

CONCORDANCE STUDY ON MAJOR STR MULTIPLEX KITS AND THE EFFECT OF USING MULTIPLE KITS ON THE CHINA NATIONAL DNA DATABASE

Bin Li¹, Jianye Ge², Fushun Wu¹, Lin Ye¹, Seung Sum Seo², Arthur Eisenberg², Bruce Budowle², Yuanzhong Chen³

¹Forensic Science Division, Department of Fujian Provincial Public Security, Fuzhou, China

²Institute of Applied Genetics, Department Forensic and Investigative Genetics, University of North Texas Health Science Center, Fort Worth, TX, USA

³Fujian Institute of Hematology, The Union Hospital of Fujian Medical University, Fuzhou, Fujian Province, China

The China national DNA database has become the largest database in the world in a relatively short time. The STR multiplex kits from the major commercial entities (i.e., Life Technologies and Promega Corporation) used to be the main kits for generating DNA profiles for the China DNA database. Recently, several domestic commercial multiplex kits have been developed, such as DNA TyperTM15 Plus, AGCU 17+1, and GoldenEye 20A. Because different kits may contain use different sets of PCR primers for the same loci, discordant genotypes have been observed. These discordances should be evaluated to determine the rate of a false negative or incorrect exclusion of two samples that originate from a common source and to build mechanisms to accommodate the phenomenon. Such mismatches also can affect kinship analysis and familial searching, as well as mixture interpretation.

To investigate the discordant rate of using multiple kits for STR typing of samples contained within the China national DNA database, 2298 unrelated individuals of the Han population in China were analyzed first with SinofilerTM and PowerPlex[®] 16 HS. Discordant types were observed for 7 samples at a total of four loci. The discordance rate for these two kits was 0.3% with 95% binomial confidence interval of [0.12%, 0.63%]. With more than 13 million profiles in China DNA database (March 31, 2012), the number of profiles that may contain a discordant type for at least one locus between the two kits can be up to 81,900, i.e., 13 million × 0.63% (with upper bound of 95% confidence interval).

Further, these 7 samples were amplified with 7 additional kits (i.e., IdentifilerTM, Identifiler[®] Direct, Identifiler[®] Plus, NGM SElectTM, DNA TyperTM15 Plus, AGCU 17+1, GoldenEye 20A). Allele drop-out was observed in 4 of the samples, including two “10” alleles at the D8S1179 locus that dropped out with all kits from Life Technologies, one allele “13” at the D5S818 locus that dropped out with all kits from Life Technologies, one allele “12” at the CSF1PO locus that dropped out with both PowerPlex[®] 16 HS and GoldenEye 20A kits. Notable peak imbalance was observed in the remaining three samples at the CSF1PO (one sample) and D18S51 loci (two samples) with PowerPlex[®] 16 HS kit and D18S51 (one sample) with GoldenEye 20A. No inconsistent allele designations (such as “10” for one kit but “11” for another kit) were observed. These seven samples will be sequenced to determine the basis for the discordant results.

A few discordant types between different kits with 2,298 individuals may seem trivial. However, when databases become the magnitude of the China database large, the number of profiles that may be involved in comparisons where discordant results become qualitatively large. A false exclusion can be of greater consequence in a database search than a false positive. The study herein allows for an assessment of the impact on database searching. ☞