrated into Mongolia in the seventeenth century (Nyambuu, 1992).

We analyzed 16 binary and 17 STR markers to study the paternal genetic variation of these four ethnic groups. The results were then compared with those analyzed for five East Asian populations (Manchu, Northern Han, Korean Chinese, Korean, and Japanese). The present study revealed that the major Mongolian ethnic groups have a relatively close genetic affinity to the northern part of East Asia. However, our study also revealed that the Khoton population has a closer relationship to Central Asia than to the other Mongolians, implying that the genetic link between Mongolia and Central Asia is not negligible.

## 2. Materials and methods

### 2.1. Population samples

A total of 669 unrelated male DNA samples were analyzed in the present study. Among them, 245 individuals belonged to four Mongolian ethnic groups: the Khalkh (85), Uriankhai (60), Zakhchin (60), and Khoton (40). All individuals were chosen on the basis of a family history in which there was no obvious admixture with individuals from other ethnic backgrounds. As control populations, we also employed 101 Manchu, 42 Northern Han, 79 Korean Chinese, 85 Korean, and 117 Japanese males. The Manchu, Northern Han, and Korean Chinese samples were collected in the northeastern part of China. Korean samples were collected in Seoul by the Catholic University of Korea. Japanese samples were collected in the Kanto region, the central part of the mainland.

Informed consent was obtained from donors at each laboratory or institution. For genetic study using these samples, we obtained permission from the ethics committee at Tokai University, Japan.

### 2.2. Y-chromosome binary markers

We typed 669 male individuals for a total of 16 binary markers on the Y chromosome. These markers were M130 (RSP4YC711), M48, YAP, M174, SRY-4064, M89, 12f2, M9, Tat, M175, M119, SRY+465, M122, M45, SRY-1532, and M17.

M130 was typed by polymerase chain reaction (PCR) amplification using the primers described in Bergen et al. (1999), followed by *Bs*I digestion. M174 was typed by PCR amplification using the primers described in Underhill et al. (2001), followed by *Bsr*I digestion. M48 and M17 were typed by the method of Qamar et al. (2002). YAP, SRY-4064, 12f2, Tat, SRY+465, and SRY-1532 were typed according to Hammer and Horai (1995), Bravi et al. (2000), Rosser et al. (2000), Zerjal et al. (1997), Shinka et al. (1999), and Santos et al. (1999), respectively.

For the in/del marker, M175, we performed PCR amplification with the primers described in Underhill et al. (2001). We labeled the 5' end of a forward primer with fluorescence (HEX). For M89, M9, M122, M119, and M45, we performed allele-specific PCR assays according to the primer designs of Su et al. (1999). These markers consist of two forward primers, which defined SNP alleles, and one reverse primer. We labeled the 5' end of each forward primer with a different fluorescent dye (FAM and HEX). The amplified products were detected using the ABI Prism 3700 DNA analyzer.

PCR reactions were carried out in a 10- $\mu$ l reaction volume, containing 1X Ampli Taq PCR buffer with 1.5 mM MgCl<sub>2</sub> (PE Applied Biosystems), 200  $\mu$ M each of

dNTP, 0.25 U AmpliTaq polymerase (PE Applied Biosystems), 1 pmol of each primer, and 6 ng of genomic DNA. The cycle conditions were as follows: 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 54–62 °C for 30 s, and 72 °C for 50 s, followed by a 5-min extension at 72 °C.

### 2.3. Y-STR genotyping

We used a total of 17 tetranucleotide STR markers on the Y chromosome. These markers were DYS389I, DYS389b, DYS390, DYS391, DYS393, Y-GATA-A7.1 (DYS460), Y-GATA-A7.2 (DYS461), Y-GATA-A10, Y-GATA-C4, Y-GATA-H4, DYS597, DYS600, DYS601, DYS603, DYS605, DYS606, and DYS607.

