

## ORIGINAL ARTICLE

Wook Kim · Dong Jik Shin · Shinji Harihara  
Yung Jin Kim

## Y chromosomal DNA variation in East Asian populations and its potential for inferring the peopling of Korea

Received: October 13, 1999 / Accepted: November 12, 1999

**Abstract** We have examined variations of five polymorphic loci (*DYS287*, *DXYS5Y*, *SRY465*, *DYS19*, and *DXYS156Y*) on the Y chromosome in samples from a total of 1260 males in eight ethnic groups of East Asia. We found four unique haplotypes constructed from three biallelic markers in these samples of East Asians. The Japanese population was characterized by a relatively high frequency of either the haplotype I-2b ( $-/Y2/T$ ) or II-1 ( $+/Y1/C$ ). These dual patterns of the distribution of Y chromosomes (I-2b/II-1) were also found in Korea, although they were present at relatively low frequencies. The haplotype II-1 was present in Northeast Asian populations (Chinese, Japanese, Koreans, and Mongolians) only, except for one male from the Thai population among the Southeast Asian populations (Indonesians, Philippines, Thais, and Vietnamese). The Japanese were revealed to have the highest frequency of this haplotype (27.5%), followed by Koreans (2.9%), Mongolians (2.6%), and mainland Chinese (2.2%). In contrast, the frequency of the haplotype I-2b was found to be 17.1% in the Japanese, 9.5% in Indonesian, 6.3% in Korean, 3.8% in Vietnamese, and 2.7% in Thai samples. These findings suggested that the chromosomes of haplotype I-2b were likely derived from certain areas of Northeast Asia, the region closest to Southeast Asia. Phylogenetic analysis using the neighbor-joining tree also reflected a general distinction between Southeast and Northeast Asian populations. The phylogeny revealed a closer genetic relationship between Japanese and Koreans than to the other surveyed Asian populations. Based on the result of the dual patterns of the haplotype distribution, it is

more likely that the population structure of Koreans may not have evolved from a single ancient population derived from Northeast Asians, but through dual infusions of Y chromosomes entering Korea from two different waves of East Asians.

**Key words** Y chromosome · *DYS287* · *DXYS5Y* · *SRY465* · *DYS19* · *DXYS156Y* · East Asians · Koreans

### Introduction

It is generally accepted that Koreans are considered a Northeast Asian group (Cavalli-Sforza et al. 1994). Although there are few archaeological data on the origins of modern Korean populations, most authors generally agree that the Korean peninsula has been occupied by humans since the Early Paleolithic age (Kim 1987). It has been suggested that the Korean population probably originated from the Tungus branch of Mongolian ethnic groups who inhabited the general area of the Altaic Mountains in Central Asia (Kim 1970). There is also evidence for a migration of Yayoi people from their original places via China and the Korean peninsula to Japan, starting around 2300 years ago (Chard 1974; Hanihara 1991). The Yayoi migration would lead us to expect a common genetic affinity in contemporary populations from Korea and Japan.

The Korean language is generally classified to the Tungus branch of the Altaic family: The Altaic family includes three subfamilies, Turkic, Mongolian, and Tungus (Ruhlen 1987). Some linguists separate Korean, Ainu, and Japanese into three isolated languages from this classification. Korean and Japanese are known to have remarkably common features in structure, vowel harmony, and lack of conjunctions in the language (Kim 1987). The close linguistic affinity suggests that Koreans and Japanese shared a common ancestral population.

Genetic studies of the variations of classic markers of protein and nuclear DNA show that Koreans tend to have a close genetic relationship with Mongolians among North-

W. Kim (✉) · D.J. Shin  
Department of Biology, Dankook University, Cheonan, Choong-Nam 330-714, Republic of Korea  
Tel. +82-417-550-3441; Fax +82-417-550-3441  
e-mail: wookkim@anse.dankook.ac.kr

S. Harihara  
Department of Biological Sciences, University of Tokyo, Tokyo, Japan

Y.J. Kim  
Department of Biology, Chungnam National University, Taejeon, Republic of Korea

east Asians (Goedde et al. 1987; Saha and Tay 1992; Hong et al. 1993). In contrast, genetic variation of mitochondrial DNA (mtDNA) indicates that the Koreans are not closely related to Mongolians, but are closely allied with Japanese and Chinese among several East Asian populations (Harihara et al. 1988; Horai et al. 1996; Hong et al. 1998). The discordance between classic genetic markers and mtDNA results in the genetic studies of the Korean population remains somewhat puzzling. Therefore further genetic markers are required to explore the origin and population history of modern Koreans.

The human Y-chromosome of the non-recombining portions has special features of a haploid and a father-to-son transmission pattern. The DNA sequence of these portions, therefore, contains a genetic record of the mutational events that occurred in their past. Haplotypes constructed with Y chromosomal alleles at multiple polymorphic loci can give us valuable information for inferring the paternal lineages and population history in humans (Ngo et al. 1986; Hammer 1994; Karafet et al. 1999).

