A13 Inference of Ethnic Origins of Forensic Unknown Y-STR Profiles in Singapore

Hang Yee Wong, MSc, Eng Seng Simon Lim, BSc, and Hui Koon Joyce Low, BSc, Health Sciences Authority, DNA Database, 7 Maxwell Road, MND Building Annex B, #05-04, S069111, SINGAPORE; and Wai Fun Tan-Siew, MSc, Health Sciences Authority, 11 Outram Road, Singapore, 169078, SINGAPORE*

After attending this presentation, attendees will understand that the cumulative variations of the Y-STR markers among the different ethnic groups might be used to infer the ethnic origins of an unknown Y-STR profile obtained from a crime scene.

This presentation will impact the forensic science community by demonstrating how the predicted ethnic origin can be a potential investigation lead for law enforcement agencies.

Singapore is a multi-ethnic country consisting of mainly Chinese, Malay, and Indian. In addition, with an increase in the foreign workforce, the percentage of other minority populations such as Caucasian, Bangladeshi, and Thai are on the rise as well. With this diverse ethnic population in view, it could be beneficial to the law enforcement agency if the ethnic group of the perpetrator can be predicted in order to provide an investigative lead. Based on the laboratory's Y-STR population database, it was observed that there are some unique and distinct differences in the Y-STR markers between the different ethnic groups. By studying these dissimilarities, the laboratory has developed an excel program to predict the ethnic group of unknown forensic Y-STR profiles.

The program was formulated based on a few assumptions and criteria. Firstly, as Y-STR is inherited as a haplotype from one generation to another, therefore haplotype comparison will be a major component in this program. Secondly, certain alleles are more common in one ethnic group compared to others, thus allowing those alleles to serve as distinctive markers for that ethnic group. Lastly, there is an unequal distribution of allele frequencies between the ethnic groups in a few loci if not all loci. Hence for calculation purposes, this program assumes each marker is an independent locus even though this defers from the haplotype inheritance pattern.

The Y-STR loci involved are DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385ab, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, Y GATA H4. Profiles of 240 individuals from each of the Chinese, Malay, Indian, Thai, Bangladeshi, and Caucasian ethnic groups were used to write this excel program. The functions of the program are to compare Y-STR haplotypes between the unknown profiles with those in the existing database, to identify alleles that are distinctive to each of the ethnic groups, and to provide arbitrary frequencies of the unknown profiles. Finally, all these aspects will be computed to rank and predict the possible ethnic group of the unknown profile.

Preliminary results based on a total of 160 Y-STR profiles demonstrated that the percentage of having the correct ethnic group being inferred is close to 88%, 66%, 89%, and 75% for Chinese, Malay, Indian, and Caucasian respectively. The percentage decreases for the three major ethnic groups when Bangladeshi and Thai are considered. This can be explained by the similarities among the Asian populations which are genetically closer in nature. An outlier ethnic group (African-American) is included in the program to serve as a control and none of the Asian ethnic groups returned African-American as the predicted ethnic origin.

This in-house excel program demonstrates the possibility of using Y- STR data to infer the most likely ethnicity of the DNA profile and to furnish such information to law enforcement agencies to serve as a potential investigative lead in Singapore.

Y-STR, Ethnic Groups, Haplotype