We labeled the 5' end of a forward primer with fluorescence (FAM or HEX), and a PCR reaction for each marker was independently carried out. PCR reactions for DYS389I, DYS389b, DYS390, DYS391, and DYS393 were performed according to published procedures with minor modifications (Kayser et al., 1997). PCR reactions for the remaining 12 markers were carried out in a total volume of 10 µl, containing 10 mM Tris–HCl, pH 8.0, 50 mM KCl, 2.0 mM MgCl<sub>2</sub>, 0.01% gelatin, 200 µM each of dNTP, 0.3 U AmpliTaq polymerase (PE Applied Biosystems), 2 pmol of each primer, and 6 ng of genomic DNA. These markers were amplified by the use of the following cycling conditions: pre-PCR of 5 min at 96 °C, 1 min at 57 °C, and 1 min at 72 °C, followed by 30 cycles of 96 °C for 45 s, 57 °C for 45 s, and 72 °C for 1 min, with a final extension at 72 °C for 5 min.

Amplicons were separated using the ABI Prism 3700 DNA analyzer, and allele sizes were calibrated with the GeneScan 500 ROX size standard. Then, fragment sizes were determined using GeneScan Analysis 3.5 and Genotyper 3.5NT software (PE Applied Biosystems). The number of repeat units in each fragment was determined by sequencing reference DNA samples with different alleles. Alleles were designated by repeat numbers. The allele length for DYS389b was obtained by subtracting the allele length of DYS389I from that of DYS389II (Qamar et al., 2002; Zerjal et al., 2002).

### 2.4. Data analysis

Y-chromosome binary haplogroups for nine populations were defined by the analysis of all 16 binary polymorphisms. The nomenclature of haplogroups followed that of the Y chromosome consortium (2002).

For Y-STR data, an unbiased estimate of haplotype diversity was calculated with the Arlequin 2.001 program (Schneider et al., 2000). A median-joining (MJ) network among Y-STR haplotypes in the binary haplogroup was constructed using the Network 2.0 and 4.0 programs (Bandelt et al., 1999). In this analysis, each STR locus was weighted according to the molecular variance in each

haplogroup, with the weight inversely proportional to the variance.

Time to the most common recent ancestor (TMRCA) of a set of chromosomes was estimated using BATWING (Wilson and Balding, 1998), assuming a demographic model of exponential growth from a constant-sized population. In this analysis, we applied the locus-specific mutation rate for the loci investigated by Kayser et al. (2000). For loci not investigated by Kayser et al. (2000), we applied a mutation rate of 0.0032, which was the average mutation rate for tetranucleotide STRs calculated from their data. The age of STR variation was also estimated using the Td approach of Zhivotovsky et al. (2004), in which we applied mutation rates ( $w$ ) of 0.00069 (Zhivotovsky et al., 2004) and 0.0032 (see above). All analyses were performed with a 25-year generation time.

The pairwise  $\Phi_{ST}$  value between each pair of populations was calculated using the sum of squared size differences option in the Arlequin program. The calculation of  $D_A$  distances based on allele frequencies was performed using the DISPAN program (Ota, 1993). Based on the  $\Phi_{ST}$  and  $D_A$  distances, we performed the multidimensional scaling (MDS) plot with the ViSta 5.6 program (Young and Bann, 1996).

## 3. Results and discussion

### 3.1. Distribution of Y-chromosomal haplogroups

Using the 16 Y-chromosome binary markers, we detected 13 haplogroups in 669 samples representing nine populations, including four Mongolian ethnic groups. Fig. 1 shows the MP tree for the detected haplogroups (Y chromosome consortium, 2002). The haplogroup diversity in each of the

