

Y chromosome homogeneity in the Korean population

Soon Hee Kim · Myun Soo Han · Wook Kim · Won Kim

Received: 15 June 2010 / Accepted: 3 August 2010 / Published online: 17 August 2010
© Springer-Verlag 2010

Abstract The distribution of Y-chromosomal variation from the 12 Y-SNP and 17 Y-STR markers was determined in six major provinces (Seoul-Gyeonggi, Gangwon, Chungcheong, Jeolla, Gyeongsang, and Jeju) to evaluate these populations' possible genetic structure and differentiation in Korea. As part of the present study, a 10-plex SNaPshot assay and two singleplex SNaPshot assays were developed. Based on the result of 12 Y-SNP markers (M9, M45, M89, M119, M122, M174, M175, M214, RPS4Y, P31, SRY465, and 47z), almost 78.9% of tested samples belonged to haplogroup O-M175 (including its subhaplogroups O3-M122: 44.3%, O2b*-SRY465: 22.5%, O2b1-47z: 8.7%), and 12.6% of the tested samples belonged to haplogroup C-RPS4Y. A total of 475 haplotypes were identified using 17 Y-STR markers included in the Yfiler kit, among which 452 (95.2%) were individual-specific. The overall haplotype diversity for the 17 Y-STR loci was 0.9997 and the discrimination capacity was 0.9387.

Pairwise genetic distances and AMOVA of the studied Korean provinces reflected no patrilineal substructure in Korea, except for Jeju Island. Thus, this survey shows that the present data of Korean individuals could be helpful to establish a comprehensive forensic reference database for frequency estimation.

Keywords Y chromosome · SNP · STR · Forensic genetics · Population structure · Koreans

Introduction

Y chromosome short tandem repeats (STRs) are powerful markers for forensic purposes (particularly in male–female DNA mixtures) and paternity testing. To estimate the match probability of the Y-STR haplotype between the compared DNA, it is a prerequisite to construct a Y-STR population database taking the subpopulation structure into account. The Y chromosome is more sensitive to genetic drift due to its fourfold smaller effective population size compared with an autosome. Therefore, it is necessary to verify that no population substructure exists before pooling Y-STR data from different regions of the country [1, 2]. Although studies have been undertaken to establish the Korean Y-STR databases [3–5], a focus on population substructure of Y-STR haplotypes within Korea has not yet drawn much attention.

Recently, as the Y chromosome single nucleotide polymorphism (SNP) haplogroup has become a valuable tool for predicting the geographical or ethnic origin of unknown forensic evidences, there has been an increasing interest about the combined data on Y-SNP haplogroups and Y-STR haplotypes [6–9]. Here, the combined Y-SNP haplogroups and Y-STR haplotypes of the Korean population consisting of six major provinces were analyzed in

Wook Kim and Won Kim contributed equally to this paper.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0501-1) contains supplementary material, which is available to authorized users.

S. H. Kim · M. S. Han
DNA Analysis Division,
National Institute of Scientific Investigation,
Seoul 158-707, South Korea

W. Kim (✉)
Department of Biological Sciences, Dankook University,
Cheonan 330-714, South Korea
e-mail: wookkim@dankook.ac.kr

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

Table 1 Distribution of Y-SNP haplogroup frequencies and forensic parameters of Y-SNP haplogroups and Y-STR haplotypes in six provinces from Korea

