



Y-chromosome haplogrouping for Asians using Y-SNP target sequencing

Sun Seong Choi, Kyung Hwa Park, Da Eun Nam, Tae Hoon Kang, Ki Wha Chung*

Department of Biological Sciences, Kongju National University, 56 Gongjudaehak-ro, Gongju 32588, Republic of Korea

ARTICLE INFO

Keywords:

Y-haplogroup

SNP

Target sequencing panel

Asian

ABSTRACT

Y-chromosomal haplogroups are sets of ancestrally related paternal lineages. We prepared a next generation sequencing (NGS)-based target sequencing panel of 85 Y-SNPs to determine Y-haplogroup of Asian populations. The Y-SNP panel was applied to several Asian ethnic groups: Korean, Chinese (Han and Dai), Japanese, Vietnamese, and Pakistani. Target capture was done using the SureSelect XT Custom capture kit (Agilent) and NGS was performed using the HiSeq 2500 (Illumina). This study could provide a fine Y-haplogroup tree for Asian groups. The Y-SNP panel was determined to be highly exact and reproducible and can be useful for the rapid determination of Asian Y-haplogroups. In particular, since there is a significant difference in the Y-haplogroup distribution among the Northeast Asian populations, applying the Y-SNP panel will help predict the ethnics among the Northeast Asian countries.

1. Introduction

Y-DNA haplogroups represent major branches of the Y-chromosome phylogenetic tree. Y-haplogroups are distributed non-randomly in the population and geographical regions, allowing us to infer the ethnic or geographical origin of unknown samples [1–3]. Because East Asian human groups are genetically close, it is difficult to distinguish race from autosomal SNP. However, certain haplogroup distributions may be different even between closely related populations, which can help predict race. Prediction of race in a foreign crime scene can narrow the scope of the suspect, which can be a powerful help in criminal investigation.

Recent advances in NGS technology have made simultaneous typing of many SNPs easy. This study prepared a target sequencing panel of 85 Y-SNPs to determine Y haplogroup of Asian populations.

2. Materials and methods

Genomic DNA was extracted from the blood samples from unrelated 56 Korean and 66 Pakistanis males using the QIAamp DNA Blood Kit (Qiagen, Hilden, Germany) after obtaining written informed consent. We selected 85 Y SNP markers which are involved in the major branches of the Y lineage or relatively common in Asian populations referred by The Y Chromosome Consortium (YCC) and ISOGG Y-DNA haplogroup tree version 2017 (<http://www.isogg.org/tree/index.html>).

The probe design was performed according to the SureDesign (<https://earray.chem.agilent.com/suredesign>), and then DNA samples

were prepared using the SureSelect DNA Sample Preparation System (Agilent Technologies, Santa Clara, CA, USA). NGS-based target sequencing was performed using a HiSeq 2500 Genome Analyzer (Illumina, San Diego, CA, USA). Paired-end sequences were mapped against reference genome UCSC hg19 (<http://genome.ucsc.edu>). The reads were mapped using BWA (<http://bio-bwa.sourceforge.net/>). Y-haplogroup was determined by the basis of Y-STR Haplotype Reference Database (YHRD, <https://yhrd.org/>) and International Society of Genetic Genealogy (ISOGG, <http://www.isogg.org/tree/index.html>).

3. Results and discussion

The Y-SNP panel was applied to Korean (n = 56) and Pakistani (n = 66) populations. Coverage rate showed an average of 97.3% ($\geq 10\times$ read depth). Mean read depth was 1294.0 ± 417.1 , and most markers were within the range of three times of \pm SD. The accuracy of the target sequencing was estimated to be 99.7% by Sanger sequencing method. Reproducibility was confirmed by double test.

In addition to Koreans and Pakistanis, polymorphic information of the Y-SNPs for several other Asian populations (Han and Dai Chinese, Japanese, and Vietnamese) could be obtained from the 1000 Genomes Browser (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes>). Haplogroup distribution was similar with previous studies for Koreans and Japanese [4,5]. A total of 48 sub-haplogroups were observed when Y-haplogroup tree was constructed for 6 Asian populations (Fig. 1). Similar distribution of Y-haplogroups was observed in Korean, Japanese and Han Chinese in Northeast Asia because of genetic similarity, but in Japan,

* Corresponding author.

E-mail address: kwchung@kongju.ac.kr (K.W. Chung).