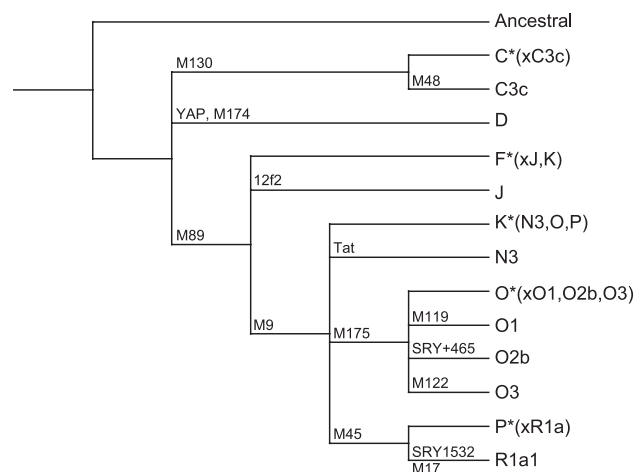


Fig. 1. Parsimony tree for 13 haplogroups detected from 16 binary markers on the Y chromosome. Haplogroup names according to the Y chromosome consortium (2002) are shown at the tips. Marker names are shown at the expected mutational points on branches. Haplogroup E, which is defined by SRY-4064, is not shown because it was not detected in the present study.

Table 1  
Y-chromosomal diversity values and haplogroup distribution.

	Population									Mean STR variance of haplogroup <sup>a</sup>
	Khalkh (n=85)	Uriankhai (n=60)	Zakhchin (n=60)	Khoton (n=40)	Manchu (n=101)	Northern Han (n=42)	Korean Chinese (n=79)	Korean (n=85)	Japanese (n=117)	
Haplogroup distribution										
C*(xC3c)	35	15	10	0	9	4	10	14	6	0.4486
C3c	13	20	18	4	8	0	0	0	0	0.1243
D	3	1	2	0	0	0	2	3	44	0.4311
F*(xJ,K)	2	0	1	0	0	0	0	0	0	–
J	2	0	1	2	0	0	1	0	0	0.3889
K*(xN3,O,P)	3	3	3	0	2	3	1	4	1	0.4121
N3	4	5	2	0	0	0	0	2	0	0.3091
O*(xO1,O2b,O3)	0	0	5	0	1	8	2	3	1	0.3079
O1	0	0	0	0	3	2	0	2	3	0.3629
O2b	0	3	2	0	34	0	38	28	42	0.2354
O3	16	4	5	1	43	22	23	29	19	0.5446
P*(xR1a)	4	5	3	0	1	3	2	0	1	0.6052
R1a1	3	4	8	33	0	0	0	0	0	0.1431
Haplogroup diversity	0.7714±0.0341	0.8119±0.0318	0.8559±0.0276	0.3141±0.0906	0.6966±0.0289	0.6841±0.0645	0.6741±0.0372	0.7510±0.0262	0.7059±0.0223	
No. of STR haplotypes	73	33	47	14	52	38	78	83	107	
STR haplotype diversity	0.9933±0.0041	0.9542±0.0165	0.9876±0.0067	0.8295±0.0489	0.9404±0.0165	0.9942±0.0071	0.9997±0.0021	0.9994±0.0019	0.9973±0.0019	

<sup>a</sup> Calculation was performed when five or more chromosomes were observed.

nine populations is shown in Table 1. It is noted that the Khoton population presented an extremely low haplogroup diversity (0.3141) compared to the other populations (0.6741 to 0.8559).

The frequency distribution of the 13 haplogroups in each population is also shown in Table 1. Haplogroups C\*(xC3c) and C3c, both of which are defined by M130, were the most common ones in Mongolians, except for Khoton. This result was consistent with the previous findings that haplogroup C was prevalent in Mongolia, Central Asia, and Siberia (Karafet et al., 2001; Tajima et al., 2002; Zerjal et al., 2002). Between the two haplogroups, C\*(xC3c) was more frequent in Khalkh and East Asian populations (Northern Han, Korean Chinese, Korean, and Japanese), whereas C3c was more frequent in the populations of western Mongolia (Uriankhai, Zakhchin, and Khoton).

Although haplogroup D, which has been known as the YAP+ chromosome, was observed at high frequency in Japanese (38%), it was found at very low frequencies in the other populations (0 to 4%). This result is consistent with the previous Y-chromosome studies of East Asia (Hammer and Horai, 1995; Tajima et al., 2002). In the present study, this haplogroup was distributed in both central and western Mongolia. This might be a signature of the genetic contribution of YAP+ chromosomes to East Asia from the northwest (Karafet et al., 2001).

