

# The Peopling of Korea Revealed by Analyses of Mitochondrial DNA and Y-Chromosomal Markers

Han-Jun Jin<sup>1,2</sup>, Chris Tyler-Smith<sup>2</sup>, Wook Kim<sup>1\*</sup>

**1** Department of Biological Sciences, Dankook University, Cheonan, Korea, **2** The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

## Abstract

**Background:** The Koreans are generally considered a northeast Asian group because of their geographical location. However, recent findings from Y chromosome studies showed that the Korean population contains lineages from both southern and northern parts of East Asia. To understand the genetic history and relationships of Korea more fully, additional data and analyses are necessary.

**Methodology and Results:** We analyzed mitochondrial DNA (mtDNA) sequence variation in the hypervariable segments I and II (HVS-I and HVS-II) and haplogroup-specific mutations in coding regions in 445 individuals from seven east Asian populations (Korean, Korean-Chinese, Mongolian, Manchurian, Han (Beijing), Vietnamese and Thais). In addition, published mtDNA haplogroup data (N = 3307), mtDNA HVS-I sequences (N = 2313), Y chromosome haplogroup data (N = 1697) and Y chromosome STR data (N = 2713) were analyzed to elucidate the genetic structure of East Asian populations. All the mtDNA profiles studied here were classified into subsets of haplogroups common in East Asia, with just two exceptions. In general, the Korean mtDNA profiles revealed similarities to other northeastern Asian populations through analysis of individual haplogroup distributions, genetic distances between populations or an analysis of molecular variance, although a minor southern contribution was also suggested. Reanalysis of Y-chromosomal data confirmed both the overall similarity to other northeastern populations, and also a larger paternal contribution from southeastern populations.

**Conclusion:** The present work provides evidence that peopling of Korea can be seen as a complex process, interpreted as an early northern Asian settlement with at least one subsequent male-biased southern-to-northern migration, possibly associated with the spread of rice agriculture.

**Citation:** Jin H-J, Tyler-Smith C, Kim W (2009) The Peopling of Korea Revealed by Analyses of Mitochondrial DNA and Y-Chromosomal Markers. PLoS ONE 4(1): e4210. doi:10.1371/journal.pone.0004210

**Editor:** Mark A. Batzer, Louisiana State University, United States of America

**Received:** September 12, 2008; **Accepted:** December 8, 2008; **Published:** January 16, 2009

**Copyright:** © 2009 Jin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study, which forms a part of the project, has been achieved with the support of Conservation Technology Research and Development project, which has been hosted by National Research Institute of Cultural Heritage of Cultural Heritage Administration, Republic of Korea. This work was supported by grant from the Koguryo Research Foundation, Republic of Korea. CTS was supported by The Wellcome Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: wookkim@dankook.ac.kr

## Introduction

An understanding of the evolutionary history of East Asian populations has long been a subject of interest in the field of human evolutionary genetics. Based on results of classical genetic markers, there is significant separation between southern and northern populations of East Asia [1]. This north-south genetic differentiation is likely to have an origin in the early peopling of the region. There have been two major models for early migration routes into East Asia. The first model postulates a southeast Asian origin, followed by a northward migration [2]. Recent genetic surveys using autosomal microsatellite markers [3] and Y-chromosomal binary markers [4] have been interpreted as supporting this model. In contrast, the second model suggests a multidirectional route: one migration through central Asia and one through southeast Asia [1,5,6]. Thus, understanding the genetic origin and history of Korea may be informative for questions concerning prehistoric migration route(s) and population expansions in East Asia.

The Korean Peninsula is located to the north of the Yellow and Yangtze Rivers of China, and bounded to the northeast by Russia. Therefore, the Koreans are geographically a northeast Asian group. Anthropological and archeological evidence suggests that the early Korean population was related to Mongolian ethnic groups who inhabited the general area of the Altai Mountains and Lake Baikal regions of southeast Siberia [7]. According to Korea's founding myths, the Ancient Chosun (the first state-level society of Korea) was established around 2,333 BC in the region of southern Manchuria but later moved into the Pyongyang area of northwest Korea. In addition, archeological evidence reveals that rice cultivation had spread to most parts of the Korean Peninsula by around 1,000–2,000 BC, introduced from the Yellow River and/or Yangtze River basin in China [8].

Studies of classical genetic markers showed that Koreans tend to have a close genetic affinity with Mongolians among East Asians [9–11]. In contrast, recent surveys of Y-chromosomal DNA variation revealed that the Korean population contained lineages typical of both southern and northern East Asian populations

[6,12,13]. The Koreans appeared to have affinities with Manchurians, Yunnan-Chinese from southern China, and Vietnamese [13].

To understand the genetic history of Korea better, more data from additional genetic markers from Korea and its surrounding regions are necessary. Mitochondrial DNA (mtDNA), like the Y chromosome, can also provide valuable information about the phylogeography of human populations due to its special features of haploidy and uniparental inheritance [14–18]. Although recent investigations of mtDNA variation in East Asia have provided valuable information for constructing a robust phylogenetic tree of mtDNA haplotypes, limited data on the Korean population are available [19–21].

In this study, we present new data on the mtDNA sequence variation of the hypervariable segments I and II (HVS-I and HVS-II) and haplogroup-specific mutations in coding regions in 445 individuals from seven East Asian populations, including Korea. In addition, mtDNA haplogroup data ( $N = 3307$ ), mtDNA HVS-I sequences ( $N = 2313$ ), Y chromosome haplogroup data ( $N = 1697$ ) and Y chromosome STR data ( $N = 2713$ ) from the literature were analyzed to elucidate wider aspects of the genetic structure of East Asian populations.

## Materials and Methods

### DNA samples and reference data

We analyzed a total of 445 individuals, collected from seven East Asian populations (Korean, Korean-Chinese (People of Korean origin now living in China), Mongolian, Manchurian, Chinese Han (Beijing), Vietnamese, and Thai). The DNA samples included subsets of the samples examined by Jin et al. [13] and Kwak et al. [22], although the exact number of subjects for each population occasionally varies between these studies. In addition, we included the following new Korean-Chinese and Mongolian samples: 51 Korean-Chinese from northern China and 47 Mongolians from Ulaanbaatar. This study was approved by the Ethics Committee and institutional review boards of Institute of Bio-Science and Technology in the Dankook University in Cheonan, and separate written informed consent was obtained for enrollment from all participants. DNA was prepared from whole blood by the standard method [23] or was extracted from buccal cells according to the procedure of Richards et al. [24].

