A CONVENIENT GUIDELINE TO DETERMINE IF TWO Y-STR PROFILES ARE FROM THE SAME LINEAGE

Jianye Ge[1], Hai Liu[2], Xiaoyang Li[2], Julio Mulero[1], Andrea Carbonaro[1], Mike Hoang[1], Marc Short[1]

[1]Human Identification Business Unit, Thermo Fisher Scientific, South San Francisco, CA, 94127, USA
[2]The Institute of Forensic Science and Technology, Henan Provincial Public Security Bureau, Zhengzhou, China

Y chromosome STR loci are used in forensic investigations primarily for identification purposes through the determination of the male lineage of a profile. Henan province of China has established a large Y-STR database for criminal investigations, which includes more than 200,000 Y-STR profiles. The database was built by first compiling the male lineages in the large cities within the province and then sampling a male representing each male lineage. The male lineages are defined by 5 generations. A large proportion of the Y-STR profiles in the database were generated using the Yfiler™ or Yfiler Plus™ kit. The additional loci in the Yfiler Plus kit as compared to the Yfiler kit results in a concomitant cumulative mutation rate increase across the loci. Therefore, in those cases when two profiles have one to a few mismatched loci, it is difficult to determine if they are from the same lineage. In other words, the boundary of related and unrelated males has become less certain especially when loci with rapidly mutating rates such as those found in the Yfiler Plus kit are taken into consideration.

In this study, 7,405 unrelated male profiles typed with the Yfiler Plus kit were manually selected from the database and a population study was conducted to compare the Yfiler loci and Yfiler Plus loci. The results showed higher power of discrimination and discriminatory capacity than the corresponding Yfiler haplotypes. Out of 1,100 pairwise matched Yfiler profiles, 986 are resolved with Yfiler Plus. Further, the distributions of the number of mismatched loci and the mismatched steps were generated for father-son, grandfather-grandson, uncle-nephew, and cousins (i.e., 1, 2, 3, and 4 meioses, respectively) by exhaustive pairwise comparison of the unrelated profiles using a dynamic programming approach. The same distributions were generated for unrelated pairs with mutation rates of the loci. With the distributions for both related and unrelated, the false negative and false positive rates of a given mismatch threshold can be inferred, which can be determined readily if two profiles are from the same male lineage with a minimum error rate. Generally, two Yfiler profiles with ≤2 mismatched loci and ≤2 steps are more likely from the same lineage than being from two unrelated lineages. Two Yfiler Plus profiles with ≤4 mismatched loci and ≤5 mismatched steps are more likely from the same lineage. For large scale Y-STR database searches, the threshold should still rely more so on the resources available for the follow-up investigations.