Haplogroups F\*(xJ,K), J, N3, and P appeared with greater frequency in Mongolians than East Asians, although the frequency of all of these haplogroups was low. Among these, haplogroup J was found in three Mongolian populations (Khalkh, Zakhchin, and Khoton) but not in East Asians, except for Korean Chinese. According to previous studies, haplogroup J is thought to represent the signatures of the Neolithic demic expansion of agriculturists, of long-distance trade via the Silk Road, or of the expansion of the Muslim world (Semino et al., 2000; Quintana-Murci et al., 2001; Zerjal et al., 2002). The presence of haplogroup J at a low frequency in Mongolia seems likely due to either of the latter two explanations, since the harsh climatic conditions of these areas would not have encouraged the spread of agriculture (Zerjal et al., 2002).

Haplogroup R1a1 was observed at the highest frequency in the Khoton population (83%) and at low frequencies in the Khalkh, Uriankhai, and Zakhchin populations (4 to 13%). This haplogroup was not observed in any East Asian populations. This haplogroup has been shown to be generally frequent in Central Asian populations (Karafet et al., 2001; Zerjal et al., 2002). Thus, the frequency distribution patterns of Y-SNP haplogroups in the Khoton population were similar to those of Central Asian populations rather than those of the other Mongolian populations.

### 3.2. Y-STR variation

Y-STR polymorphisms were analyzed to obtain a more detailed view of DNA variation in Mongolian and East

Table 2

Y-STR variance of haplogroups C (defined by M130) and O (defined by M175) lineages in each population<sup>a</sup>

Population	Mean STR variance	
	Haplogroup C	Haplogroup O
Khalkh	0.4220	0.5191
Uriankhai	0.3810	0.5210
Zakhchin	0.4094	0.7663
Khoton	–	–
Manchu	0.3306	0.4984
Northern Han	–	0.5951
Korean Chinese	0.3482	0.6025
Korean	0.2830	0.6911
Japanese	0.3578	0.5344
All populations	0.4812	0.6328

<sup>a</sup> Calculation was performed when five or more chromosomes were observed.

Asian populations. We typed 669 samples with 17 Y-STR markers and detected 504 different haplotypes, among which 451 (89%) were individual-specific. No haplotypes were shared between the SNP haplogroups. Notably, STR haplotype diversity was highest in the Korean Chinese population and lowest in the Khoton population (0.9997 and 0.8295, respectively; Table 1). Thus, both the STR data and the binary data showed that the Khoton population had less genetic diversity than the other populations.

We calculated the mean STR variance of haplogroup C (defined by M130) and O (defined by M175) lineages in each population (Table 2) because the distributions of those haplogroups between Mongolia and East Asia contrasted sharply. The results showed a propensity for the variance of haplogroup C to be higher in Mongolia (0.3810 to 0.4220) than in East Asia (0.2830 to 0.3578). In contrast, the variance of haplogroup O in East Asia seemed higher than that in Mongolia. This propensity was not very clear, however, because the value was largest in the Zakhchin (0.7663) and lowest in the Manchu (0.4984) populations.

We constructed a median-joining (MJ) network among the STR haplotypes in each haplogroup. Fig. 2 represents the networks in haplogroups C\*(xC3c), C3c, and R1a1, which were major ones in Mongolian populations. The C\*(xC3c) network was comprised of the Khalkh, Uriankhai, and Zakhchin populations as well as every East Asian population, and its pattern was relatively complex (Fig. 2a). The C3c network was predominant in Khalkh, Uriankhai, and Zakhchin, although small numbers of chromosomes in Khoton and Manchu were included (Fig. 2b). Some haplotype clusters were recognized in both the C\*(xC3c) and C3c networks, and these clusters were typically comprised of Khalkh, Uriankhai, and Zakhchin populations. Notably, we found that one cluster in the C\*(xC3c) network (the shaded area in Fig. 2a) contained haplotype 16-25-10-13 for loci DYS389b–DYS390–DYS391–DYS393. This profile corresponds to that of the “star cluster,” which was thought to be carried by likely male-line descendants of Genghis Khan, as described in Zerjal et al. (2003).