The presence or absence of the Y *Alu* polymorphic (*YAP*) insertion (*DYS287*) has proved to be a useful marker for human population studies (Batzer and Deininger 1991; Perna et al. 1992; Hammer 1994, 1995). The *YAP* insertion is likely to have originated only once, so it has the feature of a unique polymorphism. Significant differences were detected among worldwide populations in the frequencies of the *YAP* insertion (Hammer 1994; Hammer et al. 1997, 1998). Based on Southern analysis, the probe p47z detects variations at loci on the short arm of Y (*DXYS5Y*) and the long arm of X (*DXYS5X*) chromosomes (Nakahori et al. 1989; Nakagome et al. 1992). Allelic variation of the *DXYS5Y* can now be determined by polymerase chain reaction (PCR) amplification and subsequent digestion with the *StuI* enzyme, identifying a two-allele variation, *Y1* and *Y2* (Shin et al. 1998). The *Y2* allele has been reported in samples from Japanese, Korean, and Taiwanese populations, but is absent in samples representing other Asian, and European and African populations (Nakagome et al. 1992; Lin et al. 1994; Hammer and Horai 1995; Shinka et al. 1999). Recently, a new biallelic polymorphism of the *SRY* (sex-determining region Y) gene, which is raised by a C-to-T transition at nucleotide position 465, was found in the Japanese and Korean populations (Shinka et al. 1999). Shinka et al. (1999) reported that most Caucasian and Negroid males examined in their survey completely lack the *T* allele of this *SRY* gene. The genetic analysis of this *SRY* gene, denoted *SRY465* (M.F. Hammer, personal communication) could be useful for investigating male-mediated gene flow and the relationship among East Asian populations.

The *DYS19* microsatellite polymorphism is due to variation in the number of GATA repeats (Roewer et al. 1992; Santos et al. 1993). There are at least nine different alleles, including five common alleles (*A-E*) that exhibit remarkable frequency differences in worldwide populations (Hammer and Horai 1995; Santos et al. 1996; Hammer et al. 1997). The distribution of highly polymorphic alleles of pentanucleotide repeat (TAAAA)<sub>n</sub>, *DXYS156*, which maps

to human X and Y chromosomes, has reflected significantly different patterns among different human ethnic groups (Chen et al. 1994; Salem et al. 1996; Kayser et al. 1997; Karafet et al. 1998). Significantly higher levels of *DXYS156Y* diversity were found in East Asian populations compared with populations from other regions in a global survey (Karafet et al. 1998).

Here, we report results for the haplotype frequency of a set of three biallelic markers (*YAP/DXYS5Y/SRY465*) and two microsatellite loci (*DYS19/DXYS156Y*) on the Y chromosome in several East Asian populations. The origin of modern Koreans and the genetic relationship between East Asian populations based on these Y-specific markers are also discussed.

---

## Subjects and methods

### DNA samples

We analyzed a total of 1260 unrelated males from eight East Asian populations. The eight East Asian ethnic groups in this study were divided into two large groups based on the nomenclature of Nei and Roychoudhury (1993) and Hammer et al. (1997), as shown in Table 1. DNA samples were obtained from volunteers living in Korea: 412 Koreans, 5 Han Chinese, 7 Japanese, 42 Indonesians (Java), 3 Philippines (Davao/Mindoro), 13 Thais (Bangkok), and 8 Vietnamese. We also isolated DNA samples from 64 Vietnamese volunteer donors living in Hanoi. The following samples, which were collected by the coauthors, were included in our survey: 128 Chinese (86, Beijing and 42, Xian), 212 Japanese (95, Ibaraki; 70, Tokushima; and 47, Yamaguchi) and 77 Mongolians (Buryats). In addition, DNA samples provided by other investigators were as follows: 38 Han Chinese and 39 Mongolians (Buryats), by C. Tyler-Smith; 9 Han Chinese and 6 Vietnamese, by D. Labuda; 32 Japanese (Oita) by Y. Tamaki; 105 Philippines (63 Tagalog and 42 Cebuano speakers) by J.J. Miranda; and 60 samples from Thais (Chiang Mai) by H. Steger. Genomic DNAs were prepared from whole blood by the standard method (Sambrook et al. 1989). Many of the DNA samples were also extracted from buccal cells according to the procedure of Richards et al. (1993).

### PCR and electrophoresis

The *YAP* element insertion was detected by PCR amplification using flanking primers as described in Hammer and Horai (1995). The *YAP*<sup>+</sup> (~450bp) or *YAP*<sup>-</sup> fragment (~150bp) of PCR products was electrophoresed in 2% agarose gel. The allelic variation of the *DXYS5Y* locus was detected by a PCR-restriction fragment length polymorphism (RFLP) method, as described by Shin et al. (1998). There were two copies of *DXYS5* amplified by PCR: one copy on the X and the other on the Y chromosome. In addition, the PCR product amplified from the *Y2* allele has only a recognition sequence for the restriction enzyme *StuI*