In addition to our mtDNA data sets, mtDNA haplogroup data for 2862 individuals, mtDNA HVS-I sequences data for 1868 individuals, Y chromosome haplogroup data for 1697 individuals and Y chromosome STR data (ten Y-STR loci: DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 and DYS439) for 2716 individuals were retrieved from the literature [19,20,25–40] to elucidate the genetic relationship between Koreans and other East Asian populations (details in Supplementary Table S1 and Figure 1). For mtDNA haplogroup analysis, some sub-haplogroups were clustered into major haplogroups according to their phylogenetic affiliations (Supplementary Table S2). Similarly, Y chromosome haplogroups from reference data were reclassified into a common set of 13 Y-chromosomal (sub)haplogroups that captured most of the phylogenetic information to allow population comparisons (Supplementary Table S3).

### PCR amplification

PCR amplification of the HVS-I and HVS-II of mtDNA control region was performed using two primer sets as described by Yao et al. [19]: HVS-I, L15996/H16498 (nucleotide positions, 15975–15996/16517–16498); HVS-II, L29/H408 (nucleotide positions, 8-

29/429–408). Primers were designed for amplifying multiple fragments that contain haplogroup diagnostic polymorphisms in the coding regions [19,20,41]. Each set of segments was amplified in a 50  $\mu$ l reaction containing 25 ng of genomic DNA, 10 pM of each primer, 0.2 mM of dNTPs, 2.0 mM  $MgCl_2$ , 5 mM KCl, 10 mM TRIS-HCl (pH 8.3) and 1.5 U AmpliTaq DNA polymerase (Perkin-Elmer, CA, USA). The PCR amplification was carried out using a GeneAmp<sup>®</sup> PCR system 9700 thermal cycler (Applied Biosystems, CA, USA) under the conditions described in Table 1.

### mtDNA sequencing and genotyping of RFLP

After PCR amplification, each PCR product was purified using the Wizard<sup>®</sup> PCR Preps DNA Purification System (Promega, WI, USA) and then sequenced by cycle sequencing using either a MegaBase 1000 sequencer (Amersham Bioscience, USA) or an ABI PRISM<sup>™</sup> 310 Genetic Analyzer (Applied Biosystems, CA, USA) with DYEnamic ET Dye Terminator (Amersham Bioscience, USA) or BigDye<sup>™</sup> Terminator (PE Biosystems, USA), respectively. DNA sequences of the PCR amplicons were determined from both forward and reverse sequence data using the original primer pairs. The sequences from nucleotide position (np) 16024 to 16365 in HVS-I and from 73 to 340 in HVS-II were determined, since ambiguous electropherograms for 20–30 nucleotides near the primers were frequently observed.

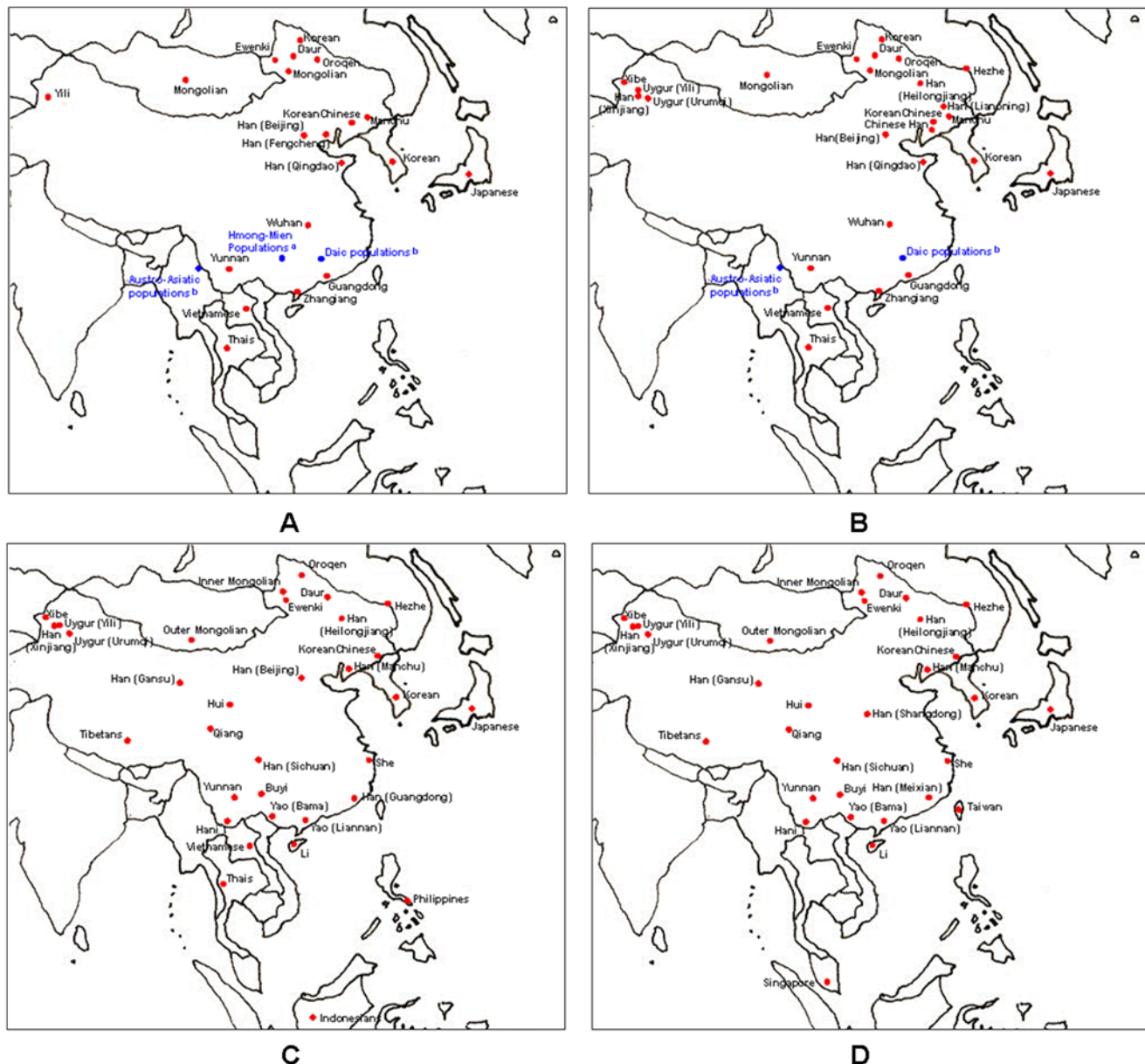
The intergenic COII/tRNA<sup>Lys</sup> 9-bp deletion was analyzed as described in Jin et al. [42]. In addition, several amplified segments, mainly in the mtDNA coding regions, were analyzed by RFLP typing and additional sequencing, as listed in Table 1.