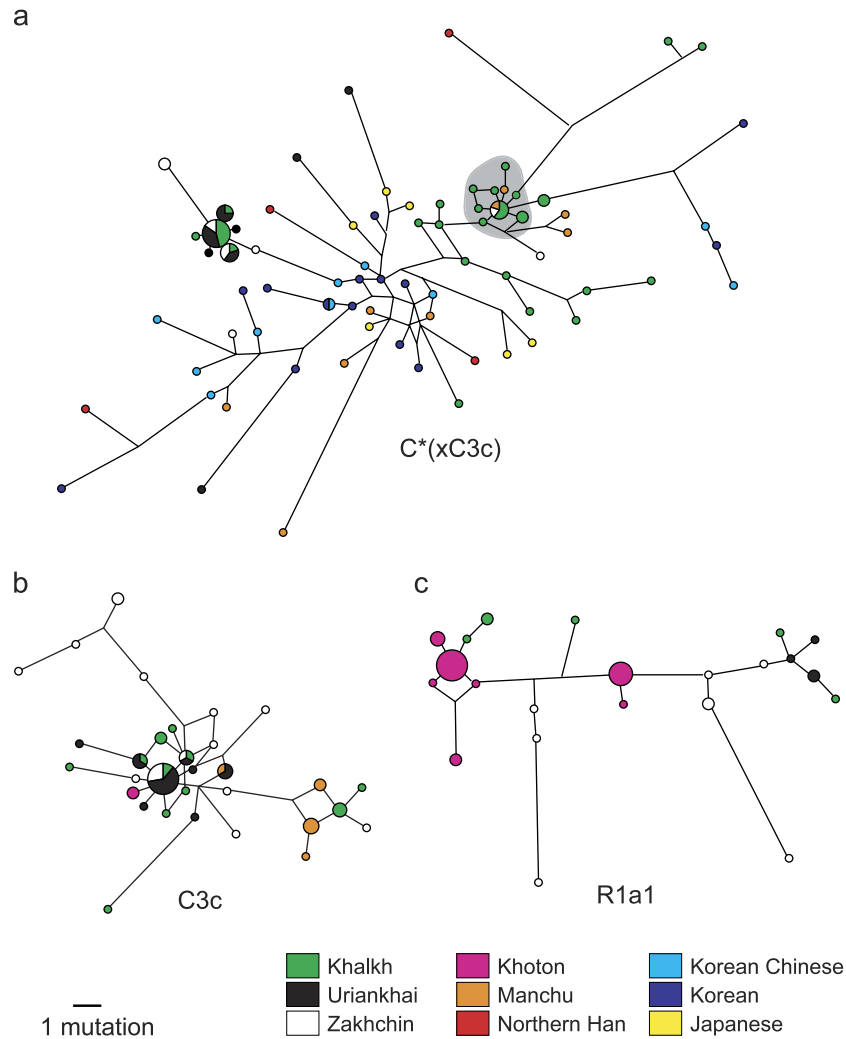


Fig. 2. Median-joining (MJ) networks of Y-STR haplotypes in three binary haplogroups. (a) C\*(xC3c); (b) C3c; (c) R1a1. Circles represent haplotypes with areas proportional to frequencies, and colors indicate the population of origin. Mutational differences are represented by lines.

The R1a1 network, which was predominant in the Khoton population, represented a relatively simple pattern with less haplotypes (Fig. 2c). It is notable that most chromosomes in the Khoton population were clustered into the same haplotype or closely related haplotypes. Generally, this pattern is expected if the population has suffered a bottleneck effect (Zerjal et al., 2002). This is compatible with the historical premise that the Khoton population was descended from a small number of individuals that migrated into their present territory in Mongolia during the seventeenth century (Nyambuu, 1992). Thus, it is suggested that the Khoton population would have experienced a recent bottleneck as a result of the premised migration event.

