

A preliminary study on the origin of Koreans based on Y-STR variation

Soon Hee Kim · Byung Won Chun · Jongwoo Jung · Brian M. Kemp · Kyoung Don Kwak · Nam Soo Cho · Jong Jin Kim · Myun Soo Han · Won Kim

Received: 19 March 2010 / Accepted: 11 May 2010 / Published online: 31 August 2010
© The Genetics Society of Korea and Springer 2010

Abstract

To investigate the origin of Koreans, we examined the 12-locus Y-chromosome short tandem repeat (Y-STR) variation in a sample of 310 unrelated males from three localities (Gochang, Andong and Geoje) in Korea and statistically analyzed the previously published four Y-STR databases ($n = 1655$) of Korean population. The median joining network of 9-locus Y-STR haplotypes inferred as haplogroup O2b-SRY₊₄₆₅ showed a “star cluster” indicative of a population expansion from a centrally positioned haplotype. The central haplotype in the “star cluster” was the most frequently occurring Y-STR haplotype among the Korean male gene pool (6%, 127 of 1965, 10,14,12,13,14,16,13,13,23, for loci DYS391, DYS389I, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, and DYS390), which was shared among all seven datasets. Based on the “star cluster” pattern from both our data (41%, 128 of 310) and those previously published (34%, 563 of 1655), we suggest that the most frequent Y-STR haplotype among the Korean male gene pool seems to be the Korean modal (ancestral) haplotype. Further study with additional Y-STR and Y-SNP data of the east Asian populations as well as Korean

population are needed to providing a genetic clue for the “star cluster” (O2b-SRY₊₄₆₅) associated with the ethnohistoric events of the Koreans.

Keywords Y-STRs; Korean population; Origin of Koreans; Star cluster; Modal (ancestral) haplotype

Introduction

Genetic, archeological and linguistic (Altaic language) evidences support the hypothesis that the proto-Koreans originated through the northern Asian-route from Mongolians who inhabited most likely in the Altai Mountain and Lake Baikal area of south-central Siberia (Saha and Tay, 1992; Han, 1995; Ruhlen, 1991). According to the mythological legend of the national foundation, a legendary king *Dangun* founded *Gojoseon* (the Old *Joseon* or *Dangun Joseon*) in 2333 BC (Ilyeon, 2007). Archeological evidences such as dolmens and bronze daggers support the view that *Gojoseon* was established in northeast region of China (contemporarily known as Manchuria) and the northern Korean peninsula. The Korean peninsula has been inhabited by the mankind (*Homo erectus*) since some Paleolithic Age (~700000 BC). However, the direct ancestor of modern Koreans (*Homo sapiens*) has been considered to be the proto-Koreans who inhabited northeast region of China and the Korean peninsula during the Neolithic (8000–1000 BC) and the Bronze Age (1500–400 BC) (National Institute of Korean History, 2004).

Studies of Y-chromosome variation have substantially contributed to forensics and population genetics (Jobling et al., 1997; de Knijff et al., 1997) but also our understanding of human origins and migration routes (Jin and Su, 2000; Xue et al., 2006). It has been known that the Y-chromosome (NRY) is inherited patrilineally in a haploid mode without recombina-

S.H. Kim · B.W. Chun · K.D. Kwak · N.S. Cho · J.J. Kim · M.S. Han
DNA Analysis Division, National Institute of Scientific
Investigation, Seoul 158-707, Korea,

W. Kim(✉) · S.H. Kim
School of Biological Sciences, Seoul National University, Seoul
151-742, Korea,
e-mail: wonkim@plaza.snu.ac.kr

J.W. Jung
Department of Science Education, Ewha Womans University,
Seoul 120-750, Korea,

B.M. Kemp
Department of Anthropology and the School of Biological Sciences,
Washington State University, Pullman, WA 99164-4236, USA.

nation. Therefore, the genetic variation of NRY is composed of the mutations accumulated from past generations. Thus, Y-chromosomal variation can provide information concerning the genetic history of a patrilineal population. Over the past 10 years, Y-chromosomal variation based on single nucleotide polymorphism (SNP) and/or short tandem repeat (STR) markers in the NRY have been analyzed in Korean and east Asian populations, suggesting that the Korean population is composed of complex components originated from northeast and southeast Asia (Kim et al., 2000; Jin et al., 2003; Kwak et al., 2005; Jin et al., 2009). These studies support the bi- and/or multidirectional route model. However, it is necessary to trace ancestral Y-chromosome lineages from the Korean gene pool in addition to these route hypotheses.