### Sequence alignment and haplogroup analyses

Sequences were aligned and compared with the revised Cambridge Reference Sequence (rCRS) [43] using the Sequencher program ver. 2000 (Gene Codes corporation, MI, USA). The results were converted into a Microsoft Excel table (Microsoft Corporation, CA, USA). The mtDNAs were classified into the (sub-)haplogroups based on HVS-I/II motifs of haplogroup specific-sequences as well as coding regions as described in recent surveys [19,20,25,44,45]. The HVS-I motif searching and haplogroup-directed comparison with closely related sequences from other databases led us to tentatively assign each mtDNA to a haplogroup. To further characterize the mtDNA lineage tested, we compared their HVS-II motif to verify the predicted haplogroup status of each mtDNA. In general, more than 95% of mtDNA lineages can faithfully be classified to specific haplogroups using HVS-I/II motifs without extra information from coding region sequences [44]. However, in the remaining cases, their (sub-)haplogroups were characterized using sequence information from some coding region sites (Table 1). After each mtDNA was assigned to the most-derived named haplogroup, the haplogroup distribution frequencies in each of seven populations were estimated. For quality assurance purposes, we performed quasi-median network analysis [46,47]. The HVS-I (np 16024–16365 np) and HVS-II (np 74–340) sequence of 445 individuals of this study have been submitted to GenBank (Accession Numbers, FJ493775–FJ494664).

### Data analyses

The genetic differentiation between different population samples and its statistical significance were assessed via  $F_{ST}$  (mtDNA HG and HVS-I/II and Y-SNPs) and  $R_{ST}$  (Y-STRs) values. The population genetic structure of the ethnic and/or regional groups was analyzed through the analysis of molecular variance (AMOVA) approach [48]. The calculations of diversity indices,  $F_{ST}$ ,  $R_{ST}$  and AMOVA were performed using the



**Figure 1. Geographic locations of the Asian populations studied.** (A) mtDNA haplogroups. (B) mtDNA HVS-I sequences. (C) Y-chromosome haplogroups. (D) Y-chromosome STRs. <sup>a</sup>Wen et al. [27]. <sup>b</sup>Li et al. [36]. doi:10.1371/journal.pone.0004210.g001

Arlequin 2.000 package [49]. Population pairwise  $F_{ST}$  and  $R_{ST}$  values were visualized by multidimensional scaling (MDS) plot analyses using SPSS 12.0 software.

Haplogroup-specific median-joining networks [50] for Y chromosome data were constructed using the NETWORK 4.2 program (www.fluxus-technology.com). Such networks were initially highly reticulated, and we reduced reticulations by first weighting the loci according to the inverse of their variance in the dataset used [51] and subsequently constructing a reduced-median network [52] to form the input of the median-joining network [53].

The admixture proportions of northeast Asian and the southeast Asian parental populations in the Korean population were estimated for mtDNA and the Y chromosome using the Admix 2.0 software [54]

## Results and Discussion

Almost all of the mtDNA lineages analyzed here could be assigned to the East Asian-specific (sub)haplogroups described recently [19,20,25,44,45], with the exception of two individuals belonging to the European mtDNA haplogroups T (Manchurian) and U5a (Mongolian) (Table 2). The gene diversity ( $H$ ), nucleotide diversity ( $\pi_n$ ), and mean number of pairwise differences of the population samples are listed in Table 3. All seven populations displayed high levels of genetic diversity ( $H > 0.99$ ), suggesting a relatively large population size and heterogeneity of each mtDNA pool. The haplogroup frequencies observed in each population are summarized in Table 2. Based on these haplogroup assignments, the Koreans share lineages with both the southern and the northern haplogroup complexes of East Asia. We first attempted

**Table 1.** Primers for mtDNA amplification, sequencing and RFLP analyses.

Primer Pair	Primer sequences (5' to 3')	Annealing Temperature (°C)	Polymorphisms at/in
L29/H408	GGT CTA TCA CCC TAT TAA CCA C CTG TTA AAA GTG CAT ACC GCC A	54	HVS-II
L4499/H5099	TGG CCC AAC CCG TCA TCT AC GGA ATG CGG TAG TAG TTA GG	60	+4831 <i>HhaI</i> (4833)
L4887/H5442	TGA CAA AAA CTA GCC CCC ATC T GCG ATG AGT GTG GGG AGG AA	60	5176 <i>AluI</i> (5178A)
L8215/H8297	ACA GTT TCA TGC CCA TCG TC ATG CTA AGT TAG CTT TAC AG	57	9-bp deletion
L10170/H10660	ACA TAG AAA AAT CCA CCC CTT ACG TTC GCA GGC GGC AAA GAC TA	59	10171-10659
L12114/H12338	CTC AAC CCC GAC ATC ATT ACC ATT ACT TTT ATT TGG AGT TGC ACC AAA ATT	58	12308
L14054/H14591	TCA CAG CAC CAA ATC TCC AC AAG CCT TCT CCT ATT TAT GG	57	14178, 14308, 14318, 14470
L14575/H15086	ACC CGA CCA CAC CGC TAA CA AGG AGG ATA ATG CCG ATG TT	57	14668, 14766, 14783, 15043
L15996/H16498	CTC CAC CAT TAG CAC CCA AAG C CCT GAA GTA GGA ACC AGA TG	60	HVS-I

PCR conditions were 94°C for 5 min, for denaturation; 94°C for 45 sec; annealing temperature shown for 45 sec, for amplification; and 72°C for 1 min, for 35 cycles; incubation at 72°C for 5 min.

doi:10.1371/journal.pone.0004210.t001

to quantitate these contributions by a detailed consideration of the distribution of each lineage.