The coalescent time of STR haplotypes and the age of STR variation for representative haplogroups [C\*(xC3c), C3c, D, O2b, O3, and R1a1] were estimated using BATWING and the Td approach. The results are summarized in Table 3. We also performed these analyses for the following selected lineages: haplogroup D in Japanese; haplogroup R1a1 in Khoton; and “star cluster” in hap-

logroup C\*(xC3c). The time estimates of haplogroup D in Japanese were relatively old, while those for R1a1 in the Khoton population were relatively recent. These results seem to well reflect the historical premises of the two

Table 3

Time estimates (in years) for selected lineages based on the diversity of 17 STRs

	BATWING estimates of TMRCA (95% CI)	Age of STR variation by Td approach	
		w=0.0032	w=0.00069
C*(xC3c)	7400 (1000–37,000)	5900±200	27,300±800
C3c	3500 (300–19,700)	1100±300	5000±1500
D	8800 (800–51,500)	4000±800	18,200±3700
O2b	6300 (600–37,000)	2300±500	10,600±2100
O3	8500 (1500–35,500)	5700±1300	26,000±6100
R1a1	3400 (600–13,900)	1600±400	7500±1700
D in Japanese	8600 (800–562,700)	3400±700	15,600±3400
R1a1 in Khoton	1600 (300–6300)	900±300	4300±1200
“Star cluster” in C*(xC3c)	1000 (100–5000)	400±200	1700±700

populations (Hammer and Horai, 1995; Nyambuu, 1992). The time estimate of the “star cluster” in haplogroup C\*(x3C3c) was estimated as  $1700 \pm 700$  years even when  $w=0.00069$  was applied in the Td approach. This result seems roughly compatible with Zerjal et al. (2003), in which those authors suggested that this lineage may represent direct ancestry from Genghis Khan.

To investigate the extent of genetic differentiation among the nine populations, we calculated the pairwise  $\Phi_{ST}$  and  $D_A$  distances based on Y-STR variation. Although the  $\Phi_{ST}$  distances were significant ( $P < 0.05$ ) for almost all pairs of populations (data not shown), the values were not significant for the pair of Uriankhai and Zakhchin ( $\Phi_{ST} = -0.0041$ ) or for that of Korean Chinese and Korean ( $\Phi_{ST} = 0.0042$ ). Using these distances, we performed the MDS analyses (Fig. 3). In both cases, East Asian populations were clustered in the right lower area of the plot. Among Mongolian populations, Khalkh, Uriankhai, and Zakhchin were placed in the upper central area, whereas Khoton was placed in the left lower area, representing an apparent genetic differentiation between Khoton and the others.