Since the 1960s Korea has experienced a continuous flow of internal migration primarily to the major cities from rural areas and between these major cities due to rapid progress of industrialization. Approximately 48% of the Korean population (about 23 million, according to the 2005 census) are concentrated in the capital region (Seoul, Incheon, and Gyeonggi province), and another 21% of Korean people reside in the five other metropolitan cities (Busan, Daejeon, Daegu, Gwangju, and Ulsan). Therefore, an extreme bias in regional distribution exists between the few densely-populated metropolitan cities and a large number of less-populated regions. To trace the ancestry from the Korean population, first of all, it is needed to investigate the genetic structure of the Korean population caused by recent migration. Previous studies of the Korean gene pool (Korean Y-STR databases) (Kim et al., 2008; Chun et al., 2005; Park et al., 2005; Hwang et al., 2007) were based on samples mainly obtained from densely-populated areas, which neglected potential differences of genetic structure of the Korean population expected from the regional bipolarization. Among the less-populated regions, the de-populated areas, which are characterized by a negative migration rate, would be expected to preserve more their original genetic structure continued from the past because the population inflow from other regions is very little. Therefore, sampling from de-populated areas provides a better opportunity to reveal patrilineal ancestor within the limited territory of current South Korea.

We examined two localities to be historically significant (Gochang and Andong) among the de-populated areas in Korea (Fig. 1). Gochang, which is located in the southwest part of Korea, is representative of the Korean prehistoric period, because the dolmens (tombs in *Gojoseon* and the Bronze Age) are most densely distributed in this area. In general, more than 50% of the world's dolmens are distributed in South/North Korea, with the dolmen site of Gochang registered as a world cultural heritage site. Andong, which is located in the eastern

part of Korea, is representative of the Korean historic period (especially the *Joseon* Dynasty from 1392–1910). It is one of the most famous historical cities which the Korean tradition has been quite well conserved, and it is particularly named as the home of Korean Confucianism. Lastly, we chose Geoje located in the southeast part of Korea because it is the largest island near the mainland of Korea and, thus, would be expected to show some degree of genetic drift after possible gene influx from the mainland.

In this study, we generated 12-locus Y-STR haplotypes from 310 unrelated males in Gochang, Andong, and Geoje and compared the results with those of the previously published Y-STR databases of the Korean population ($n = 1655$). We attempted to trace the ancestry based on the individual Y-STR variation.



Figure 1. Geographic location of the three localities (Gochang, Andong, and Geoje) analyzed in this study. A few densely-populated major cities (the capital region [Seoul, Incheon, and Gyeonggi] and metropolitan cities [Busan, Daegu, Gwangju, Daejeon, and Ulsan]) are denoted in gray or black. Abbreviations are as follows: GYG, Gyeonggi province; GAW, Gangwon province; NCC, North Chungcheong province; SCC, South Chungcheong province; NJL, North Jeolla province; SJL, South Jeolla province; NGS, North Gyeongsang province; SGS, South Gyeongsang province; JEJ, Jeju.

Table 1. Diversity of twelve-locus Y-STR haplotypes in the groups A and B.