| Haplogroup | Seoul-Gyeonggi (n=110) | Gangwon (n=63) | Chungcheong (n=72) | Jeolla (n=90) | Gyeongsang (n=84) | Jeju (n=87) | Total (%) (n=506) |
|-----------------------------|---------------------------|-------------------|-----------------------|------------------|----------------------|----------------|----------------------|
| C | 15 | 8 | 8 | 12 | 14 | 7 | 64 (12.6) |
| D | 1 | | 1 | 3 | 2 | 1 | 8 (1.6) |
| K* | 1 | | | 1 | 1 | | 3 (0.6) |
| NO* | 2 | 4 | 3 | 4 | 4 | 6 | 23 (4.5) |
| O1a | 1 | 1 | 1 | 1 | 2 | 5 | 11 (2.2) |
| O2* | | | | 2 | 3 | | 5 (1.0) |
| O2b* | 23 | 20 | 15 | 21 | 15 | 20 | 114 (22.5) |
| O2b1 | 8 | 5 | 7 | 7 | 10 | 8 | 45 (8.9) |
| O3 | 56 | 24 | 36 | 39 | 31 | 38 | 224 (44.3) |
| P | 3 | 1 | 1 | | 2 | 2 | 9 (1.8) |
| 12 Y-SNP haplogroup | | | | | | | |
| Haplogroup diversity | 0.6781 | 0.7389 | 0.6921 | 0.7383 | 0.7940 | 0.7412 | 0.7276 |
| Discrimination capacity (%) | 8.18 | 11.11 | 11.11 | 10.00 | 11.90 | 9.20 | 1.98 |
| 17 Y-STR haplotype | | | | | | | |
| No. of haplotypes | 108 | 62 | 71 | 88 | 84 | 79 | 475 |
| No. of unique haplotypes | 106 | 61 | 70 | 86 | 84 | 73 | 452 |
| Haplotype diversity | 0.9997 | 0.9995 | 0.9996 | 0.9995 | 1.0000 | 0.9971 | 0.9997 |
| Discrimination capacity (%) | 98.18 | 98.41 | 98.61 | 97.78 | 100.00 | 90.80 | 93.87 |

order to improve our understanding of the population structure of Korea.

Materials and methods

Subjects

Buccal swabs and blood samples were collected from 506 unrelated males residing in six major provinces in Korea: Seoul-Gyeonggi (YA003583, $n=110$), Jeolla (YA003583, $n=63$), Chungcheong (YA003588, $n=72$), Jeolla (YA003584, $n=90$), Gyeongsang (YA003586, $n=84$), and Jeju (YA003585, $n=87$) [10]. Appropriate informed consent was obtained from all subjects. Genomic DNA extraction was carried out by DNA IQ™ System (Promega, Madison, WI, USA). We performed proficiency testing of Collaborative Testing Service and the Y chromosome haplotype reference database (YHRD; <http://www.yhrd.org>) trials during the course of this study.

Y-SNP genotyping

As part of the present study, a 10-plex SNaPshot assay (M45, M89, M119, M122, M174, M175, M214, P31, SRY465, and 47z) and two singleplex SNaPshot assays (RPS4Y, and M9) were developed. A phylogenetic tree defined with the 12 Y-

SNP markers and scheme of 12 Y-SNP genotyping is shown in Fig. S1. The electropherograms of Y-SNP genotyping results are shown in Fig. S2. The Y-SNP haplogroups were named using the Y Chromosome Consortium nomenclature [11]. Primers for polymerase chain reaction (PCR) and single base extension (SBE) reaction were designed with the Primer 3.0 program. For M45, M89, and M175, previously described PCR primers were used [12].

Ten-plex PCR reaction was performed on a GeneAmp PCR System 9700 in a 25 μ l of 1 ng DNA, 1 \times PCR buffer II (Applied Biosystems, Foster City, CA, USA), 1.5 mM MgCl₂, 200 μ M of each dNTPs, 0.08–0.4 μ M of each primer (Table S1), bovine serum albumin (0.16 mg/ml), and 1 U AmpliTaqGold DNA polymerase. The cycling con-

Table 2 Haplogroup frequencies of the Korean population compared with those published Korean

| Haplogroup | Frequency (%) | | |
|------------|-----------------------|-----------------------------|---------------------------|
| | This study (n=506) | Hong et al. [14] (n=154) | Xue et al. [16] (n=43) |
| O3 | 44.3 | 42.2 | 39.5 |
| O2b* | 22.5 | 14.3 | 14.0 |
| C | 12.6 | 14.9 | 16.3 |
| O2b1 | 8.9 | 5.8 | 14.0 |
| Others | 11.7 | 22.7 | 16.3 |