**Table 1.** Allele frequencies of five Y-chromosome-specific loci in eight East Asian populations

Population (n)	No. (%) of																						
	DXY55Y allele					SRY465 allele					DYS19 allele					DXYS156Y allele							
	Y1	Y2	Y1	Y2	C	T	Z	A	B	C	D	E	F	8	9	11	12	13	14	15	Null		
<b>Northeast Asians:</b>																							
Chinese (180)	176 (97.8)	180 (100.0)	0	0	167 (92.8)	13 (7.2)	0	1 (0.6)	18 (10.0)	61 (33.9)	89 (49.3)	10 (5.6)	1 (0.6)	6 (3.3)	2 (1.1)	24 (13.3)	55 (30.6)	52 (28.9)	14 (7.8)	0	0	27 (15.0)	
Japanese (251)	69 (27.5)	208 (82.9)	43 (17.1)	0	172 (68.5)	79 (31.5)	0	13 (5.2)	24 (9.6)	110 (43.8)	64 (25.5)	40 (15.9)	0	4 (1.6)	1 (0.4)	92 (36.7)	75 (29.9)	40 (15.9)	20 (8.0)	0	0	19 (7.6)	
Koreans (412)	12 (2.9)	400 (97.1)	386 (93.7)	26 (6.3)	292 (70.9)	120 (29.1)	0	12 (2.9)	76 (18.4)	180 (43.7)	120 (29.1)	23 (5.6)	1 (0.2)	0	0	65 (15.8)	178 (43.2)	116 (28.2)	49 (11.9)	4	4	0	
Mongolians (116)	3 (2.6)	113 (97.4)	116 (100.0)	0	113 (97.4)	3 (2.6)	0	0	54 (46.6)	20 (17.2)	34 (29.3)	7 (6.0)	1 (0.9)	0	0	37 (31.9)	68 (58.6)	8 (6.9)	2 (1.7)	0	0	1 (0.9)	
<b>Southeast Asians:</b>																							
Indonesians (42)	0	42 (100.0)	38 (90.5)	4 (9.5)	28 (66.7)	14 (33.3)	0	1 (2.4)	3 (7.1)	30 (71.4)	3 (7.1)	5 (11.9)	0	0	0	2 (4.8)	25 (59.5)	12 (28.6)	3 (7.1)	0	0	0	
Philippines (108)	0	108 (100.0)	108 (100.0)	0	107 (99.1)	1 (0.9)	3 (2.8)	3 (2.8)	16 (14.8)	48 (44.4)	30 (27.8)	8 (7.4)	0	0	0	4 (3.7)	53 (49.1)	42 (38.9)	5 (4.6)	2	2	2	
Thais (73)	1 (1.4)	72 (98.6)	71 (97.3)	2 (2.7)	68 (93.2)	5 (6.8)	1 (1.4)	3 (4.1)	19 (26.0)	33 (45.2)	10 (13.7)	7 (9.6)	0	0	0	3 (4.1)	55 (75.3)	6 (8.2)	9 (12.3)	0	0	0	
Vietnamese (78)	0	78 (100.0)	75 (96.2)	3 (3.8)	70 (89.7)	8 (10.3)	0	1 (1.3)	13 (16.7)	27 (34.6)	19 (24.4)	18 (23.1)	0	1 (1.3)	0	14 (17.9)	25 (32.1)	32 (41.0)	4 (5.1)	2	2	0	

Figures in parentheses in the Table body are percentages

(Fig. 1A). The C-to-T transition mutation at nucleotide position 465 of the *SRY* gene was determined by an allelic-specific PCR amplification using flanking primers (M.F. Hammer, personal communication): Upper primer, 5'-GGCACCTTTCAATTTTGTC-3'; assay primer, 5'-GTTGTCCAGTTGCACTTCA-3'; lower primer, 5'-CAGGCTCACTTCTGGATG-3'. The mutant *T* allele of the *SRY465* resulted in a 523-bp PCR product amplified by the assay and lower primers (Fig. 1B). In contrast, the wild-type *C* allele of the *SRY465* can be detected by amplification using the upper and lower primers (883-bp PCR product). The PCR reaction was performed in a total volume of 15 µl containing 25 ng of genomic DNA, 5 pM each primer, 0.2 mM dNTPs, 2.0 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM Tris-HCl (pH 8.3) and 1.0 U AmpliTaq DNA polymerase (Perkin-Elmer, Foster, CA, USA). The cycling conditions were as follows: 94°C for 2 min, and then 30 cycles at 94°C for 30 s, 54°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min.

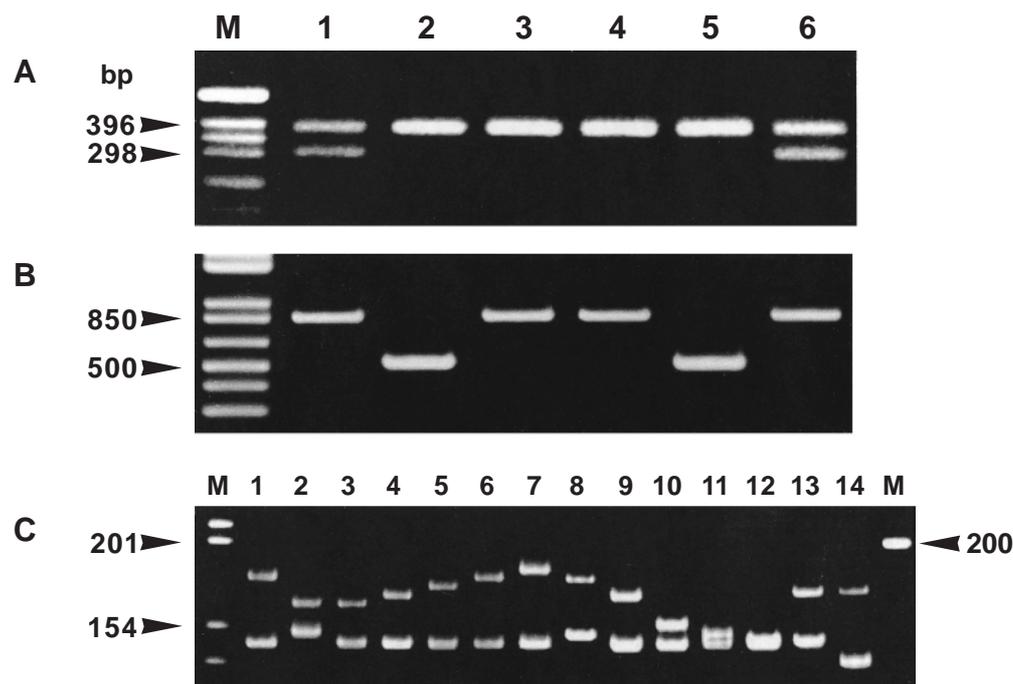
Allelic variation in the number of GATA repeats at the *DYS19* locus was analyzed as described by Hammer and Horai (1995). The PCR reaction and electrophoresis for the detection of *DXYS156Y* locus were performed as described in Karafet et al. (1998). We determined the alleles of the *DXYS156Y* locus by naming them according to the number of copies of the pentanucleotide repeat, using the control DNAs which were kindly supplied by M.F. Hammer and T. Karafet. The Y chromosome alleles were named *Y8-Y15* [ie, for (TAAAA)<sub>8</sub>-(TAAAA)<sub>15</sub>] and a null variant (*Y-null*) in this survey (Fig. 1C).

