PREDICTION OF ASIA SUB-POPULATION USING GAUSSIAN BAYES CLASSIFIER WITH ALLELE FREQUENCY

Eu-Ree Ahn, Ki-Min Seong, and Hyun-Chul Park
Forensic DNA Division, National Forensic Service

Allele frequency is the frequency of allele at a locus in a particular population. The allele frequency is different for each population group unless the Hardy-Weinberg principle is violated. Populations can be classified based on the difference of allele frequency for each locus. The Asia is historically and geographically close, and gene flow has frequently occurred. Therefore, it is not easy to distinguish between neighboring Asia countries with genetic compositions. However, when Asia continent was divided into sub-populations according to region where each country was located, the allele frequency showed a distinct difference between each sub-population. In this study, we tried to predict Asia sub-population using the gaussian bayes classifier with allele frequency of autosomal STR. We divided the Asian region into 6 sub-groups as follows; Northern Asia, Eastern Asia, South-East Asia, Southern Asia, Central Asia, and Western Asia. The allele frequency of 21 Asia countries included in each sub-group was used to calculate likelihood ratio, and power of discrimination for each locus was used as the prior probability. The posterior probability was calculated based on the likelihood ratio and prior probability.