The highest (23.8%) frequency in the Korean mtDNA pool was observed for haplogroup D4, which is widespread in northern East Asia and especially in the Korean-Chinese (21.6%), and Manchurians (20.0%). In total, haplogroup D lineages including the subhaplogroups (D4, D4a, D4b, D5, and D5a) accounted for 32.4% of the Korean mtDNA pool. In addition, the Koreans present moderate frequencies of (sub)haplogroup A (8.1%) and (sub)haplogroup G (10.3%) lineages, mostly prevalent in northeast Asia and southeast Siberia [20,55–57]. Other Siberian and Mongolian-prevalent haplogroups from the C, Y and Z lineages make up less than 4% of the Korean mtDNA pool. Haplogroups A5a and Y2 are found almost exclusively in Korea but were present at extremely low frequencies. In total, these northern haplogroups account for ~60% of the mtDNA gene pool of the Koreans. In addition, southeast Asian-prevalent mtDNA lineages of (sub)haplogroups B (14.6%), M7 (10.3%), and F (9.7) are also found at moderate frequencies in the Korean population (Table 2). These findings suggest that more than 30% of the Korean mtDNA pool is attributable to maternal lineages with a more southern origin. We also found the haplogroup M7a1 exclusively in the Korean population. This result is consistent with previous reports that haplogroup M7a is restricted to Japan and south Korea [18,20]. Thus, the distribution pattern of mtDNA haplogroups leads us to consider that the peopling of Korea is likely to have involved multiple sources.

We then investigated the mtDNA and Y-chromosomal relationships between the East Asian populations, using both the new and published data. In these analyses mtDNA haplogroups, mtDNA HVS-I sequences, Y-SNPs and Y-STRs were compared (Supplementary Tables S1, S2, S3). Pairwise  $F_{ST}$  (mtDNA haplogroup, mtDNA HVS-I sequences and Y-SNPs) and  $R_{ST}$  (Y-STRs) values between East Asian populations were calculated (Supplementary

Table S1). The  $F_{ST}$  distances of mtDNA markers (mtDNA haplogroups and HVR-I sequences) of Korean populations showed close relationships with Manchurians, Japanese, Mongolians and northern Han Chinese but not with southern Asians (Supplementary Tables S4 and S5; Figure 2A, B). In the MDS plots, the Korean samples lay entirely within the cluster of northern populations.

In contrast, the results of Y chromosome analyses (based on Y-SNPs and Y-STRs) of Korean populations revealed closer relationships with both northeast and southeast Asian populations (Supplementary Tables S6 and S7; Figure 2C, D). Like the mtDNA distances, Y-chromosomal distances from Manchurian, Japanese and northern Han Chinese populations were usually not significantly greater than zero, but some distances from southern Han populations (e.g. Yunnan Han, Y haplogroups; Meixian Han, Y-STRs) or other southern populations (e.g. Vietnamese, Y haplogroups) were also not significantly above zero (Supplementary Tables S6 and S7), as noted previously [13]. In the MDS plots, the Korean samples lay at the border between the northern and southern clusters, rather than within the northern cluster (Figure 2C, D). In order to investigate Y-chromosomal relationships in more detail, we visualized STR haplotypes within a common predominantly northern haplogroup (C\*) and southern haplogroup (O3) using networks [50] constructed with the seven Y-STRs common to all datasets (Figure 3). These networks did not show striking geographical structure, so we calculated, for each Korean haplotype, the distance to the closest northern and southern haplotype. In both haplogroups, the mean distance to the southern haplotypes was lower than to the northern haplotypes (C\* Korean-north 5.0 steps, Korean-south 4.5 steps; O3 Korean-north 3.5 steps, Korean-south 2.2 steps). This finding is particularly striking for haplogroup C\* because it is far more prevalent in the north (Figure 3A).

The genetic differences between the Koreans and other East Asians were examined by AMOVA (Table 4). When samples were

**Table 2.** Distribution of mtDNA haplogroup frequencies in 7 East Asian populations.

Haplogroup	Korean-Chinese	Mongolian	Manchurian	Han (Beijing)	Vietnamese	Thais	Korean
A			3	1			3
A4	4	2	1	1	1		6
A5	1				1		5
A5a							1
B				2		1	
B4		2	2		3		7
B4a	2	1	1		1		11
B4b	2			2	1		
B4b1							4
B4c							1
B5a	1				1	3	2
B5b	1	1					2
C	1	8	1		2	4	3
C3		2					
D	1					2	1
D4	11	5	8	5	7	1	44
D4a	3			2			3
D4b		1					3
D5	2		1	3	1		6
D5a			1	2	1		3
F	1		3	3	1		
F1a	1	3	2	4	10		8
F1b	1	3	2	2		8	8
F1c	1		1				
F2							2
F2a		1					
G		2		1			1
G1a			3		1		1
G2	1			1			7
G2a	1	5	1	2	6		
G3		1			1		4
M	3	1			1	5	1
M7a1							7
M7b		1		1		3	
M7b1			2			2	1
M7b2	2					1	4
M7c		1	1		1		1
M7c1	1		1	1	1	1	6
M8a		1	1		1	1	2
M9a	1		1		1		3
M10	2				1	3	1
M11	1						1
N							1
N9a	5	2	2	3	3	1	12
R						2	
R11	1						
T			1				
U5a		1					
Y1		1					1
Y2							1
pre-Z		1					

**Table 2.** cont.

Haplogroup	Korean-Chinese	Mongolian	Manchurian	Han (Beijing)	Vietnamese	Thais	Korean
Z		1	1	4	1	2	1
n	51	47	40	40	42	40	185
Total	445						