#### DNA sequencing of SRY465

We verified an *SRY465* mutation with DNA sequencing analysis. Each PCR product (Fig. 1B) of the *SRY465-C* and *T* allele was then further cloned by ligation into pCRII vector plasmids, using a TA cloning kit (Invitrogen, San Diego, CA, USA). Sequencing was accomplished by the chain termination method (Sanger et al. 1977), using a Sequenase kit (US Biochemical, Cleveland, OH, USA) according to the supplier's protocol.

#### Statistical analyses

The allele frequencies of five polymorphic Y-specific loci were calculated by counting genes from the observed phenotypes. The mean haplotype diversity was calculated for each group as  $h = n(1 - \sum x_i^2)/(n - 1)$ , where  $n$  represents the number of chromosomes sampled and  $x_i$  is the frequency of the  $i$ th haplotype (Nei 1987). The PHYLIP Package program (Felsenstein 1993) was analyzed to compute genetic distances for use in distance matrix programs and for constructing maximum likelihood phylogeny estimates based on Y chromosome haplotype frequencies. Genetic distance was estimated by Cavalli-Sforza's Chord distance ( $4\bar{D}$ ) (Cavalli-Sforza and Bodmer 1971). Neighbor-joining (NJ) (Saitou and Nei 1987) was used to construct branching diagrams from matrices of



**Fig. 1A–C.** Electrophoretic analyses of allelic classes at each of three Y-specific markers. **A** Detection of the allelic variation for the locus *DXY55Y* in polymerase chain reaction (PCR) products. Electrophoresis in a 2% agarose gel of restriction fragments of the PCR product produced by digestion with *Stu*I. Lane M, DNA size markers; lanes 1 and 6, Y2 allele; lanes 2–5, Y1 allele. **B** Detection of the *SRY465* variation in PCR products. Electrophoresis in a 2% agarose gel of amplified PCR fragments produced by allele-specific flanking primers

of *SRY465*. Lane M, DNA size markers; lanes 1, 3, 4, and 6, C allele; lanes 2 and 5, T allele. **C** Detection of *DXYS156Y* alleles on an 8% native polyacrylamide gel. Electrophoretic patterns of *DXYS156Y* and *DXYS156X* alleles in 14 males. Each lane shows the PCR product amplified with specific flanking primers from a single male. Lane M, DNA size markers; lanes 1 and 6, Y14/X7; lane 2, Y11/X8; lane 3, Y11/X7; lanes 4, 9, and 13, Y12/X7; lane 5, Y13/X7; lane 7, Y15/X7; lane 8, Y14/X8; lane 10, Y9/X7; lane 11, Y8/X7; lane 12, Y-null/X7; lane 14, Y12/X5

pairwise distances.

## Results

### Variation at the YAP (DYS287) locus

The *YAP* element was present at the highest frequencies in the Japanese (27.5%) and at much lower frequencies in the Koreans (2.9%), Mongolians (2.6%), and mainland Chinese (2.2%) (Table 1). The *YAP*<sup>+</sup> chromosomes were not found in the Southeast Asian populations surveyed here, except in one Thai male. These results are generally consistent with previous reports that showed the *YAP* element to be highly polymorphic in Japan and Tibet among Asian populations (Hammer 1994; Hammer and Horai 1995; Hammer et al. 1997). However, it should be noted that the *YAP*<sup>+</sup> chromosomes were also distributed at low frequencies in Northeast Asians in our survey (Table 1).

### Variation at the DXYS5Y locus

Japanese (17.1%) have the highest frequency of the Y2 allele, followed by Indonesians (9.5%), Koreans (6.3%), Vietnamese (3.8%), and Thais (2.7%) (Table 1). It has previously been reported that the Y2 chromosomes have a

limited distribution pattern. They are present in samples from populations of Japan, Korea, and Taiwan, but absent in samples representing other Asian and European, African, and Oceanian populations (Nakagome et al. 1992; Lin et al. 1994; Shinka et al. 1999). The frequency of the Y2 allele in Koreans (6.3%) found here was lower than those in Korean samples detected by Nakagome et al. (1992) and Shinka et al. (1999), which were 9.8% and 30.8%, respectively. In this study, the Y2 allele was also found in Southeast Asians, with the exception of Philippines, although it has been detected at a low frequency in their populations.