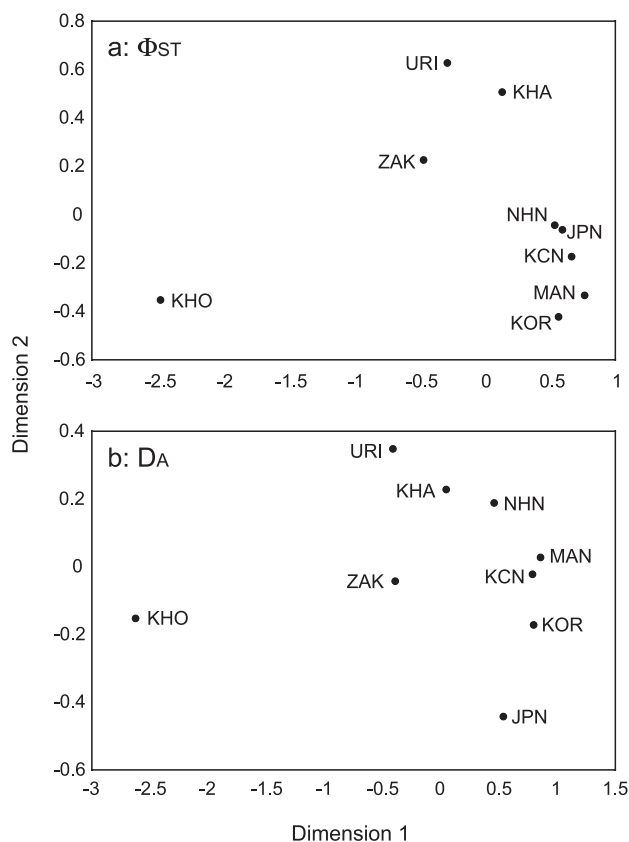


Fig. 3. Multidimensional scaling (MDS) plot for nine populations based on the genetic distances using Y-STR polymorphisms. (a)  $\Phi_{ST}$  distances. Stress values=0.2045. (b)  $D_A$  distances. Stress value=0.1312. The abbreviations are as follows: KHA, Khalkh; URI, Uriankhai; ZAK, Zakhchin; KHO, Khoton; MAN, Manchu; NHN, Northern Han; KCN, Korean Chinese; KOR, Korean; JPN, Japanese.

### 3.3. Genetic features of Mongolian ethnic groups

The present study showed that three Mongolian populations, Khalkh, Uriankhai, and Zakhchin, bore close relationships to the Siberian populations. The frequency distributions of the binary haplogroups showed that these populations possessed a higher frequency of haplogroup C lineage than any other haplogroup, which is generally frequent in the northern part of East Asia (Karafet et al., 2001; Tajima et al., 2002). It is noted that the Khalkh population mainly inhabits the central part, whereas the Uriankhai and Zakhchin populations are distributed in the western part of Mongolia. Thus, the predominance of haplogroup C lineage in both the northern and western parts suggests that it may be regarded as a common genetic feature throughout Mongolia.

Our study also showed that, among the three Mongolian populations, the Uriankhai and Zakhchin possessed a particularly close genetic relationship to each other. Although the closeness was not clearly shown in the MDS plot, the  $\Phi_{ST}$  distance between these populations presented a nonsignificant value. It is worth mentioning that the LD pattern obtained from our previous study on X chromosomes also showed great similarities between the Uriankhai and Zakhchin populations (Katoh et al., 2002). These findings produce two possible explanations of the history of these two populations. One is that they originated from a recent common ancestor, and the other is that they experienced a population admixture. Historical and ethnographical studies have shown some evidence leaning toward the former explanation (Nyambuu, 1992), while we cannot rule out the possibility of the latter. Therefore, further study will be needed in order to clarify in detail the relationship between these populations.

The frequency distributions of the binary haplogroups showed that the Khoton population had a higher frequency of haplogroup R1a1 than did any other population. According to previous studies, this haplogroup shows a high frequency in Central Asia and a relatively low frequency in the Caucasus, the Middle East, and East Asia (Semino et al., 2000; Karafet et al., 2001). Zerjal et al. (2002) speculated that this haplogroup came from the expansion of early nomadic groups in Central Asia. In addition, the MJ network in the Khoton population showed a unique topology that would represent a recent bottleneck. These results suggest that the Khoton population has a history that differs from that of the other Mongolian populations. Notably, previous anthropological studies have suggested that the Khoton population was most closely related to the Kirghiz, Kazakh, and Uzbek populations of Central Asia, which are in turn typical representatives of the Southern Siberian populations (Batsuuiri, 1977). Historical studies have suggested that the Khoton population originated from a small number of individuals of Turkish origin who recently migrated into the present territory of Mongolia (Nyambuu, 1992). Our relatively recent time estimates of

the R1a1 lineage in the Khoton population seem to be in good agreement with these anthropological and historical studies.

In this study, we analyzed 33 polymorphic markers on the Y chromosome to investigate the genetic variation in four Mongolian populations. Our study showed that the major Mongolian ethnic groups have a relatively close genetic affinity to populations in the northern part of East Asia, while analyses of the Khoton population revealed also a genetic link between Mongolia and Central Asia. It is noted, however, that these findings come from analyses of only four Mongolian populations. Therefore, further study, investigating a larger number of Mongolian populations, is required before we can understand in detail the evolutionary and migratory processes in Mongolia in relation to those of other populations.