Population	No. of different haplotypes	No. of unique haplotypes	HD	DC (%)	Sources
Group A					
Gochang (<i>n</i> = 115)	101	90	0.9974	87.83	This study
Andong (<i>n</i> = 97)	83	74	0.9946	85.57	This study
Geoje (<i>n</i> = 98)	79	67	0.9941	80.61	This study
Total (<i>n</i> = 310)	(246) ^a	(213) ^a	(0.9968) ^a	(79.35) ^a	
Group B					
Korean1 (<i>n</i> = 526)	420 ^b	359 ^b	0.9985 ^b	79.85 ^b	Kim et al., 2008
Korean2 (<i>n</i> = 259)	230 ^d	213 ^d	0.9985 ^d	88.80 ^d	Chun et al., 2005
Korean3 (<i>n</i> = 301)	262 ^c	239 ^c	0.9983 ^c	87.04 ^c	Park et al., 2005
Korean4 (<i>n</i> = 569)	473	426	0.9985	83.13	Hwang et al., 2007
Total (<i>n</i> = 1655)	(1179) ^c	(987) ^c	(0.9985) ^c	(71.24) ^c	

HD Haplotype diversity, DC Discrimination capacity.

^a Statistical values were calculated from a total of 310 males.

^{b,c} Statistical values were calculated for 12-locus Y-STR haplotypes in this study.

^d Statistical values recalculated in this study were different from the previously published values (Chun et al., 2005).

^e Statistical values were calculated from a total of 1655 males.

Materials and Methods

DNA samples and Y-STR genotyping

After obtaining informed consent and information about the birthplace of the donor, we examined Y-chromosomes of 310 unrelated males from three localities in Korea (Fig. 1): Gochang county (*n* = 115), Andong city (*n* = 97), and Geoje city (*n* = 98). Genomic DNA was isolated from buccal swabs or blood samples with the QIAamp DNA mini kit (Qiagen, Germany). Twelve Y-STRs were typed using the PowerPlex® Y System (Promega, USA) following the manufacturer's instructions: DYS391, DYS389I, DYS439, DYS389II, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, and DYS385a/b. PCR products were electrophoresed on a 310 Genetic Analyzer (Applied Biosystems, USA). The Y-STR allele designation was based on the number of repeat motifs according to the guidelines of the International Society of Forensic Genetics (Gusmão et al., 2006). Proficiency testing of Collaborative Testing Service (CTS) and the Y-chromosome Haplotype Reference Database (YHRD) (<http://www.yhrd.org/>) were carried out for quality control.

Statistical analyses

In addition to the data generated in this study, 12-locus Y-STR haplotypes of 1655 males from Korean population were retrieved from the literature (Kim et al., 2008; Chun et al., 2005; Park et al., 2005; Hwang et al., 2007). The datasets of three localities are referred as “group A”, being derived from less-populated areas. The previously published Y-STR databases

of the Korean population were mainly derived from major cities of large population size and we refer them as “group B” likely representing populous areas in Korea.

Haplotype diversity was calculated by the formula $H = n/(1 - \sum x_i^2)/(n - 1)$, where *n* is the number of individuals, and *x_i* is the frequency of each haplotype (Nei, 1987). Discrimination capacity was determined by dividing the number of different haplotypes seen in a given population by the total number of individuals. Shared haplotypes within population were used to determine the degree of the genetic differentiation. Analysis of molecular variance (AMOVA) based on 10-locus Y-STRs (excluded for DYS385a/b) was carried out with Arlequin v3.0 software (Excoffier et al., 2005).

A median joining (MJ) network (Bandelt et al., 1999) was constructed from 9-locus Y-STR data (excluded for DYS389II and DYS385a/b) using the Network 4.5.1.0 (<http://www.fluxus-engineering.com/sharenet.htm>). MJ network was constructed by weighting 13 (DYS438), then 12 (DYS392, DYS393); 11 (DYS437); 9 (DYS389I, DYS19, and DYS390); 8 (DYS391) and 1 (DYS439) (Decker et al., 2008). The DYS389II locus was omitted because the loci were interdependent, a mutation in segment DYS389I automatically created a change in the total length, expressed by DYS389II. The bilocal DYS385a/b was excluded from this analysis because the two alleles were not assigned to individual loci.