### Variation at the SRY465 locus

The *SRY465-T* allele was distributed in all the East Asian populations, although it was observed at extremely low frequency in the Mongolians (2.6%) and Philippines (0.9%) (Table 1). The *T* allele was frequent in our samples of the Indonesians (33.3%), Japanese (31.5%), and Koreans (29.1%). In addition, it was also observed in the Vietnamese (10.3%), Chinese (7.2%), and Thais (6.8%). The frequency of the *T* allele in our sample of Japanese (31.5%) is almost the same as that (32.1%) reported previously (Shinka et al. 1999). However, the *T* allele frequency in Koreans surveyed here (29.1%) is lower than that for Koreans (51.3%) reported by Shinka et al. (1999). The high

frequency of the *T* allele in their survey could be attributable to their small sample size (39 Koreans). They also found that most Caucasian and Negroid males examined in their survey completely lacked the *T* allele of the *SRY465*.

#### Variation at the DYS19 locus

In this survey, all five common alleles (*A–E*) and the two rare alleles (*Z* and *F*) were identified, as shown in Table 1. The *C* allele was frequent in most East Asian samples, whereas the *D* and *B* alleles appeared to be more common in the samples of Chinese and Mongolians, respectively (Table 1). This regional variation may be explained by isolation and genetic drift (Ciminelli et al. 1995). However, the predominance of the *C* allele in our samples of most East Asians is similar to that previously reported in other surveys of Asian populations (Hammer and Horai 1995; Santos et al. 1996; Hammer et al. 1997; Kim et al. 1998).

#### Variation at the DXYS156Y locus

A total of eight alleles of the *DXYS156Y* were detected in this study (Table 1). Although the highest frequency allele varied among different ethnic groups of East Asians, the *Y11–Y14* alleles were found at relatively high frequencies. Furthermore, the rare alleles, *Y8*, *Y9*, and *Y15*, were also observed at extremely low frequencies in limited populations of East Asians. A notable finding in our survey was that the Chinese and Japanese were revealed to have a null variant of the *DXYS156Y*, at frequencies of 15.0% and 7.6%, respectively (Table 1). The *X7* allele (140bp) is known to be X chromosome-specific, because it was usually present in two copies in females and in a single copy in males (Karafet et al. 1998). In fact, the *Y-null* variant appeared to have no copy from the PCR amplification of the *DXYS156* (Fig. 1C, lane 12). A single copy of the *X7* allele was only amplified from X chromosome, suggesting that the *Y-null* variant was not amplified from Y chromosome. The *Y-null* variant was usually detected in samples representing the *X7* allele.

#### DYS287/DXYS5Y/SRY465 combination haplotypes

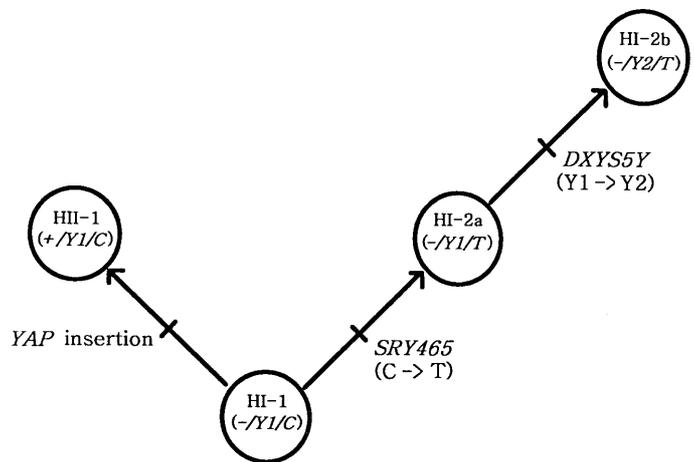
A total of four haplotypes were found with respect to the *DYS287*, *DXYS5Y*, and *SRY465* combination (Table 2). Significant regional differences in the distribution patterns of the haplotypes were observed among the ethnic groups of East Asia. The diversity values (*h*) were 0.174, 0.709, 0.482, 0.101, 0.501, 0.018, 0.157, and 0.192 for the Chinese, Japanese, Korean, Mongolian, Indonesian, Philippine, Thai, and Vietnamese samples, respectively. The highest diversity value was found in the Japanese sample (0.709). The Koreans (0.482) and Indonesians (0.501) represented moderate variability when compared with other East Asians surveyed here. Based on the result of the haplotype frequencies, a very low genetic diversity was reflected in the Chinese, Mongolian, Philippine, Thai, and Vietnamese

**Table 2.** A comparison of Y chromosome combination haplotype frequencies in eight East Asian populations

Population ( <i>n</i> )	No. (%) of individuals with haplotypes <sup>a</sup>			
	HI-1 (-/ <i>Y1</i> / <i>C</i> )	HI-2a (-/ <i>Y1</i> / <i>T</i> )	HI-2b (-/ <i>Y2</i> / <i>T</i> )	HI-1 (+/ <i>Y1</i> / <i>C</i> )
<b>Northeast Asians</b>				
Chinese (180)	163 (90.6)	13 (7.2)	0	4 (2.2)
Japanese (251)	103 (41.0)	36 (14.3)	43 (17.1)	69 (27.5)
Koreans (412)	280 (68.0)	94 (22.8)	26 (6.3)	12 (2.9)
Mongolians (116)	110 (94.8)	3 (2.6)	0	3 (2.6)
<b>Southeast Asians</b>				
Indonesians (42)	28 (66.7)	10 (23.8)	4 (9.5)	0
Philippines (108)	107 (99.1)	1 (0.9)	0	0
Thais (73)	67 (91.8)	3 (4.1)	2 (2.7)	1 (1.4)
Vietnamese (78)	70 (89.7)	5 (6.4)	3 (3.8)	0

Figures in parentheses in the Table body are percentages

<sup>a</sup>Combination haplotype: *YAP/DXYS5Y/SRY465*



**Fig. 2.** Y chromosome tree for the origin and evolutionary relationship of *YAP/DXYS5Y/SRY465* haplotypes. When the haplotype HI-1 is considered the probable ancestral, the haplotypes are created by mutation steps

samples (less than  $h = 0.200$ ).