Table 3 Pairwise R_{ST} values (below diagonal) and p values (above diagonal) between six provinces from Korea

| | Seoul-Gyeonggi | Gangwon | Chungcheong | Jeolla | Gyeongsang | Jeju |
|----------------------------|----------------|----------|-------------|----------|------------|----------|
| Seoul-Gyeonggi ($n=110$) | – | 0.15315 | 0.57430 | 0.98178 | 0.72716 | 0.03336* |
| Gangwon ($n=63$) | 0.00613 | – | 0.43421 | 0.51302 | 0.32017 | 0.02218* |
| Chungcheong ($n=72$) | –0.00209 | –0.00041 | – | 0.76705 | 0.29997 | 0.34700 |
| Jeolla ($n=90$) | –0.00722 | –0.00170 | –0.00480 | – | 0.96406 | 0.11246 |
| Gyeongsang ($n=84$) | –0.00361 | 0.00188 | 0.00193 | –0.00754 | – | 0.05465 |
| Jeju ($n=87$) | 0.01143* | 0.01784* | 0.00122 | 0.00672 | 0.01045 | – |

* $p < 0.05$

ditions were 95°C for 5 min followed by 35 cycles of amplification at 95°C for 30 s, 62°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 7 min. PCR products were purified by QIAquick gel extraction kit (Qiagen, Valencia, CA, USA) after confirmation of PCR amplicons by 2% agarose gel electrophoresis.

Ten-plex SBE reaction was performed on a GeneAmp PCR System 9700 in 10 μ l with 4 μ l of purified PCR product, 5 μ l of SNaPshot reaction mix (Applied Biosystems), and 1 μ l of SBE primer mix (0.01–1.2 μ M of each primer, Table S2). The SBE primer mix was diluted in 200 mM ammonium sulfate to minimize primer-dimer artifacts. The cycling parameters consisted of 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. The unreacted fluorescent ddNTPs were removed by addition of 1 U shrimp alkaline phosphatase to the SBE product and incubation at 37°C for 45 min followed by 80°C for 15 min. Samples were analyzed by capillary electrophoresis on an ABI Prism 3100 Genetic Analyzer in which 1 μ l of SBE product was mixed with 20 μ l of Hi-Di formamide and 0.3 μ l of LIZ120 internal size standard. Automated allele calls were made using GeneMapper ID v3.2.

To define haplogroup C-RPS4Y and K-M9, samples initially resulting as haplogroup Y* ($n=64$) and F* ($n=3$) using 10-plex SNaPshot assay were further analyzed by the singleplex SNaPshot assay based on each primer pair of RPS4Y and M9, respectively (Fig. S1). Cycling parameters were initial denaturation at 95°C for 5 min followed by 35 cycles of amplification at 95°C for 30 s, 60°C for 30 s, 72°C for 30 s, followed by a final extension at 72°C for 7 min. Conditions of the SBE reaction and detection of the RPS4Y and M9 by capillary electrophoresis were the same as the aforementioned 10-plex SNaPshot genotyping assay.

Y-STR genotyping

Y-STRs (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, YGATA H4, DYS437, DYS438, and DYS448) were amplified with the AmpFISTR Yfiler kit (Applied Biosystems) and analyzed by capillary electrophoresis using an ABI Prism 3730 Genetic Analyzer and GeneMapper ID v3.2. Alleles were named according to the International Society of Forensic Genetics recommendations [1].

Table 4 YHRD (release 34; 89,237 MinHt) search of the six most frequent minHt of the Korean population