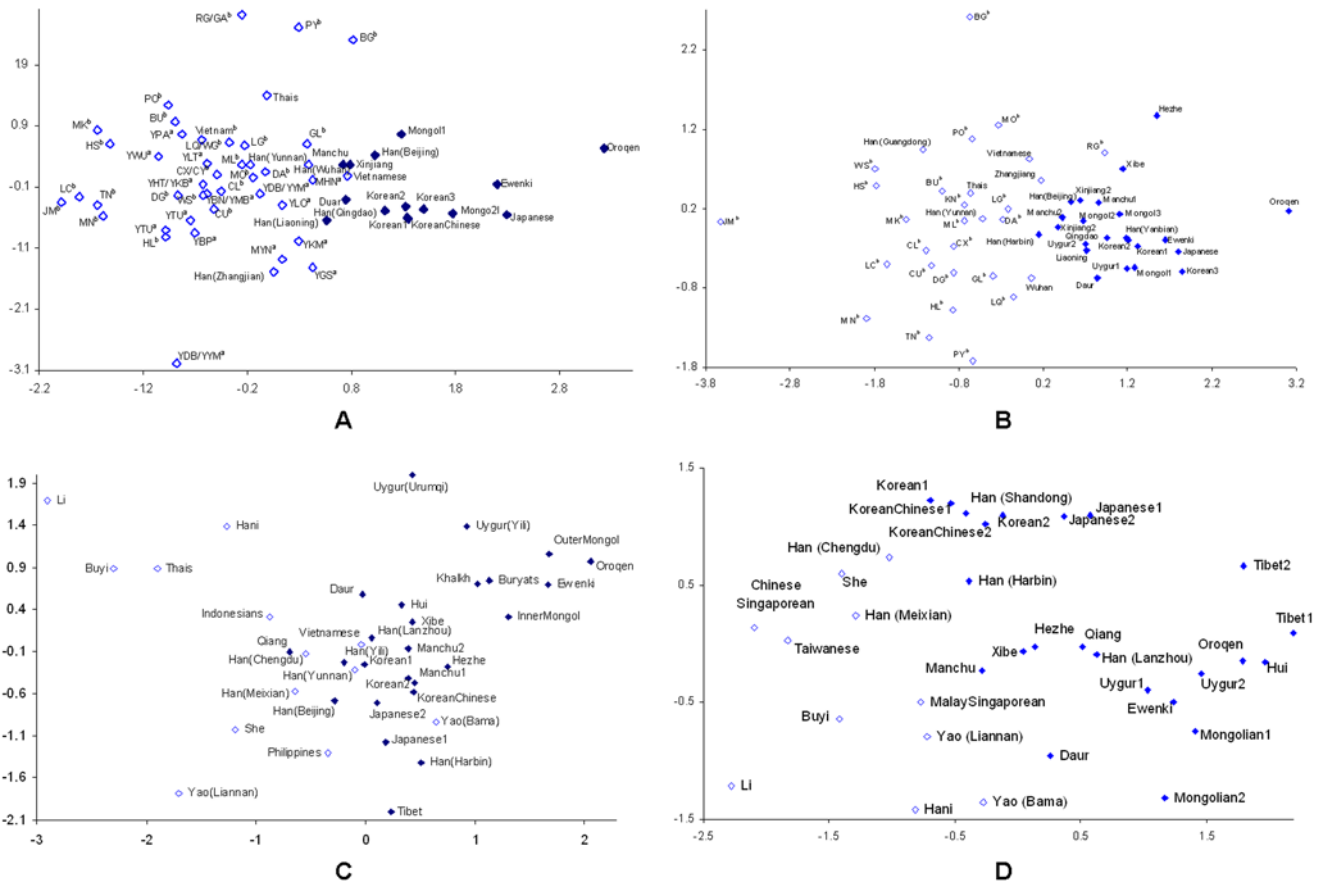
doi:10.1371/journal.pone.0004210.t002

**Table 3.** Diversity indices of mtDNA in seven east Asian populations.

	Haplogroup data	Sequence data (HVS-I/II <sup>a</sup> )		
	Gene diversity	Gene diversity	Pairwise difference	Nucleotide diversity
Korean	0.9239+/-0.0132	0.9988+/-0.0007	10.07+/-4.62	0.039+/-0.020
Korean-Chinese	0.9357+/-0.0219	0.9992+/-0.0041	10.21+/-4.74	0.039+/-0.020
Mongolian	0.9454+/-0.0172	0.9991+/-0.0046	10.80+/-5.00	0.042+/-0.021
Manchurian	0.9462+/-0.0221	0.9974+/-0.0063	10.88+/-5.05	0.042+/-0.022
Han (Beijing)	0.9526+/-0.0135	1.0000+/-0.0056	11.38+/-5.27	0.044+/-0.022
Vietnamese	0.9152+/-0.0290	0.9919+/-0.0079	9.66+/-4.52	0.037+/-0.020
Thai	0.9269+/-0.0214	1.0000+/-0.0056	11.53+/-5.33	0.045+/-0.023

<sup>a</sup>HVS-I: np 16024–16365; HVS-II: np 73–340.