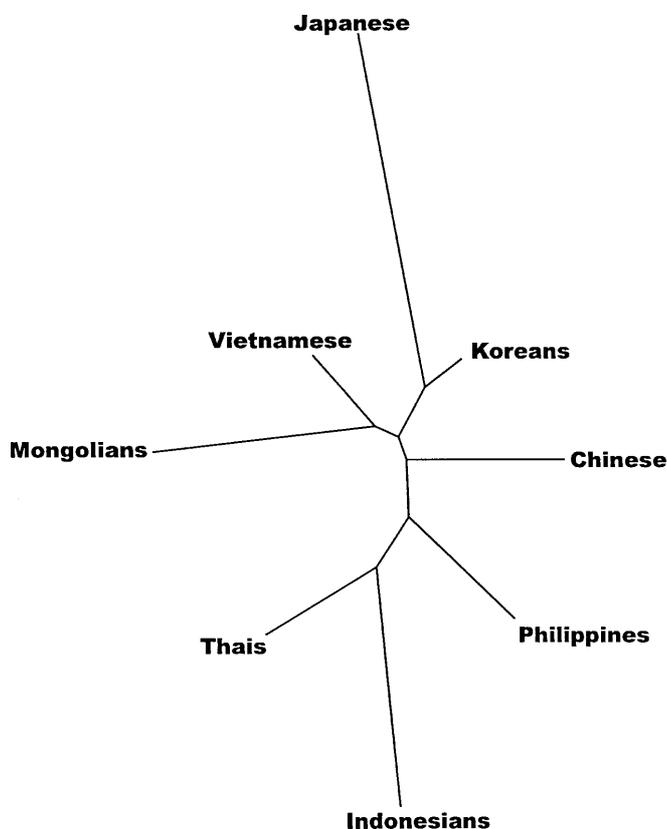
An obvious nonrandom association was found between the *DXYS5Y* and *SRY465*. The *Y2* allele is completely associated with the *T* allele (Fig. 2). The haplotype I-1 is the most frequent among East Asian populations, which is a common type anywhere in the world (Hammer et al. 1998). The Japanese and Korean populations appeared to have four unique haplotypes. Furthermore, the frequency of each haplotype was more evenly distributed in Japan. East Asian populations demonstrated a different pattern of haplotypes. The haplotype II-1 (+/*Y1*/*C*) was found only in the Northeast Asian populations, except for this finding in one male Thai from the Southeast Asian populations. The Japanese were revealed to have the highest frequency of the haplotype II-1 (27.5%), followed by Koreans (2.9%), Mongolians (2.6%), and mainland Chinese (2.2%). In contrast, the haplotype I-2b (-/*Y2*/*T*) was mostly distributed among the Japanese (17.1%), Indonesian (9.5%), Korean (6.3%), Vietnamese (3.8%), and Thai samples (2.7%). Therefore the Japanese population was characterized by a

relatively high frequency of either the haplotype I-2b or II-1 (Table 2). These dual patterns of infusions of Y chromosomes (I-2b/II-1) were also found in Korea, although they were present at relatively low frequencies.

#### Population tree analysis of Y chromosome halotypes

We used the neighbor-joining (NJ) tree to analyze the Cavalli-Sforza's Chord genetic distance resulted in a total of 72 combination haplotypes (data not shown). Phylogenetic analysis, based on the frequencies of the 72 combination haplotypes also reflected a considerable distinction between Southeast and Northeast Asian populations (Fig. 3). A notable exception to this pattern was the finding of the Vietnamese population. They were more likely clustered with Northeast Asians (e.g., Koreans), probably because of distinct founder populations and genetic drift. The Japanese population appeared to have a closer genetic relationship to Koreans than to the other surveyed Asian populations.

#### Disucussion



**Fig. 3.** Neighbor-joining (NJ) tree for eight East Asian populations. This network was produced by the NJ method (Saitou and Nei 1987), based on Cavalli-Sforza's Chord genetic distance ( $4\bar{D}$ ) (Cavalli-Sforza and Bodmer 1971). The genetic distances are proportional to the branch lengths

The pattern of Y chromosome variation presented here provides convincing genetic evidence for significant differences among East Asian populations. The haplotype II-1 (+/Y1/C) was generally found in Northeast Asian ethnic groups (Table 2). In contrast, the  $YAP^+$  chromosomes were not found in Southeast Asian populations surveyed here, except in one Thai male. The Thai carrying the  $YAP^+$  chromosome is likely due to Northeast Asian paternal admixture. These results are consistent with earlier reports showing that most East Asian ethnic groups have very low frequencies of the  $YAP^+$  chromosome, with the exception of the high frequencies of  $YAP^+$  chromosomes in Japan and Tibet. The presence of  $YAP^+$  alleles restricted to Northeast Asian ethnic groups lends support to the hypothesis of the northern origin of  $YAP^+$  chromosomes (Hammer et al. 1997, 1998).