### Acknowledgements

We would like to express our appreciation to all of the blood donors who made this study possible. Special thanks go to P. Oyungerel, N. Chuluunjav, D. Altansukh, and M. Tomizawa for their help in collecting samples. We also thank L.A. Zhivotovsky for his help concerning the statistical analyses of STR data. This research has been supported by the Special Coordination Funds for the Promotion of Science and Technology (SCF) and by a MEXT Grant-in-Aid for the 21st Century COE Program from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

### References

- Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Batsuuri, J., 1977. Anthropological characteristics of the ethnic group “Khoton”. *Proceedings of the Institute of General and Experimental Biology*, vol. 12. Mongolian Academy of Sciences, Ulaanbaatar, Mongolia, pp. 111–127.
- Bergen, A.W., et al., 1999. An Asian–Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* 63, 63–80.
- Bravi, C.M., Bailliet, G., Martínez-Marignac, V.L., Bianchi, N.O., 2000. Origin of YAP<sup>+</sup> lineages of the human Y-chromosome. *Am. J. Phys. Anthropol.* 112, 149–158.
- Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton.
- Hammer, M.F., Horai, S., 1995. Y-chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* 56, 951–962.
- Karafet, T., et al., 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 69, 615–628.
- Katoh, T., et al., 2002. Genetic isolates in East Asia: a study of linkage disequilibrium in the X chromosome. *Am. J. Hum. Genet.* 71, 395–400.
- Kayser, M., et al., 1997. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Leg. Med.* 110 (125–33), 141–149.
- Kayser, M., et al., 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y-chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 66, 1580–1588.
- Nyambuu, K., 1992. *Introduction to the Ethnography of Mongolia*. Mongolian State Press, Ulaanbaatar.
- Ota, T., 1993. *DISPAN: Genetic and Phylogeny Analysis*. Pennsylvania State University, University Park, PA.
- Qamar, R., et al., 2002. Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.* 70, 1107–1124.
- Quintana-Murci, L., et al., 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am. J. Hum. Genet.* 68, 537–542.
- Rosser, Z.H., et al., 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* 67, 1526–1543.
- Santos, F.R., et al., 1999. The central Siberian origin for native American Y-chromosomes. *Am. J. Hum. Genet.* 64, 619–628.
- Schneider, S., Roessli, D., Excoffier, L., 2000. *Arlequin: a Software for Population Genetics Data Analysis*. Ver 2.000. Genetics and Biometry Laboratory, University of Geneva.
- Semino, O., et al., 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290, 1155–1159.
- Shinka, T., et al., 1999. Genetic variations on the Y-chromosome in the Japanese population and implications for modern human Y-chromosome lineage. *J. Hum. Genet.* 44, 240–245.
- Su, B., et al., 1999. Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* 65, 1718–1724.
- Tajima, A., et al., 2002. Three major lineages of Asian Y-chromosomes: implications for the peopling of east and southeast Asia. *Hum. Genet.* 110, 80–88.
- Tumen, D., 1992. *Anthropology of the contemporary population of Mongolian People’s Republic*. PhD Thesis. MV Lomonosov Moscow State University, Moscow.
- Underhill, P.A., et al., 2001. The phylogeography of Y-chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43–62.
- Wilson, I.J., Balding, D.J., 1998. Genealogical inference from microsatellite data. *Genetics* 150, 499–510.
- Y chromosome consortium, 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12, 339–348.
- Young, F.W., Bann, C.M., 1996. A visual statistics system. In: Stine, R.A., Fox, J. (Eds.), *Statistical Computing Environments for Social Researches*. Sage Publications, New York, pp. 207–236.
- Zerjal, T., et al., 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* 60, 1174–1183.
- Zerjal, T., Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Tyler-Smith, C., 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* 71, 466–482.
- Zerjal, T., et al., 2003. The genetic legacy of the Mongols. *Am. J. Hum. Genet.* 72, 717–721.
- Zhivotovsky, L.A., et al., 2004. The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74, 50–61.