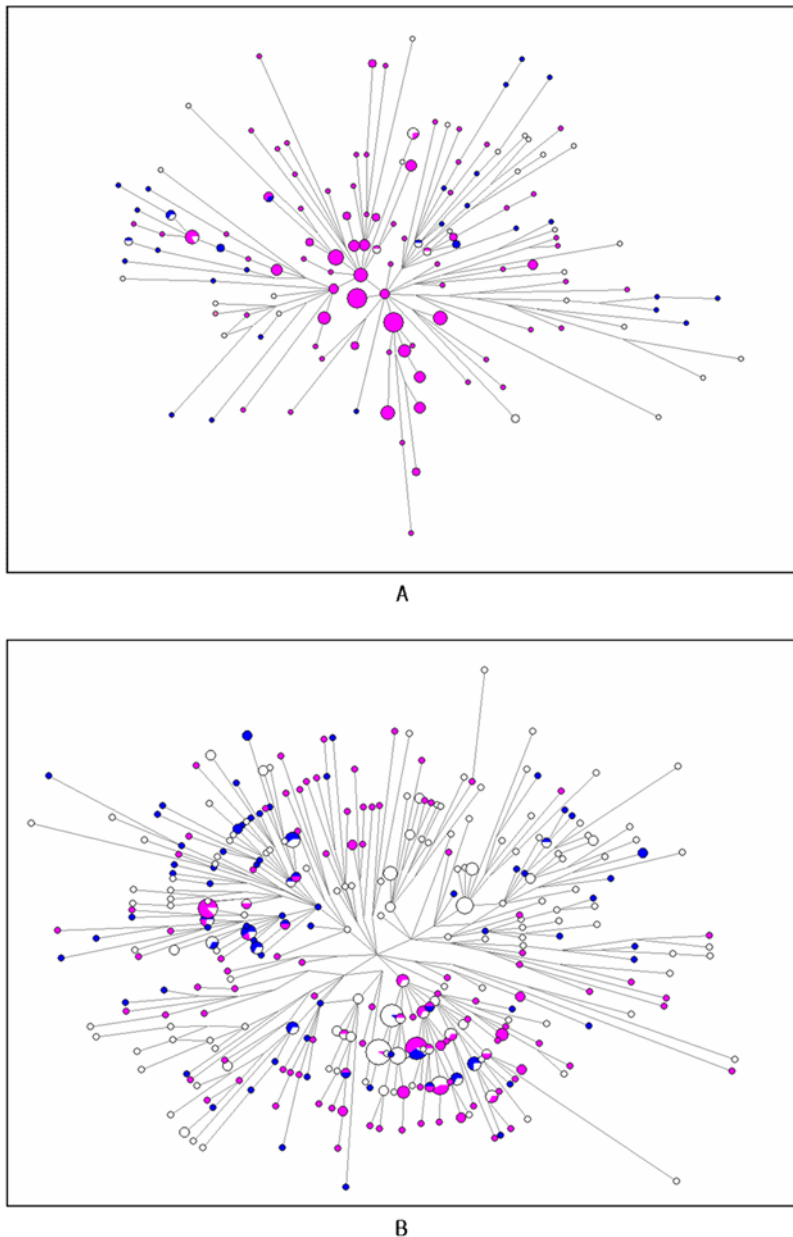
doi:10.1371/journal.pone.0004210.t003



**Figure 2. Multidimensional scaling (MDS) plot based on (A)  $F_{ST}$  distances of mtDNA haplogroups (stress = 0.21). (B)  $F_{ST}$  distances of mtDNA HVS-I sequences (stress = 0.19). (C)  $F_{ST}$  distances of Y chromosome haplogroups (stress = 0.21). (D)  $R_{ST}$  distances of Y chromosome STRs (stress = 0.19).** (closed diamond: North Asians; opened diamonds: South Asians). <sup>a</sup>Wen et al. [27]. <sup>b</sup>Li et al. [36].

doi:10.1371/journal.pone.0004210.g002





**Figure 3. Median-joining network for east Asian (A) Network of 7 Y-STRs (DYS19, DYS389b, DYS389i, DYS390, DYS391, DYS392 and DYS393) variation within Haplogroup C. (B) Network of 7 Y-STRs within Haplogroup O3.** Circle areas are proportional to haplotype frequency. Lines represent the mutational differences between haplotypes. The network corresponds the following colors: purple- far north Asian populations (Daur, Ewenki, Han (Xinjiang), inner Mongolians, Oroqen, outer Mongolians, Uygur (Yili), Uygur (Urumqi) and Xibe); blue- Koreans; white- far south Asian populations (Buyi, Han (Guangdong), Han (Sichuan), Han (Yunnan), Hani, Indonesians, Li, philippines, She, Thais, Vietnamese, Yao (Bama) and Yao (Liannan)).

doi:10.1371/journal.pone.0004210.g003

grouped into northeast Asians and southeast Asians (excluding Koreans), a highly significant difference was found between the two groups with all markers. Thus there is significant genetic differentiation within the region, and we could then compare each group separately with the Koreans. With mtDNA, Koreans were not significantly different from either group when HVRI sequences were compared, although they were distinct from the southeast Asians in the haplogroup comparisons. With the Y chromosomes, they were again not distinct from either group when haplogroup comparisons were made, but were distinct from the southeast Asians in the STR-based comparison (Table 4).

Our study documents the genetic relationships of the Koreans with their neighboring populations in unprecedented detail. Two major findings emerge. First, the Koreans are overall more similar to northeast Asians than to southeast Asians. This conclusion would be expected from the general correlation between genetic variation and geography observed for human populations, and is supported here by an examination of individual mtDNA haplogroups (Table 2), genetic distances between populations derived from mtDNA or Y-chromosomal data (Figure 2), and the apportionment of genetic diversity between different groups of populations (Table 4). Second, the conclusions from mtDNA and

**Table 4.** AMOVA Results.

Markers	Grouping	Percentage of Variance ( $p$ -value)		
		Among groups	Among population within groups	Within populations
mtDNA haplogroups	Korean vs. NEAs	-0.03 (0.20332)	1.32 (<0.00001)	98.71 (<0.00001)
	Korean vs. SEAs	2.29 (<0.00001)	2.47 (<0.00001)	95.23 (<0.00001)
	Korean vs. NEAs vs. SEAs	2.16 (<0.00001)	2.51 (<0.00001)	95.33 (<0.00001)
	NEAs vs. SEAs	3.22 (<0.00001)	2.98 (<0.00001)	93.81 (<0.00001)
mtDNA HVRI sequences	Korean vs. NEAs	-0.39 (0.98436)	1.29 (<0.00001)	99.11 (<0.00001)
	Korean vs. SEAs	-0.23 (0.66373)	1.72 (<0.00001)	98.51 (<0.00001)
	Korean vs. NEAs vs. SEAs	2.18 (<0.00001)	2.04 (<0.00001)	95.79 (<0.00001)
	NEAs vs. SEAs	0.26 (<0.00001)	1.58 (<0.00001)	98.16 (<0.00001)
Y-chromosome haplogroups	Korean vs. NEAs	-0.21 (0.46237)	9.43 (<0.00001)	90.78 (<0.00001)
	Korean vs. SEAs	1.34 (0.17889)	10.78 (<0.00001)	87.89 (<0.00001)
	Korean vs. NEAs vs. SEAs	2.89 (<0.00001)	10.35 (<0.00001)	86.77 (<0.00001)
	NEAs vs. SEAs	4.60 (<0.00001)	10.95 (<0.00001)	84.45 (<0.00001)
Y-chromosome STRs	Korean vs. NEAs	3.36 (0.08016)	6.25 (<0.00001)	90.39 (<0.00001)
	Korean vs. SEAs	7.58 (0.00293)	2.48 (<0.00001)	89.94 (<0.00001)
	Korean vs. NEAs vs. SEAs	4.99 (0.00098)	5.65 (<0.00001)	89.36 (<0.00001)
	NEAs vs. SEAs	5.40 (0.00098)	6.82 (<0.00001)	87.78 (<0.00001)

$P$  values are obtained by 10,000 permutations.  
doi:10.1371/journal.pone.0004210.t004

Y-chromosomal analyses differ. Sex-biased admixture is common in human expansions such as that of Bantu-speaking farmers in Africa [58], the spread of the Han ethnic group in China [59] or the post-Columbian peopling of the Americas [60]. The effects in Korea are more subtle, but show a larger male than female contribution from southern East Asia to the population of Korea, most clearly revealed by the admixture estimates, where a 35% contribution from the south was estimated for mtDNA, compared with a 83% contribution for the Y chromosome (Table 5).