| MinHt | Freq. (n) | Haplogroup | YHRD ^a | | | | | | | | |
|--------------------------------------------|---------------|------------------|-------------------|-------|--------|--------|-------|--------|--------|--------|--------|
| | | | Kor | Jap | Sin | Adm | Aus | E-Eu | W-Eu | Other | Total |
| Matching haplotypes (n) | | | 3,706 | 1,846 | 10,404 | 10,073 | 1,855 | 11,051 | 18,755 | 29,750 | 87,440 |
| MinHt1: 16-14-29-23-10-13-13-10 ,18 | 16 | O2b ^a | 98 | 17 | | 1 | | 1 | | | 117 |
| MinHt2: 16-14-29-23-10-13-13-10 ,17 | 14 | O2b ^a | 74 | 8 | | 1 | | | | | 83 |
| MinHt3: 16-14-29-23-10-13-13-10 ,19 | 8 | O2b ^a | 81 | 7 | 2 | 1 | 1 | | 1 | | 93 |
| MinHt4: 15-14-30-22-10-13-13-10,20 | 5 | O2b1 | 22 | 56 | 1 | 2 | | | | | 81 |
| MinHt5: 14-12-28-25-10-14-12-13,19 | 5 | O3 | 27 | 2 | 17 | | | | | | 46 |
| MinHt6: 15-12-29-23-10-12-12-11,19 | 5 | O3 | 40 | 1 | | | | | | | 41 |

Kor Korean, Jap Japanese, Sin Sino-Tibetan, Adm Admixed, Aus Austronesian, E-Eu Eastern European, W-Eu Western European.

The three most frequent minHts (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) without DYS385a/b (minHt 1–3) in Korean males were identical and shown in bold letters. The six population samples studied here are included in the YHRD (uploaded to release 32)

^a Some samples omitted DYS385a/b are not included in this data

Data analysis

Haplogroup, haplotype, and allele frequencies were estimated by direct counting. Haplogroup, haplotype, and gene diversities were calculated as previously described [13]. Discrimination capacity was determined by dividing the number of different haplogroups or haplotypes by the number of sampled individuals. Pairwise genetic distances (R_{ST}) and analysis of molecular variance (AMOVA) based on 15 Y-STR markers (excluded DYS385a/b) were calculated using Arlequin 3.11 [14].

Results and discussion

In all the individuals studied for the 12 Y-SNP markers, a total number of 10 Y-SNP haplogroups was observed from 506 Korean male samples. The frequency distribution of the 10 Y-SNP haplogroups in six major provinces of Korea is listed in Table 1. The overall haplogroup diversity was 0.7276, ranging from 0.6781 in Seoul-Gyeonggi to 0.7940 in Gyeongsang.

Most haplogroups in the Korean population were found in one of the two major Asian Y-SNP haplogroups, O-M175 (78.9%) and C-RPS4Y (12.6%). Among the five subhaplogroups of O-M175 surveyed here, O3-M122, O2b*-SRY465, and O2b1-47z constituted 44.3%, 22.5%, and 8.9%, respectively. The distribution of the four most frequent Y-SNP haplogroups in the Korean provinces was shown in Fig. S3 and compared with previously published Korean population (Table 2).

The Y-SNP haplogroups and Y-STR haplotypes obtained from the six major provinces of Korea were submitted to the YHRD and are searchable via the accession numbers (Table S3). A total of 475 haplotypes were identified using 17 Y-STR markers included in the Yfiler kit, among which 452 (95.2%) were found once. Eighteen, two, and three Y-STR haplotypes were shared by two, three and four individuals, respectively. All provinces exhibited a high haplotype diversity (>0.99), ranging from 0.9971 in Jeju Island to 1.0000 in Gyeongsang (Table 1). The overall haplotype diversity was 0.9997, which represented value similar to the previous Yfiler data set of the Korean population (0.9996, $n=526$) [5]. The overall discrimination capacity was 0.9387.

The R_{ST} between all possible pairs of six Korean provinces were non-significant ($p>0.05$), except in two comparisons of Jeju vs. Seoul-Gyeonggi, and Jeju vs. Gangwon (Table 3), indicating little genetic difference. Jeju was also lowest in haplotype diversity (0.9971) among the six provinces studied here (Table 1), probably as a result of genetic drift acting on this isolated island. AMOVA confirmed that there were no statistically significant genetic differences among the provinces ($R_{ST}=0.00169$, $p=$

0.25119), indicating that almost all (99.83%) of the genetic variation was due to variation within the provinces, with the non-significant variation (0.17%) being due to the differences among the provinces. These results indicate no patrilineal substructure in Korea, except for Jeju Island. A previous study on the coancestry coefficients estimated for the autosomal STR makers also did not support substantial heterogeneity among the geographic subpopulations of Korea [15]. The genetic homogeneity that likely arose by a shared national history (say almost 5,000 years) in the relatively small territory of the Korean peninsula seems to be the main characteristic of the population structure of Korea regarding the Y chromosome.