Results and Discussion

A total of 246 haplotypes were obtained from 310 unrelated

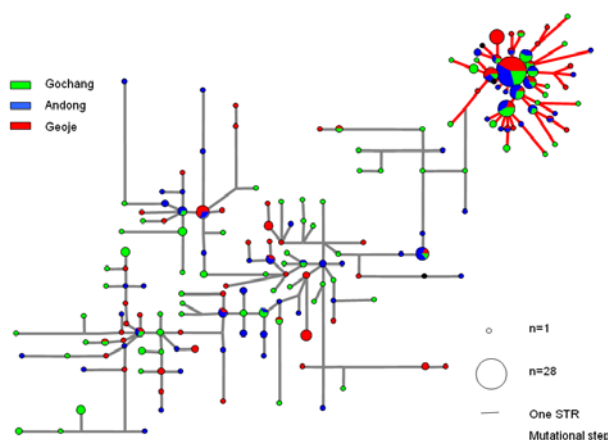


Figure 2. MJ network based on the nine-locus Y-STR (DYS391, DYS389I, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, and DYS390) haplotypes of the group A ($n = 310$). Group A haplotypes include those from Gochang, Andong and Geoje. Circles represent Y-STR haplotypes with the area proportional to frequency and colored according to different populations. The complex reticulations were removed to simplify MJ network result.

males of the group A by analyzing with 12-locus Y-STR. The haplotypes generated in this study have been submitted to YHRD (Gochang County, YA003646; Andong City, YA003644; Geoje City, YA003645). Overall haplotype diversity of the group A ($HD = 0.9968$) was slightly lower than that of the group B ($HD = 0.9985$) (Table 1). Gene flow caused by recent population movement to the major cities as well as large population size probably added further haplotype diversity to the group B. Among 246 different haplotypes, most haplotypes (87%, 213 of 246) were present in a single individual and the remaining 33 frequent haplotypes were shared within, between, and among three local populations (Table 2).

Matches of the 33 frequent haplotypes to the four Korean Y-STR databases and the YHRD 3.0 (Release 29) (Willuweit et al., 2007) are also listed in Table 2. Among the 33 frequent haplotypes, nine (Ht1–Ht9), four (Ht10–Ht13), and seven (Ht14–Ht20) haplotypes were specific to Gochang, Andong and Geoje, respectively, whereas, two (Ht21 and Ht22), two (Ht23 and Ht24), and five (Ht25–Ht29) haplotypes were shared between Gochang–Andong, Gochang–Geoje, and Andong–Geoje, respectively (Table 2). Lastly, four (Ht30–Ht33) haplotypes were shared among Gochang–Andong–Geoje. Three (Ht30–Ht32) of the four 12-locus Y-STR haplotypes shared among Gochang–Andong–Geoje belonged to the most frequent 10-locus Y-STR haplotype (with exception of DYS385a/b) in this study (Table 2). These haplotypes were one-step or two-step neighbors with the only difference among them at DYS385a/b (having highest mutation rate among 12-locus Y-STR).

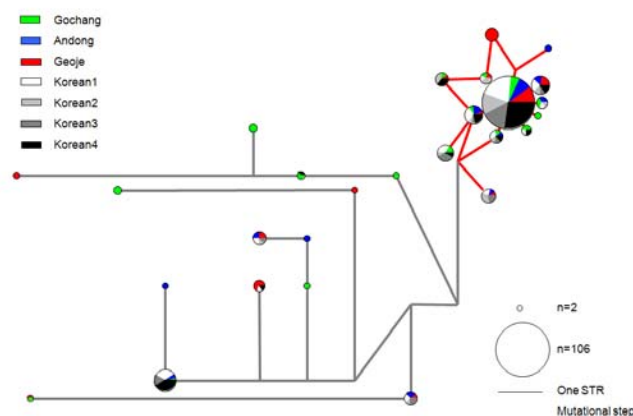


Figure 3. MJ network based on the 33 frequent nine-locus Y-STR (DYS391, DYS389I, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, and DYS390) haplotypes of the males (total, $n = 259$: group A, $n = 97$; group B, $n = 162$) from the group A and their matched haplotypes to group B as shown in Table 2.

In general, the high levels of genetic differentiation or variation among three local populations in Korea (that is, the high frequency of population specific haplotypes [Ht1–Ht20]) may be caused by genetic drift, which is characteristic of relatively small-sized populations (Karafet et al., 2002; Zhou et al., 2008). However, in this study, in spite of the high levels of genetic differentiation, the group A also displayed unusually high levels of genetic homogeneity (that is, the high frequency of shared haplotypes between and among the three local populations [Ht21–Ht33]). AMOVA confirmed the genetic homogeneity among the group A samples ($R_{ST} = -0.00233$, P value = 0.71950). Dispersal of Y-STR based closely related lineages in Korea might be one possible cause for the above findings.