The Japanese and Korean populations, however, reflected an infusion of different Y chromosome lineages from new waves of immigrants, which has resulted in an increase of haplotype I-2a (-/Y1/T) or I-2b (-/Y2/T) (Table 2). Recently, Shinka et al. (1999) reported that the  $T$  allele of the  $SRY465$  was not found in most Caucasian and Negroid males examined in their survey. Based on the frequency and distribution of the  $SRY465-T$  allele, it is inferred that these Y chromosomes originated in East Asia. Subsequently, a recent wave of immigrants carrying the  $SRY465-T$  allele from their original places has resulted in the predominant distribution of the  $SRY465-T$  chromosomes in Japan and Korea. Hammer and Horai (1995) inferred that the  $YAP$  element and  $Y2$  alleles were markers of Jomon and Yayoi male lineages, respectively. The chromosomes carrying the haplotype I-2b derived from the haplotype I-2a are direct descendants of the haplotype I-2a (Fig. 2). Therefore, it is implied that the Yayoi people carrying the haplotype I-2b were more recently introduced than those with the haplotype I-2a into Korea and Japan. Our results are also consistent with the hypothesis that the haplotype I-2b originated in China and tracks paternal migrations to Japan during the Yayoi period (Nakagome et al. 1992; Lin et al. 1994; Hammer and Horai 1995; Hammer and Zegura 1996). The high frequency of the  $Y2$  allele in Japan may be explained by the founder effect. The high Japanese diversity levels (0.709) in this survey provide evidence for the hybridization theory resulting from a combination of an ancient population (Jomon people) subject to a more recent admixture (Yayoi people) (Hanihara, 1991; Hammer and Horai 1995; Omoto and Saitou 1997). The moderate Korean variability (0.482) may be due to the homogenizing influence of migrations from East Asia.

The distribution patterns of the haplotype I-2b frequency would lead us to expect that this haplotype was likely derived from central or south China an area which is remote from north China and Southeast Asia. Recently, Lin et al. (1994) suggested that the  $Y2$  allele may have originated from an ancestral population of Henan in central east China. Melton et al. (1998) hypothesized a scenario for the early colonization of Taiwan and northeastern coastal migrations in Asia. They theorized that, during 6000–4000 BC, Neolithic proto-Austronesian speakers, spreading from

early centers of rice cultivation in central and south China, expanded to coastal China and across the Formosa Straits. This range expansion also occurred to a certain extent into the northern regions of China. Meanwhile, proto-Austronesian languages were subsequently obliterated in mainland Asia. Our findings of the distribution pattern of the Y2 allele are consistent with the suggestion of Melton et al. (1998). The occurrence of the Y2 allele varies somewhat with distance from the coastal areas of central China, with the exception of findings in Indonesia (Table 2). The relatively moderate frequency of the Y2 allele in Indonesians (9.5%) may have been caused by a bottleneck effect. Based on our results of the 72 combination haplotypes (data not shown), all of the Southeast Asian samples carrying the Y2 allele were found to be the  $-Y2/T/C/12$  haplotype, with the exception of that in one Vietnamese male ( $-Y2/T/D/12$ ). This  $-Y2/T/C/12$  haplotype was also present at the highest frequency among the chromosomes of the Y2 allele in Japan and Korea. These results implied that most East Asians carrying the Y2 allele originated from common ancestors. Additional genetic markers and populations from diverse regions in Asia (e.g., China) are required to verify this hypothesis.

On the other hand, the haplotype I-1 is the most frequent among East Asian populations, and seems to be an ancestral type anywhere in the world (Hammer et al. 1998). For example, the high frequency of the haplotype I-1 indicates that very few chromosomes were differentiated in the Chinese, Mongolian, Philippine, and Thai populations (<90%) (Table 2): They are populations in which most of their Y chromosomes may not be discriminated. Therefore we need to use more biallelic Y-markers to differentiate more chromosomes in East Asia.

Phylogenetic analysis also reflected a considerable difference between Southeast and Northeast Asian populations (Fig. 3). A notable exception to this pattern was the finding in the Vietnamese. They were more likely clustered with Northeast Asians, probably because of distinct founder populations and genetic drift. An alternative explanation for this unexpected affinity is that a recent range expansion from south China resulted in this phylogenetic pattern in Vietnam. In this survey, most of the Vietnamese sample was derived from Hanoi, the region closest to south China. Interestingly, phylogeny based on 30 microsatellites has revealed that Koreans were closely related to two Chinese populations, from Manchu and Yunnan (Chu et al. 1998). Yunnan is in southwest China, bordering on Vietnam, Laos, and Burma. The relatively high prevalence of the *DXYS156Y-null* variant in a sample of northeast Chinese implied that these northern Chinese regional populations may be excluded from being significant sources of Koreans (Table 1). The phylogeny also indicated that the Japanese population appeared to have a closer genetic relationship to Koreans than to the other Asian populations surveyed. This result is consistent with earlier surveys that showed the closest genetic affinity between the mainland Japanese and Koreans (Hammer and Horai 1995; Horai et al. 1996; Kim et al. 1998). The Korean population, however, was found to have a

more homogeneous pattern of genetic structure than the Japanese population, perhaps under the influence of different sizes of founding populations and range expansions.

**Acknowledgments** We would like to thank M.F. Hammer for providing unpublished information and advice and T.M. Karafet for sharing her data during the course of this work. We especially thank Y.B. Park and F.R. Santos for their suggestions and comments on the manuscript.

This study was supported by the Basic Science Research Institute Program (BSRI-97-No. 4440) of the Ministry of Education, Republic of Korea.

