EXPANDING THE CODIS CORE LOCI IN THE UNITED STATES, HOW MANY AND WHICH LOCI MAKE THE MOST SENSE

Jianye Ge, Bruce Budowle, Arthur J. Eisenberg
University of North Texas Health Science Center, Department of Forensic and Investigative Genetics, Institute of Investigative Genetics, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107

Approximately one year ago the Federal Bureau of Investigation (FBI) established a CODIS Core Loci Working Group to review and recommend, if deemed appropriate, changes to the original 13 Core CODIS Loci. The Working Group recently released its recommended changes (Hares, FSI Genetics 2011 in press; PMID: 21543275) to the core STR loci for typing samples to be entered into CODIS. Reviewing, expanding, and replacing the core STR loci is lauded to improve capacity and efficiency with Database searches.

A table containing the ranked list of the new CODIS core loci was divided into 2 Sections. Section A contains the required 20 STR loci (comprised of the original CODIS core set minus TPOX, the new European recommended loci, Penta E, and DYS391) plus the amelogenin marker. Section B contains a list of 4 additional loci that could be added, in the order specified, but only after the loci in Section A have been included in a kit. The motivation is to enhance the core set to: (1) reduce the likelihood of adventitious matches because the CODIS database has, and will continue to substantially increase in size; (2) increase international compatibility for better data sharing; and (3) increase discrimination power for missing persons cases. Given these goals more scrutiny is needed to consider whether a list as large as 20 or more STRs meets the desired goals and whether alternate STRs, not included in the list, may be more effective for developing investigative leads using database searches. While there is some gain in identity resolution by increasing the core to all 20 loci, this effect has to be weighed against a loss in sensitivity. As more loci are added to a multiplex there is some loss in sensitivity of detection. Unless some enhancement in multiplex design is achieved, it is unlikely that a single 20+ loci megaplex will be as sensitive as current multiplex kits. Since sensitivity of detection is one of the most important criteria, it is likely that if so many loci are required then two multiplexes may be needed for casework samples even if one megaplex is developed for reference samples. The potential increased resource strain on laboratories must also be considered. Given the direction of casework towards typing more challenged samples (such as low quantity and/or degraded samples), then those STR loci that can be converted to mini-STRs should be considered the most desirable. Under this criterion, some of the other larger loci such as FGA may not be desirable. The effect of increasing the number of core loci on database searching will be discussed; simulations on data sets of 100 million profiles have been performed with the new core loci and various subsets. Is there any statistical support that the addition of 20+ loci is required to significantly improve the efficiency for single source profiles? What is the potential effect on searches with mixture profiles given their quality and other CODIS mixture upload requirements? If enhancing analysis in missing persons cases is the goal, then again mini-STRs would be more desirable since human remains are often compromised in both quality and quantity. One of the most likely used kinship analyses with an offender and arrestee database is familial searching. In this application it may be better to select 15 autosomal loci and combine those with six highly discriminating Y STRs. The Y STRs serve as an effective filter and if included with the reference profile will minimize the laboratory work necessary to reduce the candidate familial search list to a manageable number. The inclusion of a single Y STR (i.e., DYS391) would have no practical value for this purpose and its inclusion for verifying amelogenin inconsistencies may be of limited value.

This presentation will weigh the different criteria for selecting loci based on casework requirements, functionality using simulations and considering mixtures and mutation, and what would be best suited for missing persons and familial searching. A suggested strategy is considered so that decision makers and stakeholders have sufficient information when constructing new requirements for CODIS.