We constructed a MJ network among the 9-locus Y-STR variation (excluded for DYS389II and DYS385a/b) of the group A dataset (Fig. 2) to test this possibility. The resulting MJ network displayed a “star cluster (i.e., a cluster of closely related lineages)” (upper right part in Fig. 2), where all haplotypes existed as 1–5 mutational steps from the central haplotype with a frequency of 41% (128 of 310). In contrast, the remaining network haplotypes were relatively complex and connected loosely with a frequency of 59% (182 of 310).

We also constructed a MJ network among the variation of the 33 frequent 9-locus Y-STR haplotypes (total, 259 of 1965: group A, 97 of 310; group B, 162 of 1655) in Table 2, to simply visualize the correlation between the groups A and B (Fig. 3). The MJ network formed the core “star cluster”, where all haplotypes showed 1–3 mutational steps from the central haplotype which was shared among all seven Korean datasets. Therefore, we can infer that the haplotypes of the core “star cluster” have dispersed throughout different parts of the peninsula.

Table 2. Thirty three frequent twelve-locus Y-STR Haplotypes ($n = 97$) shared within, between, and among the Group A and their match to the Group B and YHRD.

		No. of frequent Y-STR haplotypes (No. of males) in :							
Ht#	Y-STR Haplotype (Inferred O2b-SRY ₄₆₅ haplogroup)	Group A			Group B				YHRD (n=23780)
		Gochang (n=115)	Andong (n=97)	Geoje (n=98)	Korean 1 (n=526)	Korean 2 (n=259)	Korean3 (n=301)	Korean4 (n=569)	
Ht1:	10,12,12,27,11,15,14,13,12,24,13-15	(2)							
Ht2:	10,13,12,30,10,14,15,11,14,23,11-17	(2)							
Ht3:	10,14,11,30,10,14,15,11,14,23,11-20	(2)						(1)	
Ht4:	10,14,12,29,13,14,16,13,14,23,10-17 (O2b)	(2)							
Ht5:	10,14,12,30,13,14,15,13,13,22,10-18 (O2b)	(2)			(4)	(1)	(4)	(1)	(8)
Ht6:	10,14,13,29,13,14,16,13,13,23,10-18 (O2b)	(2)			(1)			(1)	(1)
Ht7:	11,14,12,29,13,14,16,13,13,23,10-17 (O2b)	(2)				(1)			(1)
Ht8:	10,14,11,29,10,14,15,11,15,24,10-20	(3)							
Ht9:	11,13,13,30,10,14,17,11,13,25,13-18	(3)							
Ht10:	10,12,12,28,11,15,14,14,12,24,12-18		(2)						(1)
Ht11:	10,13,12,29,10,15,15,12,12,23,12-17		(2)						
Ht12:	10,14,12,30,12,14,17,13,13,23,9-18 (O2b)		(2)						
Ht13:	11,14,12,29,13,14,16,13,13,23,10-10 (O2b)		(2)		(1)				(1)
Ht14:	10,12,10,29,9,15,15,13,12,24,12-19			(2)	(1)			(1)	
Ht15:	10,12,10,30,9,15,15,13,12,24,12-19			(2)					
Ht16:	10,12,13,27,10,14,17,13,12,25,14-18			(2)					
Ht17:	10,14,12,30,13,14,16,13,13,23,10-18 (O2b)			(2)				(2)	
Ht18:	11,14,10,30,10,14,15,13,14,25,12-19			(2)					
Ht19:	10,14,12,30,13,14,17,13,13,24,10-18 (O2b)			(3)					
Ht20:	10,14,12,29,13,14,17,13,13,24,10-18 (O2b)			(4)					
Haplotypes shared within each local population		20 (45)			7 (19)				5 (12)
Ht21:	10,12,11,29,10,15,15,12,12,23,11-19	(1)	(1)		(6)		(3)	(8)	(8)
Ht22:	10,13,12,28,13,14,16,13,13,23,10-17 (O2b)	(1)	(1)		(2)	(1)	(1)	(1)	(5)
Ht23:	10,14,12,29,13,14,15,13,13,24,10-18 (O2b)	(1)		(1)		(3)		(3)	
Ht24:	11,14,10,30,11,14,14,16,13,23,11-13	(1)		(1)					
Ht25:	10,14,11,29,13,14,16,13,13,23,10-17 (O2b)		(2)	(3)	(3)	(2)	(1)	(2)	(5)
Ht26:	10,12,12,28,11,15,14,14,12,25,13-19		(2)	(2)	(2)	(1)	(1)		(5)
Ht27:	11,12,12,28,10,14,15,14,13,23,13-13		(2)	(1)	(2)	(1)	(2)		(17)
Ht28:	10,13,13,27,13,14,15,13,13,23,10-18 (O2b)		(1)	(1)	(2)	(3)	(2)		(2)
Ht29:	10,14,12,29,13,14,16,13,13,24,10-18 (O2b)		(1)	(1)	(1)	(2)			(1)
Haplotypes shared between two local populations		9 (24)			8 (55)				7 (43)
Ht30:	10,14,12,29,13,14,16,13,13,23,10-19 (O2b)	(3)	(6)	(2)	(7)	(6)	(9)	(11)	(16)
Ht31:	10,14,12,29,13,14,16,13,13,23,10-17 (O2b)	(2)	(1)	(3)	(8)	(3)	(3)	(5)	(11)
Ht32:	10,14,12,29,13,14,16,13,13,23,10-18 (O2b)	(1)	(1)	(4)	(7)	(5)	(4)	(11)	(10)
Ht33:	10,14,12,29,13,14,15,13,13,23,10-19 (O2b)	(1)	(3)	(1)	(4)	(1)	(1)	(3)	(4)
Haplotypes shared among three local populations		4 (28)			4 (88)				4 (41)
Total: 33 (259)		Group A: 33 (97)			Group B: 19 (162)				16 (96)