The predominant genetic relationship with northern East Asians is consistent with other lines of evidence. Xue et al. [31] reported that the northern East Asian populations started to expand in number before the last glacial maximum at 21-18 KYA, while the southern populations all started to expand after it, but then grew faster, and they suggested that the northern populations expanded earlier because they could exploit the abundant megafauna of the "Mammoth Steppe," while the southern populations could increase in number only when a warmer and more stable climate led to more plentiful plant resources such as tubers. By this criterion, the Koreans, expanding at about 30 KYA [31] also resemble other

northern populations. Historical evidence suggests that the Ancient Chosun, the first state-level society, was established in the region of southern Manchuria and later moved into the Pyongyang area of the northwestern Korean Peninsula. Based on archeological and anthropological data, the early Korean population possibly had an origin in the northern regions of the Altai-Sayan and Baikal regions of Southeast Siberia [7,8,61].

What could be the origin of the male-biased southern contribution to Korean gene pool illustrated, for example, by haplogroups O-M122 (42.2%) and O-SRY465 (20.1%) [29]. Recent molecular genetic analyses and the geographical distribution of haplogroup O-M122 lineages, found widely throughout East Asia at high frequencies (especially in southern populations and China), have suggested a link between these Y-chromosome expansions and the spread of rice agriculture in East Asia [62-64]. In general, Y-chromosomes might be spread via a process of demic diffusion during the early agricultural expansion period [65,66]. If this interpretation were substantiated, the spatial pattern of Y-haplogroup O would imply a genetic contribution to Korea through the spread of male-mediated agriculture. Large-scale genetic analyses thus begin to reveal some of the complexities of the peopling of Korea, and further studies of individual autosomal loci or genomewide genotyping and sequencing are expected to provide further insights.

**Table 5.** Admixture estimates of Northeast Asians and Southeast Asians in Korean populations.

Markers	Parental contributions	
	Northeast Asians (SD <sup>a</sup> )	Southeast Asians (SD <sup>a</sup> )
MtDNA haplogroups	0.65 (0.25)	0.35 (0.25)
Y-chromosome haplogroups	0.17 (0.14)	0.83 (0.14)
Mt-HG & Y-HG	0.48 (0.21)	0.52 (0.21)

<sup>a</sup>Standard Deviation.  
doi:10.1371/journal.pone.0004210.t005

## Supporting Information

### Table S1 Asian populations studied

Found at: doi:10.1371/journal.pone.0004210.s001 (0.05 MB XLS)

### Table S2 mtDNA-haplogroup distributions in East Asian populations

Found at: doi:10.1371/journal.pone.0004210.s002 (0.05 MB XLS)

### Table S3 Y-haplogroup distribution in East Asian populations

Found at: doi:10.1371/journal.pone.0004210.s003 (0.03 MB XLS)



**Table S4**  $F_{ST}$  distances of mtDNA haplogroups in east Asian populations (non-significant values are underlined). <sup>a</sup>Present work; <sup>b</sup>Kivisild et al. [20]; <sup>c</sup>Yao et al. [19]; <sup>d</sup>Lee et al. [30]; <sup>e</sup>Wen et al. [27]; <sup>f</sup>Kong et al. [25]; <sup>g</sup>Li et al. [36].

Found at: doi:10.1371/journal.pone.0004210.s004 (0.22 MB XLS)

**Table S5**  $F_{ST}$  values of mtDNA HVR-I Sequences in east Asian populations (non-significant values are underlined). <sup>a</sup>Present work; <sup>b</sup>Maruyama et al. [26]; <sup>c</sup>Kivisild et al. [20]; <sup>d</sup>Kong et al. [25]; <sup>e</sup>Yao et al. [19]; <sup>f</sup>Zhang et al. [28]; <sup>g</sup>Li et al. [36]; <sup>h</sup>Powell et al. [37]

Found at: doi:10.1371/journal.pone.0004210.s005 (0.07 MB XLS)

**Table S6**  $F_{ST}$  distances of Y chromosome haplogroups in east Asian populations (non-significant values are underlined). <sup>a</sup>Hong et al. [29]; <sup>b</sup>Xue et al. [31].

Found at: doi:10.1371/journal.pone.0004210.s006 (0.05 MB XLS)

**Table S7**  $R_{ST}$  distances in east Asian populations using ten Y chromosome STRs (non-significant values are underlined). <sup>a</sup>Hara et

al. [34]; <sup>b</sup>Huang et al. [40]; <sup>c</sup>Hwang et al. [35]; <sup>d</sup>Yan et al. [38]; <sup>e</sup>Zhang et al. [33]; <sup>f</sup>Zhang et al. [39]; <sup>g</sup>Yong et al. [32]; <sup>h</sup>Xue et al. [31].

Found at: doi:10.1371/journal.pone.0004210.s007 (0.05 MB XLS)

## Acknowledgments

We would like to thank H.J. Bandelt for crucial comments and advice on the mtDNA study. We are grateful to all volunteers for providing DNA samples. We also thank K.D. Kwak and S.B. Hong for technical assistance.

## Author Contributions

Conceived and designed the experiments: HJJ CTS WK. Performed the experiments: HJJ. Analyzed the data: HJJ CTS WK. Contributed reagents/materials/analysis tools: WK. Wrote the paper: HJJ CTS WK.

