East Asian mtDNA haplogroup determination in Koreans: Haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis

Hwan Young Lee[1], Ji-Eun Yoo[1], Myung Jin Park[1], Ukhee Chung[1], Chong-Youl Kim[1,2] and Kyoung-Jin Shin[1,2]

[1] Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea

[2] Human Identification Research Center, Yonsei University, Seoul, Korea

Recently, a high-quality mtDNA control region sequence database for 593 Koreans was established through a redundant approach to data generation and analysis. For data quality control, Korean mtDNAs (99.8%) were also classified into various East Asian mtDNA haplogroups or subhaplogroups based on previously reported patterns of shared haplogroup-specific or haplogroup-associated polymorphisms in the control region. However, many haplogroup-diagnostic SNPs are located in mtDNA coding regions, and in some haplogroups, scoring of coding region SNPs is required for exact haplogroup determination due to the lack of information in their control region sequences. In addition, the ideal approach to haplogroup determination is the direct confirmation of diagnostic coding region SNPs. Accordingly, 21 SNP markers and 1 deletion motif from the coding region were selected, and three multiplex systems applying single base extension methods were developed in the present study. Using two of the multiplex systems, all 593 Korean mtDNAs were allocated into 15 East Asian major haplogroups: M, D, D4, D5, G, M7, M8, M9, M10, M11, R, R9, B, A, and N9. Using the other multiplex reaction, haplogroup D4 was further determined into six subhaplogroups: D4a, D4b D4e, D4g, D4h, and D4j. Using these techniques, the absence of major systematic errors in the data was confirmed. In addition, control region mutation motifs important for the assignment of East Asian mtDNA haplogroups and subhaplogroups were identified by collating informative control region SNPs on the basis of coding region SNP information. As the small amplicon sizes used in the three multiplex systems are expected to produce good results in degraded samples, the efficiency of the systems were tested in 101 samples from skeletal remains obtained from Korean War (1950-1953) victims.

Table 1. Selected SNPs and primers for the multiplex PCR amplification

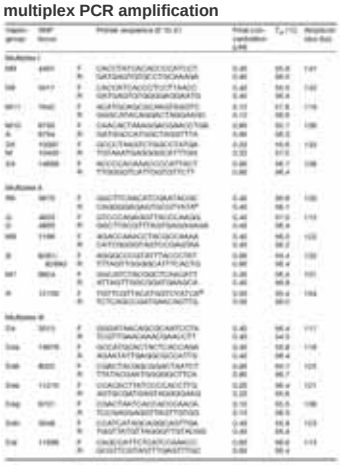


Table 2. Single base extension primers for the detection of 22 mtDNA SNPs

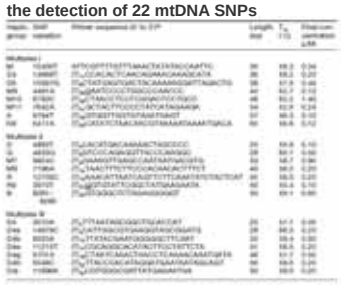


Figure 1. Electropherograms of the multiplex I and multiplex II reactions obtained from a sample belonging to the haplogroup D5 (A) and a sample belonging to the haplogroup B (B). Electropherogram of the multiplex III from samples belonging to D4a and D4b, respectively (C)

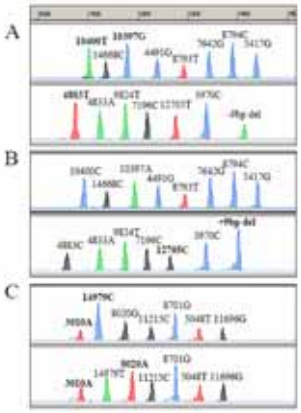


Figure 2. Electropherograms of the three multiplexes from skeletal remain sample, B042, belonging to D4g

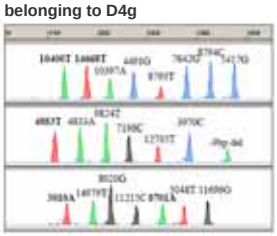


Figure 3. mtDNA Haplogroup frequencies observed in 694 Koreans

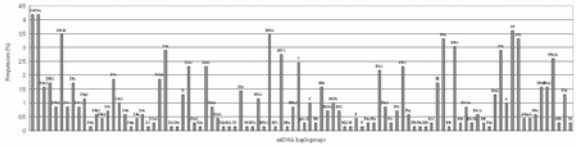
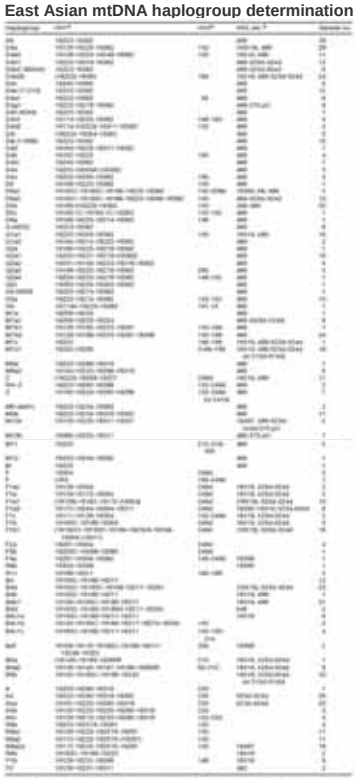


Table 3. Control region mutation motifs for East Asian mtDNA haplogroup determination



In conclusion, the data show that the identification of control region mutation motifs and the molecular dissection of haplogroups can be achieved by coding region SNP analysis. This study shows that East Asian mtDNA haplogroup determination is efficiently carried out using haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis, and that East Asian mtDNA data quality control and error identification can be easily performed using the reliable control region mutation motifs. Also, the three multiplex systems produced good results even in degraded samples and are a promising tool for forensic and evolutionary genetics involving East Asian mtDNA haplogroup determination.

Published in Electrophoresis (2006) 27:4408-4418