CHARACTERISTICS OF 23 COMMONLY USED STR LOCI AND U.S. POPULATION DATA WITH THE RECENTLY ANNOUNCED EXPANDED CODIS CORE LOCI

John M. Butler, Carolyn R. (Becky) Hill, David L. Duewer, Margaret C. Kline, Kristen L. O'Connor
National Institute of Standards and Technology, Applied Genetics Group, Gaithersburg, MD 20899-8314

Current commercially available short tandem repeat (STR) typing kits provide capabilities for analysis of 23 different autosomal loci including the 13 CODIS (Combined DNA Index System) STR loci [1,2] plus D2S1338, D19S433, Penta D, Penta E, D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33. Many of these additional STR loci are part of the expanded European Standard Set adopted in 2009 [3]. In the past two years, Promega Corporation and Applied Biosystems have released new STR kits to enable coverage of these additional loci. In 2010, Qiagen also began supplying STR typing kits in some parts of the world.

Earlier this year the National DNA Index System (NDIS), which currently contains the 13 CODIS STR loci, exceeded 10 million DNA profiles [4]. The large amount of legacy data in the U.S. makes it difficult to move to completely different typing systems, such as single nucleotide polymorphisms or a non-overlapping set of STR loci, because previously collected samples would have to be re-genotyped at the new markers. In April 2011, the NDIS Custodian announced a plan to expand the CODIS core loci in the United States to 20 required STR loci plus amelogenin with 4 optional loci [5]. The additional required loci include amelogenin, D2S1338, D19S433, D12S391, D1S1656, D2S441, D10S1048, Penta E, and the Y-chromosome locus DYS391. TPOX, the least polymorphic STR locus of the current CODIS core loci, has been made optional.

Data from almost 1,000 U.S. population samples were evaluated across the 23 STR loci using Applied Biosystems (Identifiler, MiniFiler, NGM, NGM SElect), Promega (PowerPlex 16, PowerPlex ESI 17, PowerPlex ESX 17, PowerPlex 18D), and Qiagen (ESSplex, IDplex) kits. Allele ranges and locus characteristics for the 23 common STR loci will be discussed. The probability of identity with different sets of loci will be illustrated to help assess the benefits of adding loci to the current 13 CODIS core loci.

References:

[1] Budowle, B., et al. (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification*, pp. 73-88. Madison, Wisconsin: Promega Corporation. Available at http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/2nd-eshi-oral-presentations/.

[2] Butler, J.M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *J. Forensic Sci., 51*, 253-265.

[3] Schneider, P. (2009). Expansion of the European Standard Set of DNA database loci – the current situation. *Profiles in DNA* 12(1): 6-7.

[4] http://www.fbi.gov/about-us/lab/codis/ndis-statistics

[5] Hares, D.R. (2011). Expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* (in press), doi:10.1016/j.fsigen.2011.04.012