## References

- Batzer MA, Deininger PL (1991) A human-specific subfamily of *Alu* sequences. *Genomics* 9:481–487
- Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. WH Freeman, San Francisco
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chard C (1974) Northeast Asia in prehistory. University of Wisconsin Press, Madison
- Chen H, Lowther W, Avramopoulos D, Antonarakis SE (1994) Homologous loci *DXYS156X* and *DXYS156Y* contain a polymorphic pentanucleotide repeat (TAAAA)<sub>n</sub> and map to human X and Y chromosomes. *Hum Mutat* 4:208–211
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763–11768
- Ciminelli BM, Pompei F, Malaspina P, Hammer MF, Persichetti F, Pignatti PF, Palena A, Anagnou N, Guanti G, Jodice C, Terrenato L, Novelletto A (1995) Recurrent simple tandem repeat mutations during human Y chromosome radiation in Caucasian subpopulations. *J Mol Evol* 41:966–973
- Felsenstein J (1993) PHYLIP: Phylogeny Inference Package, Version 3.5p Joseph Felsenstein and the University of Washington, Seattle
- Goedde HW, Paik YK, Lee CC, Benkmann HG, Kriese L, Bogdanski P, Winkler M (1987) Red cell and serum protein polymorphisms in three population groups of South Korea. *Gene Geogr* 1:177–188
- Hammer MF (1994) A recent insertion of an *Alu* element on the Y chromosome is a useful marker for human population studies. *Mol Biol Evol* 11:749–761
- Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376–378
- Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15:427–441
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. *Genetics* 145:787–805
- Hammer MF, Zegura SL (1996) The role of the Y chromosome in human evolutionary studies. *Evol Anthropol* 5:116–134
- Hanihara K (1991) Dual structure model for the population history of the Japanese. *Japan Review* 2:1–33
- Hanihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, Ellepola SB, Ishida T, Omoto K (1988) Mitochondrial DNA polymorphism among five Asian populations. *Am J Hum Genet* 43:134–143
- Hong SS, Horai S, Lee CC (1998) Distribution of the 9-bp deletion in COII/tRNA<sup>Lys</sup> intergenic region of mitochondrial DNA is relatively homogeneous in east Asian populations. *Korean J Biol Sci* 2:259–267
- Hong SS, Suh JA, Chae JJ, Goh SH, Kim YS, Kim UK, Namkoong Y, Lee CC (1993) Frequency distribution of alleles at *D1S80* and *apo-B* 3' VNTR loci in Korean population. *Mol Cells* 3:457–453
- Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Hanihara S, Park KS, Omoto K, Pan IH (1996)

- mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59:579–590
- Karafet T, Knijff PD, Wood E, Ragland J, Clark A, Hammer MF (1998) Different patterns of variation at the X- and Y-chromosome-linked microsatellite loci *DXYS156X* and *DXYS156Y* in human populations. *Hum Biol* 70:979–992
- Karafet T, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, Knijff PD, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831
- Kaysner M, Caglia A, Corach D, Fertwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Knijff PD, Roewer L (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–133
- Kim K (1970) People and language. In Korea — its people and cultures. Hakwon-sa, Seoul
- Kim NK (1987) The world's major languages. Croom Helm, London
- Kim W, Shin DJ, You SA, Kim YJ (1998) Y-specific DNA polymorphisms of the *YAP* element and the locus *DYS19* in the Korean population. *J Hum Genet* 43:195–198
- Lin SJ, Tanaka K, Leonard W, Gerelsaikhan T, Dashnyam B, Nyamkhishig S, Hida A, Nakahori Y, Omoto K, Crawford MH, Nakagome Y (1994) A Y-associated allele is shared among a few ethnic groups of Asia. *Jpn J Hum Genet* 39:299–304
- Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet* 63:1807–1823
- Nakagome Y, Young SR, Akane A, Numabe H, Jin DK, Yamori Y, Seki S, Tamura T, Nagafuchi S, Shiono H, Nakahori Y (1992) A Y-associated allele may be characteristic of certain ethnic groups in Asia. *Ann Hum Genet* 56:311–314
- Nakahori Y, Tamura T, Yamada M, Nakagome Y (1989) Two 47z [*DXYS5*] RFLPs on the X and the Y chromosome. *Nucleic Acids Res* 17:2152
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 10:927–943
- Ngo KY, Vernaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am J Hum Genet* 38:407–418
- Omoto K, Saitou N (1997) Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am J Phys Anthropol* 102:437–446
- Perna NT, Batzer MA, Deininger PL, Stoneking M (1992) *Alu* insertion polymorphism: a new type of marker for human population studies. *Hum Biol* 64:641–648
- Richards B, Skoletsky J, Shuber AP, Balfour R, Stern RC, Dorkin HL, Parad RB, Witt D, Klinger KW (1993) Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2:159–163
- Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* 89:389–394
- Ruhlen M (1987) A guide to the world's languages. Stanford University Press, California
- Saha N, Tay JSH (1992) Origin of the Koreans: a population genetic study. *Am J Phys Anthropol* 88:27–36
- Saitou N, Nei M (1987) The neighbor-joining methods: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Salem AH, Badr FM, Gaballah MF, Paabo S (1996) The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am J Hum Genet* 59:741–743
- Sambrook F, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ (1996) Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism *DYS19*. *Hum Genet* 97:309–313
- Santos FR, Pena SDJ, Epplen JT (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum Genet* 90:655–656
- Shin DJ, Kim YJ, Kim W (1998) PCR-based polymorphic analysis for the Y chromosome loci *DYS19* and *DXYS5Y* (47z) in the Korean population. *Korean J Biol Sci* 2:281–285
- Shinka T, Tomita K, Toda T, Kotliarova SE, Lee JW, Kuroki Y, Jin DK, Tokunaga K, Nakamura H, Nakahori Y (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