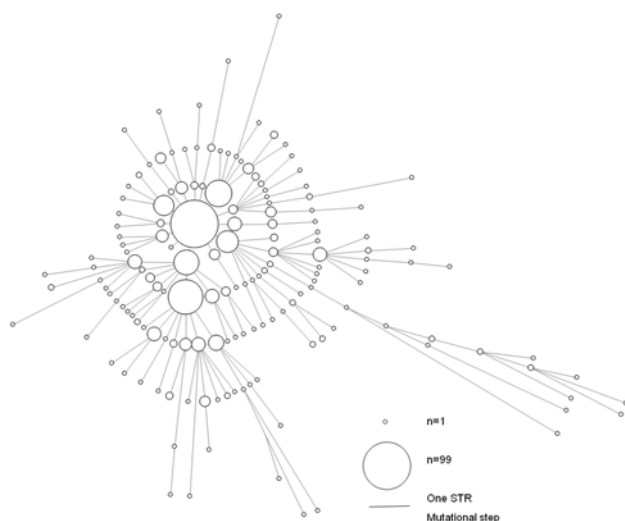


Figure 4. MJ network based on the nine-locus Y-STR (DYS391, DYS389I, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, and DYS390) haplotypes inferred as haplogroup O2b-SRY₊₄₆₅ ($n = 563$) of the group B. Group B haplotypes include those from Korean1–4.

In the present study, we inferred haplogroups from Y-STR haplotypes according to the guide for estimating SNP-delineated haplogroups from Y-STR haplotypes (Nonaka et al., 2007). Among the group A, all haplotypes inferred as the haplogroup O2b, which was defined by one of phylogenetic Y-SNP markers (i.e. SRY₊₄₆₅), were identical to all “star cluster” haplotypes (41%, 128 of 310) depicted in Figure 2. Among the group B, 34% haplotypes (563 of 1655) were estimated as O2b-SRY₊₄₆₅ (data not shown). We also constructed a MJ network based on the 9-locus Y-STR haplotypes (563 of 1655) inferred as O2b-SRY₊₄₆₅, and confirmed that the haplotypes inferred as O2b-SRY₊₄₆₅ formed the “star cluster” in the group B too (Fig. 4).