## References

- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. Princeton: Princeton University Press.
- Turner CG (1990) Major features of sundadonty and sinodonty, including suggestions about east Asian microevolution, population history and late Pleistocene relationships with Australian aboriginals. *Am J Phys Anthropol* 82: 295–317.
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, et al. (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95: 11763–11768.
- Su B, Xiao J, Underhill P, Deika R, Zhang W, 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.
- Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 10: 927–943.
- Karafet T, Xu L, Du R, Wang W, Feng S, et al. (2001) Paternal population history of east Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69: 615–628.
- Kim K (1970) People and language. In: Korea - Its People and Cultures. Seoul: Hakwon-sa Ltd.
- Choi ML, Rhee SN (2001) Korean archaeology for the 21st century: from prehistory to State formation. *Seoul J Kor Studies* 14: 117–147.
- Goedde HW, Paik YK, Lee CC, Benkmann HG, Kriese L, et al. (1987) Red cell and serum protein polymorphisms in three population groups of south Korea. *Gene Geogr* 1: 177–188.
- Saha N, Tay JSH (1992) Origin of the Koreans: a population genetic study. *Am J Phys Anthropol* 88: 27–36.
- Hong SS, Suh JA, Chae JJ, Goh SH, Kim YS, et al. (1993) Frequency distribution of alleles at *D1S80* and *apo-B 3'* VNTR loci in Korean population. *Mol Cells* 3: 457–453.
- Kim W, Shin DJ, Harihara S, 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.
- Jin HJ, Kwak KD, Hammer MF, Nakahori Y, Shinka T, et al. (2003) Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum Genet* 114: 27–35.
- Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77: 6715–6719.
- Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. *Gene* 238: 211–230.
- Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet* 33: 266–275.
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4: 598–612.
- Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41: 539–564.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70: 635–651.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, et al. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19: 1737–1751.
- Tanaka M, Cabrera VM, González AM, Larruga JM, Takeyasu T, et al. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14: 1832–1850.
- Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, et al. (2005) Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. *Int J Legal Med* 119: 195–201.
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a Laboratory Manual, 2nd edn. New York: Cold Spring Harbor Laboratory Press.
- Richards B, Skoletsky J, Shuber AP, Balfour R, Stern RC, et al. (1993) Multiplex PCR amplification from the *CFTR* gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2: 159–163.
- Kong QP, Yao YG, Liu M, Shen SP, Chen C, et al. (2003) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 113: 391–405.
- Maruyama S, Minaguchi K, Saitou N (2003) Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. *Int J Legal Med* 117: 218–225.
- Wen B, Li H, Gao S, Mao X, Gao Y, Li F, et al. (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22: 725–734.
- Zhang YJ, Xu QS, Zheng ZJ, Lin HY, Lee JB (2005) Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. *Forensic Sci Int* 149: 267–269.
- Hong SB, Jin HJ, Kwak KD, Kim W (2006) Y-chromosome Haplogroup O3-M122 Variation in East Asia and Implications for the Peopling of Korea. *Korean J Genetics* 28: 1–8.
- Lee HY, Yoo JE, Park MJ, Chung U, Shin KJ (2006) Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. *Int J Legal Med* 120: 5–14.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q (2006) Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 172: 2431–2439.
- Yong RY, Lee LK, Yap EP (2006) Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. *Forensic Sci Int* 159: 244–257.
- Zhang Q, Yan J, Tang H, Jiao Z, Liu Y (2006) Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. *Leg Med (Tokyo)* 8: 300–305.
- Hara M, Kido A, Takada A, Adachi N, Saito K (2007) Genetic data for 16 Y-chromosomal STR loci in Japanese. *Leg Med (Tokyo)* 9: 161–170.
- Hwang JH, Lee H, Lee YA, Han MS, Kim W, et al. (2007) Haplotypes for 12 Y-chromosomal STR loci in a Korean population (the central region). *Forensic Sci Int* 168: 73–84.
- Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, et al. (2007) Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol* 134: 481–488.
- Powell GT, Yang H, Tyler-Smith C, Xue Y (2007) The population history of the Xibe in northern China: a comparison of autosomal, mtDNA and Y-chromosomal analyses of migration and gene flow. *FSI Genetics* 1: 115–119.
- Yan J, Tang H, Liu Y, Jing Y, Jiao Z, et al. (2007) Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. *Leg Med (Tokyo)* 9: 196–202.
- Zhang Y, Zhang H, Cui Y, Cui H, Xu Q (2007) Population genetics for Y-chromosomal STRs haplotypes of Chinese Korean ethnic group in northeastern China. *Forensic Sci Int* 173: 197–203.
- Huang TY, Hsu YT, Li JM, Chung JH, Shun CT (2008) Polymorphism of 17 Y-STR loci in Taiwan population. *Forensic Sci Int* 174: 249–254.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, et al. (2003) Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67: 391–411.
- Jin HJ, Choi JW, Shin DJ, Kim JM, Kim W (1999) Distribution of length variation of the mtDNA 9-bp motif in the intergenic COII/tRNA<sup>Lys</sup> region in East Asian populations. *Korean J Biol Sci* 3: 393–397.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.

44. Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP (2004) Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in China. *Mol Biol Evol* 21: 2265–2280.
45. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22: 339–345.
46. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* 71: 1150–1160.
47. Brandstätter A, Klein R, Duftner N, Wiegand P, Parson W, et al. (2006) Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). *Int J Legal Med* 120: 310–314.
48. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
49. Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis, Ver. 2.000, Genetics and Biometry Lab, Department of Anthropology, University of Geneva.
50. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
51. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, et al. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70: 1107–1124.
52. Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141: 743–753.
53. Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L (2008) Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet* 82: 873–882.
54. Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: expansion to any number of parental populations. *Mol Biol Evol* 18: 672–675.
55. Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, et al. (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118: 63–76.
56. Derenko MV, Lunkina AV, Malyarchuk BA, Zakharov IA, Tsedev Ts, et al. (2004) Restriction polymorphism of mitochondrial DNA in Koreans and Mongolians. *Russian J Genet* 40: 1562–1570.
57. Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, et al. (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69: 67–89.
58. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, et al. (2008) Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A* 105: 1596–1601.
59. Wen B, Li H, Lu D, Song X, Zhang F, et al. (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431: 302–305.
60. Marrero AR, Bravi C, Stuart S, Long JC, Pereira das Neves Leite F, et al. (2007) Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. *Hum Hered* 64: 160–171.
61. Han YH (1995) Altai and the Korean Peninsula from the Perspective of Archaeological Remains. Seoul: Korea National Museum.
62. Wells S (2002) *The Journey of Man: A Genetic Odyssey*. Princeton: Princeton University Press.
63. Underhill PA (2003) Inferring human history: clues from Y-chromosome haplotypes. *Cold Spring Harb Symp Quant Biol* 68: 487–493.
64. Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, et al. (2005) Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet* 77: 408–419.
65. Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, et al. (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51: 47–58.
66. Sokal RR, Thompson BA (1998) Spatial genetic structure of human populations in Japan. *Hum Biol* 70: 1–22.