The most frequent minHts (minHt 1~3), which were shared by 16, 14, and eight Korean males, belonged to haplogroup O2b*, although the most frequent haplogroup in Korea was O3 (Table 4). The next most frequent minHts (minHt 4~6), which were shared by five Korean males, belonged to haplogroup O2b1 or O3. The six most frequent minHts were searched for in the YHRD database (release 34) among 87,440 minHts in a set of 659 populations. All six most frequent minHts (minHt 1~6) of the Korean population matched with the Japanese population samples (1,846 minHts) whereas only 3 minHts (minHt 3~5) matched with the Sino-Tibetan population samples (10,404 minHts; Table 4).

In summary, R_{ST} and AMOVA of the six Korean provinces studied here show that there are no genetic differences among these provinces, indicating Y chromosome homogeneity in the Korean population, except for Jeju Island. The Y-SNP haplogroups and Y-STR haplotypes of Korean individuals could be helpful to establish a comprehensive forensic reference database for frequency estimation.

Acknowledgments The authors would like to thank all volunteers for providing DNA samples. They also thank K.C. Kim and N.Y. Kim for technical assistance and data analysis of this survey. The present research was conducted by the research fund of Dankook University in 2009.

References

1. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (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 Leg Med* 120:191–200
2. Roewer L, Willuweit S, Krüger C, Nagy M, Rychkov S, Morozowa I, Naumova O, Schneider Y, Zhukova O, Stoneking M, Nasidze I (2008) Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. *Int J Leg Med* 122:219–223
3. Chun BW, Shin SC, Kim YJ, Kim KS, Choi DH, Kim KH, Kim JY, 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

4. Hwang JH, Lee H, Lee YA, Han MS, Kim W, Jin I, Cho NS (2007) Haplotypes for 12 Y-chromosomal STR loci in a Korean population (the central region). *Forensic Sci Int* 168:73–84
5. Kim SH, Kim NY, Hong SB, Cho NS, Kim JJ, Han MS, Kim W (2008) Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic Sci Int Genet* 2:e9–e10
6. Nonaka I, Minaguchi K, 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
7. Mizuno N, Kitayanma T, Fujii K, Nakahara H, Yoshida K, Sekiguchi K, Yonezawa N, Nakano M, Kasai K (2010) A forensic method for the simultaneous analysis of biallelic markers identifying Y chromosome haplogroups inferred as having originated in Asia and the Japanese archipelago. *Forensic Sci Int Genet* 4:73–79
8. Robino C, Crobu F, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, Piazza A, Inturri S, Torre C (2008) Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. *Int J Leg Med* 122:251–255
9. Ferri G, Tofanelli S, Alù M, Taglioli L, Radheshi E, Corradini B, Paoli G, Capelli C, Beduschi G (2010) Y-STR variation in Albanian populations: implications on the match probabilities and the genetic legacy of the minority claiming an Egyptian descent. *Int J Leg Med*. doi:10.1007/s00414-010-0432-x
10. Willuweit S, Roewer L, International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): update. *Forensic Sci Int Genet* 1:83–87
11. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838
12. Sanchez JJ, Børsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137:74–84
13. Nei M (1987) *Molecular evolutionary genetics*. Columbia University, New York
14. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
15. Lee JW, Lee HS, Hwang JJ (2002) Statistical analysis for estimating heterogeneity of the Korean population in DNA typing using STR loci. *Int J Leg Med* 116:153–160
16. Hong SB, Jin HJ, Kwak KD, Kim W (2006) Y-chromosome haplogroup O3-M122 variation in East Asia and its implications for the peopling of Korea. *Korean J Genet* 28:1–8
17. Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurler ME, Yang H, Tyler-Smith C (2006) Male demography in East Asia: a north–south contrast in human population expansion times. *Genetics* 172:2431–2439