Because the most frequent haplotype was located in the central position within the “star cluster”, and all haplotypes are supposedly to be derived from it (Jobling and Tyler-Smith, 2003), this finding suggest that the most frequent haplotype seems to be the modal (ancestral) haplotype. A previous survey of Y-STR variation associated with the five most common surnames in Korea also revealed the “star cluster” and proposed that the most frequent haplotype within this cluster might be a candidate modal (ancestral) haplotype originated before the time of the surname introduction (i.e. the Three Kingdoms of Silla, Koguryeo and Baekje, 57 BC–AD 668) (Kim et al., 2009).

On the issues regarding the origin of the Finns, the existence of two different “star clusters” of the Finnish gene pool provided the genetic evidence for the dual origins of Finns by the founder effect (Kittle et al., 1998). Some specific cases of the “star clusters” were revealed and proposed that these clusters have spread by social selection of the likely patrilineal descent of Genghis Khan (Zerjal et al., 2003), Giocangga (the

grandfather of Manchu leader Nurhaci) (Xue et al., 2005), and the semi-mythological dynasty of early medieval Ireland (the *Ui Neill*) (Moore et al., 2006). Because O2b-SRY₊₄₆₅ (which is carried by the “star cluster” in this study), is unusually frequent in Manchu (34%, 34 of 101), Korean Chinese (48%, 38 of 79) and Korean (33%, 28 of 85), but the rarity in the northern Han (0%, 0 of 42) (Katoh et al., 2005), the occurrence of O2b-SRY₊₄₆₅ in the peninsula may have been caused by a population expansion probably associated with the ethno-historic events, rather than recurrent gene flow from China.

The existence of the single “star cluster” within the Korean Y-STR gene pool in this study implies the “single origin of Koreans”, probably representing a genetic legacy of a single ancestor. However, majority of present-day Korean Y-STR gene pool is the non-“star cluster” haplotypes (Fig. 2) which are supposedly to be derived from diverse ancestors, possibly as a result of the migrations and gene flow from neighboring populations over time. Despite of the implication of the “single origin of Koreans”, intrinsic substructure consisting of two components (the minor “star cluster” and the major non-“star cluster”, Fig. 2) also indicates the previously suggested “dual origins of Koreans” (Jin et al., 2003).

In summary, MJ network of the Y-STR haplotypes from three localities revealed the “star cluster” (inferred as O2b-SRY₊₄₆₅) with a frequency of 41% (128 of 310). MJ network of the Y-STR haplotypes inferred as O2b-SRY₊₄₆₅ from the previous datasets of Korean population also revealed the “star cluster” with a frequency of 34% (563 of 1655). These findings suggest that the most frequent 9-locus Y-STR haplotype among the Korean male gene pool (6%, 10,14,12,13,14,16,13,13,23, for loci DYS391, DYS389I, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, and DYS390) may be the Korean modal (ancestral) haplotype. If we get additional data such as the Y-STR haplotypes combined with Y-SNP among the east Asian populations as well as Korean population, the “star cluster” (O2b-SRY₊₄₆₅) may provide a genetic clue to addressing issues concerning timing, origin, and prehistoric migration route of the Koreans.

Acknowledgment This work was supported by the Korea Science and Engineering Foundation (No. 2009-0084149) and in part supported by the second stage of the Brain Korea 21 project.

References

- Bandelt HJ, Forster P and Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37–48.
- Chun BW, Shin SC, Kim YJ, Kim KS, Choi DH, Kim KH, Kim JY and Kang HS (2005) Allele frequencies and haplotypes of the STR loci of the PowerPlex Y-system in southern populations from Korea. *Forensic Sci. Int.* 148: 225–231.