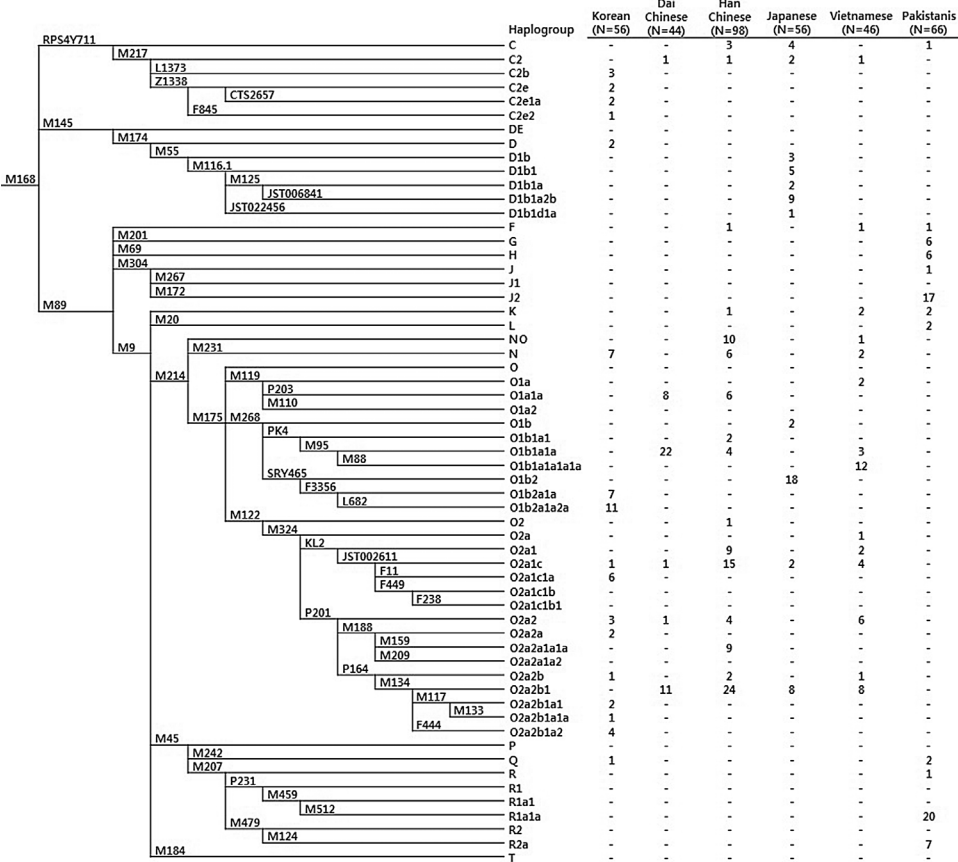


Fig. 1. Phylogenetic tree of Y-chromosome haplogroups observed in Asian populations. The observed numbers are shown at the right side of each subhaplogroup. The Y-SNP marker names are shown along the branches.

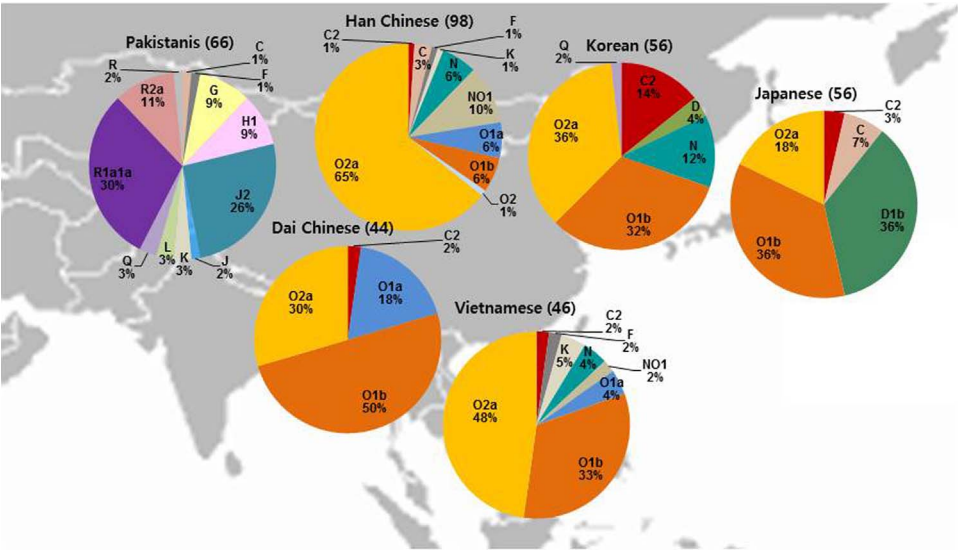


Fig. 2. Comparison of Y-haplogroup frequencies among several Asian populations. Haplogrouping was achieved by application of the Y-SNP panel for Koreans and Pakistanis, whereas haplogrouping for the two Chinese groups, Japanese, and Vietnamese was done by the data from the 1000 Genomes Browser (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>).

frequency of haplogroup D was significantly higher (36%) than that of Koreans (4%) and Chinese (0%). Frequencies of O1b haplogroup was increased in the order of Chinese (6%), Koreans (32%) and Japanese (36%). Conversely, O2a showed the reverse distribution (65% > 36% > 18%). Vietnamese showed different frequencies in the several subhaplogroups. Particularly, haplogroups of K (5%) and F (2%) were observed only in Vietnamese. Dai Chinese, a minority race, showed very simple structure with 98% of O haplogroup. As expected, the distribution of Y-haplogroups in Pakistanis was very different from that of the haplogroups in East Asian races. The frequencies of R1a1a, R2a and J2 subhaplogroups were high (30%, 11%, and 26%), whereas, O haplogroup was not observed at all

(Fig. 2).

In this study, the designed NGS-based Y-SNP panel was determined to be highly exact and reproducible and can be useful for the rapid determination of Asian Y-haplogroups. In particular, there is a significant difference in the Y-haplogroup distribution among the Northeast Asian populations.

Conflict of interest statement

None.

Acknowledgment

This study was supported by a grant from the National Forensic Service (NFS), Republic of Korea (2017-Gene-03).

References

- [1] J. Chikaroni, P.A. Underhill, L.L. Cavalli-Sforza, Y chromosome diversity human expansion, drift, and cultural evolution, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 20174–20179.
- [2] J.M. Butler, Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis, *Forensic Sci. Rev.* 15 (2003) 91–111.
- [3] M.A. Jobling, Y-chromosomal SNP haplotype diversity in forensic analysis, *Forensic Sci. Int.* 118 (2001) 158–162.
- [4] M.J. Park, H.Y. Lee, N.Y. Kim, et al., Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded DNA, *Forensic Sci. Int. Genet.* 7 (2013) 75–81.
- [5] E. Ochiai, K. Minaguchi, P. Nambiar, et al., Evaluation of Y chromosomal SNP haplogrouping in the HID-ion AmpliSeq™ identity panel, *Legal Med. (Tokyo)* 22 (2016) 58–61.