- Decker AE, Kline MC, Redman JW, Reid TM and Butler JM (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci. Int. Genet.* 2: e31–35.
- de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, et al. (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Legal Med.* 110: 134–149.
- Excoffier L, Laval G and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47–50.
- Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations of the use of Y-STRs in forensic analysis. *Int. J. Legal Med.* 120: 191–200.
- Han YH (1995) Altai and the Korean Peninsula from the perspective of archaeological remains. Korean National Museum, Seoul.
- Hwang JH, Lee H, Lee YA, Han MS, Kim W, Jin I and Cho NS (2007) Haplotypes for 12 Y-chromosomal STR loci in a Korean population (the central region). *Forensic Sci. Int.* 168: 73–84.
- IIYeon (2007) SamGuk YuSa. Translated by Kim WJ, Minumsa Press, Seoul.
- Jin HJ, Kwak KD, Hammer MF, Nakahori Y, Shinka T, Lee JW, Jin F, Jia X, Tyler-Smith C and Kim W (2003) Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum. Genet.* 114: 27–35.
- Jin HJ, Tyler-Smith C and Kim W (2009) The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS ONE* 4: e4210.
- Jin L and Su B (2000) Natives or immigrants: Modern human origin in east Asia. *Nat. Rev. Genet.* 1: 126–133.
- Jobling MA, Pandya A and Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* 110: 118–124.
- Jobling MA and Tyler-Smith C (2003) The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* 4: 598–612.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL and Hammer MF (2002) High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum. Biol.* 74: 761–789.
- Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungerel G, Chae GT, Han H, Jia GJ, Tokunaga K, et al. (2005) Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene* 346: 63–70.
- Kim SH, Jung JW, Cho NS, Hwang JH, Yoo SY, Seong KM, Han MS and Kim W (2009) Y-STR genetic structure of the most common surnames in Korea. *Genes & Genomics* 31: 243–249.
- Kim SH, Kim NY, Hong SB, Cho NS, Kim JJ, Han MS and Kim W (2008) Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic Sci. Int. Genet.* 2: e9–10.
- Kim W, Shin DJ, Harihara S and Kim YJ (2000) Y chromosomal DNA variation in east Asian populations and its potential for inferring the peopling of Korea. *J. Hum. Genet.* 45: 76–83.
- Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D and Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* 62: 1171–1179.
- Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, Krawczak M, Tyler-Smith C and Kim W (2005) Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. *Int. J. Legal Med.* 119: 195–201.
- Lee JW, Lee HS and Hwang JJ (2002) Statistical analysis for estimating heterogeneity of the Korean population in DNA typing using STR loci. *Int. J. Legal Med.* 116: 153–160.
- Moore LT, McEvoy B, Cape E, Simms K and Bradley DG (2006) A Y-chromosome signature of hegemony in Gaelic Ireland. *Am. J. Hum. Genet.* 78: 334–338.
- National Institute of Korean History (2004) History of Korea, <http://contents.history.go.kr/>.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nonaka I, Minaguchi K and Takezaki N (2007) Y chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann. Hum. Genet.* 71: 480–495.
- Park MJ, Lee HY, Yoo JE, Chung U, Lee SY and Shin KJ (2005) Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans. *Forensic Sci. Int.* 152: 133–147.
- Ruhlen M (1991) *A guide to the world's languages*. Stanford University Press, Stanford.
- Saha N and Tay JS (1992) Origin of the Koreans: a population genetic study. *Am. J. Phys. Anthropol.* 88: 27–36.
- Willuweit S, Roewer L and International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): update. *Forensic Sci. Int. Genet.* 1: 83–87.
- Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, Xu J, Du R, Fu S, Li P, et al. (2005) Recent spread of a Y-chromosomal lineage in Northern China and Mongolia. *Am. J. Hum. Genet.* 77: 1112–1116.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, et al. (2006) Male demography in east Asia: A north-south contrast in human population expansion times. *Genetics* 172: 2431–2439.
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, et al. (2003) The genetic legacy of the Mongols. *Am. J. Hum. Genet.* 72: 717–721.
- Zhou R, Yang D, Zhang H, Yu W, An L, Wang X, Li H, Xu J and Xie X (2008) Origin and evolution of two Yugur sub-clans in northwest China: a case study in paternal genetic landscape. *Ann. Hum. Biol.* 35: 198–211.